

An Enquiry Concerning the Passions:
A critical study of Hume's *Four Dissertations*

Amyas John Henry Merivale

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

School of Philosophy, Religion, and the History of Science

June 2014

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

©University of Leeds and Amyas Merivale

Acknowledgements

I am enormously indebted to Helen Steward, my principal supervisor at Leeds, for her tireless attention to the detail of this work and the correction of countless errors along the way. Without her advice and support, moreover, the beast would have been considerably harder to tame, if it had not run away from me altogether. Immeasurable thanks are due to Peter Millican, my long-standing mentor and supervisor at a distance. We agree with each other to such an extent that I now find it difficult to recall where his ideas end and mine begin; both in the planning and the execution of this thesis he has been a generous and invaluable collaborator.

I am grateful also to Matthew Kieran, my secondary supervisor at Leeds, for his assistance with the later chapters on aesthetics. In addition, various of the ideas herein have benefited greatly from discussions with James Arnold, Filippo Contesi, and Samuel Rickless.

Finally, to Esther: for making this period such a happy one, and for her curiosity, patience, and support.

Abstract

Hume's first work, *A Treatise of Human Nature*, has traditionally received the lion's share of scholarly attention, at the expense of his later and more polished texts. The tide has started to shift in recent years, with the result that Hume's two *Enquiries*—his mature investigations of the understanding and morals—are now recognised as important works in their own right (though most commentators still continue to prefer the *Treatise*). With regard to Hume's work on the passions, however, Book 2 of the *Treatise* still commands all of the attention.

In this thesis, I defend two important claims. The first is that Hume has a mature philosophy of emotion, significantly different—indeed, significantly *improved*—from that of the *Treatise*. Most strikingly, it is anti-egoist and anti-hedonist about motivation, where the *Treatise* had espoused a Lockean hedonism and egoism. In parts it is also more cognitivist, and although Hume remains as opposed to moral rationalism as he ever was, his arguments in support of this opposition are very different.

The second claim is that Hume's mature philosophy of emotion is to be found, not in the *Dissertation on the Passions*, but rather in the full set of *Four Dissertations* in which this work first appeared, including also the *Natural History of Religion*, *Of Tragedy*, and *Of the Standard of Taste*. The passions, I argue, form the unifying theme of this collection, which is in effect Hume's *Enquiry concerning the Passions*. I maintain that they are profitably studied together on this understanding, and my thesis is offered as the first such study.

Conventions

References to Hume's texts throughout use the following abbreviations, followed by Book, part, section, and paragraph numbers (as appropriate):

T *A Treatise of Human Nature*

L *A Letter from a Gentleman to his Friend in Edinburgh*

E *An Enquiry concerning Human Understanding*

M *An Enquiry concerning the Principles of Morals*

D *Dialogues concerning Natural Religion*

N *The Natural History of Religion*

P *A Dissertation on the Passions*

Tr *Of Tragedy*

ST *Of the Standard of Taste*

Other essays are abbreviated in a similar fashion, but only after they have been introduced by name, so that it will be obvious which essay is being cited. Page numbers are to the editions that most Hume scholars would now expect (Selby-Bigge's editions of the *Treatise* and *Enquiries*, Miller's edition of the *Essays*); full references are in the bibliography of primary sources at the end. The quotations themselves are taken from the editions, prepared by Peter Millican and me, available at <http://www.davidhume.org/>. I adopt the same referencing and abbreviation conventions as are used on that site.

Quotations and references from other primary sources are from the editions listed in the bibliography. The page numbers are from modern editions, wherever these were easily available, but I cite these works using the original publication year, since I consider it desirable to be able to see at a glance when the thing in question was first said. The works often went through several editions in their authors' lifetimes, receiving additions and revisions in the process. My rule is to use the date of the *earliest* edition in which the quoted passage appeared. This has the effect that the same work is sometimes referred to with a different year. Butler's *Sermons*, for example, are typically referred to as "Butler (1726)"; but when I am quoting from Butler's preface, which was first added in the second edition, I use "Butler (1729)". A glance at the bibliography of primary sources will quickly reveal that these refer to different editions of the same text.

Contents

Introduction	1
0.1 An enquiry concerning the passions	1
0.2 Four/three/five/four dissertations	3
0.3 Editing the <i>Treatise</i>	6
0.4 Epicureans, Stoics, Platonists,	10
0.5 ... and Sceptics	13
I The Butler Moment	19
1 Some Late Philosophers in England	21
1.1 Hutcheson against Clarke	21
1.2 Hutcheson against Hobbes and Mandeville	23
1.3 Butler against Hobbes and Clarke	26
1.4 Shaftesbury against Hobbes and Locke	30
1.5 Egoists, Sentimentalists, Rationalists	35
2 Motivation	39
2.1 The Butler paragraph	39
2.2 Hedonism in the <i>Treatise</i>	42
2.3 The dignity of human nature	45
2.4 The two <i>Enquiries</i>	47
2.5 Desire in the <i>Dissertation on the Passions</i>	51

3	Hume’s Classification of the Passions	55
3.1	Impressions and sentiments	55
3.2	Three more distinctions	58
3.3	The Butler paragraph (again)	61
3.4	Calm and violent	65
3.5	Some peculiar feelings	69
4	Association, Comparison, Sympathy	73
4.1	The double relation theory	74
4.2	Comparison	77
4.3	Sympathy	79
4.4	More applications of sympathy	83
4.5	The indirect passions	88
5	Original Instincts	93
5.1	Object-directed desires and original instincts	93
5.2	Later traces of sympathy and comparison	97
5.3	Ambition, vanity, and parental love	100
5.4	Sympathy and benevolence	105
5.5	Sympathy and the rules of art	108
6	The Development of Hume’s Thought	113
6.1	Anti-egoism in the <i>Treatise</i>	113
6.2	Egoism, careless egoism, and anti-egoism	117
6.3	Butler, not Hutcheson	120
6.4	Going back to the source	124
6.5	Why so much resistance?	129
II	The Third Enquiry	133
7	Superstition and the Passions	135
7.1	Reasons and causes	136

7.2	Superstition and true religion	139
7.3	Hope and fear	144
7.4	The origin of polytheism and monotheism	147
7.5	The miraculous and the momentous	149
8	Some Comparisons	153
8.1	Hobbes on the origin of religious belief	153
8.2	Comparisons between monotheism and polytheism	155
8.3	The bull-rush argument	158
8.4	Hume's agnosticism and Hume's atheism	160
8.5	The three <i>Enquiries</i>	163
9	The Intentionality of the Passions	167
9.1	Cognition and the passions	167
9.2	The representative quality argument	171
9.3	Simple and complex	176
9.4	The double-relation passions	178
9.5	The double relation theory (again)	183
10	Reason and Passion	189
10.1	Malebranche against the passions	190
10.2	Clarke against the passions	194
10.3	Hutcheson against the rationalists	197
10.4	Hume against the rationalists	202
10.5	The representative quality argument (again)	206
11	The Chemistry of the Passions	211
11.1	Two overlapping theories	211
11.2	Du Bos	214
11.3	Fontenelle	217
11.4	Hume's change of mind	219
11.5	Hume's later account	222

12 The Conversion Principle	227
12.1 Swallowing up: passions or emotions?	228
12.2 Movements in the spirits	231
12.3 Conversion	233
12.4 Before and after	236
12.5 A subtler interpretation	240
13 Passion and Value	245
13.1 The sceptical principle	245
13.2 The true judges	248
13.3 The threat of circularity	252
13.4 Getting into the circle	254
13.5 Passing the buck	257
14 The Science of Criticism	261
14.1 Relevant virtues	261
14.2 A textual problem	264
14.3 Art and the passions	267
14.4 The rules of art	271
14.5 Many standards	274
Conclusion	277
Bibliography of Primary Literature	281
Bibliography of Secondary Literature	285

Introduction

0.1. An enquiry concerning the passions

In 1739 and 1740, Hume published the three Books of his *Treatise of Human Nature*, the first on the understanding, the second on the passions, and the third on morals. Although it is now widely considered a philosophical masterpiece, and is the primary source of Hume’s present reputation, it was not well received by his contemporaries, and Hume himself was very dissatisfied with it. Having had greater success with a subsequent set of *Essays, Moral and Political* (1741-42), Hume later returned to the subject matter of Books 1 and 3 of the *Treatise*, publishing the first edition of his *Enquiry concerning Human Understanding* in 1748, and the first edition of his *Enquiry concerning the Principles of Morals* in 1751.¹

This naturally prompts the question of what happened to Book 2. Where, we might ask, is Hume’s *Enquiry concerning the Passions*? Hume never published a work with this title, but the standard view, in so far as there is one, would seem to be that the work standing in the appropriate relations is his *Dissertation on the Passions*, first published in 1757.²

¹This is a slight (but harmless) simplification. The necessary qualifications are (i) that the *Enquiry concerning Human Understanding*—or “first” *Enquiry*, as it is usually called for short—was initially published under the title of *Philosophical Essays concerning Human Understanding*, and renamed only after publication of the “moral” or “second” *Enquiry* (second in time, that is, but not in name); and (ii) that section 8 of the first *Enquiry*, *Of Liberty and Necessity*, has its origins in Book 2 of the *Treatise* (T 2.3.1-2, pp. 399-412), rather than Book 1.

²The analogy was made explicitly by L. A. Selby-Bigge in the introduction to his edition of the two *Enquiries* (1893/1975, p. viii), and has not been contradicted since. I have previously endorsed this view myself (Merivale 2009, p. 185); I will recant presently.

If Books 1 and 3 of the *Treatise* have generally overshadowed their *Enquiry* successors, Book 2 has positively eclipsed the *Dissertation on the Passions*. Nor is it hard to see why: while the two *Enquiries* immediately strike the reader as quite substantial reworkings of their *Treatise* predecessors, containing much that is different and new, the *Dissertation* can easily look like nothing more than, in Selby-Bigge’s words, “verbatim extracts from Bk. II of the *Treatise*, with some trifling verbal alterations”.³ And so, while Hume’s mature investigations of the understanding and of morals have recently started to emerge from the shadow of the *Treatise* as subjects worthy of independent study,⁴ Hume’s mature philosophy of emotion continues to languish in the dark.⁵ This thesis is an attempt to bring it into the light.

Part of my case for thinking that Hume’s mature philosophy of the passions is worthy of study in its own right, is that Selby-Bigge was wrong about the *Dissertation*. It is true that a lot of this later work is accurately described as Selby-Bigge describes the whole, but it is my view that not all of the alterations are trifling or merely verbal. More than this, however, I will also be urging that the initial premise—concerning the textual source of Hume’s mature thought on this topic—is mistaken. The true successor to Book 2 of the *Treatise* is not the truncated *Dissertation on the Passions*, but rather the complete set of *Four Dissertations* in which this work first appeared, including also the *Natural History of Religion*, *Of Tragedy*, and *Of the Standard of Taste*.⁶

That the *Four Dissertations* can be viewed as Hume’s *Enquiry concerning the Passions* is thus the central claim of this thesis. It is shorthand for two closely

³Selby-Bigge (1893/1975, pp. xx-xxi).

⁴See Peter Millican on the first *Enquiry* (2002a,b, 2006), and Jacqueline Taylor on the moral *Enquiry* (2002, 2009, forthcoming).

⁵For example, of the 126 articles published in *Hume Studies* between 2001 and 2010 (volumes 27-36), 52 quote from Book 2 of the *Treatise*, while only 3 quote from the *Dissertation on the Passions*. The first of these quotes just six words in a footnote (Cunningham 2005, p. 251, n1); the second quotes a couple of short clauses alongside parallel *Treatise* passages (Postema 2005, pp. 266-7); the third is my own article (Merivale 2009).

⁶With this in mind, an addendum to note 5 is in order: from the same 126-article sample, I count only 14 quoting any one of Hume’s *Four Dissertations*. Most of these quotations are from *Of the Standard of Taste*, easily the best known of the four.

related proposals, which I will term the *Unity Thesis* and the *Difference Thesis* respectively:

The Unity Thesis The *Four Dissertations* form a unified set, and are profitably studied as such; what unifies them, in particular, is the subject matter of the passions.

The Difference Thesis The philosophy of emotion contained in these works is substantially and importantly different from Hume's earlier view, as presented in Book 2 of the *Treatise*.

In the remainder of this introduction I will put a little flesh on the bones of these two claims: the Unity Thesis in §§0.2 and 0.3, and the Difference Thesis in §§0.4 and 0.5.

0.2. Four/three/five/four dissertations

The *Four Dissertations* came into being only after a couple of false starts. The story here is of interest in its own right, and also has a bearing on my Unity Thesis. Hume recounted the main particulars in a letter to his publisher William Strahan, dated the 25th January, 1772:⁷

I am told by a Friend, that Dr Millar said to him, there was a Bookseller in London, who had advertised a new Book, containing, among other things, two of my suppress'd Essays. These I suppose are two Essays of mine, one on Suicide another on the Immortality of the Soul, which were printed by Andrew Millar about seventeen Years ago, and which from my abundant Prudence I suppress'd and woud not now wish to have revived. I know not if you were acquainted with this Transaction. It was this: I intended to print four Dissertations, the natural History of Religion, on the Passions, on Tragedy, and on the metaphysical Principles of Geometry. I sent them

⁷For an interesting, if somewhat speculative, reconstruction of the full story that Hume here alludes to, see E. C. Mossner (1950). For a more cautious account, see Tom L. Beauchamp (2007, pp. xxii-xxiv).

up to Mr Millar, but before the last was printed, I happend to meet with Lord Stanhope, who was in this Country, and he convinced me, that either there was some Defect in the Argument or in its perspicuity; I forget which; and I wrote to Mr Millar, that I would not print that Essay; but upon his remonstrating that the other Essays woud not make a Volume, I sent him up these two, which I had never intended to have publishd. They were printed; but it was no sooner done than I repented; and Mr Millar and I agreed to suppress them at common Charges, and I wrote a new Essay on the Standard of Taste, to supply their place. (HL 2, pp. 252-3)

Having thus brought these four works together, at least in part for pragmatic reasons, Hume then quickly separated them in the next edition of his *Essays and Treatises on Several Subjects* (1758). In this edition, *Of Tragedy* and *Of the Standard of Taste* were appended to part 1 of the *Essays, Moral, Political, and Literary*, while the *Dissertation on the Passions* and the *Natural History of Religion* were placed following the first and second *Enquiries* respectively, a rearrangement that persisted in all subsequent collections.

From this short history of these four works, one might be tempted to infer that they never really belonged together in Hume's mind, and hence that my Unity Thesis is simply wrong-footed from the start. Fortunately, however, this inference does not hold up under scrutiny. First, it is only *Of the Standard of Taste* that was a late addition, and nothing suggests that Hume didn't initially conceive the other three as a set. Nor is there anything to suggest that, faced with the need for a new fourth, Hume didn't then compose this late addition precisely with the other three in mind.⁸

⁸The same cannot be said of Hume's first choices for filling the gap, *Of Suicide* and *Of the Immortality of the Soul*. These two suppressed essays were written some time earlier, and were presumably offered for inclusion simply because something was needed, and they were to hand. Their irreligious content chimes with the *Natural History* in that one very general respect, and the defence of suicide would have struck a note of accord with the dedication that was ultimately prefixed to (some copies of) the *Four Dissertations*, which praised John Home's tragedy *Douglas*, itself highly controversial for its sympathetic or even positive portrayal of the suicide of its main character. Home's play did not première until 1756, however, *after* Hume had thought of including the scandalous essay in the set. In all, the specially written essay on taste is—or so I will argue—a much better fit with the other three.

Secondly, Hume's subsequent separation of these works can be explained by other factors, without supposing it was only pragmatism that brought them together in the first place. Shortly after publication of the *Four Dissertations*, Hume sent a copy to an acquaintance in Germany, writing in a cover letter that "[s]ome of these Dissertations are Attempts to throw Light upon the most profound Philosophy: Others contain a greater Mixture of polite Literature, & are wrote in a more easy Style & Manner."⁹ There is only one plausible way of reading this: the *Natural History* and the *Dissertation* contain the profound philosophy, while *Of Tragedy* and *Of the Standard of Taste* contain the greater mixture of polite literature.

Thus it would have made sense to Hume, when considering how these works might be incorporated into his *Essays and Treatises* (with the core structure of that collection already in place), to include the "profound" works alongside the two *Enquiries*, with the other two being included with the *Essays, Moral, Political, and Literary*. Add to this the simple point about the respective lengths of these works (which, however trivial, must surely have been a factor), and there is a compelling case for the separation, one that does not depend on anything internal pushing the *Four Dissertations* apart.

Perhaps less plausible, but still to my mind well within the realms of possibility, is that the subject of the passions was intended to be the common theme even according to the original plan, which featured a concluding dissertation on the metaphysical principles of geometry, before this was replaced by *Of the Standard of Taste*. This abandoned work sadly does not survive, but presumably it was at least in part a revision of *Treatise* Book 1, part 2 (*Of the ideas of space and time*, pp. 26-68), a topic which hardly features in the first *Enquiry*, save for three paragraphs in the context of a discussion of scepticism about abstract reasoning (E 12.18-20, pp. 156-8).

⁹This letter is in the Lilly Library of Indiana University. I have not read it, and owe my knowledge of its existence, and this quotation, to Immerwahr (1994, pp. 237-8). The distinction that Hume had in mind here was likely that between "accurate" and "easy" philosophy, as drawn in section 1 of the first *Enquiry*; the phrase "easy style and manner" is taken straight from this section (E 1.5, p. 8).

One might have thought it safe to assume that geometry would have no bearing on Hume's thought about the passions, but perhaps surprisingly this assumption is in fact false. There are two quite substantial sections in Book 2 of the *Treatise* devoted to discussing the effects on the passions of distance and contiguity in space and time (T 2.3.7-8, pp. 427-38). In the *Dissertation on the Passions*, these two sections are reduced to just a single sentence, which merely states one of the main observations: "What is distant, either in place or time, has not equal influence with what is near and contiguous" (P 6.18, p. 29). It seems to me entirely possible that the abandoned dissertation on geometry combined the exclusively metaphysical discussion of Book 1 with the psychological discussion of Book 2 that the *Dissertation on the Passions* itself almost entirely omits. At least, if this had been the case, the original four would have hung together very nicely; while if Hume had not discussed the relationship between geometry and the passions in this last dissertation, it is difficult to imagine what else could have saved it from sticking out like a sore thumb. (This point is all the more striking when we consider how very clearly the first three dissertations are connected; see §0.3 below.)

My Unity Thesis holds that Hume's *Four Dissertations* are *in fact* unified by the subject matter of the passions, whether by accident or by design. It also maintains that they are profitably studied together on this basis. This second claim, of course, stands or falls with the thesis as a whole. The proof of the pudding, as they say, is in the eating. In order to whet the appetite, however, I shall argue in the next section for *deliberate* unity in the construction of this short-lived set.

0.3. Editing the *Treatise*

When we look at the third and final part of *Treatise* Book 2 (*Of the will and direct passions*), and how Hume edited and reordered this material for the *Dissertation on the Passions*, we find very clear evidence that Hume himself initially conceived of the *Four Dissertations*—or at least the first three of them—as a unified set.

The beginnings of this argument have already been made by John Immerwahr,¹⁰ but as we will see he seems to have understated the case.

There are ten sections in part 3 of Book 2 of the *Treatise*. The first two are on liberty and necessity. This being a particularly important topic, and closely related to causation, Hume understandably revisited it as soon as possible, in section 8 of his first *Enquiry*, immediately following the section 7 discussion of necessary connection.¹¹ The third section of part 3, *Of the influencing motives of the will*, contains Hume's famous discussion of the relationship between reason and passion, a discussion that reappears in section 5 of the *Dissertation on the Passions* (albeit with some substantial cuts; see §10.5). It is the remaining seven sections that matter for the present argument.

The first thing to note (and this is the point Immerwahr makes) is that Hume reorders the material from these seven sections, moving his discussion of hope and fear from the *end* of *Treatise* Book 2 (T 2.3.9, pp. 438-48) to the *start* of the *Dissertation* (P 1, pp. 3-6), leaving the *Dissertation* to end instead with a discussion of the causes of the violent emotions (P 6, pp. 26-9; formerly T 2.3.4, pp. 418-22). In particular, this latter discussion focuses almost exclusively on (what is now widely known as) Hume's *conversion principle*. The substance of this principle does not matter for now (I examine it in chapters 11 and 12). What matters is simply to point out that it is the very principle Hume goes on to apply, in *Of Tragedy*, to the puzzling phenomenon of the pleasure that we take in tragic drama. In a similar way, hope and fear are the main emotions that connect the *Dissertation on the Passions* to the *Natural History of Religion*, since they are at the heart of Hume's account of the origin of religious belief (see chapter 7).

Given these two clear links, there is an obvious explanation for Hume's re-arrangement, if we assume that he was thinking of the *Four Dissertations* as a

¹⁰Immerwahr (1994).

¹¹The link between Hume's treatments of causation and of free-will is surely obvious. It has been insisted on particularly forcefully by Millican (2007a, 2009a, 2011), in opposition to the sceptical realist interpretation of Hume on causation (Millican argues that Hume's defence of compatibilism depends crucially on a "thin" account of necessity).

unified set. For in this set, the *Dissertation on the Passions* is sandwiched in between the *Natural History of Religion* and *Of Tragedy*. There is thus a natural progression of ideas from the first of these dissertations to the second (via hope and fear), and from the second to the third (via the conversion principle). It is difficult to imagine what else could have motivated Hume's rearrangement, if not an awareness of precisely these connections.

This reordering of material from Book 2, part 3 is not the only thing that points to deliberate unity. There are also some very telling facts regarding what Hume *left out* from his earlier work (which Immerwahr does not mention). As well as shifting the penultimate section of Book 2, which included the account of hope and fear, to the start of the *Dissertation on the Passions*, Hume also removed altogether the final section, *Of curiosity, or the love of truth* (T 2.3.10, pp. 448-54). I can find no evidence that Hume came in later life to doubt his account of curiosity, and accordingly assume that he omitted it merely so as to keep things relatively short. But there is a particular reason why this section should have marked itself out as a suitable candidate for dropping, should such a candidate be needed, and it is also to be found in the *Natural History of Religion*. In the second section of this dissertation, Hume writes:

It must necessarily, indeed, be allowed, that, in order to carry men's attention beyond the present course of things, or lead them into any inference concerning invisible intelligent power, they must be actuated by some passion, which prompts their thought and reflection; some motive, which urges their first enquiry. But what passion shall we here have recourse to, for explaining an effect of such mighty consequence? *Not speculative curiosity surely, or the pure love of truth.* That motive is too refined for such gross apprehensions; and would lead men into enquiries concerning the frame of nature, a subject too large and comprehensive for their narrow capacities. No passions, therefore, can be supposed to work upon such barbarians, but the ordinary affections of human life; the anxious concern for happiness, the dread of future misery, the terror of death, the thirst of revenge, the appetite for food and other necessaries. *Agitated by hopes and fears of this*

nature, especially the latter, men scrutinize, with a trembling curiosity, the course of future causes, and examine the various and contrary events of human life. And in this disordered scene, with eyes still more disordered and astonished, they see the first obscure traces of divinity. (N 2.5, pp. 38-9; my emphases)

Here, Hume not only places hope and fear at the centre of his account of the origin of religious belief, but also explicitly *discounts* the love of truth. It is understandable, then, that when he came to rework this material in the *Dissertation on the Passions* he chose to begin with the former passions, and cut out the section on the latter altogether.

A similar point can be made with regard to the connection between the second and third dissertations. Section 6 of the *Dissertation on the Passions*, containing Hume's mature discussion of the causes of the violent emotions, derives from sections 4-8 of *Treatise* Book 2, part 3. From this original, Hume reproduces almost all of section 4, which contains his extended treatment of the conversion principle. The single paragraph reproduced from section 5, meanwhile, is the sole paragraph from that section that concerns this same principle (T 2.3.5.2, pp. 422-3). As for sections 6-8, these provide only the last six of the *Dissertation* section's eighteen paragraphs (P 6.13-18, pp. 28-9), five of which are just one or two sentences long, among the shortest in the whole work (and the sixth isn't much longer).¹²

As with his account of curiosity, I can find no evidence that Hume was dissatisfied with the abandoned material from sections 5-8, and accordingly I presume that these cuts were made largely in the interests of space. Assuming that brevity was a general consideration, however, there is still the particular question of why he chose *this* material to cut. At least a part of the answer to this question now seems obvious: the reason for focusing on the conversion

¹²The *Dissertation on the Passions* ends with a concluding paragraph, which Beauchamp counts as number 19 of the final section in his critical edition (2007); for ease of cross-reference, Millican and I have done the same with our on-line edition. But it is clear that this is not intended as the final paragraph of section 6, but as a conclusion to the whole dissertation, for which reason I count only 18 paragraphs in this section.

principle, at the expense of all the other material, is surely the importance of this principle in Hume's account of tragic pleasure given in the following dissertation. I suggest, however, that there may also have been a second consideration in play. Sections 7 and 8 of Book 2, part 3, are the two sections concerning the influence of space and time on the passions. In line with my speculation in the previous section, it seems at least possible that this material was cut from the *Dissertation on the Passions* because it appeared instead in the original fourth dissertation on the metaphysical principles of geometry.

However plausible this last speculation may be, the known facts are these: Hume made significant cuts from *Treatise* Book 2, part 3, preserving in the *Dissertation on the Passions* only the material that was relevant to the *Natural History of Religion* and *Of Tragedy*, while at the same time moving this material so that it appeared immediately following and preceding these other two dissertations respectively. When we see the *Four Dissertations* as a deliberately unified set, with the passions as the unifying theme, all of this makes perfect sense. If we do not, there is no other obvious motivation for these edits.

0.4. Epicureans, Stoics, Platonists, ...

I move now from my Unity Thesis to my Difference Thesis. While the former is relatively self-explanatory, the latter naturally prompts the question, *what differences?* There are several, in my view, and rather than merely list them all, I will attempt instead to sum up the thrust of the most important ones. In order to do this, meanwhile, it is necessary to introduce my preferred framework for viewing the early modern debate in which Hume was participating. It is well in any case to have an excuse to introduce this framework, since it will inform much of the discussion to follow.

The philosophers of the seventeenth and eighteenth centuries can be grouped and labelled in all sorts of ways, for all sorts of different purposes. For understanding Hume, however, at least as regards the passions and morals, one framework stands out as particularly helpful, since it is one that he himself adopts. In *The*

Epicurean, the first of his four essays on happiness, Hume announces his intention to “deliver the sentiments of sects, that naturally form themselves in the world, and entertain different ideas of human life and of happiness”, giving to each “the name of the philosophical sect, to which it bears the greatest affinity” (Ep n1, p. 138). The ancient schools chosen are Epicureanism, Stoicism, Platonism, and Scepticism. While nothing like these schools and their card-carrying representatives still existed in the eighteenth century, of course, Hume and his contemporaries were familiar with (and often made reference to) the classical background to their own views and debates, and there is a legitimate sense in which we can think of these four traditions as still very much alive at that time.¹³

In chapter 1, when introducing the British debate in the hundred or so years leading up to Hume, I will divide the key players into *egoists* (Hobbes, Locke, Mandeville), *sentimentalists* (Shaftesbury, Butler, Hutcheson), and *rationalists* (Clarke, Wollaston, Burnet). In chapter 10, extending my gaze overseas, I will add Malebranche to the list of rationalists; and we might reasonably include Descartes in this camp as well. These three groups, I submit, correspond to the Epicureans, Stoics, and Platonists respectively, as presented by Hume in his four essays on happiness.

The classical Epicureans were hedonists, who held that pleasure was our chief or sole good. In the early modern period, this manifested itself in the popular doctrine of psychological egoism, the view that all motivation is reducible to self-love, the desire for personal happiness. Typically this took a hedonist form, with happiness being understood as pleasure and the absence of pain. In matters of religion, meanwhile, the Epicureans were frequently branded as atheists, because of their materialism and their denial of providence and intelligent design. Hobbes was (at least perceived to be) an Epicurean in all of these senses. Locke defies such easy classification, since his metaphysical and moral Epicureanism was combined with a more orthodox (and Platonic) attitude to religion.

¹³For a recent study of Epicureanism in the early modern period, see Wilson (2008); for Stoicism, see Brooke (2012); for Platonism, see Hedley and Hutton (2008); and for Scepticism, see Popkin (2003).

The classical Stoics, by contrast, held that virtue was our main or only good. In the seventeenth and eighteenth centuries, this became the doctrine of sentimentalism or psychological altruism, defended vigorously in opposition to Hobbes and the other Epicureans. We have genuinely benevolent sentiments, according to this view, and happiness and virtue coincide in the exercising of them. The Stoics, both ancient and modern, also emphasised the beauty and order of the universe, and in religion were keen proponents of the argument for intelligent design. Given their emphasis on order, however, they were generally hostile to the idea of miracles and *particular* providence, and at the more freethinking end of the spectrum (i.e. Shaftesbury) were thus attracted to deism.

At some level of abstraction, the Stoics and the Platonists of the early modern period belong together, united by the common cause of refuting what they took to be the irreligious and immoral principles of Epicurus and his followers. With Hume, however, if not before, the difference between these two traditions becomes increasingly important. For while Hume had a great respect for Stoic thought even when he disagreed with it (and he agreed with several aspects of it too), the Platonists stood for everything that he hated. He saw the Stoics as moderate empiricists, with an interest in the open-minded study of human nature and the natural world, where the Platonists were dogmatic rationalists and religious mystics.

Where the Epicureans placed pleasure at the heart of their moral and motivational philosophy, and the Stoics placed virtue, the early modern Platonists had instead an Augustinian God (corresponding to Plato's form of the good). God was for them our sole good and the author of all our happiness. This tradition also had at its heart Descartes' real distinction between mind and body, and a belief in the absolute superiority—both metaphysical and moral—of the former over the latter. With Malebranche, as we will see in chapter 10, Cartesian philosophy received a further injection of Augustine's neo-Platonism in the form of the doctrine of Original Sin, and an increased aversion to the body and its lustful passions, which turn us away from God and our one true good.

The Platonists were also rationalists, with a firm conviction in the powers of *a priori* reason both to establish the existence of God, and to regulate our base and bodily passions in the service of our creator. Their favourite proofs for the existence of God were ontological or cosmological, whereas the Stoics, as already noted, preferred the *a posteriori* design argument based on the observable order and beauty of the universe. Accordingly, the Platonists were not averse to the idea of miracles and of a particular providence interrupting the natural order of events; indeed, they often saw these as proofs of divine authority that paved the way to revelation. In terms of the distinction that Hume sets up in his essay *Of Superstition and Enthusiasm*, Platonism tends towards superstition, while Stoicism tends towards enthusiasm. As I will argue in chapter 7 (chiefly §7.2), Hume was openly hostile to Platonic superstition, but considerably more amenable to Stoic enthusiasm.

Hume's opposition to the Platonist tradition was with him from the start, and is a consistent and unambiguous feature of his earlier and later philosophy. What is less clear, however, is whether he leans more towards Epicureanism or more towards Stoicism in his positive views. Norman Kemp Smith famously championed the Stoical interpretation, insisting particularly strongly on the influence of Hutcheson. James Moore vigorously opposed this reading, placing Hume instead in the Epicurean tradition and naming Hobbes as a much more important influence.¹⁴ I believe that there is an element of truth—and of falsehood—on both sides of this argument, as I will now explain.

0.5. ... and Sceptics

Central to my understanding of Hume is the identification of him first and foremost with neither the Epicureans nor the Stoics, but the *Sceptics*. Hume himself explicitly associates with this group, for example when he describes the philosophy of the *Treatise* as “very sceptical” (A 27, p. 657), when he endorses a

¹⁴Kemp Smith (1941/2005), Moore (1995).

“mitigated scepticism” in the conclusion of the first *Enquiry* (E 12.25-6, pp. 162-3), or when he famously offers a “sceptical solution” to his “sceptical doubts” about induction (titles to E 4 and E 5, pp. 25, 40). He also implicitly sides with the Sceptics in the aforementioned essays on happiness, by giving them the last and (by far) the longest word, and having them endorse some recognisably Humean ideas; perhaps most notably the sceptical principle that I will discuss in chapter 13, that “there is nothing, in itself, valuable or despicable, desirable or hateful, beautiful or deformed; but that these attributes arise from the particular constitution and fabric of human sentiment and affection” (Sc 8, p. 162).

Further evidence of Hume’s Sceptical credentials comes from his fondness for the Roman Sceptic, Cicero.¹⁵ The four essays on happiness themselves, indeed, must have been written with Cicero’s *De Finibus Bonorum et Malorum* in mind. In this latter work, an Epicurean, a Stoic, and a Peripatetic (i.e. an Aristotelian) converse and argue about happiness or the ultimate goal of life, while Cicero narrates and takes on the role of the Sceptic, criticising all three. There are certainly differences here: Hume wrote four separate monologues, rather than a dialogue; Cicero’s Peripatetic is replaced by a Platonist in Hume’s set; and Hume has his Sceptic present substantial positive views—albeit in a suitably undogmatic spirit—where Cicero, on behalf of that school, plays a purely critical role in his discussion. But the structural parallels are nevertheless clear and surely deliberate.¹⁶

Hume’s *Dialogues concerning Natural Religion* also have a Ciceronian model, namely *De Natura Deorum*, a dialogue in which an Epicurean, a Stoic, and a Sceptic debate the nature of the gods. Here the Sceptical side of the argument

¹⁵In his autobiographical *My Own Life*, Hume admits that it was Cicero he was secretly reading when his family thought he was studying to become a lawyer (MOL 3, p. xxxiii). For more on Hume’s Ciceronian influences, see Jones (1982).

¹⁶The replacement of Cicero’s Peripatetic with a Platonist in Hume’s set calls for some explanation, but it is not hard to find one. Descartes and the other instigators of early modern philosophy were motivated in a large part by a rejection of Aristotle and the Aristotelianism that had dominated the preceding centuries. By the time Hume came on the scene, therefore, the Peripatetic philosophy was essentially dead (certainly as far as Hume was concerned), while the Platonic tradition—thanks to Descartes and his followers—was still very much alive.

is held up by Cotta, with Cicero himself represented as an impartial observer; though Velleius (the Epicurean) remarks early on that Cotta and Cicero are both disciples of Philo, and have learnt from him to be sure of nothing. Philo was a Sceptic, and the last head of Plato's Academy before its destruction in the first century BC.¹⁷ He is of course the namesake of the Sceptic in Hume's *Dialogues*, and while it is controversial whether Hume agrees with everything this character says, there is no doubt that this is where his sympathies lie, at least broadly speaking. This is further confirmation of his strong ties to this group.

The other two characters in Hume's *Dialogues*, Cleanthes and Demea, represent the Stoical and Platonist traditions respectively. Cleanthes is named after Zeno's successor as head of the Stoic school in Athens. According to Mossner's plausible hypothesis, the modern Stoic, Butler, was the model for Hume's character.¹⁸ There is no obvious real-life candidate for Demea to be named after. Dorothy Coleman suggests etymology as a likelier inspiration: "Demea" comes from the Greek "demos", meaning people, and Hume's character, Demea, represents popular religion at the time.¹⁹ For the modern counterpart, meanwhile, Mossner suggests Clarke, who certainly seems as good a candidate as any;²⁰ though Malebranche strikes me as the better model for the Platonist in the essays on happiness, and it may be noted that Demea also quotes Malebranche in support of his mysticism (D 2.2, pp. 141-2).

In current Hume scholarship, the question of whether or in what sense Hume was a Sceptic—notably with regard to his famous argument about induction—is a hot topic, with several people now favouring *non-sceptical* readings.²¹ As it happens, I favour a more sceptical reading myself,²² but my identification of

¹⁷For the second half of its existence, the Academy had been run predominantly by Sceptics, hence the synonymy of "academical" and "sceptical" philosophy, as in, for example, the title of Hume's first *Enquiry*, section 12.

¹⁸Mossner (1936).

¹⁹Coleman (2007, p. xi).

²⁰Mossner (1936).

²¹Garrett (1997), Noonan (1999), Owen (1999), Beebee (2006).

²²In line with Millican (2007c, 2009b, 2012).

Hume with the Sceptics here is not intended to foreclose that debate. It is perfectly consistent to suppose that Hume was a Sceptic in the present sense—i.e. an inheritor of the Sceptical tradition of Philo and his pupil Cicero—while denying that he was a Sceptic about induction in the now controversial sense. And surely it will not be denied, even by defenders of non-sceptical readings, that Hume’s engagement with this topic is to be situated within the Sceptical tradition, for Hume himself explicitly presents it in this context, calling his account a “sceptical” solution to “sceptical” doubts, as already noted.

Indeed, calling Hume a Sceptic in the present sense entails very little about his positive views, for it is the prerogative of the Sceptics to pick and choose doctrines from any of the other schools, depending on whoever seems to have the better of that particular argument. This is exactly what we find Hume doing. Although he has very little time for the Platonists, as I have said, his positive views, where they have clear antecedents, are drawn from the Epicurean and the Stoical traditions, as well as more directly from the Sceptics. As I noted at the end of the previous section, however, there is some question as to whether he drew more from the Epicureans or more from the Stoics; more from Hobbes, Locke, and Mandeville, or more from Shaftesbury, Butler, and Hutcheson.

The trick to settling this question is to avoid an atemporal answer. For it is also the prerogative of the Sceptics, crucially, to *change their minds*. Kemp Smith’s evidence for his Stoical interpretation drew predominantly on the moral *Enquiry*, and he read (or re-read) the earlier *Treatise* through this lens. Moore’s case for the Epicurean reading, meanwhile, was based exclusively on the *Treatise*. The right reaction to this debate is precisely the obvious one: Hume started life as an Epicurean Sceptic, but after the *Treatise* he became more Stoical (with regard to morals and motivation, that is; with regard to religion his Scepticism was Epicurean throughout his life). This, in a nutshell, is the core of my Difference Thesis.

My thesis is divided into two parts. In Part 1, *The Butler Moment*, I tell the story of what I take to be *the* major change in Hume's later philosophy of emotion, namely a move from Lockean hedonism and egoism about motivation to Butlerian anti-hedonism and anti-egoism. I suggest that the reason for this reversal was the obvious one, namely that Hume encountered and was persuaded by Butler's arguments. (Thus I disagree with Kemp Smith on two counts: not only did Hume's Stoicism develop *after* the *Treatise*, but it derived more from Butler than from Hutcheson. His anti-Platonism, it is true, was with him from the start, and borrowed heavily from Hutcheson; but that is a different matter.)

In Part 2, *The Third Enquiry*, I embark on my critical study of the *Four Dissertations* proper, looking at the major topics of this work in roughly the order in which they appear (skipping over the aspects that came up already in the discussion of Part 1). I devote two chapters to each of the dissertations, and thus these chapters are intended to be read in pairs. As I go, however, I will be drawing particular attention to the links between these four texts, in prosecution of my Unity Thesis, and thus none of these pairs is confined to discussing just one of the dissertations. I will also be pointing to a number of additional ways in which these works differ from the *Treatise*.

The importance of my central claims notwithstanding, I take it that the *Four Dissertations* are of general interest in their own right, even when the views that they present do not differ substantially from those put forward in the *Treatise*, and even when the points being made do not resonate particularly strongly with others in the vicinity. The Unity Thesis and the Difference Thesis are central to this study, and provide its chief motivation; but the study itself does not end with them. My aim, then, is to build up a *coherent* and—as far as possible within the constraints of space—a *complete* interpretation of the *Four Dissertations*, with an emphasis on how Hume's ideas *developed* after the *Treatise*.

Part I

The Butler Moment

Chapter 1

Some Late Philosophers in England

In order to understand Hume’s philosophy of emotion, and in particular how it developed over time, it is necessary to have a sense of British philosophy in the hundred or so years leading up to his work. To this end, the present chapter offers an introduction to some of the key philosophers and debates of this period. In particular, I want to emphasise the importance of the egoism/anti-egoism debate at this time. Perhaps because the anti-egoists are now generally considered to have won this debate, while the battle between the rationalists and the anti-rationalists still rages, this latter dispute often overshadows the former nowadays, even in historical discussions. This leads to a serious distortion of Hume and his interests, which I hope—in this chapter and the next—to go some way towards correcting.

This chapter is also intended to set the scene for the central claim of Part 1 of this thesis: that, in between writing the *Treatise* and the *Four Dissertations*, Hume changed sides on precisely this egoism/anti-egoism controversy. When he wrote the *Treatise*, he was a hedonist and an egoist. Shortly afterwards, however, he read and was persuaded by Butler’s anti-egoist arguments, and subsequently became one of the keenest and clearest *opponents* of his own earlier view.

1.1. Hutcheson against Clarke

Our story of Hume’s influences begins, because I will be telling it largely in reverse, with Hutcheson’s second work. First published in 1728, this was a pair of two

treatises, *An Essay on the Nature and Conduct of the Passions* and *Illustrations on the Moral Sense*. In the preface to these two treatises, Hutcheson tells us that the *Illustrations on the Moral Sense* “had never seen the Light, had not some worthy Gentlemen mistaken some things about the moral Sense alleg’d to be in Mankind” (1728, p. 7). The gentlemen in question were Gilbert Burnet, Samuel Clarke, and William Wollaston, and the view that they brought to Hutcheson’s attention was moral rationalism.

Hutcheson took it that a moral theory must explain two distinct phenomena: moral *motivation* and moral *approbation*. On the one hand, we are moved to perform kind and generous actions; on the other hand, we approve of these motives, both in ourselves and in others. These two facts give rise to a distinction between *exciting* and *justifying* reasons:

When we ask the Reason of an Action we sometimes mean, “*What Truth shews a Quality in the Action, exciting the Agent to do it?*” Thus, why does a *Luxurious* Man pursue *Wealth*? The Reason is given by this Truth, “Wealth is useful to purchase Pleasures.” Sometimes for a Reason of Actions we shew the *Truth expressing a Quality, engaging our Approbation*. Thus the *Reason* of hazarding Life in just War, is, that “it tends to preserve our honest Countrymen, or evidences publick Spirit:” The *Reason* for *Temperance*, and against *Luxury* is given thus, “Luxury evidences a selfish base Temper.” The former sort of Reasons we will call *exciting*, and the latter *justifying*. (1728, p. 138)

Hutcheson’s view, simply put, is that the source of moral motivation is a selfless or disinterested benevolence, and that the source of moral approbation and disapprobation is what he calls a *moral sense*, a disposition to be immediately pleased by benevolent motives, and displeased by the contrary.

Hutcheson first presented this view in his *Inquiry into the Original of Our Ideas of Beauty and Virtue* (1725). In a series of letters to the *London Journal*, Burnet then criticised Hutcheson for failing to provide “*a sufficient Foundation*” for morality: “*tho’ the Conclusions were generally True and Right in themselves,*”

Burnet complained, “*and were capable of Demonstrative Proof, yet he seemed to me to have left them unsupported*” (Burnet & Hutcheson 1735, p. iii).¹ More specifically, Burnet raised a sceptical worry with regard to Hutcheson’s moral sense as the source of approbation:

I saw indeed, there was some such thing in humane Nature. But... I could not be sure, it was not a *deceitful* and *wrong Sense*. The *Pleasure* arising from the Perceptions it afforded, did not seem sufficient to convince me that it was *right*... I wanted therefore some further *Test*, some more certain Rule, whereby I could judge whether my *Sense*, my *moral Sense* as the Author calls it, my *Taste* of Things, was *right*, and agreeable to the *Truth* of Things, or not. (Burnet & Hutcheson 1735, pp. 9-10)

In order to answer this worry, Burnet appealed to the moral rationalism of Clarke and Wollaston,² according to which the things approved of by Hutcheson’s moral sense are indeed *reasonable* or *agreeable to truth*.

I will say more about moral rationalism, and about Hutcheson’s and Hume’s opposition to it, in chapter 10. It suffices for now to have made a note of the debate and some of its participants. For the remainder of this chapter, however, and indeed for the first part of this thesis, it is another view and another eighteenth-century debate that is of interest.

1.2. Hutcheson against Hobbes and Mandeville

The argument between Hutcheson and the moral rationalists has had a profound influence on the subsequent development of moral philosophy (largely through Books 2 and 3 of Hume’s *Treatise*, and then Kant’s work on the rationalist side). At the time, however, this was by no means the main debate in this area. Before

¹This book was a reprint of the correspondence which appeared in the *London Journal* in 1725, to which a preface and postscript by Burnet were added; see Peach (1970) for confirmation of the date.

²Clarke (1706), Wollaston (1722). Burnet’s appeal is in Burnet & Hutcheson (1735, p. iv); he also cites in the same context Cumberland (1672).

Hutcheson turned to attack Clarke and Wollaston, having been prompted by Burnet's objections, the main target of sentimentalism was not rationalism, but *egoism*: the "selfish hypothesis" (as Hume would later call it; M App2.6, p. 298) that all motivation is reducible to self-love.

That egoism was Hutcheson's main target is clear even in the *Illustrations* themselves, which home in on the rationalist challenge only by first setting up Hutcheson's positive story as the default alternative to the selfish systems of Epicurus and Thomas Hobbes:

There are two Opinions on this Subject entirely opposite: The one that of the old *Epicureans*, as it is beautifully explained in the first Book of *Cicero, De finibus*; which is revived by Mr. *Hobbes*, and followed by many better Writers: "That all the Desires of the *human Mind*, nay of all *thinking Natures*, are reducible to *Self-Love*, or *Desire of private Happiness*: That from this Desire all Actions of any Agent do flow." ...

The other Opinion is this, "That we have not only *Self-Love*, but *benevolent Affections* also toward others, in various Degrees, making us desire their Happiness as an *ultimate End*, without any view to private Happiness: That we have a *moral Sense* or Determination of our Mind, to *approve every kind Affection* either in our selves or others, and all publickly useful Actions which we imagined do flow from such Affection, without our having a view to our *private Happiness*, in our Approbation of these Actions."

These two Opinions seem both intelligible, each consistent with itself. The former seems not to represent human Nature as it is; the other seems to do it. (1728, pp. 134, 136)

Only then does Hutcheson introduce moral rationalism as a threat to his anti-egoist sentimentalism, and undertake to refute it. By his own admission in the preface (quoted on page 22 above), he might never have done so, had Burnet not issued the challenge.

While it is clear even in the *Illustrations* that Hutcheson intends his sentimentalism to be first and foremost a rival to *egoism* rather than rationalism, it is yet clearer in Hutcheson's first work, the *Inquiry into the Original of Our Ideas of*

Beauty and Virtue, in which, according to the title page of its first edition, “the principles of the late Earl of Shaftesbury are Explain’d and Defended, against the Author of the *Fable of Bees*” (1725, p. 199). Interestingly, although Shaftesbury is generally credited with having founded the sentimentalist tradition, we will see below that he was no anti-rationalist. Far from it, he thought that reason was required to justify and correct our sentiments. What Shaftesbury *was* opposed to, however, was egoism. I will return to Shaftesbury later (see §1.4). First, it is time to meet the aforementioned author of the *Fable of the Bees*, Bernard Mandeville.

In 1705, Mandeville published, anonymously and without a preface, a short allegory in verse called *The Grumbling Hive, or Knaves Turned Honest*. The story begins by describing a large hive of industrious and prosperous bees, individually selfish, vain, and deceitful, but on the whole thriving. Greed fuelled the industry and trade of this hive, so that in time luxury rose “To such a Height, the very Poor / Liv’d better than the Rich before” (1705, p. 11). But some particularly immoral bees then started to preach against all the dishonesty, and pray for virtue. Jove, angered by their hypocrisy, decided that the best punishment would be to grant them their wish. It was no sooner done than the hive collapsed: with everyone now a model of virtue, there was no more work for lawyers or prison guards; every bee settled its bar tab and resolved never to drink again, so the public houses closed; all being content with plain clothes, tailors were forced to shut up shop; and so on, the moral of the story being that vice is beneficial or even necessary for a happy and prosperous state.

In 1714, Mandeville published the tale again, together with a short *Enquiry into the Origin of Moral Virtue*, and a much lengthier set of remarks on the poem, all under the title of *The Fable of the Bees: Or, Private Vices, Publick Benefits*. The motivational theory espoused in this *Enquiry* is unambiguously egoist. Mandeville begins, in the very first sentence, by stating that “[a]ll untaught Animals are only solicitous of pleasing themselves, and naturally follow the bent of their own Inclinations, without considering the good or harm that

from their being pleased will accrue to others” (1714, p. 27). He goes on to describe man as an “extraordinary selfish and headstrong” animal (p. 28), and to claim that the origin of virtue is essentially flattery. No one can be moved to give up their own good for the good of another, he asserts, without some recompense for themselves, but conjectures that cunning politicians realised that praising virtuous behaviour would satisfy our natural vanity; thus “the Moral Virtues are the Political Offspring which Flattery begot upon Pride” (p. 37).

There is no mention of Shaftesbury in the first edition of the *Fable*, but in the second edition (1723) Mandeville appended an essay called *A Search into the Nature of Society*, which is a sustained attack on Shaftesbury’s philosophy. “The attentive Reader, who perused the foregoing part of this Book,” Mandeville there tells us, “will soon perceive that two Systems cannot be more opposite than his Lordship’s and mine. His Notions I confess are generous and refin’d: They are a high Compliment to Human-kind, and capable by the help of a little Enthusiasm of Inspiring us with the most Noble Sentiments concerning the Dignity of our exalted Nature: What Pity it is that they are not true” (p. 372). This, then, was the challenge that prompted Hutcheson to his defence of Shaftesbury in the *Inquiry* of 1725, as discussed above.

1.3. Butler against Hobbes and Clarke

A year after the first appearance of Hutcheson’s *Inquiry*, Joseph Butler published his *Fifteen Sermons* (1726), which includes an attack on egoism in a similar spirit to Hutcheson’s. Butler does not mention Mandeville, but like Hutcheson he attributes the view to Epicurus and Hobbes (in the preface added to the second edition; 1729, p. 42). Butler likewise insists that we have genuinely benevolent desires, and that they are the source of our virtuous behaviour. And his appeal to what he calls the principle of *conscience*, at the heart of his moral theory, is by his own (later) admission very nearly just Hutcheson’s moral sense by another name:

[W]e have a Capacity of reflecting upon Actions and Characters, and making them an Object to our Thought: And on doing this, we naturally and

unavoidably approve some Actions, under the peculiar view of their being virtuous and of Good-desert; and disapprove others, as vicious and of Ill-desert... It is manifest great Part of common Language, and of common Behaviour over the World, is formed upon Supposition of such a Moral Faculty; whether called Conscience [the term Butler himself typically prefers], moral Reason, moral Sense, or divine Reason; whether considered as a Sentiment of the Understanding, or as a Perception of the Heart, or, which seems the Truth, as including both. (1736, pp. 309)

What this quotation also illustrates, however, is that Butler’s sentimentalism—unlike Hutcheson’s—was not intended to *oppose* moral rationalism. Butler switches the ordinary use of the terms, speaking of a *sentiment* of the understanding, and a *perception* of the heart, while being indifferent between calling the moral faculty “moral Reason” or a “moral Sense”. More explicitly, in the preface added to the second edition of his sermons, Butler distinguishes “two Ways in which the Subject of Morals may be treated”: “One begins from inquiring into the abstract Relations of things: the other from a Matter of Fact, namely, what the particular Nature of Man is, its several parts, their Oeconomy or Constitution” (1729, p. 37). He says that both methods lead to the same thing, and has nothing bad to say about the former (which is Clarke’s approach); he merely remarks that his own work proceeds mainly in the latter way (*ibid.*). The common enemy of the sentimentalists before Hume was not Clarke, therefore, but Hobbes.

Butler contributed one point to this anti-egoist debate that is particularly important here, for it was the appreciation of this point—as I will argue in the next chapter—that was the pivotal moment in the development of Hume’s thought between the *Treatise* and his later work. Butler distinguishes self-love from every other desire, on the grounds that “[t]he Object the former pursues is somewhat internal, our own Happiness, Enjoyment, Satisfaction”, while the latter—which he calls “particular” appetites or passions—all pursue “this or that particular external Thing” *for its own sake*, independently of its contribution to our happiness (1726, p. 111). His argument for this crucial distinguishing feature of the particular passions is simple but effective:

That all particular Appetites and Passions are towards external *Things themselves*, distinct from the *Pleasure arising from them*, is manifest from hence; that there could not be this Pleasure, were it not for that prior suitability between the Object and the Passion: There could be no Enjoyment or Delight from one thing more than another, from eating Food more than from swallowing a Stone, if there were not an Affection or Appetite to one thing more than another. (1726, p. 111)

Egoism—particularly in its hedonist form, which reduces all desire to the general appetite for pleasure and aversion to pain—is an attractively simple view of motivation. At a time when it was very popular, and perhaps even the mainstream view, Butler was the first to point out that it gets things fundamentally *back to front*: most things are not desired because they are pleasant; rather, they are pleasant because we desire them.³

Butler’s “particular passions”—desires for objects themselves, independently of the pleasure that they bring—will feature heavily in the discussion of Hume to follow, and it will be convenient to have a name for them. Rather than adopt Butler’s terminology, which is not very descriptive, I will refer to them instead as *object-directed desires*. “Object” is used here, of course, in the broadest possible sense, as anything whatsoever that may be desired; the point of the terminology is that it is the *object* we desire, and *for its own sake*, rather than for the happiness or pleasure that it brings.

Though Butler did not object to moral rationalism, it is worth remarking that he and Hutcheson were opposed to many of the same people—and most

³It would be overstating the case to say that Butler decisively refuted psychological egoism with this point, though I take it that his argument is decisive as far as it goes. The dialectical situation is perhaps best summed up thus: There is a wealth of empirical evidence that, on the face of it at least, suggests we are not purely selfish. Egoists have a standard way of reinterpreting this evidence in accordance with their hypothesis (crudely, we help other people because of the warm fuzzy feeling that it gives us). What Butler’s point conclusively shows is that this reinterpretation cannot be right: it gets the relationship between pleasure and desire the wrong way round (we wouldn’t get that warm fuzzy feeling unless we cared about other people in the first place). Perhaps the egoists can come up with another reinterpretation of *prima facie* unselfish motivation, in line with their cynical hypothesis. As things stand, however, Butler’s argument leaves the anti-egoist in an overwhelmingly stronger position with regard to the evidence. For further discussion, see Blackburn (1998, ch. 5), Sober (2000).

notably Clarke—on the epistemology of religion. Clarke was the most prominent defender of the *a priori* cosmological argument for the existence of God as the first cause, and Butler was the most prominent defender of the *a posteriori* design argument. The two came to blows about the success of Clarke’s argument (very amicably and politely, I might add) in a private exchange that was then appended to the fourth edition of Clarke’s Boyle lectures (1716).

Notwithstanding their opposition to Clarke’s brand of rationalist theism, however, Butler and Hutcheson were both devout Christians. For them, furthermore, our benevolent nature—which they so insisted on against the egoists—was not only the foundation of morality, but also evidence for the goodness of God, the author of that nature. Thus Hutcheson writes:

The *present Constitution* of our *moral Sense* determines us to approve all *kind Affections*: This Constitution the Deity must have foreseen as *tending* to the *Happiness* of his Creatures; it does therefore evidence *kind Affection* or *Benevolence* in the Deity[.] (1728, p. 153)

Butler, on a related note, maintains:

That God has given us a moral Nature, may most justly be urged as a Proof of our being under his moral Government... For, our being so constituted, as that Virtue and Vice are thus naturally favoured and discountenanced, rewarded and punished respectively as such, is an intuitive Proof of the Intent of Nature, that it should be so; otherwise the Constitution of our Mind, from which it thus immediately and directly proceeds, would be absurd. (1736, pp. 179, 180)

Butler’s argument is embedded in a much larger argument or battery of arguments, first to the effect that God is a governor, rewarding and punishing us for our actions, and then to the effect that he is a *moral* governor, rewarding and punishing specifically according to desert. The arguments for this second conclusion point out the various ways in which nature is set up so as to benefit the virtuous and harm the vicious, all of which suggest that this was the intention of

its author. One of the various ways in which this is so is our own moral nature, and our power over each other: we ourselves are naturally inclined to reward virtue and punish vice.

Butler's argument from our moral nature to theism is thus more sophisticated and less direct than Hutcheson's, and Butler has other points to appeal to in support of God's goodness should this particular one fail. Butler is after all the more committed philosopher of religion. But for present purposes what matters is the basic point of agreement: that our natural benevolence is evidence of the moral goodness of the author of our nature.

1.4. Shaftesbury against Hobbes and Locke

Butler and Hutcheson were both, in their slightly different but closely related ways, the intellectual descendants of Anthony Ashley Cooper, the third Earl of Shaftesbury. It was from Shaftesbury that they inherited their interest in the passions and affections of human nature, and their belief in the importance of these things for morality. One can also find in Shaftesbury direct precursors both to Hutcheson's moral sense (e.g. 1711, p. 177) and to Butler's conscience (e.g. 1711, pp. 208-9); though Butler, it may be noted, explicitly criticised Shaftesbury for failing to emphasise the *authority* of conscience over the other principles of the human frame or constitution (Butler 1729, p. 40). And it was Shaftesbury, of course, whom Hutcheson explicitly set out to defend against Mandeville's criticisms.

Where Hutcheson was opposed to rationalism, and Butler merely set it to one side, it should be noted that Shaftesbury on the contrary saw reason as necessary to correct and justify our moral sentiments, as for example in the following argument against partiality:

But lest any shou'd imagine with themselves that an inferior Degree of natural Affection, or an imperfect partial Regard of this sort, can supply the place of an intire, sincere, and truly moral one... we may consider first, That PARTIAL AFFECTION, or social Love *in part*, without regard to a

compleat Society or *Whole*, is in it-self an Inconsistency, and implies an absolute Contradiction... The Person, therefore, who is conscious of this Affection, can be conscious of no *Merit* or *Worth* on the account of it. It has no Foundation or Establishment in *Reason*[.] (1711, p. 205)

Here Shaftesbury clearly rejects a certain sentiment on the grounds that it is inconsistent, and has no foundation in reason. Once again, I emphasise that sentimentalism was intended first and foremost as an alternative to *egoism*, not rationalism; and Shaftesbury's principal targets were Thomas Hobbes and John Locke.

Few things are uncontroversial in philosophy, and even the traditional egoist interpretation of Hobbes has been called into question.⁴ This is not the place, however, to enter into any detailed examination of Hobbes's views; it suffices to report that he was thought to be an egoist at the time, being named as such by Shaftesbury, Hutcheson, and Butler. Whatever may be said on the other side, it is easy to see where these sentimentalists got their interpretation from. In the first instance, Hobbes seems to have espoused a hedonist account of motivation:

This motion, in which consisteth *pleasure* or *pain*, is also a *solicitation* or provocation either to draw *near* to the thing that pleaseth, or to *retire* from the thing that displeaseth; and this solicitation is the *endeavour* or internal beginning of *animal* motion, which when the object *delighteth*, is called *appetite*; and when it *displeaseth*, it is called *aversion*... So that *pleasure*, *love*, and *appetite*, which is also called desire, are *divers names* for divers considerations of the *same thing*. (1650a, pp. 31-2)

And in the second place, his definitions of pity and charity (i.e. benevolence) have a strong appearance of egoism:

Pity is *imagination* or *fiction* of *future* calamity to our *selves*, proceeding from the sense of *another* man's calamity. (1650a, p. 44)

⁴By Gert (1967, 2006).

There can be no greater argument to a man, of his own power, than to find himself able not only to accomplish his own desires, but also to *assist* other men in theirs: and this is that conception wherein consisteth *charity*. In which, first, is contained the *natural affection* of parents to their children, which the Greeks call *Στοργη*, as *also*, that affection wherewith men seek to *assist* those that adhere unto them. But the affection wherewith men many times bestow their benefits on *strangers*, is not to be called charity, but either *contract*, whereby they seek to purchase friendship; or *fear*, which maketh them to purchase peace. (1650a, p. 49)

In this last passage, Hobbes seems to reduce charity—which includes kindness to our friends and family—to the love of power or the exercise of it,⁵ while denying that there is even this semblance of benevolence in the case of kindness to strangers. We display good will in the latter case purely with a self-interested view to peace and security. This chimes with Hobbes’s attempt to derive morals and political society from this same egoist desire for self-preservation, and for peace as its necessary means (see 1650b, pp. 81-6; 1651, pp. 86-111).

The sentimentalists mentioned Locke less often than Hobbes in this connection, but if anything Locke’s hedonism and egoism is even clearer than Hobbes’s. For Locke, desire is moved by “happiness and that alone” (1690, p. 258), and happiness is understood hedonistically: “*Happiness* then in its full extent is the utmost Pleasure we are capable of, and *Misery* the utmost pain” (ibid.). Indeed, the whole *point* of pleasure, according to Locke, is to excite us to action, and without it our wills would be entirely inert:

[T]o excite us to these Actions of thinking and motion, that we are capable of, [the Author of our being] has been pleased to join to several Thoughts, and several Sensations, a *perception* of *Delight*. If this were wholly separated from all our outward Sensations, and inward Thoughts, we should have no reason to prefer one Thought or Action, to another; Negligence, to Attention; or Motion, to Rest. (1690, p. 129)

⁵This is how Butler, at least, interprets the passage (1726, p. 52, n2).

It is worth noting the striking contrast here with Butler's motivational psychology. While for Locke there could be no desire without pleasure, for Butler there could be no pleasure without desire:

[T]he very idea of an interested Pursuit, necessarily pre-supposes particular [i.e. object-directed] Passions or Appetites; since the very Idea of Interest or Happiness consists in this, that an Appetite or Affection enjoys its Object... Take away these Affections, and you leave Self-love absolutely nothing at all to employ itself about; no End or Object for it to pursue, excepting only that of avoiding Pain. (Butler 1729, p. 20)

This reflects Butler's crucial insight that the egoists' account of motivation gets things fundamentally back to front.⁶

For Locke, furthermore, the distinction between moral good and evil consists in the following or not following of the divine law (1690, p. 352), and he insists that moral motivation derives solely from the prospect of rewards and punishments (if not in this life, then at least in the next), "it being impossible to set any other motive or restraint to the actions of a free understanding agent but the consideration of good and evil; that is, pleasure or pain that will follow from it" (c. 1686-8, p. 301). This Christian egoism may seem fairly obviously crude to us now, but it was not short of defenders in the eighteenth century. John Clarke of Hull espoused it in opposition to both Samuel Clarke's rationalism and Hutcheson's sentimentalism, and later Robert Clayton did the same in criticism of Hume's moral *Enquiry*.⁷

In failing to mention Locke alongside Hobbes, Hutcheson and Butler may simply have been following Shaftesbury (not to mention that Hobbes was an

⁶Butler seems to have gone too far in this quotation: many physical pleasures (e.g. those of taste) are surely not dependent on any antecedent desire or appetite. The pleasure of relieving hunger is one thing; the pleasure of a delicious fruit is quite another. Hume himself was more cautious in this regard: "Were there no appetite of any kind antecedent to self-love, that propensity could *scarcely* ever exert itself; because we should, in that case, have felt *few and slender* pains or pleasures" (M App2.12, pp. 301-2; my emphases). See also §6.2.

⁷Clarke (1726), Clayton (1753). John Clarke was master of the grammar school in Hull. So as to distinguish him clearly from Samuel Clarke, and from Samuel's nephew (also called John), Hutcheson referred to him as "*Clarke of Hull*" (1728, p. 6).

easy target, being widely thought of as the foremost enemy of religion). For Shaftesbury, meanwhile, Locke had been a close family friend and tutor, which may explain his reticence in this regard.⁸ But from a letter written to a student whom he had taken under his wing, we learn Shaftesbury's true thoughts:

In general truly it has happened, that all those they call *Free-Writers* now-a-days, have espoused those Principles, which Mr. HOBBS set a foot in this last Age. Mr. LOCKE, as much as I honour him on account of other Writings (*viz.* on Government, Policy, Trade, Coin, Education, Toleration, &c.) and as well as I knew him, and can answer for his Sincerity as a most zealous *Christian* and Believer, did however go in the self same Track... 'Twas Mr. LOCKE, that struck the home Blow: For Mr. HOBBS's Character and base slavish Principles in Government took off the Poyson of his Philosophy. 'Twas Mr. LOCKE that struck at all Fundamentals, threw all *Order* and *Virtue* out of the World, and made the very *Ideas* of these (which are the same as those of GOD) *unnatural*, and without Foundation in our Minds. (1716, pp. 38-9)

As I have hedged regarding Hobbes, so I may hedge regarding Locke.⁹ It does not matter for present purposes whether Shaftesbury was right in his interpretation of either of these writers; it suffices to observe that this egoist interpretation was common at the time, and that Hume himself read both Hobbes and Locke in this way, these being his two explicit examples, alongside Epicurus, of defenders of the "selfish system of morals" (M App2.3, p. 296).

⁸See Gill (2006, pp. 77-82).

⁹The question with Locke is not whether he was an egoist (no one that I know of has doubted this); the interpretative difficulty in his case arises from the fact that he seems to have been a rationalist as well: "I am bold to think, that *Morality is capable of Demonstration*, as well as Mathematicks: Since the precise real Essence of the Things moral Words stand for, may be perfectly known; and so the Congruity, or Incongruity of the Things themselves, be certainly discovered, in which consists perfect Knowledge" (Locke 1690, p. 516). One (partial) resolution of this tension is to say that Locke started out as a rationalist, and became more of an egoist in his later work (while still retaining, inconsistently, traces of his earlier rationalism); see Von Leyden (1954/2002), Aaron (1971). Another is to say that rationalism is Locke's account of moral *obligation*, while egoism is his account of moral *motivation*; see Colman (1983), Darwall (1995). For a recent discussion, and defence of a third view according to which the two strands are more closely connected, see Sheridan (2007).

1.5. Egoists, Sentimentalists, Rationalists

As Hume was growing up, there were thus three main traditions of thought in Britain regarding morals and motivation: the egoists (Hobbes, Locke, Mandeville), the sentimentalists (Shaftesbury, Butler, Hutcheson), and the rationalists (Clarke, Wollaston, Burnet). The latter two groups, however, only formally became opponents with the publication of Hutcheson's *Illustrations* in 1728; and even then Butler remained neutral, preferring the sentimentalist approach but not ruling out the possibility of a rationalist foundation of morals as well. The rationalist/anti-rationalist debate was thus comparatively recent and parochial: the main battle was between egoism and anti-egoism.

It is no secret that Hume disagreed with more or less every aspect of Clarke's philosophy: the *a priori* cosmological argument for the existence of God, the truth of (and the evidence for) Christian revelation, the dependence of morality on religion, physical and moral necessity, moral rationalism. Hume was certainly no rationalist. On the other side, it is also common knowledge that Hume was positively influenced, in at least some particulars, by all of the other major thinkers introduced in this chapter: Hobbes, Locke, Shaftesbury, Mandeville, Hutcheson, and Butler. With the exception of Hobbes, these are none other than the "late philosophers in *England*" that Hume cites with approval as having "begun to put the science of man on a new footing" (T Intro.7, p. xvii).¹⁰ As we have seen, however, these thinkers were divided on the crucial issue of motivation.

¹⁰It is perhaps interesting that Hobbes is not included in Hume's list. Russell has hypothesised that the *Treatise* was modelled on Hobbes's *Human Nature* and *De Corpore Politico* (1650a,b), and conjectured that Hobbes is not on the list precisely because his influence on Hume was so great, and Hume thought it unwise to draw attention to that fact (Russell 2008, pp. 61-6, 66-7). For scepticism regarding this conjecture, see Harris (2009), who suggests instead that "Hobbes is not included on the list because he was, precisely, *not* one of those who put the science of man on a new footing—that is, because his science of man is grounded on an *a priori* commitment to materialism and to mechanistic physics, and not the Baconian inductivism that Newton had had such success with" (p. 40). A third possibility suggests itself to me, albeit a less profound one: that Hobbes didn't make the list simply because his work was no longer all that recent. *Leviathan* was published almost 80 years before the *Treatise* (1651), while Locke's *Essay*—the earliest work from the authors that Hume does name—had appeared only 49 years previously (1690). Hobbes, however important his influence on Hume, was no longer a "late" philosopher.

And so we are brought inevitably to the question of whether Hume was an egoist or an anti-egoist, a hedonist or an anti-hedonist. Given the importance of this issue in the early modern debate, and specifically among the people that influenced Hume most, it is hardly to be expected that he would have taken no stance on the matter. As I have already advertised, this is the locus of the single most important aspect of my Difference Thesis, and of the central claim of these first six chapters: my position is that Hume was a hedonist and an egoist in the *Treatise*, but an anti-hedonist and an anti-egoist in his later work (and that it was Butler who changed his mind). It would be hard to overstate the significance of this change. It is after all a *complete reversal* on one of *the* major debates at the time. Differences do not come any bigger than this, and if it was the only thing that I could point to in support of my Difference Thesis (which it isn't), I would still consider my position sufficiently established.

In the next chapter, I will present the core of my argument for this difference. The story does not end here, however, for a reversal of this magnitude is certain to leave its mark on other aspects of Hume's philosophy of emotion. In chapters 3, 4, and 5, sure enough, I both strengthen and expand my position, by pointing to a number of additional modifications that Hume went on to make in the light of this initial change. In all of these chapters, my approach is simply to lay out the positive evidence for my view, touching on potential arguments to the contrary only insofar as these are directly relevant to the narrative. In chapter 6, I then respond to some alternative interpretations that try to smooth over the differences between the *Treatise* and the later works.

There is a general theme to the discussion to come. I take it that in every case my interpretations of the relevant passages are the most obvious and straightforward ones. Attempts to minimise the appearance of substantial change in Hume's thought, meanwhile, seem to me invariably to involve contrived and ingenious ways of resisting the most natural readings of the text. But there is nothing inherently implausible in the story that I am going to tell. There is no reason to think that Hume wouldn't initially agree with Hobbes and Locke

about motivation (since it is uncontroversial that he agreed with them about other things), and there is no reason to think that he wouldn't subsequently be persuaded by Butler's arguments to the contrary (for these arguments are indeed persuasive). When we find that the texts immediately appear to support this perfectly credible story, therefore, there is no cause to go looking for complications below the surface.

Chapter 2

Motivation

The present chapter is concerned with Hume's theory of motivation, and how it changed after he wrote the *Treatise*. But it is also about motivation in another way, for in addition I will be offering a conjecture as to what prompted Hume's change of mind. As I have advertised, I believe that the view endorsed in the *Treatise* is essentially a Lockean hedonism and egoism. After the *Treatise*, however, Hume became instead a firm adherent of Butler's object-directed motivational psychology. My conjecture, straightforwardly enough, is that Hume changed sides on this debate because he read Butler's *Sermons*, and was persuaded that Butler was right.

2.1. The Butler paragraph

The bulk of the *Treatise of Human Nature* was written between 1734 and 1737, during Hume's stay at La Flèche in France. Butler's *Analogy of Religion* was first published in 1736, and so cannot have had much of an influence on Hume's early thought. We can be fairly certain that Hume read it soon after it was published, however, and before Books 1 and 2 of the *Treatise* came out in January 1739. For in December 1737, he wrote thus to Henry Home (later Lord Kames):

Your thoughts and mine agree with respect to Dr Butler, and I would be glad to be introduced to him. I am at present castrating my work, that is, cutting off its nobler parts; that is, endeavouring it shall give as little

offence as possible, before which, I could not pretend to put it into the Doctor's hands. (HL 1, p. 25)

At about this time, Kames wrote a letter of introduction to Butler on Hume's behalf. Hume had hoped to present the letter to Butler before asking his opinion of the *Treatise*, but alas the meeting never took place, as we learn from a surviving letter from Hume to Kames in March 1738:

I shall not trouble you with any formal compliments or thanks, which would be but an ill return for the kindness you have done me in writing in my behalf, to one you are so little acquainted with as Dr Butler; and I am afraid, stretching the truth in favour of a friend. I have called upon the Doctor, with a design of delivering him your letter, but find he is at present in the country. I am a little anxious to have the Doctor's opinion. My own I dare not trust to; both because it concerns myself, and because it is so variable, that I know not how to fix it. (HL 1, p. 25)

I conjecture that it was also around this time that Hume first read Butler's *Sermons*, being prompted to do so by his good opinion of the *Analogy*. The first edition of the *Sermons* appeared in 1726, with a second in 1729, so it is possible that Hume had come across them earlier. But I can think of little reason why a set of sermons by a young English clergyman would have struck him at this time as a particularly worthwhile read.¹ A third edition appeared in 1736, coinciding with the *Analogy*, and I suspect that this is what first brought them to Hume's attention. At the very least, if he had read them before then, he nevertheless

¹Hume was obviously familiar with Clarke's Boyle lectures, and hence potentially also with Butler's correspondence with Clarke, which was appended to the fourth edition of these lectures (1716). This correspondence was anonymous on Butler's side, however, so even if Hume had read this exchange and been impressed by Clarke's antagonist, he may well not have known that the unnamed "Gentleman from Glocestershire" was also the author of the *Sermons*. A more likely route into Hume's consciousness would be through the approving mention of the *Sermons* themselves in the preface to Hutcheson's *Essay and Illustrations*, first published in 1728; but even this is just a passing remark, noting merely that Butler "has done so much Justice to the wise and good Order of our Nature" (1728, p. 9). Hume's nephew, who inherited his uncle's library, had a copy of the 1736 edition of the *Analogy*, but apparently no edition of the *Sermons* (Norton and Norton 1996, p. 80); and there is no mention of the *Sermons* in M. A. Stewart's essay on Hume's early intellectual development (2005). I take it, therefore, that my conjecture is consistent with the known facts.

failed to appreciate the force and significance of Butler’s anti-egoist arguments. This much is confirmed, as we will shortly see, by the text of the *Treatise*.

If Hume first read the *Sermons* shortly after reading the *Analogy*—just before Books 1 and 2 of the *Treatise* were published, but after the bulk of them had already been written—then we have a plausible explanation for the inclusion of this intriguing paragraph in the penultimate section of Book 2:

Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. These passions, properly speaking, produce good and evil, and proceed not from them, like the other affections. (T 2.3.9.8, p. 439)

This paragraph is a clear statement of the crucial feature of Butler’s “particular” passions—or *object-directed desires*, as I am calling them—familiar from the previous chapter: that they are the *causes* of pleasure and pain, rather than their effect. It is possible, I suppose, that Hume hit on this idea independently, but it seems much more likely that he got it from Butler (not to mention that Hume explicitly attributes the point to Butler in his later work; see §2.4 below). In any case, because I will want to refer to this paragraph often, it will be convenient to dub it *the Butler paragraph*.

The text of the first two volumes of the *Treatise* strongly suggests that the Butler paragraph was a late addition, an afterthought, and something the consequences of which Hume did not yet fully appreciate. For one thing, it is entirely isolated: of all the 397 paragraphs of Book 2 (and 555 of Book 1), this is the only one to make any mention of Butler’s point.² For another, it is at odds with Hume’s official *Treatise* doctrine. For the official doctrine of Book 2, as I will

²The only mention of Butler himself, meanwhile, is in the footnote in the introduction that I quoted in §1.5, which names Butler—alongside Locke, Shaftesbury, Mandeville, and Hutcheson—as one of the “late philosophers in *England*” to have “begun to put the science of man on a new footing” (T Intro.7, p. xvii). Butler’s inclusion in this list could likewise easily have been a late addition.

now argue, is Lockean hedonism, notwithstanding this (probably last-minute) concession to Butler.

2.2. Hedonism in the *Treatise*

At the start of *Treatise* Book 2, Hume divides the passions into two groups, the “direct” and the “indirect” (T 2.1.1.4, pp. 276-7). I will say more about this distinction in the following chapters. For now, my present argument requires just two points about the direct passions, which are easy to establish: (i) that they all arise from pleasure and pain (this is indeed part of their very definition); and (ii) that they include, among others, the passions of desire and aversion. Thus, when he first introduces the class, Hume writes: “By direct passions I understand such as arise immediately from good or evil, from pain or pleasure... And under the direct passions [I comprehend] desire, aversion, grief, joy, hope, fear, despair and security” (T 2.1.1.4, pp. 276-7).

Later on, in the section devoted exclusively to these direct passions (the same section that includes the Butler paragraph), Hume begins by reminding the reader of these important points:

’Tis easy to observe, that the passions, both direct and indirect, are founded on pain and pleasure, and that in order to produce an affection of any kind, ’tis only requisite to present some good or evil. Upon the removal of pain and pleasure there immediately follows a removal of love and hatred, pride and humility, desire and aversion...

The impressions, which arise from good and evil most naturally, and with the least preparation are the direct passions of desire and aversion, grief and joy, hope and fear, along with volition. (T 2.3.9.1-2, p. 438)

Not long after this, and just before the Butler paragraph, Hume is then explicit in spelling out the hedonist implication of these claims:

Desire arises from good consider’d simply, and aversion is deriv’d from evil. The will exerts itself, when either the good or the absence of the evil may be attain’d by any action of the mind or body. (T 2.3.9.7, p. 439)

The hedonism of this passage is inescapable, when we remember that “good” and “evil” are just synonyms for “pleasure” and “pain” respectively (as they were for Locke). This is a fact that the paragraphs already quoted amply confirm. The Butler paragraph, indeed, confirms it most clearly of all: “good and evil, or in other words, pain and pleasure” (T 2.3.9.8, p. 439). And for good measure, here is another clear statement of hedonism from several sections earlier: “’Tis from the prospect of pain or pleasure that the aversion or propensity arises towards any object” (T 2.3.3.3, p. 414).

Thus the things that Hume says about motivation prior to the Butler paragraph leave no room for the object-directed desires that suddenly and unexpectedly intrude into the *Treatise*’s official hedonist framework. To bring out this contradiction more clearly, I shall here repeat the most relevant sentences (already quoted in context above) one after the other, in the order in which they appear in the text:

By direct passions I understand such as arise immediately from good or evil, from pain or pleasure. ... ’Tis from the prospect of pain or pleasure that the aversion or propensity arises towards any object... ’Tis easy to observe, that the passions, both direct and indirect, are founded on pain and pleasure... Upon the removal of pain and pleasure there immediately follows a removal of... desire and aversion... The impressions, which arise from good and evil most naturally, and with the least preparation are the direct passions of desire and aversion... Desire arises from good consider’d simply, and aversion is deriv’d from evil. ... *Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. ... These passions, properly speaking, produce good and evil, and proceed not from them, like the other affections.*

As this clearly shows, the Butler paragraph sits very uneasily alongside the rest of the *Treatise*. In striking contrast, however, Butler’s object-directed desires are—as we will see in the remaining sections of this chapter—at the very *heart* of

Hume's later motivational psychology, which is consistently and explicitly anti-egoist throughout. Needless to say, this is a *substantial difference*, concerning moreover what was one of the most important debates at the time.

Before turning to Hume's later anti-egoism, it may be helpful to place my view on these matters in the context of the existing secondary literature. I am by no means alone in thinking that Hume was a hedonist and an egoist in the *Treatise*, or at least that he almost entirely was, but for the Butler paragraph.³ A particularly clear statement of this view is given by Stephen Darwall:

“Tis from the prospect of pain or pleasure,” Hume writes, “that the aversion or propensity arises towards any object.” Good and evil (“or in other words, pain and pleasure”) “consider’d simply” give rise to desire and aversion respectively. And “the WILL exerts itself, when either the good or the absence of the evil may be attain’d by any action of the mind or body.”

Hume's theory of action thus not only employs the traditional idea that the will invariably aims at the good. It interprets that idea hedonistically and egoistically. Desires and aversions arise from the prospect of pleasure or pain, respectively, for the agent.⁴

Darwall goes on to concede that the Butler paragraph is in tension with this hedonism, and effectively offers Hume a dilemma: he must either reinterpret these object-directed desires in hedonistic terms, or revise his theory of action and the will.

The hedonist interpretation of the *Treatise* is not without opponents.⁵ Though controversial, however, it has had plenty of supporters besides Darwall, and might even be the majority view.⁶ No one, furthermore, has ever doubted that Hume

³The full story is a little more complicated, and I am simplifying for now. In fact the Butler paragraph is not the only hint of anti-egoism in the *Treatise*, although it is the most evident. See chapter 6.

⁴Darwall (1993, p. 423).

⁵Notably Garrett (2007). My aim in this chapter (and the following chapters) is to lay out the positive evidence for my view; I turn to defensive matters, responding in particular to Garrett's arguments, in chapter 6.

⁶Glathe (1950, pp. 67-8, 76), Karlsson (2006, pp. 246-7), Cohon (2008b, pp. 31-5, 2010, p. 200), Magri (2008, p. 190).

was an anti-egoist in the moral *Enquiry*. This is hardly surprising, since, as we will see in §2.4, his anti-egoism in this work could not have been clearer or more explicit. One might have thought, therefore, that the central claim of this first part of my thesis would be an extremely familiar point (if a little controversial, given the mild controversy surrounding the egoist interpretation of the *Treatise*). Such is the general lack of interest in Hume's mature philosophy of emotion, however, and in the development of his thought in this area, that what should have been an obvious difference between the earlier and later Hume—or at least an obvious *apparent* difference—is almost never remarked upon.⁷

A few things are distinctive about my position. The first is at least the *emphasis* on this difference, even if the respective interpretations of Hume's earlier and later philosophy on which it rests are in themselves nothing new. The second is the identification of Butler as the source of Hume's later view, and the cause of his change of mind. And the third, which will emerge over the next few chapters, is the placement of this change in a larger story about the development of Hume's mature philosophy of emotion in general. As we will see, this substantial change of mind had ramifications for several other aspects of Hume's thought in this area, and accordingly prompted a number of further modifications.

2.3. The dignity of human nature

The crucial change in Hume's philosophy of motivation occurred very soon after the publication of the *Treatise*. I have conjectured that Hume in fact read Butler's *Sermons* just *before* his own first work went to print, in time to insert the Butler paragraph. Given that this paragraph contradicts the official hedonism of Books 1 and 2, however, it seems that the import of Butler's crucial point had not yet fully sunk in. However that may be, the penny had at least started to drop as

⁷*Almost* never: for some exceptions, see McGilvary (1903, p. 272), Árdal (1966, p. 70), Magri (2008, p. 193). McGilvary and Árdal, however, argue that Hume was *not* an egoist in the *Treatise* (on which alternative possibility, see chapter 6), and explicitly have no interest in Hume's later view. Magri seems to agree with me about the difference, but mentions it only in passing, being concerned in his article with other things.

early as 1741, as we may discover from an essay that was then published, *Of the Dignity or Meanness of Human Nature*.⁸ In this essay, Hume distinguishes between two “sects... in the learned world” (DM 1, p. 80), one consisting of those “who are inclined to think favourably of mankind”, the other of those who prefer to “give us a mean opinion of our nature” (DM 2, p. 81). Egoists are among the second sort of thinkers, and in arguing against their view Hume appeals—for the very first time in this context—to Butler’s key point about the priority of desire over pleasure:

In my opinion, there are two things which have led astray those philosophers, that have insisted so much on the selfishness of man. In the first place, they found, that every act of virtue or friendship was attended with a secret pleasure; whence they concluded, that friendship and virtue could not be disinterested. But the fallacy of this is obvious. The virtuous sentiment or passion produces the pleasure, and does not arise from it. I feel a pleasure in doing good to my friend, because I love him; but do not love him for the sake of that pleasure. (DM 10, pp. 85-6)

Since this paragraph mentions the first of two points against the egoists, it would be remiss to ignore the second, which is a particular objection to Mandeville. The argument is that vanity (the desire at the heart of Mandeville’s egoist account of moral motivation) cannot prompt us to virtuous action unless we already have a love of virtue for its own sake:

[V]anity is so closely allied to virtue, and to love the fame of laudable actions approaches so near the love of laudable actions for their own sake, that these passions are more capable of mixture, than any other kinds of affection; and it is almost impossible to have the latter without some degree of the former. Accordingly, we find, that this passion for glory is always warped and varied according to the particular taste or disposition of the mind on which it falls. NERO had the same vanity in driving a chariot,

⁸At the time it was titled *Of the Dignity of Human Nature*; Hume changed the name in the 1770 edition.

that TRAJAN had in governing the empire with justice and ability. To love the glory of virtuous deeds is a sure proof of the love of virtue. (DM 11, p. 86)

Of the Dignity or Meanness of Human Nature is a short essay intended for lighter reading than Hume's more substantial philosophical texts. In the first three editions of this essay (1741, 1742, and 1748), however, Hume mentions his desire to pursue these matters in greater depth:

I may, perhaps, treat more fully of this Subject in some future Essay. In the mean Time, I shall observe, what has been prov'd beyond Question by several great Moralists of the present Age, that the social Passions are by far the most powerful of any, and that even all the other Passions receive from them their chief Force and Influence. Whoever desires to see this Question treated at large, with the greatest Force of Argument and Eloquence, may consult my Lord *Shaftsbury's* Enquiry concerning Virtue. (1741, p. 620)

In the fourth edition (1753), this promissory note is deleted. The key event between 1748 and 1753 that prompted the deletion, of course, was the publication (in 1751) of the first edition of Hume's *Enquiry concerning the Principles of Morals*, the very work that he had promised in the earlier essay. While the *Treatise* had espoused a hedonistic egoism, in line with Hobbes, Locke, and Mandeville, Hume changed his mind very soon afterwards. In the 1741 set of *Essays, Moral and Political*, he was instead agreeing with Butler and Shaftesbury, and explicitly anticipating the extended arguments of the 1751 *Enquiry* to this effect.

2.4. The two *Enquiries*

We will turn to the moral *Enquiry* presently. Before this work appeared, however, there were, in 1748 and 1750, two editions of the *Enquiry concerning Human Understanding*. In section 1 of these two editions, there was a lengthy footnote intended to illustrate the possibility of establishing clear and decisive results in the

science of human nature. Two such discoveries were chosen. The first, credited to Hutcheson, was that moral distinctions are founded on sentiment rather than reason; the second, credited to Butler, was the crucial point about the priority of desire over pleasure:

It has been prov'd, beyond all Controversy, that even the Passions, commonly esteem'd selfish, carry the Mind beyond Self, directly to the Object; that tho' the Satisfaction of these Passions gives us Enjoyment, yet the Prospect of this Enjoyment is not the Cause of the Passion, but on the contrary the Passion is antecedent to the Enjoyment, and without the former, the latter could never possibly exist[.] (E 1748)⁹

This note was deleted in the third edition (1756), and never reappeared. One obvious explanation for its removal is the same as for the removal of the promissory note in *Of the Dignity or Meanness of Human Nature*, namely the appearance of the *Enquiry concerning the Principles of Morals*, which examined both anti-rationalism and anti-egoism in considerably more depth.

The student who approaches Hume's moral philosophy today is likely to imagine that rationalism was a more important target for Hume than egoism. For one thing, current moral philosophers seem only ever to be concerned with Hume's anti-rationalist arguments, and this bias has perhaps affected Hume scholarship as well.¹⁰ For another, the texts themselves may seem to support this view: students

⁹This note is not reproduced in Selby-Bigge's edition of the two *Enquiries*; it can be found, however, in Beauchamp (2000, p. 232) and Millican (2007b, p. 177). I am indebted to Millican for drawing my attention to this note, and for first suggesting to me the importance of Butler as a source of Hume's later anti-egoism.

¹⁰In Radcliffe's *Companion to Hume* (2008)—a 528-page volume including 8 essays on the passions and morals—the index entry for egoism points to one solitary page. The page in question is not even from one of the essays on the passions or morals, but from Buckle's introductory article putting Hume in his historical context, in which Hume's opposition to Hobbes, in agreement with Shaftesbury and Hutcheson, is briefly noted (Buckle 2008, p. 34). Norton and Taylor's *Cambridge Companion to Hume* (2009) does a little better, mostly thanks to their own contributions. Norton's, on the foundations of morality in Hume's *Treatise*, begins by summarising the debates leading up to that work, including in particular the debate between the sentimentalists and the egoists. When we come to Hume himself, however, this introduction seems strangely out of place, for that debate has suddenly been replaced with the more parochial debate between Hutcheson and the rationalists. Norton himself seems not to have noticed that these were separate issues, for he speaks indiscriminately of "the dispute regarding the origin

are presented with a copy of the *Treatise*, and a copy of the *Enquiries*; the former is fiercely anti-rationalist, with hardly any hint of anti-egoism (indeed, if I am right it is even an egoist work itself), while the latter begins with the debate between sentimentalists and rationalists, reserving its anti-egoist arguments for an appendix at the end.

This reaction, however forgivable, is quite mistaken. Egoism was at least as important a target for Hume as rationalism, if not more important. To get a more accurate reflection of its significance, we must do several things. First, we must increase the weighting for the anti-egoist work: Hume only ever published one edition of *Treatise* Book 3, but *ten* editions of the moral *Enquiry*. Second, we must include the additional material: the deleted note from the early editions of the first *Enquiry*, and the essay *Of the Dignity or Meanness of Human Nature*. Last but not least, it must be noted that what we now know as the second appendix of the moral *Enquiry*, which contains Hume's most sustained attack on egoism, was only relegated to this position at the very end of his life, for the posthumous 1777 edition. In all other editions—which is to say, *in every edition published in Hume's lifetime*—it in fact had pride of place at the very start of section 2, *Of Benevolence*.¹¹ The first appendix, in contrast, which contains Hume's anti-rationalist arguments (where section 1 merely introduces and postpones the debate) was *always* an appendix.

The facts, therefore, are these: After one edition of the *Treatise*, a publication which the author himself sorely regretted, Hume wrote an essay against *egoism*,

of moral distinctions" (2009, p. 286; my emphasis). His article focuses exclusively on the *Treatise*, which focus, being shared by most Hume scholars, presumably helps to explain the corresponding focus on anti-rationalism at the expense of anti-egoism. Taylor is the exception that proves the rule: her article spends more time on this aspect of Hume's moral philosophy (2009, pp. 316-8), but then she has long been urging a preference for the moral *Enquiry* over Book 3 of the *Treatise*. Penelhum's article in the same collection, on Hume's moral psychology, notes Hume's anti-egoism just once in passing, 23 pages in (Penelhum 2009, p. 260).

¹¹This fact, incidentally, explains a curious remnant of the earlier structure in the 1777 edition: in a footnote in the second appendix, Hume writes that "we shall have occasion frequently to treat of [the sentiment of benevolence] in the course of this enquiry" (M App2.5, n60, p. 298); an odd thing to say in an appendix, but of course an entirely natural thing to say right at the start of a work. The remnant suggests that Hume did not fully think through this last-minute rearrangement.

not rationalism, using this as the occasion to promise a fuller examination of morals. In an early footnote in the first *Enquiry* (deleted after the moral *Enquiry* appeared), Hutcheson's anti-rationalism and Butler's anti-egoism were mentioned side by side. They were side by side again in the moral *Enquiry*, the second section of which, before this was moved to an appendix, began as follows:

There is a principle, supposed to prevail among many, which is utterly incompatible with all virtue or moral sentiment... This principle is, that all benevolence is mere hypocrisy, friendship a cheat, public spirit a farce, fidelity a snare to procure trust and confidence; ...

There is another principle, somewhat resembling the former; which has been much insisted on by philosophers, and has been the foundation of many a fair system; that, whatever affection one may feel, or imagine he feels for others, no passion is, or can be disinterested; that the most generous friendship, however sincere, is a modification of self-love[.] (M App2.2, pp. 295-6)

After a number of preliminary objections to this egoist view, Hume once again appeals to Butler's object-directed motivational psychology in his defence of the alternative:

There are bodily wants or appetites, acknowledged by every one, which necessarily precede all sensual enjoyment, and carry us directly to seek possession of the object... In the same manner, there are mental passions, by which we are impelled immediately to seek particular objects, such as fame, or power, or vengeance, without any regard to interest; and when these objects are attained, a pleasing enjoyment ensues, as the consequence of our indulged affections... In all these cases, there is a passion, which points immediately to the object, and constitutes it our good or happiness; as there are other secondary passions, which afterwards arise, and pursue it as a part of our happiness, when once it is constituted such by our original affections...

Now where is the difficulty in conceiving, that this may likewise be the case with benevolence and friendship, and that, from the original frame of

our temper, we may feel a desire of another's happiness or good, which, by means of that affection, becomes our own good, and is afterwards pursued, from the combined motives of benevolence and self-enjoyment?

(M (M App2.12-3, pp. 301-2)

Just like Butler, Hume makes clear that our pleasure in the important cases is a *consequence* of our desires being satisfied, and hence that those desires cannot themselves be hedonistic.

2.5. Desire in the *Dissertation on the Passions*

It would seem to be the standard view that, when writing the *Dissertation on the Passions*, Hume did nothing more than select a few passages from Book 2 of the *Treatise*, reorder them, and make some minor stylistic changes. My Difference Thesis, of course, claims that Hume's revisions were much more substantial. It is true that only a very small proportion of the *Dissertation* contains entirely new material, and that many of the modifications to the old material are purely stylistic. But though they are small in number, the few additions and non-stylistic changes that Hume did make are enormous in implication. What Hume does with the Butler paragraph is an excellent case in point.

Let us start by looking again at the original Butler paragraph in the *Treatise*, together with some of its preceding context:

'Tis easy to observe, that the passions, both direct and indirect, are founded on pain and pleasure, and that in order to produce an affection of any kind, 'tis only requisite to present some good or evil. ...

When good is certain or probable, it produces JOY. When evil is in the same situation there arises GRIEF or SORROW.

When either good or evil is uncertain, it gives rise to FEAR or HOPE, according to the degrees of uncertainty on the one side or the other.

DESIRE arises from good consider'd simply, and AVERSION is deriv'd from evil. The WILL exerts itself, when either the good or the absence of the evil may be attain'd by any action of the mind or body.

Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. These passions, properly speaking, produce good and evil, and proceed not from them, like the other affections. (T 2.3.9.1, 2.3.9.5-8, p. 438-9)

Recall, in particular, that “good” and “evil” are used in the *Treatise* simply as synonyms for “pleasure” and “pain”, and thus that the penultimate paragraph just quoted constitutes a clear endorsement of hedonism; an account which is then promptly contradicted by the Butler paragraph.

With this in mind, let us now take a look at what happened to these paragraphs when Hume came to rewrite them in the *Dissertation on the Passions*:

Some objects produce immediately an agreeable sensation, by the original structure of our organs, and are thence denominated GOOD; as others, from their immediate disagreeable sensation, acquire the appellation of EVIL. Thus moderate warmth is agreeable and good; excessive heat painful and evil.

Some objects again, by being naturally conformable or contrary to passion, excite an agreeable or painful sensation; and are thence called *Good* or *Evil*. The punishment of an adversary, by gratifying revenge, is good; the sickness of a companion, by affecting friendship, is evil.

When good is certain or probable, it produces JOY. When evil is in the same situation, there arises GRIEF or SORROW.

When either good or evil is uncertain, it gives rise to FEAR or HOPE, according to the degrees of uncertainty on the one side or the other.

DESIRE arises from good considered simply, and AVERSION is derived from evil. The WILL exerts itself, when either the good or the absence of the evil may be attained by any action of the mind or body. (P 1.1-2, 4-6, p. 3)

To a careless eye, not much has changed. The last three paragraphs are word for word repeats (spot the extra comma), and the second paragraph is just an abridged or slightly modified version of the Butler paragraph, moved a little way up the page. But if we look again more closely, we can see that, with a very few carefully crafted alterations, Hume has effected a *considerable* change in meaning.

The key is in the first of the paragraphs just quoted (which is, appropriately enough for my first argument in support of the Difference Thesis, the very first paragraph of the *Dissertation on the Passions*). Here Hume defines good and evil as pertaining to the *objects* that *give rise to* agreeable and disagreeable sensations: as the *external causes* of pleasure and pain, then, rather than as *identical* to these feelings. Similarly in the second paragraph (the amended version of Butler paragraph from the *Treatise*): *objects* are said to be good or evil, when they *excite* pleasure or pain (in virtue of gratifying or thwarting an object-directed desire). This seemingly small terminological difference has a considerably important consequence: the very same words that constituted, in the *Treatise*, an endorsement of hedonism, *no longer carry that meaning*. Desire is now said to arise from the consideration of *objects*, rather than the consideration of pleasure; and similarly the will is now said to exert itself when the *object* can be attained, rather than the pleasure. The author of the *Dissertation*, therefore, was not a hedonist. He was, as he had been ever since 1741, a clear and consistent supporter of Butler's object-directed motivational psychology.

Analogous to this particular point about desire and aversion, there is a general point about the direct passions as a whole. At the start of the *Treatise*, we may recall, Hume defined the direct passions—"desire, aversion, grief, joy, hope, fear, despair and security" (T 2.1.1.4, p. 277)—as those that "arise immediately from good or evil, from pain or pleasure" (T 2.1.1.4, p. 276). Hume's definition of these passions in the *Dissertation* is on the face of it much the same: he says that they "arise from a direct pursuit of good and aversion to evil" (P 2.1, p. 7). But again, since good and evil are no longer identical to pleasure and pain, but are instead applied to the objects that give rise to these feelings, the very same words now have a very different meaning. And Hume's mature definition of the

direct passions, unlike his earlier one, is consistent with the inclusion of object-directed desires in this set. I have drawn this closely related conclusion here and now, for the sake of clarity, but I will draw it again in the next chapter—perhaps a little more carefully, and at any rate in the context of the relevant existing debate—when I turn to examine Hume’s classification of the passions.

As noted in the introduction, Selby-Bigge dismissed the *Dissertation on the Passions* long ago as containing nothing more than “verbatim extracts from Bk. II of the *Treatise*, with some trifling verbal alterations”.¹² More recently, Terence Penelhum has expressed a similar sentiment, saying that the later work “is merely a brief résumé of the arguments of Book II of the *Treatise*”.¹³ Though it is seldom said explicitly (the *Dissertation* is seldom mentioned at all), this is presumably the standard view. As the present example illustrates, it is certainly understandable. But this hasty reaction has apparently discouraged commentators from looking more closely at the *detail* of what Hume says in the earlier and the later works. When we do this, we can see that the differences are in fact considerable. By altering just a few words, Hume changed sides on one of *the* major debates at the time, abandoning his hedonism in favour of Butler’s object-directed view. Or rather (as we saw in the previous two sections), the change of sides had already taken place. But the first four paragraphs of the *Dissertation*, when contrasted with their corresponding *Treatise* passages, provide a clear and independent corroboration of the developmental story that I have told, and a refutation of the view that there are no substantial differences between these two texts; and all this just on the first page of the later work. As we will see over the course of the thesis, this was by no means the only change that Hume made.

¹²Selby-Bigge (1893/1975, pp. xx-xxi).

¹³Penelhum (1975, p. 110).

Chapter 3

Hume's Classification of the Passions

The aim of this chapter is to locate the passions within Hume's general taxonomy of the perceptions of the mind, and to introduce his distinctions among the passions themselves. It will emerge that Hume's terminology and classificatory remarks are not exactly the same in the earlier and later works, and that many of these changes can be traced back to a greater awareness of Butler's object-directed desires and of the need to make room for them. The present chapter is concerned only with Hume's general remarks; when we turn to the particulars, in the next two chapters, yet more differences will emerge.

3.1. Impressions and sentiments

The very first distinction that Hume draws in his science of the mind, both in the *Treatise* and in his later work, is that between “feeling and thinking”, or equivalently that of the mind's perceptions into “impressions and ideas” (T 1.1.1.1, p. 1; E 2.1-3, pp. 17-8). This distinction, I trust, is well known, and I have little to add to our existing understanding of it. One thing, however, is not commonly remarked upon, and is of particular relevance here: where in the *Treatise* (and particularly in Books 1 and 2) Hume almost invariably labelled our felt perceptions “impressions”, in his later writing (and to an extent in Book 3 of the *Treatise*) there is an increasing preference for the term “sentiment”.

When Hume draws the distinction in section 2 of the first *Enquiry*, “impression” is still his official choice:

The less forceful and lively [perceptions] are commonly denominated THOUGHTS or IDEAS. The other species want a name in our language... Let us, therefore, use a little freedom, and call them IMPRESSIONS[.] (E 2.3, p. 18)

But while the term “sentiment” appears not once in the first section of the *Treatise* (where “impression” appears 46 times), it appears 5 times in section 2 of the first *Enquiry* (where “impression” appears 10 times), and in the *Enquiry* as a whole it outnumbers the other word by almost 2 to 1.¹

There can be no doubt that some of the time, at least, Hume intends “sentiment” as just a synonym for “impression”. A couple of examples suffice to prove the point (and any of the 5 from section 2 of the first *Enquiry* would do):

These faculties [memory and the imagination] may mimic or copy the perceptions of the senses; but they never can entirely reach the force and vivacity of the original sentiment. (E 2.1, p. 17)

[W]hen we analyse our thoughts or ideas, however compounded or sublime, we always find, that they resolve themselves into such simple ideas as were copied from a precedent feeling or sentiment. (E 2.6, p. 19)

Furthermore, the theory of the “double relation of impressions and ideas” from Book 2 of the *Treatise* (e.g. T 2.1.5.5, p. 286)—which I will look at briefly in the next chapter, and more thoroughly in chapter 9—has become, in the *Dissertation on the Passions*, that of the “double relation of *sentiments* and ideas” (e.g. P 2.22, p. 11; my emphasis). Again, the switch is not complete: the theory is later referred to once by its former name (P 3.1, p. 18). But the new name occurs three times, where in the *Treatise* it was never used.

¹The exact figure is 63 to 33. The trend here is quite general: in the *Treatise* as a whole, “impression” wins by 616 to 236 (a ratio of more than 5 to 2), while in the *Four Dissertations* and the two *Enquiries* combined, “sentiment” wins by an enormous 362 to 50 (a ratio of more than 7 to 1). In the two *Enquiries* alone the ratio is only fractionally smaller (261 to 37). These statistics do not measure exactly what we want: sometimes “impression” is used in its ordinary sense (e.g. “[t]he inference is by no means just, that, because a system of religion has made no deep impression on the minds of a people, it must therefore have been positively rejected by all men of common sense”; N 12.19, p. 73), and sometimes “sentiment” is used to mean *opinion* (see page 57 below). Ideally, the data would be sifted and these instances discounted. Nevertheless, the figures are telling.

In a sense, the change here is superficial, as all terminological changes are. “Impression” is after all a technical term for Hume, meaning no more than what he has defined it to mean, and “sentiment” is later used in a similarly artificial sense. But, as anyone who has tried to coin a technical term themselves knows, connotations do matter. And in any case, the fact that Hume increasingly adopted a new term, either deliberately or unconsciously, demands its explanation; and that explanation must presumably take the form of pointing to some connotations of the later word that would have been attractive.

I can think of two such connotations, which, if I am right that these are what attracted Hume, point to or at least highlight two noteworthy features of his thought in this area. The first point is the obvious one: that with this choice of word Hume more closely aligned himself, verbally at least, with the earlier “sentimentalists” Shaftesbury and Hutcheson (Butler does not use the term very much; Hutcheson uses it more than Shaftesbury). This reflects his general move in this direction, and in particular his joining these thinkers in the campaign against egoism (recall chapter 2).

The second connotation of the word “sentiment” is perhaps the more interesting, and this is its *intellectual* or *cognitive* connotation. Though Hume describes impressions as *feelings* rather than thoughts, the word “sentiment” is nicely ambiguous or midway between these two things. Johnson’s dictionary defines it as “thought, notion, opinion” (1755), without any mention of the feeling sense of the word at all (though that is surely an oversight). Kames makes nice use of the double meaning in his definition of a sentiment as a “thought prompted by passion” (1762, vol. 1, p. 311; see also vol. 2, p. 741). And Hume himself sometimes uses the word in its opinion sense, for example in the first *Enquiry*:

It is confessed, that the colour, consistence, and other sensible qualities of bread appear not, of themselves, to have any connexion with the secret powers of nourishment and support. For otherwise we could infer these secret powers from the first appearance of these sensible qualities, without the aid of experience; contrary to the sentiment of all philosophers, and contrary to plain matter of fact. (E 4.21, p. 37)

And though this reasoning may contradict the systems of many philosophers, in ascribing necessity to the determinations of the will, we shall find, upon reflection, that they [all mankind] dissent from it in words only, not in their real sentiment. (E 8.21, p. 92)

Something like Hume's impression/idea distinction survives in current debates about the nature of emotions, dividing the "feeling" or "non-cognitivist" views from the "judgment" or "cognitivist" ones.² Given Hume's identification of the passions with *impressions* of a certain sort—see §3.2 below—it is tempting to align him with the modern-day non-cognitivists.³ This temptation, however, should be resisted. As we will see later (chapter 9), Hume's view has clear cognitivist elements, and overall occupies something of a middle ground between these two rival schools. Accordingly it seems to me that Hume's later choice of the word "sentiment" is extremely apt, and perhaps less likely than "impression" to invite crude non-cognitivist misreadings.

3.2. Three more distinctions

Having distinguished impressions from ideas, Hume proceeds in the *Treatise* to divide impressions themselves into "those of SENSATION and those of REFLEXION" (T 1.1.2.1, p. 7). As we will see shortly, there is a difficulty regarding the way in which Hume draws this distinction. Nevertheless, the intention behind it is perfectly clear. Impressions as a whole include "all our sensations, passions and emotions" (T 1.1.1.1, p. 1), and the point of the sensation/reflection distinction is simply to draw a line at the comma, with "sensations" on one side, and "passions and emotions" on the other. Impressions of sensation are obviously meant to be sensations (hence the name); impressions of reflection, meanwhile, are said to be our "passions, desires, and emotions" (T 1.1.2.1, p. 8).

²See Nussbaum (2001) and Solomon (2004a) on the cognitivist side, and Prinz (2004) on the non-cognitivist side. Solomon's collection (2004b) also contains articles by Nussbaum and Prinz, and is a good source of further articles and references.

³Thus Pitcher (1965), Solomon (1977), Deigh (1994), Zemach (2001).

The examples of sensations at the start of the *Treatise* are “heat or cold, thirst or hunger, pleasure or pain of some kind or other” (T 1.1.2.1, p. 8). In the context this must mean *bodily* pleasure or pain, for at the start of Book 2, when Hume repeats the distinction in new terms (impressions of sensation are now called “original”, and those of reflection are now called “secondary”), the former are said to comprise “all the impressions of the senses, and all *bodily* pains and pleasures” (T 2.1.1.1, p. 275; my emphasis). The latter, as before, are said to be “the passions, and other emotions resembling them”. (ibid.).

Extensionally, then, there can be no doubt about what Hume intended by this sensation/reflection distinction. Unfortunately, however, the *way* in which the distinction is drawn in the *Treatise* is problematic, for it presupposes the hedonism which I have already argued dominates this early work:

Impressions may be divided into two kinds, those of SENSATION and those of REFLEXION. The first kind arises in the soul originally, from unknown causes. The second is derived in a great measure from our ideas, and that in the following order. An impression first strikes upon the senses, and makes us perceive heat or cold, thirst or hunger, pleasure or pain of some kind or other. Of this impression there is a copy taken by the mind, which remains after the impression ceases; and this we call an idea. This idea of pleasure or pain, when it returns upon the soul, produces the new impressions of desire and aversion, hope and fear, which may properly be called impressions of reflexion, because derived from it. (T 1.1.2.1, pp. 7-8)

Here Hume seemingly *defines* impressions of reflection (including desire and aversion) as those that arise from an antecedent idea of pleasure or pain. At the very least, if this is not intended as their definition, it is nevertheless a general thesis that is apparently supposed to apply to them all. Having familiarised ourselves with Hume’s later views, however, we can see that Butler’s object-directed desires are counterexamples to this definition or general thesis. They are not derived from pleasure and pain, but rather pleasure and pain are derived from them. Furthermore, as we will see in chapter 5, they arise in the soul originally, from unknown

causes, much like the sensations or “original” impressions with which all passions and desires are here contrasted. When Hume wrote this paragraph, I therefore submit, he was not yet aware of the existence of these original, object-directed desires. In line with my conjecture at the start of chapter 2, he had not yet read Butler's *Sermons*; or if he had, he certainly had not absorbed Butler's crucial point about the priority of desire over pleasure and pain.

Having distinguished impressions from ideas, and divided impressions into original and secondary, Hume's next move in Book 2 of the *Treatise* is to further subdivide the secondary or reflective impressions into the *calm* and the *violent*. When drawing this third distinction, Hume uses “emotion” as the general term (i.e. to refer to *all* secondary impressions, both calm and violent), while reserving “passion” specifically for the *violent* emotions:

The reflective impressions may be divided into two kinds, *viz.* the *calm* and the *violent*. Of the first kind is the sense of beauty and deformity in action, composition, and external objects. Of the second are the passions of love and hatred, grief and joy, pride and humility... properly call'd *passions*... [I]n general the passions are more violent than the emotions arising from beauty and deformity... I... shall now explain those violent emotions or passions, their nature, origin, causes, and effects. (T 2.1.1.3, p. 276)

On the face of it, at least, the point of this distinction is to separate the moral and critical sentiments (i.e. moral approbation and disapprobation, and the sentiments of beauty and deformity) from passions in the more natural sense, i.e. what we would today call “emotions”. Hume doesn't mention the moral sentiments explicitly, but I take it that these are what he means by “the sense of beauty and deformity *in action*” (my emphasis). I will examine this distinction in more detail in §3.4 below, but we will see no cause to modify this straightforward first impression.

With just one more distinction, our reconstruction of Hume's *Treatise* taxonomy will be complete. This is the division of the passions (the violent secondary impressions) into *direct* and *indirect*:

When we take a survey of the passions, there occurs a division of them into *direct* and *indirect*. By direct passions I understand such as arise immediately from good or evil, from pain or pleasure. By indirect such as proceed from the same principles, but by the conjunction of other qualities. This distinction I cannot at present justify or explain any farther. I can only observe in general, that under the indirect passions I comprehend pride, humility, ambition, vanity, love, hatred, envy, pity, malice, generosity, with their dependants. And under the direct passions, desire, aversion, grief, joy, hope, fear, despair and security. (T 2.1.1.4, pp. 276-7)

Like Hume, I will not presently explain this distinction any further. There are important interpretative questions surrounding it, but it will be easier to answer these over the next two chapters, when we move from the current level of generality to discuss more particular matters. For now, it suffices to make a note of the distinction, and to point out that—like the distinction between impressions of sensation and impressions of reflection—the way in which Hume draws it again presupposes hedonism. The passions on *both* sides of this divide, Hume says, “arise... from good or evil, from pain or pleasure”, the difference being that the direct passions do so “immediately”, while the indirect do so “by the conjunction of other qualities”. (The other qualities that Hume has in mind here will be examined in the next chapter.) This was in effect my argument in §2.2, but we are now able to place it in the context of Hume’s taxonomy as a whole.

3.3. The Butler paragraph (again)

With this context in mind, let us turn again to the Butler paragraph:

Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. These passions, properly speaking, produce good and evil, and proceed not from them, like the other affections. (T 2.3.9.8, p. 439)

As I pointed out in §2.2, classing these object-directed desires as *direct* contradicts the definition of the direct passions as those that arise immediately from pleasure and pain. I shall now rub the point in: calling them passions of *any* kind also contradicts the definition of impressions of reflection (as impressions that arise from ideas of pleasure and pain, one way or another); and placing the bodily appetites such as hunger in the same group contradicts the sensation/reflection distinction. For impressions of sensation, we may recall, include all bodily pains and pleasures, with hunger and thirst being named as explicit examples at T 1.1.2.1 (p. 8); bodily appetites, being impressions of sensation, cannot also be passions (impressions of reflection).

Existing attempts to reconstruct Hume's taxonomy of impressions and passions all focus on the *Treatise*, ignoring Hume's later works, and are also curiously blind to these inconsistencies. In one of the earliest attempts at reconstruction, Kemp Smith treats the object-directed desires as a third category, distinct from both the direct and the indirect passions, thereby being faithful to Hume's definition of the direct passions, but ignoring his claim that object-directed desires are examples of such passions.⁴ In one of the most recent attempts, by contrast, James Fieser respects this claim, counting object-directed desires among the direct passions, but ignores the contradictory definition.⁵ Either proposal seems a viable and consistent *revision* of the *Treatise* taxonomy, but neither can claim to be the view that is actually presented in that work. As far as the *Treatise* is concerned, there is nothing to choose between them.

Fortunately Hume's later work breaks the tie, for attention to the *Dissertation on the Passions* shows that Hume himself revised the *Treatise* taxonomy into something consistent; as a result, no doubt, of coming to appreciate the importance of Butler's object-directed desires, and of the need to make room for them. He might have gone either way, I suppose, but as it happens he opted for

⁴Kemp Smith (1941/2005, pp. 164-9). To be more precise, Kemp Smith does briefly acknowledge the contradictory claim in a footnote (p. 165, n2); but he ignores the contradiction.

⁵Fieser (1992, pp. 10-1).

Fieser’s revision. He no longer uses the technical terms “direct” and “indirect”, but the distinction is still clearly present in the *Dissertation*, as indicated by the opening sentence of section 2:

Besides those passions above-mentioned, which arise from a direct pursuit of good and aversion to evil, there are others which are of a more complicated nature, and imply more than one view or consideration. (P 2.1, p. 7)

The passions “above-mentioned”, i.e. the ones discussed in section 1 of the *Dissertation*, are desire and aversion (including the object-directed desires such as benevolence and anger), joy and sorrow, hope and fear. These are precisely the direct passions of the *Treatise*, including the object-directed desires.⁶ They are said to arise from a direct pursuit of good and aversion to evil. As noted before (§2.5), “good” and “evil” are no longer synonymous with “pleasure” and “pain” in the *Dissertation*, but are instead applied to the *objects* that *give rise to* pleasure and pain. The present definition of the “above-mentioned”—i.e. direct—passions is thus verbally similar but substantially different from that given in the *Treatise*; in exactly the same way as we have seen with Hume’s account of motivation. And the new definition is consistent with the inclusion of object-directed desires in the group. For ease of exposition, I already rehearsed this argument at the end of the previous chapter, but it is well to have repeated it here in the context of discussing Hume’s classification of the passions in general.

As to the other passions—those “of a more complicated nature”—these correspond to the *Treatise* category of indirect passions, though Hume happens not to use that label any more. The immediate examples in the *Dissertation* are pride, humility, love, and hatred, which are named as indirect passions at the start of *Treatise* Book 2 (T 2.1.1.4, p. 276-7). These four passions are the main

⁶To be thorough, I should note that the *Treatise* list also includes despair and security, which are not mentioned in section 1 of the *Dissertation*. Despair and security are the contraries of hope and fear respectively, and so must presumably still belong in the same group, although Hume happens not to mention them. Hume has very little to say about these passions anyway; for the little that he does say, see T 2.3.4.8 (pp. 421-2), repeated at P 6.9 (p. 27).

subject matter of section 2 of the *Dissertation*, and of most of sections 3 and 4 as well. Later on, in section 3, Hume also examines pity, malice, and envy, three more of the passions named as indirect at the start of *Treatise* Book 2 (*ibid.*). Generosity, another of Hume's *Treatise* examples, is not mentioned in the *Dissertation*, but we will see over the course of the next two chapters that there is nothing significant about that. Ambition and vanity are more interesting: these are indirect passions in the *Treatise*, but object-directed desires—and hence *direct* passions—in the moral *Enquiry* (see the brief comment below, and the detailed discussion in §5.3). Notwithstanding this particular change of mind, however, Hume clearly retained the general distinction.

One wrinkle remains to be ironed out. What about the bodily appetites such as hunger and thirst, which in the Butler paragraph Hume counted as straightforward examples of object-directed desires, but which are on the wrong side of his sensation/reflection distinction? It may be noted that Butler did not much care for the distinction between the bodily appetites and other desires, and so on the assumption that Hume added this paragraph after reading Butler, the inclusion of bodily appetites in the group is not surprising. In his later work, however, Hume himself maintained the distinction between bodily appetites and mental desires, while agreeing with Butler insofar as to say that the two are *analogous* in the key respect of their causal priority over pleasure and pain:

There are *bodily wants or appetites*, acknowledged by every one, which necessarily precede all sensual enjoyment, and carry us directly to seek possession of the object. Thus, hunger and thirst have eating and drinking for their end; and from the gratification of these primary appetites arises a pleasure, which may become the object of another species of desire or inclination, that is secondary and interested. In the same manner, there are *mental passions*, by which we are impelled immediately to seek particular objects, such as fame, or power, or vengeance, without any regard to interest; and when these objects are attained, a pleasing enjoyment ensues, as the consequence of our indulged affections. (M App2.12, p. 301; my emphases)

Insisting on the distinction between bodily appetites and mental passions, as Hume does here, is consistent with the revised Butler paragraph that appears in the *Dissertation* (P 1.2, p. 3), for this paragraph mentions only benevolence and anger as examples of object-directed desires, making no reference to the bodily appetites at all. While we are here, it is also well to observe that the paragraph just quoted establishes something I claimed a moment ago, namely that in the moral *Enquiry* Hume saw ambition and vanity (the desires for power and fame respectively) as object-directed desires, rather than indirect passions. But I will pursue this point in detail in chapter 5.

Hume's considered view, then, contrary to the Butler paragraph (and to both Kemp Smith's and Fieser's interpretations of the *Treatise*, which agree on this point), is that the bodily appetites are in a category of their own, distinct from the passions. This, in fact, was also the official position of the *Treatise*—as shown by Hume's drawing of the sensation/reflection distinction—and would have consistently been his view throughout that work, were it not for the isolated contradictory assertion in the Butler paragraph. This is therefore further confirmation of my claim that the Butler paragraph fits very uncomfortably into the rest of the *Treatise*, and was thus most likely a last-minute addition. Needless to say by now, however, object-directed desires fit very nicely into Hume's later view; indeed, they are absolutely central to it.

3.4. Calm and violent

In §3.2, I suggested that the aim of Hume's calm/violent distinction was to separate the passions in the familiar sense from the moral and critical sentiments. This seems the only natural and plausible reading of T 2.1.1.3, which I quoted at the time but will repeat here:

The reflective impressions may be divided into two kinds, *viz.* the *calm* and the *violent*. Of the first kind is the sense of beauty and deformity in action, composition, and external objects. Of the second are the passions of love and hatred, grief and joy, pride and humility... properly call'd *passions*...

[I]n general the passions are more violent than the emotions arising from beauty and deformity... I... shall now explain those violent emotions or passions, their nature, origin, causes, and effects. (T 2.1.1.3, p. 276)

I take it that this interpretation of the calm/violent distinction is, by itself, fairly innocuous and insubstantial; certainly my endorsement of it is intended to be. Louis Loeb, however, has defended it in the context of a more substantial point, namely that Hume's views on morality—*contra* Pall Árdal—can be understood without reference to his account of the passions. Loeb saw the calm/violent distinction as driving a wedge between the passions (the subject matter of Book 2) and the moral sentiments (the subject matter of Book 3). More specifically, where Árdal had argued that the moral sentiments were indirect passions, Loeb insisted that they were not *passions* at all, let alone indirect ones.⁷

The question of where the moral sentiments belong in Hume's taxonomy remains a controversial one. Before Árdal argued that they were indirect passions, Kemp Smith had classified them as direct passions, and later Thomas Hearn defended Kemp Smith's claim in opposition to Árdal. Then Loeb, as we have seen, insisted in opposition to both parties that they were not passions at all. More recently, Fieser concurred with Loeb's classification; more recently still, a version of Árdal's view has been resurrected by Rachel Cohon, although Cohon insists only that the moral sentiments are indirect *secondary impressions*, leaving open the question of whether or not they are *passions*.⁸

As I have said, my inclination on this matter is to go along with Loeb and Fieser regarding the classificatory question. There is a distinction to be drawn between the passions on the one hand, which are partial and incapable of being rightly or wrongly held (save with regard to mistakes of fact), and the moral and critical sentiments on the other, which are supposed to be derived from an impartial common point of view, and are subject to a *standard of taste* that goes beyond mere factual errors. But I would be surprised if this claim, prop-

⁷Árdal (1966), Loeb (1977).

⁸Kemp Smith (1941/2005, p. 168), Hearn (1973), Fieser (1992), Cohon (2008a, 2010).

erly understood, excited any controversy. In particular, it does not rule out the possibility that the moral sentiments have a lot in common with the passions (or perhaps with some special subset of the passions; see below). Nor—*contra* Loeb—does it undermine Árdal’s basic point that Hume’s treatment of the passions is of considerable value for understanding his treatment of morals. Árdal was surely right about this; Hume was interested in the passions for their own sake, but like everyone else at the time he was also interested in their bearing on morality, and this secondary interest is seldom out of sight.

Insofar as the moral sentiments have a particular affinity with some of the passions more than with others, it seems to me that Árdal and Cohon are closer to the truth in identifying them with the indirect passions. I have neither the space nor the need to defend this claim in any detail. I do want to suggest, however, that this identification is perhaps a little careless in one crucial respect (as this will be relevant again in an argument at the end of the next chapter). Among the indirect passions, four stand out as particularly important for Hume, namely pride, humility, love, and hatred. I propose to call these the *double-relation* passions, since Hume accounts for their origin with his theory of the double relation of sentiments and ideas (see chapters 4 and 9). Insofar as the moral sentiments have analogues among the passions, for Hume, it seems to me that these analogues are not the indirect passions in general, but only the double-relation passions in particular. It is easy to see how moral approbation might be a kind of love or pride, and how disapprobation might be a kind of hatred or humility. It is considerably harder to see what the moral sentiments might have in common with ambition, pity, malice, or any of the other indirect passions.

The carelessness of the association of the moral sentiments with the indirect passions as a whole, rather than the double-relation passions in particular, comes out in a recent argument from Cohon, to the effect that the moral sentiments do not motivate us directly.⁹ Cohon rightly observes that pride, humility, love, and hatred do not move us to act directly. We may be moved by the *prospect* of pride

⁹Cohon (2010).

or humility (as with any potential pleasure or pain), but the *present experience* of these passions, according to Hume, has no motivational effect. Love and hatred, meanwhile, move us only *indirectly*, by first prompting the desires of benevolence and anger:

[P]ride and humility are pure emotions in the soul, unattended with any desire, and not immediately exciting us to action. But love and hatred are not compleated within themselves, nor rest in that emotion, which they produce, but carry the mind to something farther. Love is always follow'd by a desire of the happiness of the person below'd, and an aversion to his misery: As hatred produces a desire of the misery and an aversion to the happiness of the person hated. (T 2.2.6.3, p. 367)

Cohon then uses these four examples—pride, humility, love, and hatred—to argue that the indirect passions *as a whole* do not move us to act directly, and concludes that the moral sentiments, being likewise indirect, must also have this property. But she seems, in this argument, to have forgotten about the other indirect passions, such as ambition, vanity, pity, and malice. These very plainly *do* move us to act directly, and so the argument is made to depend upon a false generalisation. Fortunately, the generalisation seems to be as unnecessary as it is untrue. She need only have said that the moral sentiments are analogous to the *double-relation* passions in particular, not to the indirect passions in general.

Hume's calm/violent distinction appears to resurface later in the *Treatise*, in the form of a contrast between "certain calm desires and tendencies" such as "benevolence and resentment, the love of life, and kindness to children" on the one hand (T 2.3.3.8, p. 417), and such things as "a violent passion of resentment" or particularly strong "fears, apprehensions, and aversions" on the other (T 2.3.3.9, pp. 418). On closer inspection, however, this would seem to be a different distinction. For one thing, the examples on both sides of this new line all belong on the *violent* side of the earlier division. For another, Hume is explicit in saying that the "calm desires" in this new sense are nevertheless "real passions" (T 2.3.3.8, p. 417), i.e. violent in the earlier sense.

This is not to say that the corresponding terminology is mere coincidence. Both distinctions are based on a difference in felt intensity between the sentiments on the different sides of the divide. But the similarity ends there, and I suggest that the two distinctions need to be kept apart. The later calm/violent distinction is drawn in the context of Hume's discussion of the relationship between reason and passion, and will be examined further in chapter 10.

3.5. Some peculiar feelings

Hume's distinction between impressions of sensation and impressions of reflection is an echo of Locke's distinction between ideas drawn in the same terms (Locke 1690, p. 105). While Locke's and Hume's categories of sensation coincide, however, their notions of reflection are importantly different. For Hume, as we have seen, impressions of reflection are essentially passions and emotions (though there is a complication here that I will come to presently). For Locke, on the other hand, reflective ideas are those that arise in us from "the perception of the operations of our own mind", and are of such things as "perception, thinking, doubting, believing, reasoning, knowing, willing" (1690, p. 105).

Hume seems to have had a curious blind-spot with regard to Lockean reflection as a source of ideas.¹⁰ Without ever denying it explicitly, he implicitly rules it out with his copy principle—that all our simple ideas are copied from a precedent impression (T 1.1.1.7, p. 4; E 2.5, p. 19)—and also fails to appeal to it at three crucial points in his work where one might have thought it was the obvious thing to appeal to: belief, necessary connection, and volition. In each of these three cases he instead posits an *impression* as the source of our ideas.¹¹ Since these cannot be sensory impressions, meanwhile, they are presumably to be classified as impressions of reflection. As such, they complicate the otherwise straightforward equation of impressions of reflection with emotions.

¹⁰I am grateful to Peter Millican for drawing my attention to this, and to its relevance in the present context.

¹¹See, respectively, T App.2-3 (pp. 623-5) and E 5.10-3 (pp. 47-50); T 1.3.14.22 (pp. 165-6) and E 7.28 (pp. 75-6); and T 2.3.1.2, 2.3.9.2, 2.3.9.4 (pp. 399, 438-9).

Hume's treatment of belief is notoriously problematic, even by his own admission, and this is not the place to delve into these murky waters.¹² I shall also set aside here the origin of our idea of necessary connection.¹³ Both of these would take us too far from our present concerns. The impression of the will or volition, however, does belong in a discussion of Hume on the passions. Indeed, we find it in *Hume's* discussion of the passions in Book 2 of the *Treatise*.

This impression does not feature heavily in Hume's theory. It appears in just three paragraphs in *Treatise* Book 2, none of which is repeated in his later work on the subject:

I desire it may be observ'd, that by the *will*, I mean nothing but *the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind.* (T 2.3.1.2, p. 399)

The impressions, which arise from good and evil most naturally, and with the least preparation are the direct passions of desire and aversion, grief and joy, hope and fear, along with volition. (T 2.3.9.2, p. 438)

Thus a suit of fine cloaths produces pleasure from their beauty; and this pleasure produces the direct passions, or the impressions of volition and desire. (T 2.3.9.4, p. 439)

Despite the isolated—and frankly very peculiar—nature of these remarks, it seems to be fairly common for interpreters to accept the impression of the will or volition unquestioningly into the fold, and even to make it the crucial distinctive feature of Hume's account.¹⁴ In opposition to this trend, however, I should like to suggest that these three paragraphs were *mistakes*, and to attribute to Hume instead the

¹²See Broackes (2002) for a discussion.

¹³See Millican (2007a) for an interpretation of Hume's account along Lockean lines: "it is our reflexive awareness of making the inference that leads us to the very idea of connexion" (p. 224; cf. p. 249n26).

¹⁴Thus Bricke (1984), Stalley (1986), Connolly (1987). Fieser stops short of classifying volition as a passion, but still accepts it as a secondary impression (1992, pp. 12-3).

more commonsensical view that the will is a *mental faculty*, and that volition is the *mental act* for which this faculty is responsible.

This much more natural position should certainly have been familiar to Hume, for it was precisely Locke's view:

Volition, 'tis plain, is an Act of the Mind knowingly exerting that Dominion it takes it self to have over any part of the Man, by imploying it in, or withholding it from any particular Action. And what is the *Will*, but the Faculty to do this? (1690, p. 241)

Moreover, when Hume himself discusses volition or the will—in all but the three unfortunate paragraphs quoted above—he seems pretty clearly to have the same understanding of these things in mind. For example:

Nothing can oppose or retard the impulse of passion, but a contrary impulse; and if this contrary impulse ever arises from reason, that latter faculty must have an original influence on the will, and must be able to cause, as well as hinder any act of volition. (T 2.3.3.4, pp. 414-5)

Beside these calm passions, which often determine the will, there are certain violent emotions of the same kind, which have likewise a great influence on that faculty. (T 2.3.3.9, pp. 417-8)

The chief spring or actuating principle of the human mind is pleasure or pain; and when these sensations are remov'd, both from our thought and feeling, we are, in a great measure, incapable of passion or action, of desire or volition. (T 3.3.1.2, p. 574)

Volition is surely an act of the mind, with which we are sufficiently acquainted. (E 7.20, p. 69)

I suggest, therefore, that Hume's views on volition and the will should be understood in essentially Lockean terms. It is true that he was tempted towards a different view, presumably because of his copy principle and his blind-spot regarding Lockean reflection. But the three *Treatise* passages committing him

to an impression of volition or the will should be treated as *aberrations*, not as representing the core of his position.¹⁵ It is no accident, I take it, that none of these passages reappears in the *Dissertation on the Passions*, or anywhere else in Hume's mature writing.

Before proceeding, I cannot resist a general complaint. As we have seen, existing discussions of Hume's classification of the passions all focus on Book 2 of the *Treatise*, ignoring the evidence of the later works. Not only that, but such is the reverence for the early work that even its internal imperfections are glossed over or simply ignored. Cohon, for example, has described what we find in the *Treatise* as a "careful taxonomy".¹⁶ I do not see how a close and unprejudiced reading of the text can leave one with this impression. And a reading of the later works alongside the *Treatise* reveals quite the opposite: Hume's first work presents what can only be described as a *careless* taxonomy, but one that he subsequently tidied up (once Butler's anti-hedonist arguments had had time to sink in). Trying to understand Hume's taxonomy as a static view that was already fully in place in the *Treatise* is a task that is doomed to failure. In this, as in many other things, Hume changed his mind.

¹⁵Cf. Millican (2009b): "[Hume's] equation of the will with an internal impression, no doubt motivated by his Copy Principle, seems to be a slip, as it leaves no obvious mark on his treatment of the will elsewhere. A more charitable reading would be that Hume intends 'the will' to refer to our faculty of knowingly - and 'willingly' - giving rise to actions (of the mind and body), a faculty of which we become aware, and whose idea we thus acquire, through a corresponding internal impression" (p. 6).

¹⁶Cohon (2008a, p. 160).

Chapter 4

Association, Comparison, Sympathy

In explaining the causes of the indirect passions, Hume appeals to three general principles: the comparison of ideas, the communication of sentiments (or sympathy), and a complex associative mechanism that he calls the double relation of sentiments and ideas.¹ His thinking concerning all three of these principles changed after the *Treatise*. In the present chapter, however, my Difference Thesis is temporarily on hold. My description of the double relation theory will be deliberately vague, glossing over the differences between Hume's earlier and later versions of it. These differences are not relevant to the current line of argument, and will be examined separately in chapter 9. The differences in Hume's attitude towards sympathy and comparison, meanwhile, *are* immediately relevant, but it will be simplest to tell this story chronologically. In this chapter, therefore, I will focus exclusively on what Hume had to say about these two principles in the *Treatise*, the better to see what changed when, in the next chapter, I turn to what he said about them in his later works.

I also have a secondary aim in this chapter, not directly related to my main thesis. I will argue for it overtly in the final section, but the whole chapter is intended to support it by illustration. The point is that sympathy, comparison, and the double relation belong very much together in Hume's mind, and that all three of them are crucial to his account of the indirect passions. This is in

¹More familiarly, Hume refers to the third of these things in the *Treatise* as the double relation of *impressions* and ideas; but, as we saw in the previous chapter (§3.1), he increasingly preferred the term "sentiment" over "impression" in his later work.

opposition to a tendency among recent commentators to overemphasise the double relation theory, at the expense of sympathy and comparison, and correspondingly to shrink the category of indirect passions to just pride, humility, love, and hatred. But Hume's explicit list of indirect passions in the *Treatise* includes ambition, vanity, pity, malice, envy, and generosity, as well as these more familiar four. In this chapter I mean to account for *every one* of the passions on Hume's list.

4.1. The double relation theory

The centrepiece of Hume's psychology of the passions is his theory of the double relation of sentiments and ideas. This is his account of the causes of the indirect passions of pride, humility, love, and hatred. These are the four indirect passions that I singled out in the previous chapter as particularly closely associated with the moral sentiments, and decided to call—for obvious reasons—the *double-relation* passions. In Book 2 of the *Treatise*, the double relation theory, and the four passions whose origins it accounts for, dominates both part 1 (*Of pride and humility*) and part 2 (*Of love and hatred*). In the *Dissertation on the Passions* it similarly dominates sections 2-4, which together make up about two thirds of the whole.

There is plenty to be said about this theory, and in particular about how Hume's thoughts concerning it developed over time. What I have to say in this regard, however, is not directly relevant to the story that I am telling in Part 1, about Hume's change from egoism to anti-egoism. I shall therefore postpone the full discussion until chapter 9. In the meantime, it is necessary to be familiar with a rough outline of this theory, in order to complete the picture of Hume's classification of the passions.

In outline, then, this theory comprises two principles of association, regarding sentiments and ideas respectively, together with a third principle of the "mutual assistance" of these other two: "The present theory of the passions depends entirely on the double relations of sentiments and ideas, and the mutual assistance, which these relations lend to each other" (P 4.1, p. 21). The association of ideas,

first, is familiar from Book 1 of the *Treatise* (T 1.1.4, pp. 10-3) or the first *Enquiry* (E 3, pp. 23-4). Hume sums it up as follows in the *Dissertation on the Passions*:

However uncertain and changeable our thoughts may be, they are not entirely without rule and method in their changes. They usually pass with regularity, from one object, to what resembles it, is contiguous to it, or produced by it. [Footnote: See *Enquiry concerning Human Understanding*, Sect. III.] When one idea is present to the imagination, any other, united by these relations, naturally follows it, and enters with more facility, by means of that introduction. (P 2.6, p. 7; cf. T 2.1.4.2, p. 283)

Thus this “principle” of the association of ideas is in fact *three* principles (resemblance, contiguity, and causation), all of which nevertheless have the same effect of governing the progress of our thoughts.

In Book 2 of the *Treatise*, and again in the *Dissertation on the Passions*, Hume goes on to endorse a less familiar parallel to this principle of the association of ideas, namely an analogous association of sentiments:

The *second* property, which I shall observe in the human mind, is a like association of impressions or emotions. All *resembling* impressions are connected together; and no sooner one arises, than the rest naturally follow. Grief and disappointment give rise to anger, anger to envy, envy to malice, and malice to grief again. In like manner, our temper, when elevated with joy, naturally throws itself into love, generosity, courage, pride, and other resembling affections. (P 2.7, p. 8; cf. T 2.1.4.3, p. 283)

Notice that there really is just one principle or relation governing this association, namely resemblance.

Finally, Hume adds a third principle concerning the mutual assistance of these other two: “In the *third* place, it is observable of these two kinds of association, that they very much assist and forward each other, and that the transition is more easily made, where they both concur in the same object” (P 2.8, p. 8; cf. T 2.1.4.4, pp. 283-4).

Pride and humility, for Hume, are good and bad feelings, respectively, about ourselves: “*Pride* is a certain satisfaction in ourselves, on account of some accomplishment or possession, which we enjoy: *Humility*, on the other hand, is a dissatisfaction with ourselves, on account of some defect or infirmity” (P 2.1, p. 7). Love and hatred are the corresponding good and bad feelings about other people: “*Love* or *Friendship* is a complacency in another, on account of his accomplishments or services: *Hatred*, the contrary” (P 2.2, p. 7).

Hume’s three principles operate together to produce the first of these double-relation passions in the following way. When presented with something pleasant that is mine—my beautiful house, for example—I experience both an idea of the house, and a pleasing impression of its beauty. The impression, by the association of sentiments, has an inherent tendency to prompt a feeling of pride, since pride is also a pleasant feeling. The idea, meanwhile, has the very same tendency, thanks to the association of ideas and the fact that pride is a feeling of *self*-satisfaction (and it is *my* house). Each of these principles on its own, finally, might give rise to any number of consequent perceptions in the mind: pride is not the only pleasant feeling, and I am not the only thing associated with my house. Given the mutual assistance of the two associative principles, however, pride becomes the most likely result.

It is obvious, I trust, how an exactly parallel story can be told with regard to the other double-relation passions, humility, love, and hatred. If my house is ugly, humility will be the result. If someone else’s house is in question, love and hatred will follow, depending on whether it is a pleasant or an unpleasant place to live. As I have already advertised, there is more to be said here about the detail, and in this summary I have skipped over several important and controversial points (see chapter 9 for the full discussion). For immediate purposes, however, all that is needed is to know that Hume had this theory, and to have at least an approximate idea of its shape.

4.2. Comparison

Next to this theory of the double relation, Hume's other major psychological principles in the *Treatise* are the communication of sentiments (or sympathy) and the comparison of ideas. The last of these is by far the simplest of the three, the straightforward idea being just this: we may compare our own situation with that of another, and conclude that the other person is either better or worse off than ourselves. If the comparison goes in our favour, this produces the passion of *malice*, "a joy in the sufferings and miseries of others" (T 2.2.8.1, p. 372). If the comparison goes in the other person's favour, it produces *envy*, this passion being "excited by some present enjoyment of another, which by comparison diminishes our idea of our own" (T 2.2.8.12, p. 377).

Malice and envy in these senses are kinds of pleasure and pain arising from a comparison of actual, present circumstances. There is another closely related sense of malice to be found in the *Treatise*, however, and that is the *desire* for this particular pleasure: "malice is the unprovok'd desire of producing evil to another, in order to reap a pleasure from the comparison" (T 2.2.8.12, p. 377). This understanding of malice as a *desire*, rather than a kind of pleasure, surely fits better with current usage. Nor were things different in the eighteenth century: Johnson's dictionary defines malice as "[i]ll intention to any one; desire of hurting" (1755); and Hutcheson similarly defines it as "the Desire of [another's] Misery" (1728, p. 53). But Hume seems to be indifferent between using the word to refer to the pleasure of a comparison in our favour, and to the desire for this pleasure. In fact, it is doubtful whether he had quite mastered the distinction between these two things, since he writes throughout as though there were only one passion called "malice", never once acknowledging the ambiguity.

This ambiguity was first brought to my attention by Samuel Rickless.² However, it seems to me that Rickless makes the matter more complicated than it needs to be, by insisting that there must be just *one* passion called "malice", and

²Rickless (2013).

therefore that one of Hume's senses must be dismissed as a mistake. On the basis of this assumption, he argues—rightly, to my mind—that the joy should be given priority over the desire, since this is the direct result of the comparison, while the desire is simply an application of self-love to this particular pleasure.³ But I don't see why we shouldn't just say that “malice” is ambiguous in the *Treatise*, referring to two distinct (but closely related) passions.

It is important to stress how *specific* the malicious desire is in the *Treatise*, since this specificity is unusual (and not to be found in Hutcheson, for example). Rather than a general desire for the suffering of another, it is very precisely a desire for the particular pleasure of comparison. The malicious person, for Hume, does not seek another's misery for its own sake, but rather for the sake of feeling better about themselves by contrast: to repeat, “malice is the unprovok'd desire of producing evil to another, *in order to reap a pleasure from the comparison*” (T 2.2.8.12, p. 377; my emphasis). Hume's conception of the malicious desire is thus consistent with the hedonism of the *Treatise*. Malice is not an object-directed desire for the unhappiness of another for its own sake; their unhappiness gives us a pleasure from comparison, and we desire the former only as a means to the latter.

Hume examines the principle of comparison in depth during his discussion of envy and malice (T 2.2.8, pp. 372-80). Earlier in the *Treatise*, however, he had referred ahead to this account, while deriving ambition from the same origin:

Comparison is in every case a sure method of augmenting our esteem of any thing. A rich man feels the felicity of his condition better by opposing it to that of a beggar. But there is a peculiar advantage in power, by the contrast, which is, in a manner, presented to us, betwixt ourselves and the person we command. The comparison is obvious and natural: The imagination finds it in the very subject: The passage of the thought to its conception is smooth and easy. And that this circumstance has a considerable effect in augmenting its influence, will appear afterwards in examining the nature of malice and envy. (T 2.1.10.12, pp. 315-6)

³Rickless (2013, pp. 340-1).

Though the word “ambition” does not appear in this passage, this is the name that Hume gives to the desire for power.⁴ And in referring back to this passage later on (in the discussion of pity and malice), the word is explicit: “I have observ’d in considering the nature of ambition, that the great feel a double pleasure in authority from the comparison of their own condition with that of their slaves” (T 2.2.8.14, p. 378).

Finally, comparison also plays a role in Hume’s account of the causes of respect and contempt, in the following way. The good qualities of another give rise to love, and their bad qualities give rise to hatred, according to the double relation of sentiments and ideas already examined. But good qualities may also give rise to humility, and bad qualities to pride, by virtue of a comparison with ourselves. The love and humility then mix together to form respect, as the hatred and pride mix together to form contempt (T 2.2.10, pp. 389-93).

4.3. Sympathy

Towards the end of part 1 of Book 2 of the *Treatise*, Hume introduces and endeavours to explain “that propensity we have to sympathize with others, and to receive by communication their inclinations and sentiments, however different from, or even contrary to our own” (T 2.1.11.2, p. 316). On Hume’s story, there are two stages to this communicative process. First, someone else’s passion gives rise to an idea of that passion in our minds (typically because we are in contact with them, and observe the outward signs of their emotional state); and second, that idea is then converted into the passion itself:

When any affection is infus’d by sympathy, it is at first known only by its effects, and by those external signs in the countenance and conversation, which convey an idea of it. This idea is presently converted into an impression, and acquires such a degree of force and vivacity, as to become the

⁴E.g. “Strength is a kind of power; and therefore the desire to excel in strength is to be consider’d as an inferior species of *ambition*” (T 2.1.8.4, p. 300); “If I be void of ambition, power gives me no enjoyment” (M App2.12, p. 301).

very passion itself, and produce an equal emotion, as any original affection.

(T 2.1.11.3, p. 317)

Of these two stages, it is the second that is the most important, and Hume treats the conversion of an idea into a sentiment as an instance of sympathy even when there is no real passion in the mind of the other person to get the communicative process underway.⁵ For example, Hume counts our emotional responses at the theatre as instances of sympathy, when the ideas to be converted are obviously of fictitious emotions (T 2.2.7.3, p. 369-70). And he says that we sympathise with real people in fortunate or unfortunate situations, even when they themselves have no (strong) feelings about the matter:

[W]hen a person obtains any honourable office, or inherits a great fortune, we are always the more rejoic'd for his prosperity, the less sense he seems to have of it, and the greater equanimity and indifference he shews in its enjoyment. In like manner a man, who is not dejected by misfortunes, is the more lamented on account of his patience; and if that virtue extends so far as utterly to remove all sense of uneasiness, it still farther encreases our compassion. (T 2.2.7.5, p. 370)

Finally, Hume also allows for a sympathy with the anticipated future feelings of others: "Sympathy being nothing but a lively idea converted into an impression, 'tis evident, that, in considering the future possible or probable condition of any person, we may enter into it with so vivid a conception as to make it our own concern; and by that means be sensible of pains and pleasures, which neither belong to ourselves, nor at the present instant have any real existence" (T 2.2.9.13, pp. 385-6).

The conversion itself, meanwhile, is supposed to take place in something like the following manner. Ideas and impressions differ only in their relative force and vivacity (T 2.1.11.7, p. 319), so all that is needed to convert an idea of a passion into the passion itself is an increase in this force. The idea of someone

⁵For a closer examination of the first stage, see Pitson (1996).

else's passion receives its additional vivacity from an association with the ever-present impression of self, in proportion to the strength of the association (which is why we sympathise more with people who are more closely related to us): "The stronger the relation is betwixt ourselves and any object, the more easily does the imagination make the transition, and convey to the related idea the vivacity of conception, with which we always form the idea of our own person" (T 2.1.11.5, p. 318).

In addition to his evident pride in his associative explanation of sympathy, the Hume of the *Treatise* was also extremely excited by the explanatory potential of sympathy itself. When he first introduces the principle, he writes that "[n]o quality of human nature is more remarkable, both in itself and in its consequences" (T 2.1.11.2, p. 316). And by way of immediate advertisement, he offers one such consequence, namely the uniformity of character among people of the same nation:

To this principle we ought to ascribe the great uniformity we may observe in the humours and turn of thinking of those of the same nation; and 'tis much more probable, that this resemblance arises from sympathy, than from any influence of the soil and climate, which, tho' they continue invariably the same, are not able to preserve the character of a nation the same for a century together. (T 2.1.11.2, pp. 317-6)

But the headline application, given in the section in which the principle is first introduced and explained, is to the passion of *vanity*, the desire for fame or respect.

This desire does not arise from any "original instinct" (T 2.1.11.11, p. 321), Hume argues, but is instead derived from our desire for the pleasant feeling of pride. The respect of other people is a source of pride—or at least helps to reinforce our existing pride—for two reasons. First, it literally becomes pride when communicated to us through sympathy; and second, it confirms our own good opinion of ourselves through an argument from authority:

[N]othing is more natural than for us to embrace the opinions of others... both from *sympathy*, which renders all their sentiments intimately present to us; and from *reasoning*, which makes us regard their judgment, as a kind of argument for what they affirm. These two principles of authority and sympathy influence almost all our opinions; but must have a peculiar influence, when we judge of our own worth and character... [B]eing conscious of great partiality in our own favour, we are peculiarly pleas'd with any thing, that confirms the good opinion we have of ourselves, and are easily shock'd with whatever opposes it. (T 2.1.11.10, p. 321)

The principle of authority is only mentioned briefly, however, and in this one paragraph; in the rest of the section it is Hume's favourite principle of sympathy that receives all of the attention.

Though Hume's emphasis on sympathy in this context was new, he appears to have borrowed the core of his explanation of vanity from Mandeville:

Nature has given [Creatures] an Instinct, by which every Individual values itself above its real Worth; this in us, I mean, in Man, seems to be accompany'd with a Diffidence, arising from a Consciousness, or at least an Apprehension, that we do over-value ourselves: It is this that makes us so fond of the Approbation, Liking and Assent of others; because they strengthen and confirm us in the good Opinion we have of ourselves. (1729, p. 134).

At any rate this shows that Mandeville said it first. The striking similarity in the wording additionally suggests that this is where Hume got it from. As with Hume's derivations of the malicious and compassionate desires from comparison and sympathy respectively, this reduction of vanity is consistent with his early hedonism (which is unsurprising, given the similarities between Hume and Mandeville on this point). We do not pursue fame for its own sake, but merely as a source or confirmation of pride. And pride is pursued, of course, simply because it is pleasant.

4.4. More applications of sympathy

Having introduced the principle of sympathy at T 2.1.11, and used it to account for vanity, Hume then goes on to apply it liberally throughout the rest of the *Treatise*. Indeed, so enthusiastic was he in his application of this principle that one commentator has claimed there are simply “too many [uses] to bother counting”.⁶ Undeterred, I have counted eight, including the two already examined in the previous section. These are, in the order in which they appear:

1. The uniformity of character among people of the same nation (T 2.1.11.2, pp. 316-7).
2. Vanity, or the love of praise and fame (T 2.1.11, pp. 316-24).
3. The love of family relations (T 2.2.4, pp. 351-7).
4. Respect for the rich and contempt for the poor (T 2.2.5.1-14, pp. 357-63).
5. The beauty of utility (T 2.2.5.16-20, pp. 363-5).
6. Pity or compassion (T 2.2.7.2-6, pp. 369-71).
7. Our emotional responses at the theatre (T 2.2.7.3, pp. 369-70).
8. Moral approbation and disapprobation (T 3.2.2.24, 3.3.1.10-11, pp. 499-500, 577-9).

Perhaps the most obvious and straightforward case is number 6 in this list: pity or compassion.⁷ For this isn't really an *application* of the principle of sympathy, so much as an instance of the general phenomenon. When the sentiment being communicated through sympathy is *sorrow*, the resulting sadness at the other end is called *pity*:

⁶Debes (2007a, p. 313).

⁷Hume uses these terms interchangeably in the *Treatise*, and though “compassion” is the word used in the section titles (T 2.2.7, p. 368; T 2.2.9, p. 381), “pity” is more common in the main text. By contrast, “compassion” is more common in the moral *Enquiry* and the *Dissertation*, with “pity” not actually being used in the *Dissertation* at all. I doubt, however, that there is anything significant to be read into this.

'Twill be easy to explain the passion of *pity*, from the precedent reasoning concerning *sympathy*. We have a lively idea of every thing related to us. All human creatures are related to us by resemblance. Their persons, therefore, their interests, their passions, their pains and pleasures must strike upon us in a lively manner, and produce an emotion similar to the original one; since a lively idea is easily converted into an impression. If this be true in general, it must be more so of affliction and sorrow. These have always a stronger and more lasting influence than any pleasure or enjoyment. (T 2.2.7.2, p. 369)

Immediately following this paragraph, Hume also cites sympathy as responsible for our sharing in the feelings—both positive and negative—of the characters in a tragic drama (T 2.2.7.3, p. 369-70). For clarity, I have counted this as a distinct application in my list (number 7); but there is nothing particularly special about this special case, beyond the fictional nature of the feelings that are communicated.

As we saw in §4.2 above, a favourable comparison of ourselves with someone else gives rise to malice, a *joy* in another's *misery*; while an unfavourable comparison gives rise to envy, a *uneasiness* in another's *happiness*. The sympathetic communication of a negative passion, as we have just seen, gives rise to pity, an *uneasiness* in the *sufferings* of another. These three passions set up an obvious pattern, which would be completed by a *joy* in the *happiness* of another, arising from the communication of a positive passion.

There is no reason why positive passions should not be communicated in just the same way as negative ones. For the most part, Hume does focus on the negative cases, even saying—as we saw above—that the effect is more pronounced in the case of “affliction and sorrow”, because “[t]hese have always a stronger and more lasting influence than any pleasure or enjoyment” (T 2.2.7.2, p. 369). But he does give at least one example of the happy counterpart to pity, which I already quoted in the previous section to illustrate a different point: “when a person obtains any honourable office, or inherits a great fortune, we... [rejoice] for his prosperity” (T 2.2.7.5, p. 370). He also maintains that we sympathise with “the

fictitious joy as well as every other passion” at the theatre (T 2.2.7.3, p. 369), and the communication of a positive feeling is important in accounting for our respect for the rich, as we will see shortly.

Though there is thus no doubt that Hume believed in the positive counterpart to pity, it is not obvious what name should be given to this passion, since Hume never labels it explicitly. At the start of Book 2, however, Hume’s list of indirect passions includes “envy, pity, malice, [and] generosity” side by side (T 2.1.1.4, p. 277). Though this is scant evidence to go on, my tentative conjecture is that “generosity” is therefore the word that we are looking for. Generosity, while hardly featuring at all in Book 2 or the *Four Dissertations* (at least not by name), is nevertheless frequently mentioned in Book 3 and the moral *Enquiry*. I do not know whether my conjecture would have any bearing on the interpretation of Hume’s moral philosophy, or whether any such interpretation would give us grounds for abandoning the conjecture (or simply acknowledging an ambiguity in the term). Nor do I have the space here to do anything more than observe this interesting point.

We also saw in §4.2 that Hume’s use of “malice” is ambiguous between the pleasure of comparison, and the desire for this pleasure. “Pity” is ambiguous in exactly the same way, referring sometimes to a communicated sorrow (as above), and sometimes to a corresponding desire to alleviate the cause of this pain: “pity is a desire of happiness to another, and aversion to his misery; as malice is the contrary appetite” (T 2.2.9.3, p. 382). It may be noted that this compassionate desire, with its origin from the sympathetic communication of passions, is likewise consistent with Hume’s early hedonism and egoism. Self-love is a desire for our own pleasure and the avoidance of our own pain; since other people’s pleasure and pain can be a cause of our own (through sympathy), we thus have a self-interested motive to help others.

As in the case of malice, it is thanks to Rickless that I became aware of this ambiguity. Rickless’s treatment of pity is exactly analogous to his treatment of malice: he insists that there must be just one passion of that name, and hence

that one of Hume's uses must be a mistake; and he gives preference to the sorrow over the desire, since it is the sorrow that is the direct consequence of sympathy.⁸ Again, I agree that priority should be given to the sorrow, and for the same reason, but I do not see why we should not simply allow that there are two (closely related) passions of the same name.

The third application of sympathy in the *Treatise* is to the love of family relations, and is an addendum to the theory of the double relation. According to that theory, recall, our love of others is prompted by some pleasing quality that they possess. But we love our relations, says Hume, independently of their character, and even when they do not have any particularly endearing features (T 2.2.4.2, pp. 351-2). There are two parts to Hume's explanation for this. The first part is an idea borrowed from Du Bos, namely that any strong emotional stimulation, even of a negative sort, is often preferable to boredom.⁹ By itself, this explains "that continual search after amusement in gaming, in hunting, in business; by which we endeavour to forget ourselves, and excite our spirits from the languid state, into which they fall, when not sustain'd by some brisk and lively emotion" (T 2.2.4.4, p. 352). The second part is sympathy, thanks to which the company of other people produces—through the communication of their sentiments—an agreeable enlivening of the spirits. We sympathise most, of course, with those closely related to us, making their company particularly stimulating. Thus we have the requisite sentimental input for the double relation to work on, and produce our love for these people.

Hume's fourth use of sympathy in the *Treatise* is another addendum to the theory of the double relation. Ordinarily, the causes that excite our love of other people are things related to the person in question that are also *immediately agreeable* to us; thus we have some input for the association of sentiments as

⁸Rickless (2013, pp. 340-1).

⁹Du Bos (1719, p. 5). No doubt Du Bos was not alone in making the observation; Norton and Norton, in their critical edition of the *Treatise*, cite Pascal as a likely source, before mentioning Du Bos as well (2007, volume 2, p. 846). But the link with Du Bos is the more important here, given Hume's appeal to this idea again in *Of Tragedy*, when he mentions Du Bos explicitly (Tr 3, p. 217). This may be considered a small taste of things to come in Part 2, when I turn more explicitly to the unity of the *Four Dissertations*. See chapter 11.

well as of ideas. But Hume also wants to account for the love or esteem that we have for people on account of their wealth and power, even when we expect no particular favour from them. Without this expectation, their wealth and power can be no direct source of pleasure to us; but it can please us by a sympathetic communication of the immediate pleasure that it brings *them*. Thus we have the sentiment as well as the idea, and the double relation can thereafter proceed as before (T 2.2.5.1-14, pp. 357-62).

As it is with respect for the rich and powerful, so it is with contempt for the poor and weak: from them we receive a sympathetic communication of negative emotions, which feeds into the double relation mechanism on the negative side, resulting in hatred or contempt. Hume doesn't spell this out explicitly in the *Treatise*, but it is obvious, I trust, how the exactly parallel story would run. At the start and end of his discussion Hume mentions the negative case as well (T 2.2.5.1, 2.2.5.14, pp. 357, 362), and later on he makes it even clearer that his account was intended to apply to both:

I have endeavour'd to prove, that power and riches, or poverty and meanness; which give rise to love or hatred, without producing any original pleasure or uneasiness; operate upon us by means of a secondary sensation deriv'd from a sympathy with that pain or satisfaction, which they produce in the person, who possesses them. From a sympathy with his pleasure there arises love; from that with his uneasiness, hatred. (T 2.2.9.11, p. 384)

This comment appears after Hume has argued that pity and malice give rise to love and hatred respectively, because of their similar "impulses or directions" (T 2.2.9.2, p. 381): pity and love both give rise to a desire for the happiness of the person in question, as malice and hatred both give rise to a desire for their misery.

There is a tension here, and Hume himself draws explicit attention to it: sometimes pity gives rise to love, because of the similar tendencies of these passions (both prompt a benevolent desire), but sometimes it gives rise to hatred or contempt, because of the resemblance of sentiment (both are negative emotions).

Hume argues that the difference is accounted for by variations in the force of the sympathetically communicated passion. If it is strong, then it extends beyond the immediate suffering, to take in the prospects of the person's future happiness or misery as well, and the similar tendency with love or benevolence wins. If it is weak, then it remains only with the present sorrow, and hatred or contempt is the result (T 2.2.9.12-9, pp. 385-9).

The remaining applications of sympathy in the *Treatise* are to the beauty of utility and moral approbation. These (closely related) applications will be examined in the next chapter (§§5.4 and 5.5).

4.5. The indirect passions

In the *Dissertation on the Passions*, Hume says that the indirect passions (though he does not call them by that name) are “of a more complicated nature” than the direct (P 2.1, p. 7). In the *Treatise*, he says that they “proceed from the same principles [as the direct], but by the conjunction of other qualities” (T 2.1.1.4, p. 276). We are now in a position to see clearly what Hume means by this: the other qualities in question are precisely the three principles examined above, namely, comparison, sympathy, and the double relation. The distinction between the direct and the indirect passions is based on their causal origins, and the defining feature of the indirect passions is that they have a more complicated aetiology than the direct.

At the start of *Treatise* Book 2, Hume enumerates the indirect passions thus: “under the indirect passions I comprehend pride, humility, ambition, vanity, love, hatred, envy, pity, malice, generosity, with their dependants” (T 2.1.1.4, pp. 276-7). The dependants that Hume has in mind, as becomes clear later on, are respect and contempt (T 2.2.10, pp. 389-393) and “the amorous passion” (T 2.2.11, pp. 394-6). This last is a mixture of love, complacency in beauty, and the bodily appetite for generation (see also P 3.14, p. 20). The mixture of these things is accounted for by the association of sentiments; they are all pleasant feelings. Respect and contempt, as we have seen, are accounted for in general by

means of comparison and the double relation (recall §4.2), and in the particular case of respect for the rich and contempt for the poor, sympathy also plays a role (§4.4). We have also seen above how all of the other passions in Hume's list are accounted for by one or other of these three principles.

This reconstruction of the category of indirect passions ought, therefore, to be pretty well established. Not only does it make sense of Hume's definition, it also agrees—to the letter—with his explicit list in the *Treatise*. Amongst commentators, however, there have been two kinds of opposition to it; one principled, and the other careless.

As I noted in §4.1 above, the double-relation passions (pride, humility, love, and hatred) receive more of Hume's attention—both in *Treatise* Book 2 and in the *Dissertation*—than all of the other passions put together. It is presumably because of this emphasis that a careless tendency has arisen to shrink the category of the indirect passions to the double-relation passions alone, ignoring ambition, vanity, pity, generosity, envy, and malice. For example, Cohon has written:

Hume's account [of the passions] in both the *Treatise* and the *Dissertation* concentrates on four passions he groups together under the label "indirect": pride, humility, love, and hatred. (He mentions other indirect passions, but apparently they are all combinations or simulacra of these four.)¹⁰

This mistake presumably explains Cohon's false generalisation that I complained about in the previous chapter (§3.4), when she says that the indirect passions as a whole do not motivate us immediately, when in fact there are several indirect passions that do: ambition and vanity, and also pity and malice (at least in the desire senses). But it *is* a mistake. Respect and contempt, admittedly, are combinations of the double-relation passions, as Cohon says. But ambition, vanity, generosity, pity, malice, and envy are neither combinations nor simulacra, and it is difficult to see how anyone could suppose for a moment that they were. Though alike in being indirect (i.e. in having a more complicated aetiology than the direct passions), they are no more closely related than that.

¹⁰Cohon (2008a, p. 160).

This carelessness has, I fear, become quite common.¹¹ I should confess that I myself was initially misled by Hume’s emphasis on the theory of the double relation into ignoring the indirect passions that do not arise in this way. To my embarrassment, furthermore, I went into print thus blinkered, talking about the indirect passions in an article when I really just meant the double-relation passions.¹² In my defence, nothing in my argument there hung on the point: simply replace every occurrence of “indirect passions” with “double-relation passions”, and all will be well.¹³ The same, unfortunately, cannot be said for Cohon’s article, which is precisely about the indirect passions, and the nature of Hume’s direct/indirect distinction. It is therefore in need of some rather more substantial rewriting.

While Cohon and I simply made a careless mistake, paying insufficient attention to the passions that Hume himself pays less attention to, a recent article by Rickless has presented a clear and well-reasoned case that might be appealed to in our defence. Rickless argues that pity and malice (and hence presumably also generosity and envy) *should* in fact be classified as direct passions rather than indirect, and hence that their inclusion in Hume’s explicit list of indirect passions should be treated as a mistake.¹⁴ If this were right, I would not need to be so hard on Cohon and my earlier self after all. But it seems to me that it is not right.

Rickless has two reasons for reclassifying pity and malice as direct. First, these two passions are special cases of sorrow and joy respectively: a sorrow and a joy in the sufferings of others. But joy and sorrow are both *direct* passions, and so pity and malice ought to be direct as well.¹⁵ This consideration does not seem

¹¹The error goes at least as far back as Kemp Smith (1941/2005, pp. 179-191). More recently, see e.g. Inoue (2003, p. 208), Taylor (2008, pp. 276-7).

¹²Merivale (2009).

¹³In fact, this is what I have done in §§9.4 and 9.5 below; though I have made some other changes there as well.

¹⁴Rickless (2013).

¹⁵As noted in §§4.2 and 4.4, “malice” and “pity” are both ambiguous in the *Treatise*, referring to a joy and sorrow in the sufferings of others respectively, and also to corresponding desires. As he points out himself, however, Rickless’s argument applies to both readings, for desire in general is also a direct passion.

to me to be persuasive. The direct/indirect distinction is not based on intrinsic features of the passions, but on their causal origins. There is thus no contradiction in the idea that joy is a direct passion *in general*, but an indirect passion in the special case of joy in another person's misery; in this case, uniquely, it has a more complicated aetiology involving comparison. Similarly for the general passion of sorrow with its direct origin, and the particular passion of pity, which arises in a more complicated manner through sympathy. It is perhaps a little unfortunate that Hume was not clear on this point; it seems an obvious potential source of confusion that he might have spared a few words to clear up. But since he clearly says that joy and sorrow are direct, and that pity and malice are indirect, we are left to make the best sense of these claims that we can. And on reflection, it seems that we can render them consistent.

Rickless's second consideration is that "[i]f there is anything that counts as a fixed point in [Hume's] theory, it is that indirect passions, unlike direct passions, are either produced or constituted by a double association of ideas and impressions".¹⁶ In other words, the indirect passions just *are* the double-relation passions; precisely Cohon's and my earlier careless thought. If this were right, then it obviously would restrict the class of indirect passions to just the double-relation ones, leaving pity and malice and all of the rest outside the group.

Rickless, unlike Cohon and me, did produce some direct textual evidence for his "fixed point":

That propensity, which unites us to the object, or separates us from it, still continues to operate, but in conjunction with the indirect passions, which arise from a double relation of impressions and ideas. (T 2.3.9.3, p. 439)

The natural interpretation of this sentence, Rickless maintains, entails that *all* of the indirect passions arise from the double relation.¹⁷ I concede that this does indeed seem a plausible reading. On my broader interpretation of the category of indirect passions, Hume should perhaps rather have said "*those* indirect passions,

¹⁶Rickless (2013, pp. 339-40).

¹⁷Rickless (2013, p. 340, n13).

which arise from a double relation”, indicating that there are some that do not arise in this way.

In reply, I can only say that the weight of evidence on the other side seems strongly to overbalance this one passage. There is of course Hume’s explicit inclusion of pity and malice (and envy and generosity) in his list of indirect passions, which Rickless acknowledges, but argues should be treated as a mistake. There is also the inclusion of ambition and vanity in the list, which Rickless doesn’t mention, but which he must also suppose is an error. But perhaps most striking of all are the cases of respect and contempt, the explanations of which, as we have seen, feature *all three* of Hume’s principles: sympathy, comparison, and the double relation. By drawing a line between sympathy and comparison on the one hand, and the double relation on the other, Rickless would seem to render respect and contempt unclassifiable on Hume’s system. Not to mention that sympathy also plays a role in supporting the double relation theory in the case of the love of family relations.

Existing discussions of Hume on the indirect passions have given vastly more attention to the theory of the double relation than to sympathy and comparison, and correspondingly more attention to pride, humility, love, and hatred than to the other indirect passions. This is understandable: Hume’s own discussions exhibit precisely the same bias. But we must not let this emphasis distort our understanding of the bigger picture. As I hope to have made clear, the category of the indirect passions is larger than just that of the the double-relation passions. Rickless’s fixed point needs to be dislodged.

Chapter 5

Original Instincts

The previous chapter introduced Hume's three main psychological principles concerning the causal origins of the (indirect) passions: sympathy, comparison, and the theory of the double relation. When these things fail him, however, Hume's other theoretical device is to appeal to an *original instinct*, a brute, inexplicable feature of human nature. These original instincts are, in particular, the foundation of our object-directed desires. With the rise in importance of these desires after the *Treatise*, therefore, (as argued for in chapters 2 and 3) we should expect a corresponding rise in original instincts at the expense of the other three principles. In this chapter we will see that this is exactly what happens: the double relation remains unchallenged, but comparison takes a hit, and sympathy suffers most of all. Many of the things that sympathy earlier explained are, it turns out, accounted for just as well or even better by original instincts.

5.1. Object-directed desires and original instincts

For the later Hume, as for Butler, the vast majority of our desires are for objects directly, independently of their contribution to our happiness. The only exceptions are the various applications of self-love to particular pleasing objects. Even self-love itself is a direct desire for happiness or pleasure, rather than for the happiness or pleasure that these things bring. Nevertheless, it will be useful to have in hand Hume's explicit examples of object-directed desires.

In the Butler paragraph from the *Treatise*, Hume mentions the desires of happiness to our friends and of punishment to our enemies (T 2.3.9.8, p. 439), named as “friendship” and “revenge” in the corresponding passage in the *Dissertation* (P 1.2, p. 3). In the same vein, he often speaks of “benevolence” and “anger” (e.g. T 2.2.6.3, p. 367; T 2.2.9.3, p. 382), or “gratitude” and “resentment” (e.g. T 2.3.3.8, p. 417). Sometimes the implication is that “gratitude” is a response to services rendered (T 2.2.3.3, p. 348), and “revenge” or “resentment” a response to injuries done to us (T 2.3.3.9, p. 418; T 2.2.7.1, p. 369); but in any case, the core ideas—desiring another person’s happiness or misery, for whatever reason—are the same for all of these passions.

In the introduction to the *Natural History*, Hume lists, alongside gratitude and resentment, “self-love, affection between the sexes, [and] love of progeny” (N Intro.1, p. 33). Love of progeny is named again (as the “affection of parents to children”) in the *Dissertation on the Passions*, in an important footnote that I will return to in §5.3 below (P 3.3, n4, p. 18). In the moral *Enquiry* (M App2.12, p. 301), the explicit examples are “anger” (again), alongside “vanity” (the desire for praise or fame) and “ambition” (the desire for power). Most of these examples would have had a particular resonance in the context of the egoism debate: benevolence and parental love for being such striking potential counterexamples to the egoist thesis,¹ and ambition and vanity for their roles in Hobbes’s and Mandeville’s respective accounts of (superficially) altruistic motivation (recall §§1.2 and 1.4).

As to the *origins* of these object-directed desires, Hume has very little to say, because he believes that there is very little that *can* be said. They all arise from what he calls an “original instinct” or “primary impression” of our human nature; as when, in the introduction to the *Natural History of Religion*, he contrasts them with the belief in invisible, intelligent power:

It would appear, therefore, that this preconception springs not from an original instinct or primary impression of nature, such as gives rise to self-

¹See Hutcheson (1725, pp. 112-4), Butler (1726, p. 76).

love, affection between the sexes, love of progeny, gratitude, resentment...

The first religious principles must [rather] be secondary[.] (N Intro.1, p. 33)

This effectively just means that they are brute, inexplicable facts of our frame and constitution. Hume says as much explicitly, we may recall, in the Butler paragraph from the *Treatise*: “Besides good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable” (T 2.3.9.8, p. 439). Object-directed desires, or at least the original instincts on which they are founded, had for Hume the same status in psychology as elasticity and gravity had in physics (see E 4.12, pp. 30–1). He was certainly not inclined—as Butler and Hutcheson had been before him (recall §1.3)—to explain them in terms of final causes, as being planted in our nature by a kind and intelligent creator, with a view to the happiness of his creation. Our motivational make-up is for Hume simply an observable fact, that neither has nor requires any further explanation.

The fundamental point to appreciate here, a point at the heart of everything that I have to say in this chapter, is that original instincts or object-directed desires are in implicit tension with the principles of sympathy and comparison. To illustrate this, consider hedonistic egoism, the view that our only original motivational instinct is for our own pleasure (and the avoidance of pain), and that all other desires are at bottom reducible to this one. The obvious potential counterexamples to this thesis are benevolent desires for other people’s happiness, and (slightly less obviously) malicious desires for other people’s suffering. For if it is only our *own* pain and pleasure that ultimately moves us, how is it that we can be guided—either way—by other people’s?

Sympathy and comparison, as we saw in the previous chapter, provide the egoist with potential answers to this question, in keeping with their selfish hypothesis. Sympathy gives us a joy in the happiness of others, and a sorrow in their misery, while comparison gives us a pleasure in their pain, and a pain in their pleasure. In these two ways, then, the felicity of others has an impact—either positive or negative—on our own well-being, and so our selfish desire for

the latter can explain our (derivative) desires for and against the former. But for anti-egoists, like Butler and the later Hume, who acknowledge *multiple* original desires, there is an obvious and more straightforward alternative explanation to hand: we have benevolent and malicious desires originally implanted in our nature. And it is through the satisfaction or thwarting of these desires that our pleasure or pain in the happiness and misery of others arises, rather than through sympathy or comparison.

The situation here is not black and white. The egoists certainly have no business appealing to any original instincts beyond self-love: it is precisely their view that self-love is the only one, from which all other desires are derived. But the anti-egoists are free to admit sympathy and comparison into their psychological theories, alongside the several original instincts that they posit. Suppose, therefore, that an egoist—let us call him David—was formerly very keen on these two principles, but was subsequently persuaded of Butler’s view. We might expect David to be reluctant, in later life, to abandon these principles altogether, but to place considerably less weight on them, and correspondingly more weight on object-directed desires.

Lo and behold, *this is exactly what Hume did*. We will see the details below, but as a quick preliminary point, here are some crude but very telling statistics: The word “sympathy” appears 77 times in Book 2 of the *Treatise*, and 77 times in Book 3. By contrast, it appears only 9 times in the *Four Dissertations* (3 times in the *Dissertation on the Passions*, 4 times in *Of Tragedy*, and once in each of the other two), and 26 times in the moral *Enquiry*. Even taking into account the fact that the later works are (slightly) shorter overall, that is still a staggering reduction of over 70 per cent. “Benevolence”, meanwhile, the most important object-directed desire in the context of the egoism debate, appears just 17 times in *Treatise* Book 3, but 36 times in the moral *Enquiry*. And in addition to this crude word count, let us not forget that benevolence and anti-egoism are the very *first* things that Hume discusses in this later work, following the initial introductory section.

The very evident cooling of Hume's enthusiasm for the principle of sympathy after the *Treatise* is, I take it, one of the major puzzles regarding his intellectual development. I would not be so bold as to claim to have a definite and complete explanation of this curious fact. There might have been several reasons, and unfortunately Hume left us very little direct textual evidence to work with. But his conversion to Butler's object-directed motivational psychology is certainly a *possible* explanation, and such indirect textual evidence as we have seems to me to make it a very plausible one, as I will argue below.

5.2. Later traces of sympathy and comparison

The first thing to note is that, notwithstanding the striking drop in enthusiasm, Hume does not abandon his principle of the sympathetic communication of passions altogether. Its first application in the *Treatise*, we may recall, was to the uniformity of character among people of the same country. Hume returned to this topic in a 1748 essay, *Of National Characters*, in which he argues that character is determined solely by *moral* causes, i.e. "circumstances, which are fitted to work on the mind as motives or reasons", and not at all by *physical* causes, i.e. "those qualities of the air and climate, which are supposed to work insensibly on the temper" (NC 2, p. 198).

Those who believe in the influence of physical causes have an obvious explanation for the uniformity of character among compatriots: they are in the same physical environment. Given Hume's position, this explanation was obviously not available to him, but sympathy continued to furnish him, after the *Treatise*, with a moral alternative:

The human mind is of a very imitative nature; nor is it possible for any set of men to converse often together, without acquiring a similitude of manners, and communicating to each other their vices as well as virtues. The propensity to company and society is strong in all rational creatures; and the same disposition, which gives us this propensity, makes us enter deeply into each other's sentiments, and causes like passions and inclina-

tions to run, as it were, by contagion, through the whole club or knot of companions. Where a number of men are united into one political body, the occasions of their intercourse must be so frequent, for defence, commerce, and government, that, together with the same speech or language, they must acquire a resemblance in their manners, and have a common or national character, as well as a personal one, peculiar to each individual...

If we run over the globe, or revolve the annals of history, we shall discover every where signs of a sympathy or contagion of manners, none of the influence of air or climate. (NC 9-10, pp. 202-3)

Here we have the very same idea of sympathy as a *communication* (or “contagion”) of passions, and a repeat of its first application in the *Treatise*. What we are lacking is the detailed associative account of how this communication is supposed to come about.

In the *Dissertation on the Passions*, however, we find some more of Hume’s earlier applications of this principle being repeated. And although the detailed associative account is never given explicit attention in its own right, Hume says just enough to indicate that he still held to it. Here is his very brief repetition of the earlier explanation of pity or compassion:

Compassion... is an uneasiness in the sufferings of another. It seems to spring from the intimate and strong conception of his sufferings; and our imagination proceeds by degrees, from the lively idea to the real feeling of another’s misery. (P 3.7, p. 19)

This shows that he still thought of the communicative process as involving the conversion of an idea into an impression or sentiment. And here is his succinct restatement of the *Treatise* account of the love of family relations:

A person, who is related to us, or connected with us, by blood, by similitude of fortune, of adventures, profession, or country, soon becomes an agreeable companion to us; because we enter easily and familiarly into his sentiments and conceptions: Nothing is strange or new to us: Our imagination, passing from self, which is ever intimately present to us, runs smoothly along the

relation or connexion, and conceives with a full sympathy the person, who is nearly related to self. He renders himself immediately acceptable, and is at once on an easy footing with us: No distance, no reserve has place, where the person introduced is supposed so closely connected with us. (P 3.4, p. 18)

This shows that he still considered the ever-present impression of self to be the source of the extra force and vivacity that turns the idea into the real feeling.

The *Treatise* account of the love of family relations, recall, also depended on an idea borrowed from Du Bos, namely our general desire for emotional stimulation. Our relations, whose passions affect us particularly deeply through sympathy, are stimulating in this way to a particularly high degree. Although there is no explicit repetition of this point in the *Dissertation on the Passions*, it does recur in *Of Tragedy*, albeit in a different context:

L'Abbe DUBOS, in his reflections on poetry and painting, asserts, that nothing is in general so disagreeable to the mind as the languid, listless state of indolence, into which it falls upon the removal of all passion and occupation. To get rid of this painful situation, it seeks every amusement and pursuit; business, gaming, shews, executions; whatever will rouse the passions, and take its attention from itself...

It is impossible not to admit this account, as being, at least in part, satisfactory. You may observe, when there are several tables of gaming, that all the company run to those, where the deepest play is, even though they find not there the best players. The view, or, at least, imagination of high passions, arising from great loss or gain, affects the spectator by sympathy, gives him some touches of the same passions, and serves him for a momentary entertainment. (Tr 3-4, p. 217)

In fact, Hume did change his mind with regard to tragedy, on a matter closely related to this point; but that is a topic for further down the line (see chapter 11). For the present, we may note that the point itself was not abandoned.

Hume also retained his earlier account the role of comparison in producing envy and malice:

The comparison of ourselves with others seems to be the source of envy and malice. The more unhappy another is, the more happy do we ourselves appear in our own conception. (P 3.8, p. 19)

However, in the same way that the later references to the sympathetic communication of passions are sparse and oblique, these two sentences are a pale imitation of the 20-paragraph section in the *Treatise* on envy and malice (T 2.2.8, pp. 372-80), which examines in great detail the effect that comparison has on the passions.

On the one hand, the fact that Hume retained these principles in the *Dissertation* makes the difference between the earlier and the later work in this respect less radical than it might otherwise have been. The present component of my Difference Thesis would certainly have been more dramatic if he had abandoned them altogether, or explicitly renounced them. On the other hand, however, it makes such differences as there are all the more puzzling. Though not entirely absent, these formerly central aspects of Hume's treatment of the passions are, in the later work, peripheral at best. Something ought to explain this very curious fact.

5.3. Ambition, vanity, and parental love

The applications of sympathy and comparison examined in the previous section show that these principles are not altogether absent from Hume's later work, although they do seem to be much less important. In fact, however, this is just the start. In three crucial cases—ambition, vanity, and parental love—original instincts and object-directed desires explicitly muscle in on explanatory territory that was formerly occupied by sympathy and comparison. As I pointed out at the end of §5.1 above, original instincts are in implicit tension with sympathy and comparison, and so the rise in the former is at least a *possible* explanation for the fall of the latter. I now start to present my case that it is not just possible, but probable.

When Hume gave his account of the love of family relations in the *Treatise*, based on the sympathetic communication of passions between people with a close

relationship or connection to one another, he was explicit in including parental love within the remit of this mechanism: “Thus the relation of blood produces the strongest tie the mind is capable of in the love of parents to their children, and a lesser degree of the same affection, as the relation lessens” (T 2.2.4.2, p. 352). As we saw in the previous section, Hume repeats his general account of the love of relations in the *Dissertation*. But there is a crucial difference in this later work that I passed over above: parental love is now explicitly singled out in a footnote as being accounted for *instead* by an original instinct:

The affection of parents to children seems founded on an original instinct. The affection towards other relations depends on the principles here explained. (P 3.3, n4, p. 18)

This little footnote is another one of those small changes in the *Dissertation* with large implications. For it establishes beyond any doubt that Hume saw sympathy and original instincts as *alternative* explanations, and that in at least one case he changed his mind from the former to the latter.

This crucial footnote is of course all of a piece with what we have seen already. Parental love is one of Hume’s later explicit examples of an object-directed desire. It ought therefore to be accounted for by an original instinct. And with this original instinct admitted, there is no longer any need to account for the phenomenon in terms of sympathy. Sympathy *may* remain, as I said in §5.1 about the general case; there is no reason why there shouldn’t be some overdetermination. But insofar as parental love is an *object-directed* desire, accounted for by an original instinct, sympathy can be playing no role. Any additional fondness for one’s children generated through sympathy will be a derivative of self-love. Why do parents take a pleasure in their children’s happiness? The communication of passions is one possible explanation: the children’s pleasures are conveyed to their parents via the corresponding ideas. An original and instinctive love of one’s offspring is another: parents simply *want* their children to be happy.

We saw in the previous chapter how ambition and vanity were explained, in the *Treatise*, in terms of comparison and sympathy respectively. In a nutshell,

power over others gives us a pleasure from the favourable comparison of ourselves with them, while the respect of others, communicated to us through sympathy, produces or at least reinforces our pride. As noted in §5.1 above, however, ambition and vanity are two of Hume's explicit examples of object-directed desires in the moral *Enquiry*. Here, now, is the relevant passage in full:

[T]here are mental passions, by which we are impelled immediately to seek particular objects, such as fame, or power, or vengeance, without any regard to interest; and when these objects are attained, a pleasing enjoyment ensues, as the consequence of our indulged affections. Nature must, by the internal frame and constitution of the mind, give an original propensity to fame, ere we can reap any pleasure from that acquisition, or pursue it from motives of self-love, and desire of happiness. If I have no vanity, I take no delight in praise: If I be void of ambition, power gives me no enjoyment: If I be not angry, the punishment of an adversary is totally indifferent to me. (M App2.12, p. 301)

As with parental love, these new accounts of ambition and vanity as object-directed desires go against the grain of the *Treatise* accounts based on comparison and sympathy respectively. Hume needn't abandon the earlier accounts altogether, if he is prepared to allow for the possibility of overdetermination. But he needn't keep them either. And insofar as these desires are object-directed, they must arise from an original instinct rather than from comparison or sympathy; for when they arise in these latter two ways, they are merely derivatives of self-love.

Ambition and vanity, as I noted in §5.1 above, were particularly important passions in the debate about the foundation of morals, since they had been put forward by Hobbes and Mandeville respectively as the major sources of altruistic motivation. In the concluding section of the moral *Enquiry*, indeed, Hume explicitly mentions them as forming no part of his own theory concerning the origin of morals (M 9.5, p. 271), which appeals instead to the principle of humanity or benevolence. This is of course a clear and unambiguous statement of his alignment with the sentimentalists or Stoics, in opposition to the egoists or Epicureans.

Insisting on the object-directed nature of ambition and vanity, furthermore, was an implicit attack on the Epicurean tradition: if even these two passions cannot be accounted for in terms of self-love, then the most prominent egoist accounts of altruism involve object-directed desires *anyway*. Thus there can be no good reason for refusing to acknowledge the more straightforward explanation: that we simply have object-directed desires for the happiness of others.

The case of vanity is in fact somewhat complicated by the *Dissertation on the Passions*, for in this work Hume reverts to his old *Treatise* account, apparently forgetting the change that he had already made in the moral *Enquiry*. In the *Treatise*, when deriving vanity from the principle of sympathy, Hume was explicit in denying that original instincts played any role:

Among these phaenomena we may esteem it a very favourable one to our present purpose, that tho' fame in general be agreeable, yet we receive a much greater satisfaction from the approbation of those, whom we ourselves esteem and approve of, than of those, whom we hate and despise. In like manner we are principally mortify'd with the contempt of persons, upon whose judgment we set some value, and are, in a great measure, indifferent about the opinions of the rest of mankind. But if the mind received from any original instinct a desire of fame, and aversion to infamy, fame and infamy wou'd influence us without distinction; and every opinion, according as it were favourable or unfavourable, wou'd equally excite that desire or aversion. (T 2.1.11.11, p. 231)

In the moral *Enquiry*, as we have just seen, he changed his mind, insisting that the mind *does* contain “an original propensity to fame” (M App2.12, p. 301). But then in the *Dissertation on the Passions*, when he was obviously looking at the *Treatise* again and revising this material, he seems to have forgotten what he wrote in the meantime, for he reverts to his former denial of the originality of this passion, and repeats the Mandevillean explanation:

Our opinions of all kinds are strongly affected by society and sympathy, and it is almost impossible for us to support any principle or sentiment, against

the universal consent of every one, with whom we have any friendship or correspondence. But of all our opinions, those, which we form in our own favour; however lofty or presuming; are, at bottom, the frailest, and the most easily shaken by the contradiction and opposition of others... Hence that strong love of fame, with which all mankind are possessed. It is in order to fix and confirm their favourable opinion of themselves, not from any original passion, that they seek the applauses of others. And when a man desires to be praised, it is for the same reason, that a beauty is pleased with surveying herself in a favourable looking-glass, and seeing the reflection of her own charms. (P 2.33, p. 14)

It is difficult to escape the conclusion that Hume was simply careless on this point: in the *Treatise* he explicitly denies that vanity arises from an original instinct (preferring Mandeville's explanation); in the moral *Enquiry* he explicitly says that it does (preferring Butler's story); and in the *Dissertation* he again explicitly denies it (reverting once more to the Mandeville story).

Fortunately, this oscillation is not a devastating problem, either for Hume's own mature view, or for my hypothesis concerning the development of his thought. We need only remember that a commitment to an original instinct or object-directed desire—as it might be, for fame and praise—is not itself *inconsistent* with Mandeville's idea. There is no reason why we should not have *both* an original desire for praise, *and* a derivative desire for the same, as a self-interested means of bolstering our own pride. Hume doesn't actually say this; he seems to assume that it must be one or the other, and to oscillate back and forth between the two. Taking into account both the moral *Enquiry* and the *Dissertation*, however, and doing our best to find in them a coherent view, we may surmise that this hybrid account was, or at least ought to have been, his considered view. And regarding my developmental story, it suffices that the original desire for praise is an additional feature of the later work, doubtless inspired by Butler, that was nowhere to be found in the *Treatise*.

It is also notable about the passage just quoted from the *Dissertation* that, although the *word* "sympathy" appears in the first sentence, as an echo of Hume's

view from the *Treatise*, when we get down to business it is in fact only the principle of *authority* that is doing any of the real work in this later presentation: “the favourable suffrages of the world are regarded *only as authorities*, or as confirmations of our own opinion” (P 2.40, p. 15; my emphasis). In the *Treatise*, we may recall, Hume mentioned the principle of authority only in passing, giving most of the discussion over to sympathy; in the *Dissertation*, it is the other way round.

5.4. Sympathy and benevolence

It is not only in the area of motivation that original instincts take over from sympathy and comparison as an explanatory mechanism, but also in the area of *approbation* (recall Hutcheson’s crucial distinction between motivation and approbation, introduced in §1.1). There are two applications of sympathy in the *Treatise* that I mentioned in the previous chapter but said nothing about, promising to return to them here: moral approbation and disapprobation, and the beauty of utility. In these final two sections I will look at each of these in turn.

It is well known that sympathy plays a central and indispensable role in Hume’s account of the causes of moral approbation and disapprobation in Book 3 of the *Treatise*, first with regard to the “artificial” virtue of justice (T 3.2.2.24, pp. 499-500), and then with regard to the “natural” virtues such as beneficence, charity, generosity, and moderation (T 3.3.1.10-1, pp. 577-9). These virtues all have a “tendency to the good of mankind” (T 3.3.1.9, p. 577), which is precisely what pleases us about them and raises our moral approbation. But:

[A]s the means to an end can only be agreeable, where the end is agreeable; and as the good of society, where our own interest is not concern’d, or that of our friends, pleases only by sympathy: It follows, that sympathy is the source of the esteem, which we pay to all the artificial virtues. (T 3.3.1.9, p. 577)

Hume then proceeds, in the next two paragraphs, to urge the same explanation with regard to the natural virtues as well. As he summarises the point in the conclusion of the book:

Justice is certainly approv'd of for no other reason, than because it has a tendency to the public good: And the public good is indifferent to us, except so far as sympathy interests us in it. We may presume the like with regard to all the other virtues, which have a like tendency to the public good. They must derive all their merit from our sympathy with those, who reap any advantage from them: As the virtues, which have a tendency to the good of the person possess'd of them, derive their merit from our sympathy with him. (T 3.3.6.1, p. 618)

In section 5 of the moral *Enquiry*, however, when Hume repeats an exactly analogous argument, benevolence has taken the place of sympathy:

Usefulness is only a tendency to a certain end; and it is a contradiction in terms, that any thing pleases as means to an end, where the end itself no wise affects us. If usefulness, therefore, be a source of moral sentiment, and if this usefulness be not always considered with a reference to self; it follows, that every thing, which contributes to the happiness of society, recommends itself directly to our approbation and good-will. Here is a principle, which accounts, in great part, for the origin of morality: And what need we seek for abstruse and remote systems, when there occurs one so obvious and natural? (M 5.17, p. 219)

“Good-will” is of course just another word for benevolence. And in a footnote at precisely this point, furthermore, Hume makes his position even clearer:

It is needless to push our researches so far as to ask, why we have humanity or a fellow-feeling with others. It is sufficient, that this is experienced to be a principle in human nature. We must stop somewhere in our examination of causes; and there are, in every science, some general principles, beyond which we cannot hope to find any principle more general. No man is

absolutely indifferent to the happiness and misery of others. The first has a natural tendency to give pleasure; the second, pain. This every one may find in himself. It is not probable, that these principles can be resolved into principles more simple and universal, whatever attempts may have been made to that purpose. But if it were possible, it belongs not to the present subject; and we may here safely consider these principles as original: Happy, if we can render all the consequences sufficiently plain and perspicuous! (M 5.17, n19, pp. 219-20)

Where sympathy once was, we now have an original and unaccountable instinct. It is now *benevolence*, more than sympathy, that accounts for moral approbation and disapprobation: “Thus, in whatever light we take this subject, the merit, ascribed to the social virtues, appears still uniform, and arises chiefly from that regard, which the natural sentiment of benevolence engages us to pay to the interests of mankind and society” (M 5.43, p. 230).

As we have seen, traces of Hume’s earlier commitment to sympathy remain in the *Dissertation on the Passions* and the essay *Of National Characters*. There are also traces in the moral *Enquiry*; benevolence has there taken centre stage, but it has not shunted out sympathy altogether. The latter reappears most unambiguously in Hume’s treatment of our approbation of qualities immediately agreeable to their possessor:

CHEARFULNESS carries great merit with it, and naturally conciliates the good-will of mankind. No quality, indeed, more readily communicates itself to all around...

From this influence of cheerfulness, both to communicate itself, and to engage approbation, we may perceive, that there is another set of mental qualities, which... diffuse a satisfaction on the beholders, and procure friendship and regard. Their immediate sensation, to the person possessed of them, is agreeable: Others enter into the same humour, and catch the sentiment, by a contagion or natural sympathy: And as we cannot forbear loving whatever pleases, a kindly emotion arises towards the person, who communicates so much satisfaction. (M 7.1-2, pp. 250-1)

Who would live amidst perpetual wrangling, and scolding, and mutual reproaches? The roughness and harshness of these emotions disturb and displease us: We suffer by contagion and sympathy; nor can we remain indifferent spectators, even though certain, that no pernicious consequences would ever follow from such angry passions. (M 7.21, p. 257-8)

A full account of what is going on here—explaining, in particular, whether this is just a stylistic echo of the earlier view, or whether sympathy still plays some substantial role in Hume’s later theory—is well beyond the scope of this thesis. My focus is on Hume’s later philosophy of emotion, not his later moral philosophy. But there is no denying the fact that benevolence subsequently moves in on sympathy’s territory in Hume’s moral theory; and it seems probable that this was at least in part the result of Hume’s conversion to Butler’s object-directed motivational psychology.²

5.5. Sympathy and the rules of art

The remaining application of sympathy in the *Treatise* is to the beauty of utility. Hume’s views on this matter are to be understood in the context of a debate between Hutcheson and Berkeley.³ Hutcheson, in his *Inquiry into the Original of Our Ideas of Beauty and Virtue*, argued that our perception of beauty was immediate, that it did not depend on any knowledge or reasoning, and in particular that it had nothing to do with the utility of the beautiful object:

This superior Power of Perception is justly called a Sense, because of its Affinity to the other Senses in this, that the Pleasure does not arise from

²In a recent pair of articles, Remy Debes argues that there is no real change in Hume’s later attitude to sympathy in the moral *Enquiry* (2007a,b). He claims, first, that Hume did not abandon his associationist account of sympathy; I agree, and have offered evidence for this in §5.2 above. He goes on to argue that sympathy still plays a role in the later moral philosophy, working in tandem with benevolence to generate our moral approbation and disapprobation. I have my doubts about the details of Debes’s proposal here, but let that be. For even supposing that this is right, the argument nevertheless misses a step. Assuming that sympathy continues to play some role in Hume’s later account, there is still the crucial *addition* of benevolence, which had no part in the *Treatise* explanation. This alone is a substantial difference.

³See Guyer (2002).

any Knowledge of Principles, Proportions, Causes, or of the Usefulness of the Object; but strikes us at first with the Idea of Beauty[.] (1725, p. 25)

Berkeley then objected to Hutcheson's view, insisting that the perception of beauty *did* depend on reason, and indeed on precisely the recognition of the object's utility:

Euph. Is not a thing said to be perfect in its kind when it answers the ends for which it was made?

Alc. It is.

Euph. The parts, therefore, in true proportions must be so related, and adjusted to one another, as that they may best conspire to the use and operation of the whole?

Alc. It seems so.

Euph. But the comparing parts with one another, the considering them as belonging to one whole, and the referring this whole to its use or end, should seem to be the work of reason: should it not?

Alc. It should.

...

Euph. Consequently, beauty, in your sense of it, is an object, not of the eye, but of the mind. (1732, p. 67)

In the fourth edition of his *Inquiry*, Hutcheson added a footnote replying to Berkeley's objection. Several points are made, of which just one is salient here, namely that we can find things beautiful even when we expect no use from them ourselves: "The Shapes of the Horse or the Ox may promise Use to the Owner; but is he the only Person who relishes the Beauty?" (1738, p. 208).

Hume's contribution to this debate in the *Treatise* was a hybrid theory that combined Berkeley's utility-based story with Hutcheson's appeal to a direct non-rational perception of pleasure: "the beauty of all visible objects causes a pleasure pretty much the same, tho' it be sometimes deriv'd from the mere species and appearance of the objects; sometimes from sympathy, and an idea of their utility" (T 3.3.5.6, p. 617). The addition of sympathy on the Berkelean side of this story,

however, is a marked improvement, answering Hutcheson's point that was just highlighted: for Hume, a useful object needn't be of use to the *observer* in order to give rise to a sentiment of beauty; it need only be useful to someone with whom the observer can sympathise.

Although he officially endorses a hybrid theory, however, Hume places more weight on Berkeley's account than on Hutcheson's: "Most of the works of art are esteem'd beautiful, in proportion to their fitness for the use of man, and even many of the productions of nature derive their beauty from that source" (T 3.3.1.8, p. 577; cf. T 2.1.8.2, pp. 298-9). He is similarly grudging in his concession to Hutcheson in the moral case:

Moral good and evil are certainly distinguish'd by our *sentiments*, not by *reason*: But these sentiments may arise either from the mere species or appearance of characters and passions, or from reflections on their tendency to the happiness of mankind, and of particular persons. My opinion is, that both these causes are intermix'd in our judgments of morals; after the same manner as they are in our decisions concerning most kinds of external beauty: Tho' I am also of opinion, that reflections on the tendencies of actions have by far the greatest influence, and determine all the great lines of our duty. (T 3.3.1.27, pp. 589-90)

Hume also says here that Hutcheson's moral sense explains our approbation "by particular *original* principles of human nature, which cannot be accounted for" (ibid.); and while he allows some room for such original principles, he thinks that his more complicated account in terms of sympathy and utility fits the vast majority of cases.

In the *Treatise*, then, not only did sympathy explain the beauty of utility, but utility was also the great source of beauty in art as well as morals. To repeat: "Most of the works of art are esteem'd beautiful, in proportion to their fitness for the use of man" (T 3.3.1.8, p. 577). In the moral *Enquiry*, Hume continues to hold that utility is the main source of *bodily* beauty, at least:

It will naturally be expected, that the beauty of the body, as is supposed by all ancient moralists, will be similar, in some respects, to that of the mind...

It is evident, that one considerable source of beauty in all animals is the advantage, which they reap from the particular structure of their limbs and members, suitably to the particular manner of life, to which they are by nature destined...

Broad shoulders, a lank belly, firm joints, taper legs; all these are beautiful in our species, because signs of force and vigour. Ideas of utility and its contrary, though they do not entirely determine what is handsome or deformed, are evidently the source of a considerable part of approbation or dislike. (M 6.23-5, pp. 244-5)

He also indicates that utility accounts for some of the beauty of artefacts that have an obvious purpose, such as ships and doors:

A ship appears more beautiful to an artist, or one moderately skilled in navigation, where its prow is wide and swelling beyond its poop, than if it were framed with a precise geometrical regularity, in contradiction to all the laws of mechanics. A building, whose doors and windows were exact squares, would hurt the eye by that very proportion; as ill adapted to the figure of a human creature, for whose service the fabric was intended. (M 5.1, pp. 212-3)

(There is an implicit dig at Hutcheson here, who emphasised the beauty of geometrical regularity, while denying that ideas of utility played any role; Hutcheson 1725, pp. 28-35.)

It is unclear whether Hume still thought that sympathy played the crucial role in engaging our critical approbation of useful objects or bodily features, for there is simply too little text to go on. Perhaps, as in the moral case, he would have placed more weight on the role of benevolence. But however that may be, what is more interesting about the development of Hume's thought in this area is that he no longer places anywhere near so much emphasis on utility in the explanation of beauty in general. Certainly it is *one* factor, but it is now only one among many.

In Hume's fourth dissertation, *Of the Standard of Taste*, in which he considers our critical approbation of art works in general (especially poetry and plays), there

is no mention of utility or sympathy at all.⁴ Instead, we have now a multiplicity of original principles responsible for our reaction:

It appears then, that, amidst all the variety and caprice of taste, there are certain general principles of approbation or blame, whose influence a careful eye may trace in all operations of the mind. Some particular forms or qualities, from the original structure of the internal fabric, are calculated to please, and others to displease[.] (ST 12, p. 233)

There is also a much more sophisticated account in this dissertation of the causal “noise” that interferes with these several general principles: ignorance, inexperience, prejudice, and so on. I will look at this more closely in chapters 13 and 14, when I come to examine *Of the Standard of Taste* in its own right. The *variety* of taste observable among different people is something that Hume’s somewhat crude appeals to sympathy and utility in the *Treatise* are powerless to explain.

In the meantime, the point to take away from these last two sections is that Hume’s conversion to Butler’s object-directed motivational psychology appears to have had an effect on his treatment of approbation as well as motivation. The effect on motivation is of course immediate and obvious. The case of approbation is more complicated, and there may well have been other factors in play. My examination of the matter here has of necessity been much briefer than the topic deserves. But even in this brief discussion, the parallels that we have seen are striking enough to warrant the general claim of this chapter: that in Hume’s later work, original instincts encroached very seriously on the explanatory territory of sympathy and comparison, in the case of approbation as well as motivation.

⁴An insignificant exception: Hume notes that we may have “a peculiar sympathy with the writer who resembles us” (ST 29, p. 244). This is the one occurrence of “sympathy” in this work that featured in my crude word count in §5.1 above.

Chapter 6

The Development of Hume's Thought

In chapters 2 and 3, I argued that Hume's conversion to Butler's object-directed motivational psychology came only after the bulk of the *Treatise* had been written, albeit in time to insert the Butler paragraph. The core of my case was, quite simply, that Hume was committed to egoism in this early work (since his anti-egoism in the later work is obvious and uncontroversial). It is of course crucial to my thesis that this should be so: my position is that Hume's anti-egoism was a *later* development, and one that explains a lot of the differences between the *Treatise* and the later work. As I indicated at the time, however, my egoist interpretation of the *Treatise*—though by no means original—is not without its opponents. In this final chapter of Part 1, the case for supposing that Hume was already an anti-egoist when he wrote the *Treatise* will be considered, and found wanting.

6.1. Anti-egoism in the *Treatise*

As I noted in §2.2, Stephen Darwall (among others) agrees with my hedonist and egoist interpretation of the *Treatise*.¹ In response to Darwall's interpretation, however, Don Garrett has recently objected that “although Hume clearly holds that prospective pleasure or pain can be sufficient to move the will, he never claims that it is necessary. (Nor does he treat motivation by the prospect of pleasure or

¹Darwall (1995).

pain egoistically—Appendix 2 of *An Enquiry concerning the Principles of Morals*, entitled ‘Of Self-Love,’ argues directly against psychological egoism).² This is the latest in a long-running tradition of anti-egoist interpretations of the *Treatise*, interpretations that would of course minimise or deny altogether the differences between the earlier and later Hume that I have been insisting on throughout the first part of this thesis.

Garrett’s parenthetical remark, first of all, is easily dealt with. One can hardly argue that Hume was not an egoist in the *Treatise* because of his explicit anti-egoism in the moral *Enquiry*. Such an argument would plainly presuppose the falsehood of my Difference Thesis, and so cannot be used against it. Nor would Garrett use such an obviously question-begging move; this comment of his is taken from the context of a different debate. (Darwall had not claimed that Hume later changed his mind; like so many, he more or less just ignored the later work.) Nevertheless, if anti-egoist interpreters of the *Treatise* have hitherto drawn much of their evidence from the moral *Enquiry*, their case will now be substantially weakened in the light of my proposals about how Hume’s thought changed on precisely this matter.

More hopeful is Garrett’s reinterpretation of the apparently hedonist passages of the *Treatise* as holding merely that pleasure and pain are *sufficient* for desire, not that they are *necessary*. In favour of reading the *Treatise* as endorsing only this partial hedonism, Garrett quotes the Butler paragraph, unsurprisingly, and also the following more moderate claim from Book 1: “There is implanted in the human mind a perception of pain and pleasure as the chief spring and moving principle of all its actions” (T 1.3.10.2, p. 118). As Garrett rightly notes, a *chief* spring need not be the *only* spring; and indeed, describing it in this way clearly suggests that there are others.

Let us run with Garrett’s suggestion for a while. If pleasure and pain are merely the *chief* springs of action, but not the only ones, what might the others be? Garrett himself is not explicit, but the two most obvious potential candidates

²Garrett (2007, p. 273).

in the *Treatise* are benevolence (the desire of another's happiness) and anger (the desire of another's misery). These two desires, says Hume, arise from love and hatred respectively:

The passions of love and hatred are always followed by, or rather conjoin'd with benevolence and anger... Love is always follow'd by a desire of the happiness of the person below'd, and an aversion to his misery: As hatred produces a desire of the misery and an aversion to the happiness of the person hated. (T 2.2.6.3, p. 367)

Furthermore, the connection between love and benevolence, and between hatred and anger, is explained in the *Treatise* by an original instinct, just like those that are later responsible for the object-directed desires: "the desire of the happiness or misery of others, according to the love or hatred we bear them, [is] an arbitrary and original instinct implanted in our nature" (T 2.2.7.1, p. 368).

In Hume's account of the generation of benevolence and anger, then, there is no hint of any hedonism. In the absence of evidence to the contrary, the default interpretation must surely be that these were not hedonistic desires for Hume, even in the *Treatise*. Even more than this, there is direct evidence *in favour* of this default: benevolence and anger are precisely Hume's examples of object-directed desires in the Butler paragraph.

Hume also apparently acknowledges (somewhat less explicitly) another pair of desires for the happiness or misery of another, namely *gratitude* and *resentment*, which arise in response to services rendered and injuries received respectively (see the brief remarks at T 2.2.3.3 and T 2.3.3.9; pp. 348, 418). As with benevolence and anger, the default interpretation of these two desires is also surely anti-hedonist. What is more, Hume's only remark about the origin of resentment is *explicitly* so: "When I receive any injury from another, I often feel a violent passion of resentment, which makes me desire his evil and punishment, *independent of all considerations of pleasure and advantage to myself*" (T 2.3.3.9, p. 418; my emphasis).

It is true, as I have already noted, that pity and malice either are or give rise to desires for another person's happiness or misery respectively, and in a way that is perfectly consistent with the selfish hypothesis. It is to be expected—and in perfect accord with my thesis—that Hume should devote a lot of attention to these passions in the *Treatise*, and to the principles of sympathy and comparison on which they depend. But, to put the point most forcibly against myself, have I not thus far simply been misrepresenting the *Treatise*, by examining these passions, but ignoring the apparently anti-egoist ones of benevolence, anger, gratitude, and resentment? Doesn't the presence of these latter desires in Hume's philosophy as early as 1739 show that he was never the egoist that I have made him out to be?

Another couple of passages from the *Treatise* may be added to the case against me here. In the first, Hume asserts that there are "certain instincts originally implanted in our natures, such as benevolence and resentment, the love of life, and kindness to children" (T 2.3.3.8, p. 417). Here benevolence and resentment are once again stated to be original instincts, which goes against the egoist thesis, and kindness to children is also included in this list. By "kindness to children", Hume presumably means parental love. What is more, when comparing human beings to other animals, Hume explicitly acknowledges the instinctive nature of this passion: "The affection of parents to their young proceeds from a peculiar instinct in animals, as well as in our species" (T 2.2.12.5, p. 398). In §5.3, I noted that Hume's official account of parental love in the *Treatise* depended on sympathy, treating it as just a special case of the love of family relations in general (while in the *Dissertation* it is explicitly singled out as *not* being accounted for in this way). The truth, however, is slightly more complicated: the *Treatise* derives parental love from sympathy, but *also* contains these two remarks suggesting instead that it is an original instinct.

In this way, then, there is quite a lot to be said for Garrett's interpretation of the *Treatise* as endorsing what we might call a *moderate* hedonism or egoism, as opposed to the *full-blown* egoism that I have been attributing to him thus far. On Garrett's reading, Hume maintains that self-love or the desire for our own

pleasure is the *chief* component of our motivational psychology, but not the *only* one. Already in the *Treatise* he was acknowledging the existence of anti-egoist desires. If Garrett is right—and as we have seen there is certainly evidence in his favour—then my case for substantial change appears to be significantly weakened, depending as it does on a blinkered interpretation of the *Treatise*.

6.2. Egoism, careless egoism, and anti-egoism

There are two reasons, however, why the argument of the previous section is not the devastating objection to my thesis that it might at first sight appear. In the first place, my position all along has been that Hume was committed to (full-blown) hedonism and egoism in the *Treatise*; I never said that this commitment was *consistent* with everything else that he said in this large and ambitious work. Indeed, I have insisted from the very start that it is *not*: in the Butler paragraph Hume clearly contradicts it. What I am now acknowledging is that the Butler paragraph is not the only place in which Hume contradicts himself. There is also the claim that benevolence, resentment, the love of life, and kindness to children are all original instincts (T 2.3.3.8, p. 417), and again that parental love is instinctive in human beings as well as other animals (T 2.2.12.5, p. 398), not to mention the explicit claim that resentment is a desire for another's misery independent of all considerations of our own pleasure (T 2.3.3.9, p. 418). The account of the origin of benevolence and anger, furthermore, as arising from love and hatred respectively, seems also to be implicitly in tension with it.

It does not follow from these exceptions, however, that Hume was a moderate hedonist. The evidence that he was committed to full-blown hedonism—which I presented in chapters 2 and 3—remains as forceful now as it was then. The final conclusion, taking all of the evidence into account, is not that Hume was a consistent moderate hedonist in the *Treatise*, but that he was an inconsistent full-blown hedonist. What this means is that the young Hume already had an *inkling* of the view that he later grasped more clearly, and expressed consistently. I have never denied this.

To be clear, then, I am disagreeing with Garrett on a crucial point: Hume *does* claim that pleasure is necessary for motivation. As I explained in §2.2, Hume begins Book 2 of the *Treatise* by dividing the passions into the direct and the indirect, a division that is clearly supposed to be exhaustive. And he claims that the passions in both of these classes arise from pleasure and pain (but in different ways). He explicitly places desire and aversion in the former category, as among those that “arise immediately from good or evil, from pain or pleasure” (T 2.1.1.4 p. 276). He says that “[u]pon the removal of pleasure and pain there immediately follows a removal of desire and aversion” (T 2.3.9.1, p. 438). If this is not claiming that pleasure (or pain) is necessary for motivation, I do not know what would be. And this is just a brief summary; my argument on this head has been given in more detail in chapters 2 and 3.

The second and more important reason why the previously noted exceptions do not trouble me is this. Suppose that I am after all being too harsh on the *Treatise*, and that in this work Hume is consistently endorsing a moderate hedonism, rather than inconsistently endorsing a full-blown hedonism. Suppose, that is to say, that Garrett is right and I am wrong. Even then, the concession that I would be required to make regarding my Difference Thesis is in fact negligible.

The reason for this is that moderate hedonism—according to which pleasure and pain are the *chief* spring and moving principle of the mind, though there are a handful of exceptions—is still a very long way from the rigorous *full-blown anti-hedonism* of Hume's later works. Consider Locke's claim, already quoted in §1.4, that without pleasure and pain there would be no desire:

[T]o excite us to these Actions of thinking and motion, that we are capable of, [the Author of our being] has been pleased to join to several Thoughts, and several Sensations, a *perception* of *Delight*. If this were wholly separated from all our outward Sensations, and inward Thoughts, we should have no reason to prefer one Thought or Action, to another; Negligence, to Attention; or Motion, to Rest. (1690, p. 129)

And consider, alongside this, Hume's strikingly similar claim in Book 3 of the *Treatise*:

The chief spring or actuating principle of the human mind is pleasure or pain; and when these sensations are remov'd, both from our thought and feeling, we are, in a great measure, incapable of passion or action, of desire or volition. (T 3.3.1.2, p. 547)

Here again, as in the Book 1 passage that Garrett draws attention to (T 1.3.10.2, p. 118), Hume does appear to be retreating ever so slightly from Locke's full-blown hedonism. He says only that pleasure and pain is the *chief* principle, and that without it we are incapable *in a great measure* of action or desire. The implication is that there are exceptions.

But Butler's objection to Locke, of course, was not just that there are a handful of exceptions to the hedonist account of motivation. His point was that this view is *utterly* mistaken, since it gets the causal and explanatory relationship between pleasure and desire fundamentally the wrong way round. For Butler, the *overwhelming majority* of our desires are object-directed, with hedonistic applications of self-love being the exception rather than the norm. This view is opposed to hedonism in all its forms, both moderate and full-blown.

After the *Treatise*, as we have seen, Hume was fully persuaded of Butler's point. He was no longer remotely attracted to hedonism, even in a moderate form, but fervently insisted on Butler's conclusion that it got things back to front. He was no longer endorsing Locke's claim (even in a moderated form) that without pleasure and pain there would be no desire. Rather, he insisted on Butler's alternative, that without desire there would be (almost) no pleasure or pain:

Were there no appetite of any kind antecedent to self-love, that propensity could scarcely ever exert itself; because we should, in that case, have felt few and slender pains or pleasures, and have little misery or happiness to avoid or to pursue. (M App2.12, pp. 301-2)³

Even if Garrett is right, therefore, his interpretation of the *Treatise* in fact does very little to bring that earlier work in line with the later. The *Treatise* account

³Butler implausibly claimed that there could be *no* pleasure without desire. Hume does not go that far, while still accepting the essential point. Recall chapter 1, page 33.

of motivation is hedonist, whether moderate or full-blown. The later work is radically anti-hedonist.

6.3. Butler, not Hutcheson

It seems clear that the Butler paragraph in the *Treatise* did indeed come from Butler. The point that Hume makes in this paragraph is very obviously Butler's, and in the footnote from the first two editions of the first *Enquiry* (removed after the appearance of the moral *Enquiry*), Hume explicitly attributed it to him (recall §2.4). Surprisingly, however, it has been suggested twice before—by Kemp Smith, and more recently by the Nortons—that the inspiration for this paragraph was in fact Hutcheson, rather than Butler.

In chapter 2, I conjectured that Hume read Butler's *Sermons* just before the *Treatise* was printed (and in time to insert the Butler paragraph), but after it had already been mostly written. And I suggested that this event was not only what prompted the Butler paragraph, but also—once the importance of Butler's point had sunk in—what led Hume to abandon his hedonism altogether, in favour of Butler's object-directed motivational psychology. These claims are not crucial to my Difference Thesis: the textual evidence for the change of mind is there, regardless of what caused it. Nevertheless, this background story fits very nicely with the story told by the texts themselves. And so if Hume got this point from Hutcheson rather than Butler, that ought to give me some cause for concern; for Hume was already familiar with Hutcheson when he was writing the *Treatise*.

Fortunately, however, in addition to the direct evidence that Hume got the point from Butler (i.e. it is to Butler that Hume himself attributes it in the first *Enquiry*), there is also evidence that he *can't* have got it from Hutcheson. This is for the very simple reason that *Hutcheson nowhere makes this point*. No doubt its anti-egoist implications would have been welcome to him, but Hutcheson's own motivational psychology is in fact precisely the moderate hedonism that Garrett attributes to Hume in the *Treatise*, according to which the pleasure and pain of *others* moves us directly, as well as the prospect of our own:

Desires arise in our Mind, from the Frame of our Nature, upon Apprehension of Good or Evil in Objects, Actions, or Events, to obtain for our *selves* or *others* the *agreeable Sensation*, when the Object or Event is good; or to prevent the *uneasy Sensation*, when it is evil. (1728, p. 18)

Butler's point that desires give rise to pleasure and pain, rather than the other way round, was as unfamiliar to Hutcheson as it was to the young Hume.⁴

Why, then, would someone think that Hume got this idea from Hutcheson? Norton and Norton, in their annotations to the *Treatise*, seem to hold that he did.⁵ In support of this claim, they quote the following passage from Hutcheson's *Essay*:

But we must here observe an obvious Difference among our Desires, viz. that "some of them have a *previous*, painful, or uneasy Sensation, antecedently to any *Opinion* of Good in the Object; nay, the Object is often chiefly esteemed good, only for its *allaying this Pain* or Uneasiness; or if the Object gives also positive Pleasure, yet the *uneasy Sensation* is previous to, and independent of this *Opinion* of Good in the Object." These Desires we may call *Appetites*. "Other Desires and Aversions necessarily presuppose an *Opinion* of Good and Evil in their Objects; and the Desires or Aversions, with their concomitant uneasy Sensations, are produced or occasioned by this *Opinion* or *Apprehension*." (1728, p. 67)

I can only reply that this, though perhaps superficially similar, is on closer examination quite clearly a different point. Here Hutcheson is simply noting—something which is perfectly consistent with hedonism, full-blown as well as moderate—that some desires are for the removal of a present pain (e.g. hunger and thirst, the examples that Hutcheson goes on to give), while others are for positively pleasant objects.

⁴Hutcheson's *Inquiry* (1725) predated Butler's *Sermons* by a year (1726). We can surmise that Hutcheson read them soon after, for they are mentioned in the preface to his *Essay* and *Illustrations* (1728, p. 9). But I can find no evidence that this later work took on board Butler's crucial anti-hedonist point.

⁵Norton and Norton (2007, volume 2, p. 877).

A more sustained case for the importance of Hutcheson's influence rather than Butler's on this matter was made by Kemp Smith.⁶ Kemp Smith was generally very keen on emphasising the importance of Hutcheson's influence, an influence that he took to have been with Hume right from the start; it was his overarching hypothesis "That Hume, under the Influence of Hutcheson, entered into his Philosophy through the Gateway of Morals".⁷ This hypothesis does not seem to me particularly plausible, but this is not the place to criticise it in detail.⁸ What I want to take issue with here is a particular proposal made in the context of this general claim: that the existence of object-directed desires, and consequently the denial of hedonism, was one of Hume's very *first* philosophical commitments; and, of course, in keeping with the general claim, that Hume got this idea from Hutcheson.⁹

The term "object-directed desire" is my own; Kemp Smith himself speaks of a distinction between "primary" and "secondary" passions, where the latter are founded on pleasure and pain, and the former instead *give rise to* pleasure and pain. Thus my object-directed desires are Kemp Smith's primary passions. Kemp Smith also claims that "primary" and "secondary" were Hume's own terms, and that Hume got them from Hutcheson.¹⁰ No references are given, and Fieser has objected that Hume himself did not use these terms.¹¹ In this, Fieser gives away the insufficient attention that he (like many others) has paid to Hume's later work, for Hume does use the terms; but he uses them in the moral *Enquiry* rather than the *Treatise*:

⁶Kemp Smith (1941/2005).

⁷Kemp Smith (1941/2005, p. 12).

⁸For criticism, and a defence of what strikes me as a much more plausible hypothesis—namely, that Hume approached his philosophy through an interest in causation in relation to irreligion and the free-will debate—see Millican (forthcoming).

⁹Part 3 of Kemp Smith's book is titled a "Detailed Consideration of [Hume's] Central Doctrines, Taken in What May Be Presumed to Have Been the Order of Their First Discovery" (p. 157); following some introductory remarks, object-directed desires are then one of the first things discussed (pp. 163-5).

¹⁰Kemp Smith (1941/2005, p. 165).

¹¹Fieser (1992, p. 9).

There are bodily wants or appetites, acknowledged by every one, which necessarily precede all sensual enjoyment, and carry us directly to seek possession of the object. Thus, hunger and thirst have eating and drinking for their end; and from the gratification of these *primary* appetites arises a pleasure, which may become the object of another species of desire or inclination, that is *secondary* and interested. In the same manner, there are mental passions, by which we are impelled immediately to seek particular objects, such as fame, or power, or vengeance, without any regard to interest; and when these objects are attained, a pleasing enjoyment ensues, as the consequence of our indulged affections... In all these cases, there is a passion, which points immediately to the object, and constitutes it our good or happiness; as there are other *secondary* passions, which afterwards arise, and pursue it as a part of our happiness, when once it is constituted such by our original affections. (M App2.12, p. 301; my emphases)

This of course is in the context of stating precisely Butler's anti-hedonist point; contrary to Kemp Smith, it has nothing to do with Hutcheson, and it postdates the *Treatise* by more than ten years. There is no evidence that Hume got the terms from Hutcheson while he was writing his early work.

Kemp Smith doesn't say where Hutcheson uses these terms any more than where Hume does, but my best guess is that he had in mind the following passage from Hutcheson's *Essay*:

Now since we are capable of *Reflection, Memory, Observation, and Reasoning* about the distant Tendencies of Objects and Actions, and not confined to things present, there must arise, in consequence of our *original Desires*, "*secondary Desires* of every thing imagined useful to gratify any of the primary Desires, with strength proportioned to the several original Desires, and the imagined Usefulness, or Necessity, of the advantageous Object." Hence it is that as soon as we come to apprehend the Use of *Wealth* or *Power* to gratify any of our original Desires, we must also desire them. (1728, p. 19)

Here, however, Hutcheson is once again making a different point. Hutcheson's original or primary desires—as confirmed by the paragraphs preceding that above (one of which I quoted on page 121)—are all *directly* for pleasure or the avoidance of pain (either for ourselves or for others). His secondary desires, meanwhile, are for the means to pleasant ends. This is no doubt a perfectly respectable distinction; but it is not Butler's distinction between the desire for pleasure on the one hand, and desires for objects, independently of the pleasure that they bring, on the other.

In line with his view that anti-hedonism was one of the first philosophical positions that Hume adopted, Kemp Smith of course interprets the *Treatise* as an anti-hedonist work. The only evidence of Hume's anti-hedonism that he produces from the *Treatise*, however, is the Butler paragraph.¹² Otherwise, all of his quotations—including his terminology for the distinction between hedonistic and object-directed desires—are from the moral *Enquiry*. No compelling reasons are given for supposing that Hume appreciated Butler's point before he wrote the *Treatise*.

6.4. Going back to the source

Anti-egoist interpreters of the *Treatise* have their hearts in the right place. The view emerged, in the early 1900s, out of opposition to T. H. Green's false and ungenerous interpretation of Hume as doing nothing more than taking Locke's and Berkeley's principles to their logically absurd conclusions, an interpretation that had its roots, in Hume's own day, with Thomas Reid.¹³

As well as viewing Hume (as Reid had before him) as a dogmatically negative sceptic, Green also interpreted him as a thoroughgoing Lockean hedonist,¹⁴ and it is in large part with the noble aim of refuting Green that the tradition of anti-hedonist interpretations of the *Treatise* begins. Green was in many respects

¹²He quotes this paragraph on p. 141, and again later on p. 164.

¹³Reid (1764), Green (1882a).

¹⁴Green (1882b, pp. 31-5).

Kemp Smith's principal target, not only regarding Hume's views on motivation; and Garrett, in his introduction to the recent reprint of Kemp Smith's book, commends its author for criticising "Green's unduly hedonistic reading of Hume's psychology".¹⁵

This anti-Green tradition begins, in fact, not with Kemp Smith, but with an old article by E. B. McGilvary.¹⁶ Kemp Smith explicitly leans on the arguments in this article,¹⁷ as did at least one other defender of the anti-egoist interpretation at the time.¹⁸ In Taylor's words: "With respect to Hume's moral philosophy, Green's analysis began a trend of viewing Hume as an advocate of hedonism. As Norman Kemp Smith reminds us... this trend only began to reverse in the twentieth century, with E. B. McGilvary's important 1903 article".¹⁹

While I heartily approve of the ultimate aims of these critics of Green, it seems to me that they are fighting their battle in the wrong way. Green's uncharitable interpretation of Hume—much like Reid's before him—draws its ammunition exclusively from the *Treatise*. Rather than attempt tortuous readings of the pretty obviously hedonist passages of this early work, the proper response is surely the very one that Hume himself made in the light of Reid's criticisms: to deny that the *Treatise* contains his considered view, and to look to the *Enquiries* and the *Dissertations* instead. In the famous advertisement prefixed to volume 2 of the posthumous edition of his *Essays and Treatise on Several Subjects*, Hume complained (in response to Reid) of precisely this sort of unfair treatment:

[S]everal writers, who have honoured the Author's Philosophy with answers, have taken care to direct all their batteries against that juvenile work [the *Treatise*], which the Author never acknowledged, and have affected to triumph in any advantages, which, they imagined, they had obtained over it: A practice very contrary to all rules of candour and fair-dealing, and

¹⁵Garrett (2005, p. xxviii).

¹⁶McGilvary (1903).

¹⁷Kemp Smith (1905, p. 336); Kemp Smith (1941/2005, pp. 140-2).

¹⁸Chapman Sharp (1921, p. 43).

¹⁹Taylor (2007, pp. 305-6).

a strong instance of those polemical artifices, which a bigotted zeal thinks itself authorised to employ. Henceforth, the Author desires, that the following Pieces [the *Enquiries* and *Dissertations*] may alone be regarded as containing his philosophical sentiments and principles. (Ad 1777, p. 2)

When we follow Hume's own very clear and explicit request, which most Hume scholars seem strangely not to take seriously, the defence against Green's argument becomes unnecessary; he is firing blanks. And that Hume's considered view was anti-hedonist and anti-egoist, meanwhile, is inescapable even on the most cursory reading of his later work.

McGilvary's defence of the non-hedonist reading of the *Treatise* relies in part on moves that are by now familiar: of course he quotes the Butler paragraph,²⁰ and insists, like Kemp Smith and Garrett, that it does not contradict the apparent endorsements of hedonism, for the latter assert only that pleasure and pain are the *chief* motives, not the only ones.²¹ In addition to these points, however, McGilvary has two further arguments that ought to be addressed. The first is his argument that, even with those desires that Hume says are founded on pleasure and pain, the foundation is *causal* rather than *intentional*. These desires are always *caused by* an immediate sensation of pleasure, that is to say, but need not themselves be desires *for* pleasure; they might instead be for certain objects directly.²²

Against this interpretation, one might simply quote Hume: "Tis from the prospect of pain or pleasure that the aversion or propensity arises towards any object" (T 2.3.3.3, p. 414). But to be thorough I ought also to address the evidence that McGilvary offers in support of his alternative reading, which comes from a couple of paragraphs in Book 1:

²⁰McGilvary (1903, p. 276).

²¹McGilvary (1903, pp. 277-8).

²²McGilvary (1903, p. 281). Perhaps Garrett had this idea in mind when he claimed that Hume doesn't treat motivation by the prospect of pleasure or pain egoistically in the *Treatise*; though Garrett, recall, defends this claim solely by appeal to the anti-egoism of the moral *Enquiry*.

There is implanted in the human mind a perception of pain and pleasure, as the chief spring and moving principle of all its actions. But pain and pleasure have two ways of making their appearance in the mind; of which the one has effects very different from the other. They may either appear in impression to the actual feeling, or only in idea, as at present when I mention them. 'Tis evident the influence of these upon our actions is far from being equal. Impressions always actuate the soul, and that in the highest degree; but 'tis not every idea which has the same effect...

Tho' an idle fiction has no efficacy [i.e. no power of actuating the will], yet we find by experience, that the ideas of those objects, which we believe either are or will be existent, produce in a lesser degree the same effect with those impressions, which are immediately present to the senses and perception. The effect, then, of belief is to raise up a simple idea to an equality with our impressions, and bestow on it a like influence on the passions. This effect it can only have by making an idea approach an impression in force and vivacity. (T 1.3.10.2-3, pp. 118-9)

In this passage, Hume says three things: first, that present impressions of pleasure or pain give rise to desire or aversion; secondly, that *ideas* of pleasure or pain also give rise to desire or aversion; and thirdly, that ideas have this effect only when we believe that the objects “either are or will be existent”, since belief renders them nearly as forceful and vivacious as the impressions themselves.

While the first two of these claims seem straightforward enough, the third is surely false: we do not need to believe that an object either is or will be existent in order to form a desire for it. Perhaps what Hume meant to say was that desire (or aversion) only arises when we believe that the object in question *might* be obtained (or avoided). This would fit with his claim in the introduction to the *Treatise* that “we are no sooner acquainted with the impossibility of satisfying any desire, than the desire itself vanishes” (T Intro.9, p. xvii); though that itself is also a very questionable claim.

However that may be, the important point here is that there is nothing in this passage to support McGilvary's interpretation. According to McGilvary, these

paragraphs show that Hume thought pleasure and pain (whether as impressions or as believed ideas) are the *efficient causes* of desire and aversion, but that they are not always the *objects* of desire and aversion: “Hume’s view, here expressed, is that when we are influenced by pleasure to perform an action, we always act *from* pleasure, not always *for* pleasure”.²³ True enough, Hume does say here that pleasure and pain are the *causes* of desire and aversion. But that doesn’t mean that they are not *also* the objects of desire and aversion, and nothing here suggests that Hume thought they were not. In McGilvary’s terms, Hume does indeed affirm that we always act from pleasure; but he nowhere denies that we always act *for* pleasure as well. At best, McGilvary’s contrived reading renders this passage consistent with Hume’s later anti-hedonist view; it certainly does not force that view upon us any earlier.²⁴

McGilvary’s second argument admits of a similar response. It is natural to associate sympathy with egoism, as I have already pointed out in previous chapters, for sympathy with other people’s pains and pleasures provides an obvious explanation for seemingly altruistic motivation, on the assumption that we are ultimately only motivated by *our own* pleasure and pain. If other people’s pleasures and pains effectively become our own through the mechanism of sympathy, that can explain why we want other people to be happy, even supposing a purely egoist motivational psychology. In this way, as I have said, Hume’s account of pity or compassion is not only consistent with egoism, but is also strongly supportive of it, since it provides an egoist reduction of much seemingly altruistic behaviour.

McGilvary, however, came up with an ingenious alternative reading of Hume on this point. He suggested that it was not only other people’s pleasure or pain that is communicated to us by sympathy, but also their *desires*. These desires, meanwhile, even if they are egoist to begin with, are no longer egoist

²³McGilvary (1903, p. 281).

²⁴Indeed, it is difficult to see what causal mechanism Hume could have in mind here, if not precisely the obvious intentional one: present pleasure and pain move us to continue or change what we are doing *because* we desire the former shun the latter; and ideas of these things—“the prospect of pain or pleasure” (T 2.3.3.3, p. 414)—move us for the very same reason.

when communicated sympathetically to other people. You want *your* pleasure, but when this desire is communicated to me it remains a desire for *your* pleasure; which, when I have it, is plainly not egoist.²⁵

It seems to me that McGilvary focuses his argument here in exactly the wrong place. He spends some time defending an interpretation of the sympathy mechanism according to which a sympathetically communicated desire, though it is egoist in the original mind, is not egoist in the mind to which it has been communicated. But this point is surely quite obvious. The substantial claim that needs defending, meanwhile, but which McGilvary offers literally nothing in support of, is that Hume thought of pity as a sympathetically communicated *desire*, rather than a sympathetically communicated *sorrow*. If this were so, then Hume's treatment of pity would indeed be anti-egoist. But it simply isn't what Hume says. In the *Treatise*, the talk is exclusively of "affliction and sorrow" being communicated (T 2.2.7.2, p. 369), never of desire; in the *Dissertation*, similarly, compassion is "an uneasiness in the sufferings of another" (P 3.7, p. 19), not a transmitted desire. There is no conceptual difficulty with McGilvary's suggestion; the problem is that it is *McGilvary's* rather than Hume's.²⁶

6.5. Why so much resistance?

Interestingly, just after quoting the Butler paragraph, McGilvary himself anticipated precisely my own view of the *Treatise* on this matter:

A higher criticism of the *Treatise* might try to distinguish between egoistic passages which were written first and non-egoistic passages which were afterwards inserted without proper rewriting of older passages in the interest of complete consistency.²⁷

²⁵McGilvary (1903, pp. 291-4).

²⁶In §4.4 I noted that "pity" is ambiguous in Hume, sometimes referring to a desire as well as to a sympathetically communicated sorrow. But this fact alone will not help McGilvary; nowhere does Hume say that this compassionate desire is a sympathetically communicated one. And in the absence of any such statement, the default interpretation is the obvious one, namely that this desire is the result of self-love applied to the sympathetically communicated sorrow.

²⁷McGilvary (1903, p. 277).

His reply to this possibility is hardly compelling:

But whatever may be the truth of such a view, we must remember that the recognition of the existence of instinctive passions, as opposed to passions founded on pleasure and pain, was an integral feature of the *Treatise* as it was published by Hume.²⁸

By *instinctive* passions, of course, he meant non-hedonistic, object-directed desires; those founded on original instincts. The recognition of these passions is indeed a feature of the *Treatise*. To repeat, I have never denied that this is so. What I strenuously do deny is that it is an *integral* feature. At best, it is a peripheral feature in a work that is predominantly hedonist; at worst, it is a late insertion that contradicts the assertions of full-blown hedonism made in the surrounding text.

Lest it be objected that my interpretations of any of the relevant passages in Hume are somewhat crude, I shall say explicitly that I consider this to be a *virtue* of my view. The direct evidence in favour of my position involves utterly straightforward readings of Hume's texts, taking what he says entirely at face value. If opponents would challenge this position with careful glosses on Hume's remarks, or subtle re-readings of the texts, that in itself is already a presumption against their alternative. Perhaps earlier passages can, with sufficient ingenuity, be re-interpreted in the light of later claims, or later claims in the light of earlier, so as to minimise any appearance of substantial difference. But by far the more likely interpretation is that Hume genuinely changed his mind.

For there is nothing inherently improbable in the developmental story that I have told in these chapters. It is not surprising that Hume should have begun life as a hedonist and an egoist. These were popular views at the time, held in particular by Hobbes and Locke (or at least, it was widely and plausibly held that this was Hobbes's view). No one doubts the important influence that these writers had on Hume in general, and there is no reason why this influence should not

²⁸McGilvary (1903, *ibid.*).

have extended to this point in particular. Nor is it surprising that Hume should then have changed his mind upon being confronted with Butler's arguments; for these arguments are indeed persuasive.

Nor is there anything unduly *uncharitable* about my interpretation. I have charged Hume, it is true, with endorsing a false view in the *Treatise*, and moreover with being inconsistent about it. These might seem like harsh criticisms. But once we take into account the whole picture, they are quite the contrary. *Of course* Hume was a hedonist to begin with. There is no shame in that: many very intelligent people were at the time. It is easy for us to see the falsehood of this position now, but we have the benefit of Butler's brilliant insight, not to mention Hume's own very clear presentations of it in his later work, and almost three centuries for these arguments to have sunk into the collective philosophical consciousness. As for Hume's inconsistency in the *Treatise*, this is surely very much to his credit. It shows that he was prepared to wrestle with his inherited commitments, in the light of persuasive arguments against them, and that he was able to be moved by these arguments. And ultimately, of course, he himself cleared up the inconsistencies that I have identified, and came down firmly on Butler's side, even against his own earlier self.

The significance of Butler's contribution to the egoism debate has been widely acknowledged.²⁹ But Hume deserves his place in this story too, as the first convert to Butler's cause, who then quickly succeeded him as its best and clearest defender. Though it was Butler who first had the insight, therefore, the significance of Hume's agreement should not be understated. Hume seems to have been the only other notable philosopher of the period who actually *got* Butler's point, and realised just how important it was; even Hutcheson, recall, didn't cotton on. This is all the more impressive when we realise that Hume started life on the other side of the debate, and would have been initially hostile to Butler's conclusion in a way that Hutcheson was not.

²⁹See Broad (1965, ch. 3), Nagel (1970, ch. 9), Blackburn (1998, ch. 5), Sober (2000).

I have the sense that for some scholars there is a strong emotional resistance to the idea that Hume's thought might have developed during his adult years, which goes along with a deep reverence for his early *Treatise*. And I anticipate that, where I have charged the *Treatise* with errors or inconsistencies, many will be eager to leap to its defence (perhaps preferring Garrett's moderate interpretation, for example, to my comparatively hard-line one). But quite why this attitude has arisen, or what sustains it, I do not know. If it is the love of Hume, leading to an aversion to the idea that he can ever have put a foot wrong, then I suggest that it is misplaced. The best philosophers are those who correct their mistakes; those who make none to begin with do not exist. And so when I criticise the *Treatise*, it is not with a view to throwing mud at that work, in order to make the later work shine by comparison. It is the love of Hume, and of uncovering the truth about his intellectual development, that motivates me.

Part II

The Third Enquiry

Chapter 7

Superstition and the Passions

Taking my project now to be sufficiently motivated by the preceding argument, I turn in this second part of the thesis to the study of Hume's mature philosophy of emotion as a whole, as presented in the *Four Dissertations*. I will run through the topics that Hume addresses more or less in textual order (naturally skipping over matters that have already been discussed in Part 1). This takes us through a number of important Humean topics: superstition, hope and fear, associationism, the relationship between reason and passion, the causal interaction of the passions, art and the passions, and the reply to the relativist about taste.

I begin, in this chapter and the next, with superstition and its causes (notably the passions of hope and fear). My position is perhaps best summed up by highlighting my opposition to an emphasis on the distinction between *philosophy* and *psychology*, between *reasons for belief* and *causes of belief*. As we will see, Hume was no stranger to this distinction. But it was his custom to run the discussion of these two things together—for example in his famous argument about induction, or in his criticism of belief in miracles—and the *Natural History of Religion* is no exception. This latter work is generally viewed as Hume's major contribution to religious psychology. This would be fine by itself, but the implication (and often it is more than just an implication) is that it contains nothing of interest regarding the *philosophy* of this subject. On the contrary, I will argue that this dissertation also includes a vital part of Hume's attack on the *rationality* of religious belief.

7.1. Reasons and causes

It used to be thought that Hume's interest in religion derived mainly from a desire for notoriety, and that his post-*Treatise* attacks on Christian orthodoxy were intended chiefly to excite some controversy and to boost book sales.¹ We now know better. Hume's interest in religion was with him from the start, and if his sceptical views were downplayed in the *Treatise*, this was only from prudence or a desire not to offend. This interest, furthermore, is no small or detachable aspect of his thought; on the contrary, a plausible case can be made for it being the driving force behind his whole philosophy.²

Precisely because of its centrality, however, not to mention the sensitive nature of the topic, there is no single work laying out Hume's views on religion systematically, and students of this aspect of his philosophy are obliged to garner evidence from several different texts. Keith Yandell lists the *Natural History of Religion*, the *Dialogues concerning Natural Religion*, sections 10 and 11 of the first *Enquiry* (*Of Miracles* and *Of a Particular Providence and of a Future State*), and the three essays *Of Superstition and Enthusiasm*, *Of Suicide*, and *Of the Immortality of the Soul*.³ J. C. A. Gaskin also includes—"less obviously"—the moral *Enquiry* and the *History of England* for their moral attacks on religion, and of course the *Treatise*.⁴

Given the scattered nature of this material, it is desirable to have some framework in which to place Hume's philosophy of religion, some way both of bringing the relevant texts together, and of classifying their distinct but complementary contributions to Hume's overall critique. We owe the standard framework—indeed, the *only* framework, so far as I am aware—to Gaskin.⁵ On Gaskin's

¹See for example Selby-Bigge (1893/1975, pp. xi-xii).

²See Russell (2008) and Millican (2002b, forthcoming).

³Yandell (1990, p. 4).

⁴Gaskin (2009, pp. 484-5). Gaskin notes that the *Treatise* "does not seem to us much concerned with religion [because] our sensitivities regarding what would constitute an attack on religion are much weaker than those of Hume's contemporaries" (ibid.). Those wanting to enhance their sensitivities in this regard should read Russell (2008).

⁵Gaskin (1988, 2009).

view, the key distinction is that between *reasons for belief* and *causes of belief* (where the latter may happen to be good reasons, but need not be):

Suppose we put the fundamental question thus: Why does anyone believe in God or gods, or cleave to the teachings of such theistic religions as Christianity or Islam? The answer may be given (nonexclusively) in terms of either reasons or causes, and it is under this division that Hume's examination of religion begins to look like a comprehensive critique rather than a collection of challenging but discrete sections.⁶

On this understanding, Hume's explicit discussions of religion divide into those examining the *causes* of religious belief (primarily the *Natural History of Religion*, but also perhaps the earlier essay *Of Superstition and Enthusiasm*), and those criticising the *arguments* in support of it (sections 10 and 11 of the first *Enquiry*, the posthumous essays *Of Suicide* and *Of the Immortality of the Soul*, and of course the *Dialogues concerning Natural Religion*).

The distinction between reasons and causes was certainly familiar to Hume, and he even draws it himself in the context of religious belief, in the introduction to the *Natural History*:

As every enquiry, which regards religion, is of the utmost importance, there are two questions in particular, which challenge our attention, to wit, that concerning its foundation in reason, and that concerning its origin in human nature. (N Intro, p. 33)

He goes on, seemingly, to restrict his attention in this work to the latter question only: "What those principles are, which give rise to the original belief, and what those accidents and causes are, which direct its operation, is the subject of our present enquiry" (ibid.). On the face of it, this is clear support for Gaskin's framework.

As the reader will probably have guessed, however, I have been setting this framework up with a view to knocking it down. The distinction between reasons and causes, between *philosophy* and *psychology*, seems to me an extremely

⁶Gaskin (2009, p. 485).

unhelpful one to use in carving up Hume's critique of religion. I do not believe that this distinction structured Hume's own thoughts on the matter, and in §7.2 below I will examine the alternative distinction that I believe did. Immediately, however, what troubles me about Gaskin's framework is that it encourages a blinkered interpretation of the *Natural History*.

This work obviously does provide (the main part of) Hume's account of the causes of religious belief, and so it certainly is a contribution to the study of religious psychology. In fact, Gaskin if anything understates the importance of the psychological context of the *Natural History*; in §7.3 I will go further in insisting on the relevance of Hume's work on the passions in completing his account of the origin of religious belief (a key component of my Unity Thesis). What I object to, however, is the suggestion that this work is an examination of causes *at the expense of* reasons:

[T]he account is of the causes and conditions that “naturally” produce religion (as, for example, the presence of air and water “naturally” produces rust on iron) without reference to any reasons that can be produced in favor of or against the religion in question.⁷

Rather, I want to insist that it is a work of psychology *and* philosophy, containing an important part of Hume's overall attack on the *rationality* of religious belief in addition to the discussion of its causes.

Though my focus here is obviously on the *Natural History*, the point generalises. Hume was perfectly well aware of the distinction between reasons and causes (as we have seen), but the *Natural History* is by no means unique among his works in running the discussion of these two things together. I am thinking here particularly of Hume's famous argument concerning induction, but also—more relevantly in the present context—section 10 of the first *Enquiry, Of Miracles*.

⁷Gaskin (2009, p. 483). I am focusing on Gaskin here for exegetical convenience, because I also want to highlight his framework and offer an alternative to it. But he is by no means alone in this restricted interpretation of the *Natural History*. On the contrary, this is the mainstream view. For example, “the *Natural History of Religion*... was essentially a sociological enquiry into the origin of religion” (Jenkins 1992, p. 8); “Hume's purposes in ‘The natural history’ were not fundamentally critical” (Falkenstein 2003, p. 1).

This essay suffers from precisely the opposite problem facing the *Natural History*: being placed on the side of reasons rather than causes, its part 1 argument tends to receive considerably more attention than the psychological discussion of part 2. This latter, as well as being relevant to part 1, complements Hume's account of the causes of religious belief given in the *Natural History*. At the end of this chapter (§7.5), I will also argue that the *argument* of the *Natural History* is an important complement to the *argument* in *Of Miracles*.

7.2. Superstition and true religion

Before arguing directly for my claim that the *Natural History of Religion* contributes to Hume's discussion of reasons as well as causes, it is well to have my alternative distinction for structuring Hume's critique of religion in place. The dichotomy that I propose instead is that between *truth* and *falsehood*. Since philosophers and theologians naturally disagreed about religious truth, it was established eighteenth-century practice to distinguish between "true" and "false" religion, and to demean the latter with the name of "superstition". Hume latched onto this practice, verbally at least, openly criticising "superstition" on both moral and rational grounds, while ostensibly endorsing "true religion".⁸

It is doubtful whether these endorsements of true religion were sincere, and it is also unclear how much Hume would be committed to even if they were. It is possible, however, to place some definite upper and lower bounds on the notion as it is deployed by Hume. The *least* that Hume's true religion involves, first, is the belief in an intelligent author of nature; or, at the *very* least, the belief that "*the cause or causes of order in the universe probably bear some remote analogy to human intelligence*" (D 12.33, p. 227). As an upper bound, meanwhile, what it presumably *cannot* include is the belief that this author of nature ever intervenes in the natural order once it has been set up. That true religion precludes all such particular divine intervention is suggested by the argument in support of it, which is premised precisely on the *absence* of any such interference:

⁸See Bell (1999) for a discussion of this practice, and of Hume's appropriation of it for his own irreligious ends.

Many theists, even the most zealous and refined, have denied a *particular* providence, and have asserted, that the Sovereign mind or first principle of all things, having fixed general laws, by which nature is governed, gives free and uninterrupted course to these laws, and disturbs not, at every turn, the settled order of events by particular volitions. From the beautiful connexion, say they, and rigid observance of established rules, we draw the chief argument for theism; and from the same principles are enabled to answer the principal objections against it. (N 6.2, p. 52)

Within these two constraints, however,—the belief in invisible, intelligent power, but the denial that it is among the immediate causes of particular events—there remains much room for manoeuvre. Cleanthes’ quite precise statement of (what he takes to be) true religion falls within these bounds:

[It] represents us as the workmanship of a Being perfectly good, wise, and powerful; who created us for happiness, and who, having implanted in us immeasurable desires of good, will prolong our existence to all eternity, and will transfer us into an infinite variety of scenes, in order to satisfy those desires, and render our felicity compleat and durable. (D 12.24, p. 224)

But, for all I have said so far, true religion might also be consistent with the author of nature being indifferent to good and evil (what Philo maintains is the most probable hypothesis; D 11.15, p. 212), and with the mortality of the soul and the absence of any future state (as Hume argues in *Of the Immortality of the Soul* and section 11 of the first *Enquiry*). On top of this, there is also the question, prompted by Gaskin’s rather rich interpretation of the notion,⁹ about the “proper office of religion”, which, according to Cleanthes, “is to regulate the heart of men, humanise their conduct, infuse the spirit of temperance, order, and obedience” (D 12.12, p. 220). Should this “proper office of religion” be built into Hume’s concept as well, as Gaskin maintains?

It is wrong, I think, to expect a definite answer to the question of how thick or thin Hume’s notion of true religion is. For when Hume is attacking superstition,

⁹Gaskin (1988, p. 188).

he is *deliberately* vague about the alternative; as well he might be, since true religion is in this context the face-saving escape route that he is strategically offering to his orthodox readers. The less he says about it here, the more likely such readers are to align themselves with it, and therefore to acquiesce in Hume's anti-superstitious arguments. And when Hume is not attacking superstition, meanwhile, but directly considering how much should go into true religion, the concept will necessarily be subject to much give and take, along with the flow of the arguments. This is particularly clear in the *Dialogues concerning Natural Religion*, which is after all a debate between three characters who all have very different opinions on what counts as true in matters of theology.¹⁰

However that may be, the two limitations that we can confidently place on Hume's true religion—that it includes the belief in (something like) an intelligent author of nature, but one who does not interfere with creation once it is set up—are really all we need to make the concept a useful one. For they enable us to explain both why true religion and superstition are classified together under the broader heading of religious belief, and how they chiefly differ from one another. Superstition is like true religion in as much as the superstitious also believe in invisible, intelligent power. It is *unlike* true religion, however, in that the superstitious believe that this invisible power is among the *immediate* causes of certain natural phenomena, and not just (or not even, depending on the superstition) the initial intelligent cause of the whole frame of nature.

If we replace Gaskin's distinction between reasons and causes with this distinction between superstition and true religion, a different way of framing Hume's discussions of religion emerges that is, I suggest, both less distorting of what he says and closer to his own way of seeing things. On this suggestion, Hume's discussions of religion divide into those explicitly attacking superstition on both moral and intellectual grounds (*Of Superstition and Enthusiasm*, *Of Miracles*, the

¹⁰For example, Demea speaks of “true Theists” at D 4.2 (p. 159), only to have Cleanthes claim immediately afterwards that such as Demea has in mind are really atheists without knowing it (D 4.3, p. 159). Philo once refers to “the true system of Theism” in an argument to the effect that it is inconsistent with Cleanthes' principles (D 5.2, p. 165).

Natural History of Religion, *Of Suicide*, and *Of the Immortality of the Soul*),¹¹ and those examining true religion with a view to seeing how much should go into it (*Enquiry* section 11, the *Dialogues*). The difference in content between these sets of works is also reflected by a striking difference in style: when discussing true religion, Hume is notably more cautious and guarded, using the dialogue form to distance himself from the debate and from his own controversial opinions; while in the other texts he speaks his mind directly in his own voice.

In the early essay, *Of Superstition and Enthusiasm*, Hume ostensibly criticises enthusiasm as well as superstition, both under the guise of “corruptions of true religion” or “species of false religion” (SE 1-2, p. 73). Given my intention to frame Hume’s critique of religion in terms of the superstition/true religion divide, something ought to be said about where enthusiasm fits into this picture. The obvious move, perhaps, is to include it alongside superstition, as a variation of the false religions that Hume openly attacks. But this would be to miss the irony in Hume’s essay. For although he begins by claiming that superstition and enthusiasm are both “pernicious” (SE 2, p. 73), albeit in different ways, as we proceed it quickly becomes apparent that he was not in earnest when he described the consequences of enthusiasm in this way.

The essay draws three comparisons between superstition and enthusiasm. The first is that “*superstition is favourable to priestly power, and enthusiasm not less or rather more contrary to it than sound reason and philosophy*” (SE 5, p. 75). But since Hume himself was obviously no friend to priestly power, this is really an attack on superstition that leaves enthusiasm shining by comparison. The second is that “*religions, which partake of enthusiasm are, on their first rise, more furious and violent than those which partake of superstition; but in a little time become more gentle and moderate*” (SE 7, p. 76). There is something of a criticism of enthusiasm in this, certainly, but on the whole the ensuing discussion emphasises how moderate and harmless the enthusiasts are now, in contrast with the superstitious:

¹¹In *Of Miracles* and *Of the Immortality of the Soul* Hume in fact ends by *endorsing* superstition (having first argued that it is irrational), on the grounds either of faith (E 10.40, p. 130) or of revelation (IS 1, 45, pp. 590, 598). But these endorsements are obviously ironic.

Superstition, on the contrary, steals in gradually and insensibly; renders men tame and submissive; is acceptable to the magistrate, and seems inoffensive to the people: Till at last the priest, having firmly established his authority, becomes the tyrant and disturber of human society, by his endless contentions, persecutions, and religious wars. (SE 8, p. 78)

The third and final comparison is that “*superstition is an enemy to civil liberty, and enthusiasm a friend to it*” (SE 9, p. 78). Here, even the pretence at criticising enthusiasm has disappeared altogether.

It is thus with a mild qualification that I place enthusiasm, for Hume, on the side of *true* religion rather than false, notwithstanding his ironic claim to the contrary at the start of this essay. The qualification is a consequence of the point already made about Hume’s notion of true religion being vague (indeed *deliberately* so). Because of this, we cannot definitively say whether enthusiasm should count as a part of it or not. Nor, to be clear, am I suggesting that Hume was himself an enthusiast; he wasn’t. But the superstitious—who are represented by Demea in the *Dialogues concerning Natural Religion*, and correspond to the rationalists or Platonists—are his primary target. In his works they are subject to several fierce and open attacks, on both moral and rational grounds. The enthusiasts, on the other hand, were much more acceptable to him (though he was of course sceptical of their metaphysics). Their representative in the *Dialogues* is Cleanthes, and they correspond to the sentimentalists or Stoics.

To return to the central issue, the only direct textual evidence in favour of Gaskin’s framework is the introduction to the *Natural History*, which on the surface does seem to carve things up in the same way that Gaskin does. But Hume’s apparent setting aside of the question of reasons here was most likely just a device to disarm any orthodox readers. Reading between the lines, we can see that the distinction between superstition and true religion is once again at play:

As every enquiry, which regards religion, is of the utmost importance, there are two questions in particular, which challenge our attention, to wit,

that concerning its foundation in reason, and that concerning its origin in human nature. Happily, the first question, which is the most important, admits of the most obvious, at least, the clearest solution. The whole frame of nature bespeaks an intelligent author; and no rational enquirer can, after serious reflection, suspend his belief a moment with regard to the primary principles of genuine Theism and Religion. But the other question, concerning the origin of religion in human nature, is exposed to some more difficulty. (N Intro, p. 33)

The discussion that Hume is setting aside here is not in fact the rational basis of religion *in general*, but only the rational basis of “genuine Theism”. And while he allows that the basis of *this* belief is sound, nothing at all has been said about the reasonableness of *superstitious* belief.

7.3. Hope and fear

What the above shows is that the introduction to the *Natural History of Religion* at least does not *rule out* the possibility that this work will contain an attack on the rationality of (superstitious) religious belief in addition to a discussion of its causes. When we come to look at the text directly, in §§7.4 and 7.5 below, we will see that it actually *does* contain just such an attack. Before turning to this argument, however, the text must be placed in its proper context, as part of Hume’s philosophy of emotion. It is the aim of the present section to lay out this background.

In the *Natural History* itself, Hume tells us that superstition would never have arisen but for the passions:

It must necessarily, indeed, be allowed, that, in order to carry men’s attention beyond the present course of things, or lead them into any inference concerning invisible intelligent power, they must be actuated by some passion, which prompts their thought and reflection; some motive, which urges their first enquiry. (N 2.5, p. 38)

But, Hume continues, “what passion shall we here have recourse to, for explaining an effect of such mighty consequence?” (ibid.). His answer is the passions of hope and fear: “the first ideas of religion arose not from a contemplation of the works of nature, but from a concern with regard to the events of life, and from the incessant hopes and fears, which actuate the human mind” (N 2.4, p. 38).

I will look at the details of Hume’s account in the next section. First, however, we need to acquaint ourselves with these two components of it, hope and fear.¹² The importance of these passions in Hume’s account is clear and explicit in the *Natural History* itself. As if to ensure that no one would miss the connection, Hume also originally published this work in a volume bound up with the *Dissertation on the Passions* immediately following it, a dissertation that begins with his account of the direct passions, with a particular emphasis on hope and fear: “None of these passions seem to contain anything curious or remarkable, except *Hope* and *Fear*, which, being derived from the probability of any good or evil, are mixed passions, that merit our attention” (P 1.7, p. 3). It is striking that commentaries on Hume seldom follow Hume’s own example and discuss these things together.

Before looking at what Hume says about the passions of hope and fear, we must take a step back to look at joy and grief or sorrow (Hume uses these last two terms interchangeably); for these, as we will see presently, are the ingredients of the mixed passions of hope and fear. On the one hand, joy and sorrow are simply pleasant and painful passions respectively, regarding some fact, present or past; the joy of a magnificent feast (T 2.1.5.1, p. 285), for example, or regarding the birth of a son (T 2.3.9.14, p. 441; P 1.21, p. 6), or the sorrow at losing a law-suit (ibid.). On the other hand, however, Hume also thinks of joy and sorrow

¹²These passions are also at the heart of Hume’s earlier and sketchier account in *Of Superstition and Enthusiasm*, where he attributes superstition to fear and other negative passions, and enthusiasm to hope and other positive emotions (SE 2-3, pp. 73-4). In the *Natural History*, with its focus on superstition, the emphasis is accordingly on fear rather than hope: “Agitated by hopes and fears of this nature, *especially the latter*, men scrutinize, with a trembling curiosity, the course of future causes, and examine the various and contrary events of human life. And in this disordered scene, with eyes still more disordered and astonished, they see the first obscure traces of divinity” (N 2.5, p. 39; my emphasis).

as pleasures and pains regarding *imagined* situations, possibly future. “When good is *certain or very probable*,” he tells us, “it produces joy”, and likewise for evil and sorrow (P 1.4, p. 3; my emphasis; cf. T 2.3.9.5, p. 439).

Hume’s idea that we experience joy in response to possible future pleasures as well as actual present ones seems to me somewhat puzzling. There is a precedent for the idea in Locke: “*Joy* is a delight of the Mind, from the consideration of the present *or assured approaching possession* of a Good” (1690, p. 231; second emphasis mine).¹³ This Lockean precedent, however, does not render the notion any the less problematic. I am not troubled by the idea that the anticipation of a future joy, as well as the present experience, can be or give rise to a kind of pleasure. What troubles me is the suggestion that there is one single passion—joy—for both of these cases. It seems to me only natural to distinguish *joy* from (let us say) *excitement*. The possession of a certain book will give me joy; having ordered it, I am pleasantly excited. But these are two very different feelings. Similarly with *sorrow* and (let us say) *apprehension*: loss of the book will make me sad, while consideration or anticipation of this loss instead makes me apprehensive.

Joy and sorrow are, for Hume, the ingredients of the mixed passions of hope and fear (P 1.10, p. 4). As one might expect, however, it is the *anticipatory* sort of joy and sorrow that Hume has in mind here, what might less misleadingly be termed (as above) *excitement* and *apprehension* respectively.

With this point cleared up, the next most remarkable feature of Hume’s account is the perhaps surprising extent to which *reason* is involved in the production of these passions. Hope and fear, for Hume, are not unthinking “gut reactions”, but rather felt responses to the intellectual exercise of weighing up probabilities:

¹³Locke’s definition of sorrow, by contrast, is restricted to real evil: “*Sorrow* is uneasiness in the Mind, upon the thought of a Good lost, which might have been enjoy’d longer; or the sense of a present Evil” (ibid.). But it is easy to see how one might extend the pleasant case to the unpleasant one, as Hume does.

Probability arises from an opposition of contrary chances or causes, by which the mind is not allowed to fix on either side; but is incessantly tossed from one to another, and is determined, one moment, to consider an object as existent, and another moment as the contrary. ... Suppose, then, that the object, concerning which we are doubtful, produces either desire or aversion; it is evident, that, according as the mind turns itself to one side or the other, it must feel a momentary impression of joy or sorrow [excitement or apprehension]. ... According as the probability inclines to good or evil, the passion of grief or joy [apprehension or excitement] predominates in the composition; and these passions being intermingled by means of the contrary views of the imagination, produce by the union the passions of hope or fear. (P 1.8-10, pp. 3-4; cf. T 2.3.9.10-2, pp. 440-1)

Given the extent to which reason is involved in the production of these passions, and the role that these passions then have in the generation of religious belief, it should, I hope, be immediately doubtful that Hume's account of the *causes* of this belief should be straightforwardly separable from the question of the *reasons* that might be given in support of it. At the very least, we have grounds for approaching the matter with an open mind. When we do approach the matter, in the next two sections, I will indeed argue that the two are intimately bound up with each other.

7.4. The origin of polytheism and monotheism

The origin of polytheism, Hume argues, lies in mankind's ignorance of the true causes of the several ills that face them, and in their natural tendency to "conceive all beings like themselves" (N 3.2, p. 40). Thus the "*unknown causes*" become intelligent agents (albeit invisible), with powers much greater than our own, but with "thought and reason and passion" just like us (N 3.2, p. 41). What motivates this postulation of invisible intelligence, Hume says, is not "speculative curiosity... or the pure love of truth", but rather "the anxious concern for happiness, the dread of future misery, the terror of death" (N 2.5, p. 38). Our hopes and fears

give us a very pressing need to know what causes “the various contrary events of human life” (ibid.), so that we can gain some control over them.¹⁴

As we have seen, Hume does not consider hope and fear to be immediate unthinking emotional responses. On the contrary, they are mediated by judgments of probability, i.e. of cause and effect. It is only to be expected, therefore, that *reasoning* concerning cause and effect is going to be relevant for Hume here, and that his discussion is not likely to be *purely* descriptive or psychological. Sure enough, Hume is quite clear and explicit that this first form of religious belief is *contrary to the evidence*:

Could men anatomize nature, according to the most probable, at least most intelligible philosophy, they would find, that these causes are nothing but the peculiar fabric and structure of the minute parts of their own bodies and of external objects; and that, by a regular and constant machinery, all the events are produced, about which they are so much concerned. But this philosophy exceeds the comprehension of the ignorant multitude, who can only conceive the unknown causes in a general and confused manner; though their imagination, perpetually employed on the same subject, must labour to form some particular and distinct idea of them. (N 3.1, p. 40)¹⁵

Polytheism, as far as Hume is concerned, is thus *bad science*. There is nothing *a priori* objectionable about it; the belief in invisible, intelligent power is perfectly consistent and possible. Hume makes this point particularly clear in the ambitious claim that “the whole mythological system is so natural, that, in the vast variety

¹⁴Hume does not think that we arrive at the belief in invisible, intelligent power from some desire to *explain* the world around us, but from the much more basic need for *control* (or at least the illusion of it). The contrary interpretation was suggested by Yandell (1990), but refuted very thoroughly by Ferreira (1995).

¹⁵This passage is reminiscent of one in the first *Enquiry*: “A peasant can give no better reason for the stopping of any clock or watch than to say that it does not commonly go right: But an artist easily perceives, that the same force in the spring or pendulum has always the same influence on the wheels; but fails of its usual effect, perhaps by reason of a grain of dust, which puts a stop to the whole movement” (E 8.13; p. 87; cf. T 1.3.12.5, p. 132). Superstitious believers are in the same epistemic situation as the peasant; but instead of searching for the grain of dust they allow their imaginations to conjure up an idea of some invisible supernatural agent responsible for keeping watches ticking (or at least with the power to stop them).

of planets and worlds, contained in this universe, it seems more than probable, that, somewhere or other, it is really carried into execution” (N 11.1, p. 65). We needn’t take this suggestion about the existence of gods on other planets too seriously. Toning down the rhetoric, we nevertheless have the important point that the existence of supernatural agents is, for Hume, an unproblematic *a priori* possibility. “The chief objection to it with regard to this planet,” he goes on, “is, that it is not ascertained by any just reason or authority” (N 11.2, p. 65).

In offering his account of the psychological *causes* of polytheism, therefore, Hume has been unable to refrain from saying that this particular brand of superstition is *irrational*. The argument here is presumably very obvious to many twenty-first-century readers. But, obvious or not, it is an argument against the rationality of (one form of) religious belief. Certainly Hume is pursuing a psychological investigation in the *Natural History*; but that isn’t the *only* thing he is doing.

Monotheism differs from polytheism only in the insistence that there is but one true God. Hume charts the emergence of such forms of superstition in the felt need to flatter one deity above all the others, because this particular god is supposed to be in charge either of the believers’ own nation, or of all the other gods (N 6.5, p. 53). Eventually, after sufficient exaggeration of the praises bestowed on this god, all others are seen as vain pretenders. And perhaps this supreme deity will even be credited with creating the world, in which case his votaries “coincide, by chance, with the principles of reason and true philosophy” (N 6.5, p. 54). But as long as this deity, however supreme, is still supposed to disturb “the settled order of events by particular volitions” (N 6.2, p. 52), the belief is unjustified, founded on “*irrational* and superstitious principles” (N 6.4, p. 53; my emphasis).

7.5. The miraculous and the momentous

To be clear, it is not just that the religious beliefs in question (both polytheistic and monotheistic) have, as a matter of historical fact, been founded on irrational

principles, while there might be other paths to these beliefs that are rational. The only rational path that Hume countenances to religious belief ends in true religion, which only allows for the existence of a *non-interventionist* god. The superstitious, however, believe in the existence of at least one divinity who interferes in the settled order of things, and this belief is not just *arrived at* by irrational principles, according to Hume, but is *itself* irrational, since all the evidence available points to natural causes of the contentious phenomena.

This *a posteriori* argument that there is no evidence for supernatural intervention will doubtless call to mind Hume's famous (and much more widely discussed) argument against the testimonial evidence of miracles, in section 10 of the first *Enquiry*. Perhaps it will even be thought that, when Hume claims that there is no evidence of supernatural intervention in the *Natural History*, he is merely recalling that argument; for of course testimony of miraculous events purports to be just that. The *Natural History* is not simply referring back to an argument given elsewhere, however, but offering a different argument of its own. In doing so, furthermore, it is filling an important gap that would otherwise be left in Hume's attack on religion, as I now explain.

Perhaps it is easily forgotten that the superstitious see the work of supernatural agency, not just in miraculous events, but also in natural events that are suitably *momentous*. Droughts are nothing miraculous, but they are of such devastating consequence that the superstitious are apt to see them as the result of divine wrath. It is no miracle, either, that someone should recover from a potentially (but not necessarily) fatal illness, but when a superstitious friend has been praying for this very thing, they will readily see the hands of a particular providence at work:

Even at this day, and in EUROPE, ask any of the vulgar, why he believes in an omnipotent creator of the world; he will never mention the beauty of final causes, of which he is wholly ignorant... He will tell you of the sudden and unexpected death of such a one: The fall and bruise of such another: The excessive drought of this season: The cold and rains of another. These

he ascribes to the immediate operation of providence: And such events, as, with good reasoners, are the chief difficulties in admitting a supreme intelligence, are with him the sole arguments for it. (N 6.1, p. 52)

The issue in the case of purported miracles is not whether supernatural agency was responsible, but whether the supposed miracle actually took place. Thus Hume's argument in *Of Miracles* is designed to show that reports of miraculous events are never sufficiently credible, given the antecedent unlikelihood of the event reported, to establish that it genuinely occurred. The issue in the present case, on the other hand, is not whether the event hoped for or dreaded really took place, but whether anything divine was directly responsible. Hume's argument in the *Natural History*, accordingly, does not attempt to undermine the reliability of testimony. Rather, Hume points to the observed order and regularity in the world: when we "anatomize nature" more closely, we discover that the causes of these momentous events are in fact "nothing but the peculiar fabric and structure of the minute parts of [our] own bodies and of external objects; and that, by a regular and constant machinery, all the events are produced, about which [we] are so much concerned" (N 3.1, p. 40).

Thus we see that Hume's psychological discussion of the causes of religious belief is intimately bound up with the philosophical issue of the *arguments* or *evidence* in support of it. The distinction between reasons and causes, while itself perfectly legitimate and known to Hume, just is not a distinction that can cleanly be applied to his discussions of religion in general, or to the *Natural History* in particular.

Chapter 8

Some Comparisons

In the previous chapter, I argued against viewing the *Natural History of Religion* as concerned exclusively with the causes of religious belief and not at all with its rational credentials. I also argued, in general, against framing Hume's critique of religion in terms of this distinction, preferring instead the distinction between superstition and true religion. In this chapter, I aim to provide some additional support for these claims, and in various ways to complete my picture of the first of Hume's four dissertations. I shall do this through a series of comparisons: between Hobbes and Hume, between polytheism and monotheism, between my own interpretation and some others, and finally between the earlier and the later Hume on this matter.

8.1. Hobbes on the origin of religious belief

In part 1, chapter 2 of his *Leviathan*, Hobbes briefly attributes "the greatest part of the Religion of the Gentiles in time past" to the "ignorance of how to distinguish Dreams, and other strong Fancies, from Vision and Sense" (1651, p. 18). His point, simply enough, is that vivid imaginings and daydreams may pass for reality, giving rise to a belief in all manner of fanciful things. Towards the end of chapter 11, however, and in chapter 12, he sets out in more detail what he takes to be the main causes of religious belief.

Polytheism, first, is attributed by Hobbes to the ignorance of natural causes and fear regarding an uncertain future:

[T]hey that make little, or no enquiry into the naturall causes of things, yet from the feare that proceeds from the ignorance it selfe, of what it is that hath the power to do them much good or harm, are enclined to suppose, and feign unto themselves, severall kinds of Powers Invisible; and to stand in awe of their own imaginations; and in time of distresse to invoke them; as also in the time of unexpected good successe, to give them thanks; making the creatures of their own fancy, their Gods. (1651, p. 75)

Fear, thinks Hobbes, is the inevitable upshot of our ignorance, since we are of course so very concerned for our own future. And fear must needs have its object: “therefore when there is nothing to be seen, there is nothing to accuse, either of their good, or evil fortune, but some *Power*, or Agent *Invisible*” (1651, p. 76).

The similarities with Hume’s account of the origin of polytheism are striking. There is no explicit mention by Hobbes of our natural tendency towards anthropomorphism, but some such idea seems to be implicit later on in his explanation of our behaviour towards supernatural agents:

[F]or the worship which naturally men exhibite to Powers invisible, it can be no other, but such expressions of their reverence, as they would use towards men; Gifts, Petitions, Thanks, Submission of Body, Considerate Addresses, sober Behaviour, premeditated Words, Swearing (that is, assuring one another of their promises,) by invoking them. (1651, p. 78).

Otherwise the two key components of Hobbes’s account—ignorance of natural causes, and fear for an uncertain future—are exactly the same as in Hume’s account. Hume’s discussion is much longer than Hobbes’s, and his theory is worked out more fully, with more offered by way of evidence. But the basic account itself was not new with Hume.¹

¹Before he worked out his account in more detail, Hume presented, in *Of Superstition and Enthusiasm*, an explanation of the origin of superstitious belief more closely in line with Hobbes: “The mind of man is subject to certain unaccountable terrors and apprehensions, proceeding either from the unhappy situation of private or public affairs, from ill health, from a gloomy and melancholy disposition, or from the concurrence of all these circumstances. In such a state of mind, infinite unknown evils are dreaded from unknown agents; and where real objects of terror are wanting, the soul, active to its own prejudice, and fostering its predominant inclination, finds imaginary ones, to whose power and malevolence it sets no limits” (SE 2, pp. 73-4).

Regarding monotheism, however, Hobbes and Hume differ more significantly. According to Hobbes, the belief in monotheism has a more respectable explanation, namely *curiosity* satisfied by the cosmological argument that there must be a unique first cause:

But the acknowledging of one God Eternall, Infinite, and Omnipotent, may more easily be derived, from the desire men have to know the causes of naturall bodies, and their severall vertues, and operations; than from the feare of what was to befall them in time to come. For he that from any effect hee seeth come to passe, should reason to the next and immediate cause thereof, and from thence to the cause of that cause, and plunge himselfe profoundly into the pursuit of causes; shall at last come to this, that there must be (as even the Heathen Philosophers confessed) one First Mover; that is, a First, and an Eternall cause of all things; which is that which men mean by the name of God[.] (1651, p. 77)

Hume, as we have seen, suggested instead that reason was no more the origin of monotheism than it was of polytheism. Rather, it arises out of this more primitive superstition, and in much the same way: the idolatrous “are guided to that notion, not by reason, of which they are in a great measure incapable, but by the adulation and fears of the most vulgar superstition” (N 6.5, p. 54).

For this reason, Hume devotes the first section of the *Natural History* to arguing that polytheism was prior to monotheism. This is a claim that Hobbes had no need to make or defend. In this, and in the attendant account of the origin of monotheism, we have one of Hume’s most significant original contributions to the psychology of religious belief.

8.2. Comparisons between monotheism and polytheism

It is well known that Hume’s objections to superstition were *moral* as well as intellectual, a point that I already made in passing in the previous chapter, but have not yet sufficiently emphasised. In particular, it has not escaped scholarly notice that the *Natural History*, in addition to offering an account of the causes of

religious belief, also spends some time discussing what Hume saw as its pernicious moral effects.² This is the focus of most of sections 9-12 (for the remaining part, see §8.3 below), in which Hume compares monotheism and polytheism in various respects, typically pertaining to morality, and with the former almost always coming off the worse of the two. In section 14 he then argues explicitly for the “[b]ad influence of popular religions on morality” (N 14, title, p. 81).

Given this moral dimension to the text, it is clear (even by Gaskin’s own admission) that this work is not *solely* concerned with causes. The distinction between reasons and causes, indeed, leaves no room at all for a discussion of religion’s *effects*, on morality or anything else. When we see the *Natural History* instead as part of Hume’s overall attack on superstition, however, everything falls very neatly into place; for this attack is comprehensive, challenging superstition from the point of view of both reason *and* sentiment.

If the introduction to the *Natural History* is intended to be disarming, by giving the impression that what follows will concern only causes and not reasons, the introductions to Hume’s other attacks on superstition are typically more forthright:

I flatter myself, that I have discovered an argument... which, if just, will, with the wise and learned, be an everlasting check to all kinds of superstitious delusion, and consequently, will be useful as long as the world endures. (E 10.2, p. 110)

One considerable advantage, that arises from philosophy, consists in the sovereign antidote, which it affords to superstition and false religion... [S]uperstition, being founded on false opinion, must immediately vanish, when true philosophy has inspired juster sentiments of superior powers. (Su 1, pp. 578-9)

That the corruption of the best things produces the worst, is grown into a maxim, and is commonly proved, among other instances, by the pernicious

²See Penelhum (1975, p. 173-4), Gaskin (1988, chapter 11), Webb (1991), Ferreira (1999).

effects of *superstition* and *enthusiasm*, the corruptions of true religion.
(SE 1, p. 73)

This last quotation, from the essay *Of Superstition and Enthusiasm*, also sets the tone for the comparisons between polytheism and monotheism in the *Natural History*. Hume dresses up this latter, thinly-disguised attack on monotheistic religions with the same maxim—“that the corruption of the best things begets the worst” (N 11.1, p. 65)—politely allowing that monotheism in the form of true religion is better than pagan idolatry, while enabling him to pull no punches in his criticism of its superstitious corruptions.

The importance of the *Natural History* in Hume’s overall attack on the morality of superstition should not be overlooked, any more than in his overall attack on its rationality. It is of course in the moral *Enquiry* that Hume develops his own secular account of moral distinctions, without which this attack would have no solid foundation. And it is in the moral *Enquiry* that Hume uses this foundation to argue that the “monkish virtues” of “[c]elibacy, fasting, penance, mortification, self-denial, humility, silence, solitude” are really vices (M 9.3, p. 270). But the *Natural History* provides a very important addition to this argument, by offering an explanation of why the superstitious are led to the practice of these monkish virtues:

The duties, which a man performs as a friend or parent, seem merely owing to his benefactor or children; nor can he be wanting to these duties, without breaking through all the ties of nature and morality. A strong inclination may prompt him to the performance: A sentiment of order and moral obligation joins its force to these natural ties: And the whole man, if truly virtuous, is drawn to his duty, without any effort or endeavour. ... In all this, a superstitious man finds nothing, which he has properly performed for the sake of this deity, or which can peculiarly recommend him to the divine favour and protection. He considers not, that the most genuine method of serving the divinity is by promoting the happiness of his creatures. He still looks out for some more immediate service of the supreme Being, in

order to allay those terrors, with which he is haunted. And any practice, recommended to him, which either serves to no purpose in life, or offers the strongest violence to his natural inclinations; that practice he will the more readily embrace, on account of those very circumstances, which should make him absolutely reject it. It seems the more purely religious, because it proceeds from no mixture of any other motive or consideration. (N 14.6, pp. 82-3)

Because there are, in human nature, motivations to moral behaviour that do not involve any reference to the divine (as argued in the moral *Enquiry*), such behaviour is only coincidentally motivated (if at all) by the desire of pleasing God. In order to demonstrate their devotion, therefore, and to secure divine favour, the superstitious must needs pursue actions that are in no way recommended by, or are perhaps even contrary to, our natural benevolent instincts: “if he [the superstitious man] fast a day, or give himself a sound whipping; this has a direct reference, in his opinion, to the service of God. No other motive could engage him to such austerities” (ibid.).

Where the moral *Enquiry* offers a criticism of the monkish virtues, therefore, the *Natural History* turns this into a criticism of *superstition*, by arguing that the former are the natural consequence of the latter. The connection between superstition and the monkish virtues is *observed* in the moral *Enquiry*, but it is only in the *Natural History* that it is *explained*.

8.3. The bull-rush argument

Most of Hume’s comparisons between polytheism and monotheism in the *Natural History* are moral, as we have seen above. Section 11, however, compares the two “[w]ith regard to reason or absurdity” (N 11, title, p. 65). (This was the exception that I had in mind in my qualification at the start of §8.2, page 156.) What Hume has to say about polytheism in this connection was already noted in passing in the previous chapter (§7.4, page 148): he takes it to be a straightforward *a priori* possibility, and therefore not inherently absurd, but one that happens to

be contradicted by the evidence. Yet again, monotheism comes off worst from the comparison, for it frequently lands itself in *a priori* absurdity in addition to involving claims that are contrary to experience.

Hume provocatively suggests that an “appetite for absurdity and contradiction” (N 11.3, p. 66) is almost inevitably coincident with popular theology:

If that theology went not beyond reason and common sense, her doctrines would appear too easy and familiar. Amazement must of necessity be raised: Mystery affected: Darkness and obscurity sought after: And a foundation of merit afforded to the devout votaries, who desire an opportunity of subduing their rebellious reason, by the belief of the most unintelligible sophisms. (ibid.)

Hume doesn’t mention any such “sophisms” explicitly, or spell out the arguments of “rebellious reason” against them; but the insinuation is already clear enough. There are as many *a priori* objections to monotheistic superstitions, Hume thinks, as there are absurdities associated with them.

Hume is admittedly pessimistic about the persuasive force of these *a priori* objections. In a memorable passage, he writes that “[t]o oppose the torrent of scholastic religion by such feeble maxims as these, that *it is impossible for the same thing, to be and not to be*, that *the whole is greater than a part*, that *two and three make five*; is pretending to stop the ocean with a bull-rush” (N 11.5, p. 66). But there is no denying that this bull-rush, though it only contradicts *some* forms of superstition, is another part of Hume’s general attack on the rationality of religious belief.

It seems to me that the *a posteriori* argument against superstition that Hume offers alongside his account of its causes, which targets belief in momentous but non-miraculous events (recall §§7.4 and 7.5), is intrinsically rather more interesting than the bull-rush argument of section 11. But this latter argument is of indirect interest, in providing the nail in the coffin of Gaskin’s framework and his interpretation of the *Natural History* as a purely psychological work. When we view Hume’s dissertation in this way, the bull-rush argument becomes eerily

invisible. But there it is all the same, together of course with the moral critique. When we replace Gaskin's distinction with that between superstition and true religion, meanwhile, and see the *Natural History* as just one more of Hume's several works aimed at criticising the former on both moral and intellectual grounds, we can comfortably make sense of the text in its entirety.

8.4. Hume's agnosticism and Hume's atheism

Though mine seems to be a minority view, I am not altogether alone in thinking that the *Natural History of Religion* contributes to Hume's discussion of reasons as well as causes. Recently Peter Kail has defended precisely this claim:

The prevalent view is incorrect. *NHR* [the *Natural History of Religion*] is a philosophically important and powerful component in Hume's campaign against the rationality of religious belief.³

Kail's reasons for thinking this, however, are quite different from my own. And while I welcome the conclusion, I disagree with the argument that gets him there. I am anxious, therefore, to distance myself from Kail's position.

According to Kail, the *Natural History* constitutes Hume's argument against a view that Kail calls "rational fideism": "the position that it is rational to maintain religious belief in the *absence* of evidence or arguments in its favour".⁴ The kernel of Hume's supposed argument is that some causal explanations of belief are *destabilizing*, in the sense that they give one a reason to suspend the belief in question in the absence of any evidence in support of it (for example, believing something merely because of being told by a habitual liar); and that Hume's explanation of religious belief is of this kind. Furthermore, Kail thinks that Hume's attack on religion is incomplete without an argument against the rational fideist. For although Hume argues elsewhere that there is no evidence

³Kail (2007b, p. 191).

⁴Kail (2007b, p. 195).

for religious belief, this does nothing to upset the fideist who thinks it is rational to believe even in the absence of such evidence.

There seem to me to be two objections to the suggestion that Hume was offering this argument in the *Natural History*. The first is the obvious one: that if he *was* offering it, he certainly had a curiously *secretive* way of doing so. The rational fideist is never mentioned or even alluded to by Hume as a target, nor does he ever so much as hint at the crucial premise in Kail's argument (that his explanation of the causes of religious belief is destabilizing in Kail's sense). The argument may well have been *inspired* by Hume, but Kail is unable to produce any textual evidence that shows *Hume himself* actually rehearsing it. The case is quite different, note, with the *a priori* bull-rush argument noted in the previous section, and the *a posteriori* argument that I looked at in the previous chapter; uncovering those arguments required no imagination on my part, but emerged from a simple exercise in comprehension.

The second difficulty for Kail is that this argument against rational fideism, were Hume to endorse it, would involve conceding *too much* to his opponents. For Hume does not in fact believe that there is an *absence* of evidence in the present case: as we have seen, he is quite explicit in the *Natural History* that the evidence is *against* the existence of any supernatural intervention: "Could men anatomize nature, according to the most probable, at least most intelligible philosophy, they would find, that these causes are nothing but the peculiar fabric and structure of the minute parts of their own bodies and of external objects" (N 3.1, p. 40). In Hume's attack on superstition, therefore, there is quite simply no gap for the rational fideist to move into.

Kail is not the first commentator to have read arguments into the *Natural History of Religion* that are not actually there. Gaskin himself, though a major promoter of the idea that this work is concerned exclusively with causes, does at least allow that Hume's psychological story has philosophical *implications*, in forming the basis of a reply to the argument from common consent:

[A]n adequate account of the *causes* of religious belief, though not itself philosophy, is an essential complement to [Hume's] philosophical thinking about religion. If, as Hume seems to conclude, the arguments of natural religion are bad or establish only a hesitant and highly attenuated conclusion, and if the authenticity of revelation is suspect, then an appeal will almost inevitably be made by the believer to the argument from general consent: why is it that religious belief is and always has been so very prevalent?⁵

My objection to this suggestion is exactly the same as to Kail's. While Hume *might* have used the account of the origins of religious belief given in the *Natural History* as the basis of a reply to the argument from common consent, there is no line of text in this work where he actually *does*, or even so much as hints at the possibility. Furthermore, to repeat, Hume doesn't merely maintain that the arguments in favour of superstition are suspect or unsuccessful; he also insists that there is positive evidence in the *other* direction, in the form of the observed regularities in nature.

The situation is analogous in the two essays *Of Suicide* and *Of the Immortality of the Soul*, where Hume doesn't only criticise the arguments in favour of life after death and the immorality of suicide, but offers positive arguments for the *opposite* conclusions. In the latter, for example, he writes that “[t]he *physical* arguments from the analogy of nature are strong for the mortality of the soul; and these are really the only philosophical arguments, which ought to be admitted with regard to this question, or indeed any question of fact” (IS 30, p. 596). His position is not that there is an *absence* of evidence. His position is that the evidence tells *against* superstition.

There is a general sense—and it might be this that underlies the assumption common to Kail and Gaskin that I am here opposing—that Hume tended more towards agnosticism than outright atheism. After all, isn't that the appropriate stance for a Sceptic to take? There may be some truth to this; I do not have the space or the need to delve into the matter deeply here. But whatever we ought

⁵Gaskin (1988, p. 183).

to say about this, we must *start* with the distinction between superstition and true religion. If there is any sense in which Hume was agnostic, it only concerns true religion. When it comes to superstition, as I have said, he is openly and unequivocally opposed to it all the way down. Perhaps true religion (in some attenuated form) is consistent with the known facts, and with a just sense of morals. But superstition, for Hume, is in accordance with neither.

8.5. The three *Enquiries*

Before moving on from the *Natural History of Religion*, it is time to bring out explicitly what has only been implicit in the discussion of this chapter and the previous one, namely the evidence that this dissertation offers in support of my Unity Thesis and Difference Thesis. Regarding the Unity Thesis, the point is perhaps already obvious. Hume's account of the origin of superstition depends crucially on the passions of hope and fear, which receive their explanation, in turn, in the *Dissertation on the Passions*. Thus without Hume's philosophy of emotion, his account of the causes of religious belief remains incomplete. And there is a *value* to appreciating this context too, since an attention to the role of reason in the generation of hope and fear can help to undermine the popular thought that Hume's explanation of religious belief is devoid of rational implications.

If we keep Hobbes in mind, we should not be at all surprised by the fact that Hume published the *Natural History of Religion* together with the *Dissertation on the Passions*. Quite aside from the *particular* connection between these two works, namely the passions of hope and fear, this combination was not new with Hume. In part 1 of the *Leviathan*, Hobbes offers his account of motivation and the passions (mainly in section 6) before applying it to the phenomenon of religious belief (predominantly in section 12). The *Four Dissertations* are the same in this respect, save only that Hume presents the religious application *before* the background psychological theory. In claiming that these things belong together in Hume, I am not suggesting anything idiosyncratic on his part.

The contribution that the *Natural History* makes to the Difference Thesis, meanwhile, should be equally obvious, though not from the preceding discussion. Hume did not change his mind about anything offered in this dissertation, so far as I can discover. But the simple fact is that he *added* a considerable amount that is nowhere to be found in the *Treatise*. And additions, no less than changes, constitute substantial differences. This might seem like too crude a point to bother mentioning at all. Given the extraordinarily disproportionate amount of attention that is lavished on the *Treatise*, however, at the expense of Hume's later works, it is well worth making.

Of course, the addition of the *Natural History* only counts as a difference between Hume's earlier and later philosophy of *emotion* on the assumption of my Unity Thesis. And it might be thought that I am somewhat overstating the case here. Even if it is allowed that Hume's account of the causes of religious belief has close ties to his philosophy of the passions, the fact is that the *Natural History* includes more than just this account. There is also the bull-rush argument, for example, and the moral critique of superstition. Don't these show that this dissertation cannot really be thought of as a part of Hume's *Enquiry concerning the Passions*?

As I see things, however, the fact that the *Natural History* oversteps the bounds of the philosophy of emotion in these ways only serves to *strengthen* my hypothesis, for two reasons. First, this is exactly the pattern that we see when moving from Books 1 and 3 of the *Treatise* to the two *Enquiries*, with the addition of the attack on belief in miracles in section 10 of the first *Enquiry* (not to mention the cautious discussion of true religion in section 11), and the criticism of the monkish virtues in section 9 of the moral *Enquiry*. The psychological discussion in part 2 of section 10 of the first *Enquiry* is no more out of place in a Humean enquiry concerning the *understanding*, than the philosophical aspects of the *Natural History* are in a Humean enquiry concerning the *passions*.

Secondly, these philosophical aspects of the *Natural History* complement the two *Enquiries* in important ways, notably in the argument against belief in

momentous events that parallels the more famous one concerning miracles (recall the previous chapter, §§7.4 and 7.5), and in the criticism of the bad influence of superstition on morality for which the groundwork was laid in the moral *Enquiry* (recall §8.2 above). My Unity Thesis focuses on the integration of the *Four Dissertations* amongst themselves, but what we have seen of the *Natural History* in this chapter suggests that they have a tight connection with Hume's mature philosophy as a whole. The *Four Dissertations* are not only united with each other by their shared concern for the passions; they are also closely connected to the two *Enquiries*. They are, I submit, justly viewed as Hume's *Enquiry* concerning the passions.

Chapter 9

The Intentionality of the Passions

Back in §3.1, I insisted that the passions, for Hume, are not bodily sensations, but *sentiments* (a term that is nicely ambiguous between “feeling” and “opinion”), and that we should resist the temptation to classify Hume as a crude non-cognitivist about emotions. In this chapter—before turning, in the next chapter, to Hume’s famous views on the relationship between reason and passion—I will expand on this claim. I will also examine more fully Hume’s theory of the double relation of sentiments and ideas (first introduced in §4.1), arguing that an error surrounding this theory in the *Treatise* prompted Hume to make his account of some of the passions even more cognitivist in his later work.¹

9.1. Cognition and the passions

“Cognition” has become a somewhat technical term, with many philosophers using it slightly different ways. I shall try to be as clear as possible about the sense that I attach to it. I have in mind three closely related mental powers that seem to come as a package, and which many animals have, but human beings in a particularly high degree. The central one is *conscious awareness*, both of oneself and of one’s surrounding environment (together with the ability to identify aspects of the latter as things of a certain kind; food, shelter, predator, potential

¹I first presented this argument in Merivale (2009), and my opinion has not changed since then. In some respects the present discussion is briefer; in others, I hope, it is a little clearer and more refined.

mate, and so on). The surrounding powers are the retention of this awareness in the form of *memory*, and the the imaginative ability to *anticipate*, or more generally to *speculate* about the unobserved (past and present, as well as future). This speculation may involve deductive inferences, in such animals as are capable of them, but in general I would place much greater emphasis on induction in this connection. Certainly I mean to exclude abstract (mathematical, philosophical) reasoning from my concept of cognition; this strikes me as a quite different ability, unique to human beings, and a much later evolutionary development. However that may be, it is only the kinds of cognitive powers that we have in common with other animals that matter in the context of the present discussion.

An extreme non-cognitivist view of the passions would see them simply as bodily sensations, or at any rate as akin to bodily sensations in respect of their relationship to cognition. On this view, saying that I am happy because you are here, is just like saying that I am warm because the heating is on. In both cases, there is an external stimulus, and contact with that stimulus causes a certain feeling. The room that I am in is hot, and consequently I have a feeling a warmth. Just so, I see that you are here, and consequently experience a feeling of elation; or I see a snake, and consequently experience a feeling of fear; and so on. Hume's placement of the passions alongside sensations, in the broader category of impressions or sentiments, invites the thought that he was an extreme non-cognitivist of this sort. And indeed, many have been attracted to something very like this interpretation.²

Perhaps the most immediate problem for this extreme view is that cognition seems to be necessarily or at least typically involved in the causal processes that give rise to emotions, in a way that it is not for bodily sensations. I need not be aware that the heating is on in order to feel warm; the temperature will have this effect on me regardless. But I do need to be aware that you are here for this fact to make me happy. Furthermore, *errors* in cognition can give rise to emotions, something for which there is nothing analogous in the case of bodily feelings. If I don't realise that you have left, I may continue to be glad that you are here; or I

²Pitcher (1965), Solomon (1977), Deigh (1994), Zemach (2001).

may be glad when someone tells me that you're here, although in fact you are not. Finally, we also have emotional responses to imagined possible situations. For example, I can hope that you will come, or be pleased by the thought that you might, even though you haven't come yet, and even if you don't end up coming at all. Sometimes, then, cognition is the sole cause of our emotions; and even when there is some external stimulus, cognition seems generally to be a necessary intermediary.³

All of this, however, is something Hume is in complete agreement with. We have seen the evidence for this already, but may now draw attention to it explicitly. Hume in fact holds the very strong view that cognition is *always* the immediate cause of *almost all* of the passions. (The exceptions are some—but not all—of our original, object-directed desires, which are caused, insofar as they have a cause at all, simply by our motivational make-up; see §9.4 below.) Joy and sorrow are caused by the ideas of whatever it is that we are glad or unhappy about. Hope and fear are caused by judgments of probability (recall §7.3).⁴ Pride, humility, love, and hatred are caused by associations of sentiments and ideas (recall §4.1; I will examine this aspect of Hume's view again in §9.5 below). Malice and envy are caused by comparisons of ourselves with others. And pity and generosity are caused by the sympathetic communication of passions, a process which crucially involves awareness of oneself, of the other person, and an idea of the passion being communicated. Kames, as we may recall from §3.1, defined a sentiment as a “thought prompted by passion” (1762, vol. 1, p. 311). For Hume, meanwhile, passions or sentiments—aside from the few exceptional desires mentioned above—could be accurately described as *feelings prompted by thought*. In this respect, at least, Hume was not the extreme non-cognitivist that he is so often considered to be.

³*Generally*, but maybe not always. Perhaps, for example, there is an immediate fear reaction to certain common and particularly dangerous stimuli that bypasses any conscious awareness or classification of the threat. Such a reaction would presumably be quicker than one mediated by cognition, and therefore come with an obvious evolutionary benefit. This is of course not to deny that a cognitively mediated response, though more costly, has some compensating advantages: it would function more widely and reliably, and adapt to a changing environment.

⁴Thus there is, for Hume, no immediate fear response of the sort hypothesised in note 3 above. I am inclined to think that Hume's view was probably *too* cognitivist in this respect.

It is one thing to hold that the passions are *caused by* cognition, but what about the inherent nature of the passions themselves? Opinions on this matter can vary enormously, both in kind and in sophistication. It will be sufficient for present purposes to have two fairly crude views in place for consideration, at different ends of the cognitivist/non-cognitivist spectrum. The simplicity in my statement of these views is quite deliberate: we should be wary of saying too much here on Hume's behalf, for the truth is that he never thought the matter through in any detail himself. Nevertheless, it is profitable to imagine what he might have been able to say, consistently with his explicit commitments, even if only to better acquaint ourselves with those other commitments.

Cognitive states—be they states of present awareness, memories, or speculations concerning the unobserved—are all *about* something or other (real or imagined). They are *intentional* states. When I am aware that you are here, my awareness has a certain intentional content; this content is, as we might say, *that you are here*. I take it for granted that the passions also have intentional content, specifically the same intentional content that their immediate cognitive causes have. When I am glad *because* you are here, I am also glad *that* you are here. Unfortunately, Hume is often thought to have denied this surely undeniable fact. I will argue in the next section that he did not.

More interesting in this connection is the question of *how* or *in what way* the passions have this intentional content. It is in their different answers to this question that my two somewhat crude views about the nature of the passions distinguish themselves. The view at the non-cognitivist end of the spectrum holds that the passions have intentional content only in a derivative way, by virtue of being appropriately related to the cognitive state that causes them (a state which presumably has that content in an intrinsic, non-derivative way). Perhaps merely being the effect of this cognitive state is enough, though that seems unlikely; a plausible first stab might be to add that one must also be *aware* of this causal connection. I see that you are here, I feel glad, and I realise, furthermore, that my joy is the result of your presence; thus my feeling comes to be about the fact that you are here, by a causal connection, and a conscious association.

The more cognitivist view, by contrast, holds that the passions have intentional content *in their own right*, and in exactly the same way that cognitive states do. It holds, in fact, that the passions themselves actually are special kinds of cognitive states. The difference between merely being *aware* that you are here, on this view, and being *glad* that you are here, does not consist in the addition of some separate feeling (such as the non-cognitivist view proposes). Rather, the difference is in the *manner* of my awareness: I can be aware of your presence in an emotionally neutral or indifferent way, or I can be aware of your presence in a glad way.

Translating these alternative views as well as possible into Humean terms, we may safely say that Hume's ideas are cognitive states that have intrinsic intentional content. The question, then, is whether sentiments or passions have intentional content only *derivatively*, by virtue of being appropriately related to some idea, or whether they have some *intrinsic* intentional content, in the same way that ideas do. This is the question that I want to address in the remainder of this chapter. The answer to it is not, as we will see, a simple either-or matter. In the *Treatise*, I will argue, Hume held that the intentionality of the passions was entirely derivative. In the *Dissertation on the Passions*, however, he made some steps in the direction of a more cognitivist view, insisting that *some* of the passions—namely the double-relation passions (pride, humility, love, and hatred)—have *some* intrinsic intentional content.

9.2. The representative quality argument

I said above that Hume does not deny the intentionality of the passions, but before going any further I need to justify this preliminary claim. For there is one particular (and famous) passage from the *Treatise* in which he appears to deny precisely this. Following Phillips,⁵ I will refer to this as Hume's *representative quality argument*:

⁵Phillips (2005, p. 308).

A passion is an original existence, or, if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possest with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high. 'Tis impossible, therefore, that this passion can be oppos'd by, or be contradictory to truth and reason; since this contradiction consists in the disagreement of ideas, consider'd as copies, with those objects, which they represent. (T 2.3.3.5, p. 415)

The logic of this argument is, I trust, sufficiently clear. Just in case, we can reconstruct it perhaps even more clearly as follows:

1. To be contradictory to truth and reason is to be a representation or copy of something, but “disagree” with that thing (i.e., presumably, to be an *inaccurate* representation).
2. Passions are not representations or copies of anything (let alone inaccurate ones).
3. *A fortiori*, then, passions are not contradictory to truth and reason.

The difficulty arises from the second premise: that a passion “contains not any representative quality”, or that, in having an emotion, I have “no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high”. Phillips objects to this premise by reading it as a denial of the intentionality of the passions: “Hume is wrong to think that passions cannot represent: anger, his own example, typically has cognitive content (one is angry with a person, institution, etc.)”.⁶ Baier is even more forthright, dismissing this whole paragraph as “very silly” on the grounds of this interpretation of the second premise.⁷ Many others have read the premise in the same way, some to complain about it, and some merely to observe that this was Hume’s view.⁸

⁶Phillips (2005, p. 311).

⁷Baier (1991, p. 160).

⁸Kenny (1963/2003, p. 17, n1), Millgram (1995, p. 87), Prinz (2004, pp. 11,52), Garrett (2006, p. 303), Penelhum (2009, pp. 250-1).

I can certainly see why the wording of Hume's second premise gave rise to this interpretation. In context, however, it seems to me that there is a much better interpretation available. On this alternative reading, Hume is not denying that the passions have intentional content, but rather, denying that they are *truth-apt*. This makes much better sense of the last sentence of Hume's argument, where he says that for something to be contradictory to truth and reason is for it to *disagree* with the object that it is supposed to represent. Why this talk of disagreement, if Hume is simply denying that the passions have any objects at all? The point, I suggest, is not that anger isn't directed at anything; the point is rather that it isn't a *copy*, that it isn't representation of anything *with which it might disagree*. Beliefs *are* representations of some part of the world in this sense, and because of that they can be *false* representations (and hence contrary to truth and reason). Hume is simply claiming, in this argument, that the passions are not like *that*; he is not suggesting that they are not about anything at all.

Hume restates this argument at the start of Book 3 (explicitly describing it as a repetition of the earlier argument). In this second presentation, it is even more apparent that this is what Hume meant by the troublesome premise:

Reason is the discovery of truth or falshood. Truth or falshood consists in an agreement or disagreement either to the real relations of ideas, or to real existence and matter of fact. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. Now 'tis evident our passions, volitions, and actions, are not susceptible of any such agreement or disagreement; being original facts and realities, compleat in themselves, and implying no reference to other passions, volitions, and actions. 'Tis impossible, therefore, they can be pronounced either true or false, and be either contrary or conformable to reason. (T 3.1.1.9, p. 458)

Again, the logic is surely perfectly clear, but for even more transparency we might represent it thus:

1. To be conformable/contrary to reason is to be true/false (because reason is the discovery of truth and falsehood).
2. Truth/falsehood consists in agreement/disagreement with reality.
3. Passions, volitions, and actions are not susceptible of this agreement or disagreement (so they are not truth-apt).
4. Passions, volitions, and actions, therefore, are neither conformable nor contrary to reason.

This is indeed essentially the same argument as before. There are, however, a few noteworthy differences. First, Hume adds his explicit definition of reason as the discovery of truth and falsehood, which is presumably there to motivate the claim that being conformable (or contrary) to reason is a matter of being true (or false). Secondly, he splits up his first premise into two. Where in Book 2 he said directly that for something to be contrary to reason was for it to disagree with the object that it is supposed to represent, he now inserts falsehood as a middle term: for something to be contrary to reason is for it to be false, and for something to be false is for it to disagree with the object that it is supposed to represent. This second difference surely makes my preferred interpretation of the troublesome premise inescapable.

There are two further differences not directly relevant to the present point, but which might as well be brought out for clarity. In Book 2, Hume focuses on being *contrary* to reason: passions cannot be false, and therefore cannot be contrary to reason. In Book 3, he draws the obvious complementary conclusion as well: passions cannot be *true* either, and so can be neither contrary *nor conformable* to reason. And not only passions, but volitions and actions as well, which likewise are not truth-apt. I take it that there is nothing different about the logic here, and that if any one of these inferences works, then they all do. It is simply that, by beefing up the premises in the Book 3 presentation, the conclusion is correspondingly more meaty.

I am not presently interested in the merits of Hume's representative quality argument (though I will have something to say on this head in §10.5). The point to take away from the present section is that Hume, in this argument, did not deny that the passions have intentional content.⁹ This is the only passage, furthermore, in which he even *appears* to be denying this. Elsewhere, he seems, like any normal person, to take the intentionality of the passions for granted. In section 1 of the *Dissertation* alone, for example, he talks of trembling *at* the prospect of torture (P 1.17, p. 5), of feeling anxious *upon* the account of a sick friend (P 1.19, p. 5), of being afflicted *for* the loss of a law-suit, or joyful *for* the birth of a son (P 1.21, p. 6).

To be clear, I take it that the representative quality argument does not even commit Hume to the view that the intentionality of the passions is merely derivative, the upshot of being related in some appropriate way to an idea. As it happens, I believe that this *was* Hume's view (see §9.3), but not for this reason. For suppose that one holds the alternative, more cognitivist view, according to which passions are themselves cognitive states (but emotional ones, rather than bare or indifferent ones). On this view, there is a sense—a very trivial sense—in which the passions are truth-apt and within the jurisdiction of reason: they are true or false when the corresponding bare cognitive state is true or false. But the point of Hume's argument is not to deny this. In fact, he explicitly *allows* that the passions can be contrary to reason in this trivial sense:

According to this principle, which is so obvious and natural, 'tis only in two senses, that any affection can be call'd unreasonable. First, When a passion, such as hope or fear, grief or joy, despair or security, is founded on the supposition of the existence of objects, which really do not exist. Secondly, When in exerting any passion in action, we chuse means insufficient for the design'd end, and deceive ourselves in our judgment of causes and effects. (T 2.3.3.6, p. 416)

⁹A fuller defence of this claim is given in Weller (2002); Cohon (1994, pp. 188-9) seems to agree as well, although Weller, curiously, cites her as an opponent.

The point is rather that the passions can never be contrary to reason in any further and characteristically emotional way. What Hume has in mind is obviously what is *distinctive* about the passions, as opposed to the corresponding bare cognitive states: the *feeling* or *manner* of the awareness. It is *this* aspect of emotional life, according to Hume, that has no counterpart in external reality against which it might be measured as true or false, as conformable or contrary to reason. This view (surely a very plausible one) is not only consistent with the intentionality of the passions, but also consistent with the more cognitivist understanding of this intentionality as intrinsic rather than derivative.

9.3. Simple and complex

This being said, I take it nevertheless that Hume's view of the intentionality of the passions is best understood—in the *Treatise*, at least—as a derivative rather than an intrinsic matter. This is because of his insistence that the passions are *simple* perceptions, rather than complex ones. A passion that was intrinsically intentional in nature would presumably be *complex*, consisting both of the awareness (which it has in common with bare, unemotional states), and of the feeling or manner of this awareness. But this does not seem to have been Hume's view; his position was that the essence of an emotion was feeling alone.

The simple/complex distinction is first drawn in the second paragraph of the *Treatise*, immediately following the distinction between impressions and ideas:

There is another division of our perceptions, which it will be convenient to observe, and which extends itself both to our impressions and ideas. This division is into SIMPLE and COMPLEX. Simple perceptions or impressions and ideas are such as admit of no distinction nor separation. The complex are the contrary to these, and may be distinguished into parts. Tho' a particular colour, taste, and smell are qualities all united together in this apple, 'tis easy to perceive they are not the same, but are at least distinguishable from each other. (T 1.1.1.2, p. 2)

Near the start of Book 2, meanwhile, Hume states explicitly that the passions are simple rather than complex:

The passions of PRIDE and HUMILITY being simple and uniform impressions, 'tis impossible we can ever, by a multitude of words, give a just definition of them, or indeed of any of the passions. (T 2.1.2.1, p. 277)

The specific examples are pride and humility, but he says that the same is true “of *any* of the passions” (my emphasis). At the start of part 2, furthermore, he repeats the point with regard to love and hatred:

'Tis altogether impossible to give any definition of the passions of *love* and *hatred*; and that because they produce merely a simple impression, without any mixture or composition. (T 2.2.1.1, p. 329)

Hume does allow that the passions can mix with one another, and in this way form “compound” emotions (T 2.2.3.1, 2.2.11.1, pp. 347, 394). Hope and fear, for example, are both mixtures, in different proportions, of joy and sorrow (T 2.3.9.16, pp. 442-3; recall §7.3); respect and contempt are mixtures, respectively, of love with humility, and hatred with pride (T 2.2.10.3, p. 390); love between the sexes is a mixture of affection, the pleasure of beauty, and sexual desire (T 2.2.11.1, p. 394). But these *compound* mixtures are not *complex* in the present sense:

Ideas may be compar'd to the extension and solidity of matter, and impressions, especially reflective ones, to colours, tastes, smells and other sensible qualities. Ideas never admit of a total union, but are endow'd with a kind of impenetrability, by which they exclude each other, and are capable of forming a compound by their conjunction, not by their mixture. On the other hand, impressions and passions are susceptible of an entire union; and like colours, may be blended so perfectly together, that each of them may lose itself, and contribute only to vary that uniform impression, which arises from the whole. (T 2.2.6.1, p. 365)

Thus the mixture of love and humility, for instance, does not give rise to a complex passion, but to a simple and *uniform* impression of respect; as a mixture of red and green light does not give rise to a red-green complex, but to yellow.

In the passages quoted above, where Hume says that pride, humility, love, and hatred are all simple impressions, he also insists that simplicity entails *indefinability*, since a definition (of a complex perception) is just an enumeration of its component parts.¹⁰ (Presumably, then, Hume’s reduction of the compound mixtures into their ingredients are not *definitions* in the strict sense of that term.) “The utmost we can pretend to,” Hume goes on, “is a description of them, by an enumeration of such circumstances, as attend them” (T 2.1.2.1, p. 277), i.e. their causes and effects.¹¹

9.4. The double-relation passions

It is at this point, however, that matters become rather more complicated, and rather more interesting. The first thing that we need to put in place is Hume’s distinction between the *causes* and the *objects* of the passions. He draws this distinction only in the case of the double-relation passions: “The object of pride and humility is self: The cause of the passion is some excellence in the former case; some fault, in the latter. The object of love and hatred is some other person: The causes, in like manner, are either excellencies or faults” (P 2.3, p. 7). We can perhaps get a clearer idea of what Hume is getting at with this distinction by looking at the main passions that he discusses side by side, and in a canonic form:

Joy I am glad because of something good.

Sorrow I am sad because of something bad.

Hope I am hopeful because something good may happen.

¹⁰“Complex ideas may, perhaps, be well known by definition, which is nothing but an enumeration of those parts or simple ideas, that compose them” (E 7.4, p. 62).

¹¹That the passions are simple impressions in the *Treatise* is, to my mind, beyond all question. The evidence that we have just seen is as clear and unambiguous as textual evidence can be. It is surprising, therefore, to find quite a few people denying it; Dietl (1968), Sutherland (1976), Inoue (2003), Alanen (2006). I shall not discuss these alternative interpretations here, as I have criticised them in detail elsewhere (Merivale 2009, pp. 192-5, 198-200). For now, I will simply let Hume’s explicit claims to the contrary speak for themselves.

Fear I am afraid because something bad may happen.

Generosity I am glad because you are glad.

Pity I am sad because you are sad.

Malice I am glad because you are worse off than me.

Envy I am sad because you are better off than me.

Pride I am proud *of myself* because I have/did something good.

Humility I am ashamed *of myself* because I have/did something bad.

Love I love *you* because you have/did something good.

Hatred I hate *you* because you have/did something bad.

Everything on the right-hand side of the “because” in these presentations of the passions is, for Hume, the *cause* of the sentiment on the left-hand side. In the case of the double-relation passions, however, there is also something else on the left-hand side, namely oneself or another person. This is the *object* of these passions.

As I have said, Hume only explicitly draws this object/cause distinction in the case of the double-relation passions. On his behalf, however, I am tempted to draw it in the case of desires as well, and to say that the objects of desire are indeed *objects* in the sense of this distinction, rather than causes. Thus:

Benevolence I want *you to be happy* because I love you.

Anger I want *you to be miserable* because I hate you.

Gratitude I want *you to be happy* because you did something nice for me.

Resentment I want *you to be miserable* because you did something mean to me.

If this is right, then it may further be noted that, while every desire must have an object, some have no cause (in the appropriate sense):

Parental love I want *my children to be happy*.

Ambition I want *power*.

Vanity I want *praise*.

These desires do have a cause in another sense, but that cause is simply our general sentimental make-up: I want power because I'm ambitious; I want praise because I'm vain.¹² These are the exceptions that I had in mind in §9.1 above, when I noted that some desires are not immediately caused by cognition.

I am going to make a controversial claim shortly about the double-relation passions, and their relationship with their objects. Insofar as the parallel holds, it seems to me that the same ought to apply to desires and their relationship with their objects. The only textual evidence that I have for my claim about the double-relation passions, however, comes from what Hume says specifically about these sentiments. I mention the similarity merely because it seems to me quite striking. But it is a flimsy basis on which to infer anything about Hume's understanding of desire in this regard. I will remain entirely silent on the relationship between desires and their objects, because I cannot find any text on which to base any firm conclusions on this matter.

Be that as it may, with this distinction in place I can start to approach my controversial suggestion. Since the passions are all very clearly simple perceptions in the *Treatise*, I assume, for the reasons given in the previous section, that their intentionality is a matter of standing in some appropriate relationship to an idea, and not an intrinsic feature of the feelings themselves. This goes both for the causes of the passions and for their objects. And indeed, Hume is most explicit about the simplicity of emotions in the case of the double-relation passions, the only ones (except perhaps desires) that have objects as well as causes.

In the *Dissertation on the Passions*, however, Hume's view appears to be quite significantly different. I assume he still thought the intentionality that the

¹²I am thinking of vanity here as it occurs in the moral *Enquiry*. In *Treatise* Book 2 and the *Dissertation*, vanity does have a cause in the appropriate sense: we want praise because it reinforces our good opinion of ourselves. Recall §5.3.

passions derive from their *causes* was an extrinsic, relational matter, for there is nothing to suggest otherwise. In the case of the double-relation passions and the intentionality associated with their *objects*, however, matters are different. Recall, first, the two *Treatise* passages quoted in the previous section, about the simplicity and indefinability of these passions:

The passions of PRIDE and HUMILITY being simple and uniform impressions, 'tis impossible we can ever, by a multitude of words, give a just definition of them, or indeed of any of the passions. (T 2.1.2.1, p. 277)

'Tis altogether impossible to give any definition of the passions of *love* and *hatred*; and that because they produce merely a simple impression, without any mixture or composition. (T 2.2.1.1, p. 329)

And let us add to this, for good measure, his claim that the “sensations” associated with pride and humility “constitute their very being and essence” (T 2.1.5.3; p. 286). Now compare these very strikingly different claims in the *Dissertation on the Passions*:

Pride is a certain satisfaction in ourselves, on account of some accomplishment or possession, which we enjoy: *Humility*, on the other hand, is a dissatisfaction with ourselves, on account of some defect or infirmity.

Love or *Friendship* is a complacency in another, on account of his accomplishments or services: *Hatred*, the contrary. (P 2.1-2, p. 1)

Inclined as I am to take things at face value, and open to the possibility that Scotsmen can change their minds, these seem pretty clearly to me to be precisely the sorts of definitions that Hume had earlier maintained were not possible. Pride is a satisfaction in ourselves, and humility is a dissatisfaction in ourselves; love is a complacency in another, and hatred a displeasure in another. And where in the *Treatise* the satisfaction or dissatisfaction *alone* was said to be the essence of pride and humility, in the *Dissertation* we are told that what is “essential to pride” is “to *turn our view on ourselves* with complacency and satisfaction” (P 2.4, p. 7; my emphasis).

To spell this out, my suggestion is that, on Hume's mature view, the double-relation passions are now *complex* perceptions, consisting both of an awareness of self or another *and* the manner of this awareness (a satisfaction or dissatisfaction). They are themselves cognitive states, ways of being aware of oneself or someone else. It may be that feelings of joy, for example, are not *intrinsically* about the good things that prompt them. If so, it would be possible—logically, at least, if not psychologically—to have a feeling of joy not prompted by anything, and not about anything. In the same way, it may be that feelings of pride are not intrinsically about the good qualities or actions that give rise to them, in which case it would be possible to have such a feeling in the absence of the usual cause: simply to feel proud, without that pride being because of any particular thing that we have or have done. But it is *not* possible, on the present suggestion, to have a feeling of pride that is not intrinsically about oneself. Feeling proud just *is* thinking of oneself in a certain way (a positive way). And similarly for the other double-relation passions.¹³

One advantage of this interpretation of Hume's later account of the double-relation passions is that it corrects what was a very strange and counterintuitive feature of Hume's position in the *Treatise*. Hume's later view, on my face-value interpretation of what he says, is surely *right*. It *is* impossible to feel love or hatred that is not love or hatred *for someone*, and it *is* impossible to feel pride or humility that is not a satisfaction or dissatisfaction *with oneself*. I can see grounds for debate about the intentionality that the passions derive from their causes, and whether this is intrinsic or a matter of being appropriately related to an idea. With the intentionality that the double-relation passions derive from their objects, however, it is very difficult to see any appeal whatsoever in Hume's earlier position.

I am by no means alone in this reaction. Indeed, the counterintuitiveness of this aspect of the *Treatise* account is quite a common source of complaint. Here, for example, is Árdal's clear statement of the objection:

¹³Insofar as the parallel with desire holds, it would be similarly impossible to have a desire that is not a desire *for something*. It may however be possible to have a desire for something *for no reason*; or perhaps better, *simply because that's the sort of person that one is*.

Hume thinks of the relation of pride and its object as a contingent relation. One might be proud and yet not think of oneself... But, contrary to Hume's view, one must insist that it would be logically absurd to suggest that a man might have the passion of pride, and, at the same time, that the object of this pride... is another and not the person himself. Hume, who in most places appears to think of pride as a form of self-valuing, ought to have seen that 'to think highly of oneself because of y' and 'to be proud of y' are two ways of saying almost the same thing, and that the relation to oneself is a logical aspect of pride without which it could not be pride at all.¹⁴

I sympathise very much with Árdal's criticism. But I sympathise more with Hume, whose own clear and express wishes have since been ignored by generations of philosophers, who "have taken care to direct all their batteries against that juvenile work, which the Author never acknowledged, and have affected to triumph in any advantages, which, they imagined, they had obtained over it" (Ad 1777, p. 2). Hume's mature statement of his considered position, meanwhile, continues to be ignored.

9.5. The double relation theory (again)

Hume may or may not have had something like Árdal's objection in mind when he made this revision to his *Treatise* account of the double-relation passions. I can find nothing to suggest that he did, but it seems a perfectly plausible motivation. However, there is another potential reason for the change as well, and we can say with rather more confidence that this would have been an effective consideration for Hume. The double-relation passions are of course so named because Hume endeavours to account for them with his theory of the double relation of sentiments and ideas. This theory is almost certainly the aspect of Hume's philosophy of emotion that he was most proud of, both when he wrote Book 2 of the *Treatise* and when he came to rework it into the *Dissertation on the Passions*, for it dominates both of these works. But there is a very significant

¹⁴Árdal (1966, pp. 23-4).

(and frequently overlooked) problem with the *Treatise* presentation of this theory, and it stems from the fact that the double-relation passions are not intrinsically about their objects. By changing his mind on this matter, however, Hume thereby fixed his theory.

It is perhaps easiest to bring out this point by explaining first how the double relation theory works according to the *Dissertation* account, with the double-relation passions being intrinsically about their objects, in the same way that ideas are intrinsically about their content. In the causes of pride, for example, Hume discovers two crucial features: first, they are *good* things, that give rise to pleasure, or joy, or some such positive emotion; and second, they are in some way associated with *oneself* (more than with anyone else, or no one at all). These things, when we encounter or think about them, therefore have a natural tendency to make us think about ourselves, because of the association of ideas. They also make us feel good, and feeling good about something has a natural tendency to make us feel good about other things as well, because of the association of sentiments. Put these two things together, and add to them the mutual assistance of these associative principles, and the more or less inevitable result is that these things make us feel *good about ourselves*.

The important point here is that, for this causal story to make sense, pride and the other double-relation passions all have to be on the far end of *two* kinds of relation: a relation of ideas *and* a relation of sentiments. That is to say, they have to be *complex* perceptions, comprising both an idea component (awareness of self or another) and a sentiment component (satisfaction or dissatisfaction). If pride, in and of itself, is simply a feeling of objectless satisfaction, then the whole thing falls apart. For a start, there is then nothing to distinguish it from love; but more importantly there is no reason why it should be excited specifically by things that are related to *oneself*. And similarly for the other three.

That at least is the basic idea. Now for the gory detail. In the *Treatise*, the idea of self is not an intrinsic component of the passion of pride, but is merely related to it externally. As we will see, Hume is not perhaps entirely clear on the

nature of this relationship, but what he mostly says is that the idea is a *causal consequence* of the passion. For example:

The first idea, that is presented to the mind, is that of the cause or productive principle. This excites the passion, connected with it; and that passion, when excited, turns our view to another idea, which is that of self. Here then is a passion plac'd betwixt two ideas, of which the one produces it, and the other is produc'd by it. (T 2.1.2.4, p. 278)

Or again, a few sections later:

[N]ature has given to the organs of the human mind, a certain disposition fitted to produce a peculiar impression or emotion, which we call *pride*: To this emotion she has assign'd a certain idea, *viz.* that of *self*, which it never fails to produce. (T 2.1.5.6, p. 287)

But if the idea of self is the *effect* of pride, then there is no association of ideas that *ends* with this emotion. A potential cause of pride, being appropriately related to me, may cause me to think of myself by the association of ideas. But now that I am thinking about myself, how is the feeling of pride supposed to follow? The feeling of pride, Hume says, isn't an *effect* of the idea of self, it's a *cause* of it. So once I'm thinking of myself, I've already gone a step too far. The story sounds vaguely all right when you don't look at it too closely, but when you start to think about the detail—about the exact chain of events in the overall process—the whole thing falls apart.

We might propose to get around this problem by saying instead that the idea of self produces the feeling of pride, rather than the other way round. There is even some oblique textual evidence that Hume was tempted by this view himself (although the statements quoted above to the opposite effect are much clearer and more direct). For example, when first drawing the distinction between the causes and objects of pride and humility, Hume writes that “tho' that connected succession of perceptions, which we call *self*, be always the object of these two passions, 'tis impossible it can be their CAUSE, or be sufficient *alone* to excite

them” (T 2.1.2.3, p. 277-8; my emphasis). This naturally suggests that it *is* sufficient when joined with other circumstances. Perhaps more clearly, Hume later writes:

In order to excite pride, there are always two objects we must contemplate, *viz.* the *cause* or that object which produces pleasure; and self, which is the real object of the passions. (T 2.1.6.5, p. 292)

This gives the impression that Hume thought pride was produced by a combination of its cause (in his technical sense) and its object.

On this alternative reconstruction of the theory in the *Treatise*, we do get a little further. The details break down as follows: first, the cause produces both the idea of self (by the association of ideas), and a separate pleasure; and next, the separate pleasure has a tendency to produce pride (by the association of ideas), while the idea of self also has a tendency to produce pride (for reasons unknown). Putting the two things together at the second stage, pride is the more or less inevitable result.

This is certainly an improvement on our first attempt. But the situation is far from ideal. Not only is it inconsistent with Hume’s remarks about pride producing the idea of self (rather than the idea of self producing pride), but it also doesn’t square with everything that Hume says about the workings of the theory itself. The theory is supposed to depend, we may recall, not only on the two principles of association, but also on their mutual assistance. The problem with this second reconstruction is that there can be no such collaboration, because the association of ideas has already done its job at the first stage, before the association of sentiments gets to work at stage two. The principle of mutual assistance states that the two associative principles “very much assist and forward each other, and that the transition is more easily made where they both concur in the same object” (T 2.1.4.4, p. 284). But on the present proposal they *don’t* concur in the same object: the association of ideas prompts the idea of self, while the association of sentiments prompts the entirely distinct feeling of pride.

To return to where we started, there is only one way in which the two associative principles can concur in the same object, and that is if the object that they both concur in is a *complex* perception, part sentiment and part idea. That is precisely the view suggested most naturally by Hume's definitions of the double-relation passions in the *Dissertation*, as feelings of *self-satisfaction* and the like. This view is both more plausible in its own right, and makes better sense of the double relation theory. I submit that it was for this second reason, at least, if not also for the first, that Hume changed his mind when he came to revisit his account, and to think through these matters more clearly.

Chapter 10

Reason and Passion

Undoubtedly the most famous thing that Hume wrote about the passions is that “[r]eason is, and ought only to be [their] slave” (T 2.3.3.4, p. 415). This remark is a victim of its own success. Moral philosophers since have been so taken with the project of either confirming or refuting it, that the discussion in which it appears has become increasingly divorced from its original context. For example, Charles Pigden’s recent collection devoted to this topic contains only two passing references to Samuel Clarke, and no mention at all of Nicolas Malebranche, though these were Hume’s principal targets.¹ Its contributors seem content to study Hume’s argument without so much as a glance in the direction of the views that it was intended to refute.

In this chapter, I mean to inject some much-needed historical context back into this discussion, examining first the views of Malebranche and Clarke, and then of Hutcheson. Having done so, making sense of Hume will be a comparatively straightforward matter. I also have another point to make in support of my Difference Thesis. The substance of Hume’s view of the relationship between reason and passion did not change after the *Treatise*, so far as I can see, but the way in which he argues for it did. In particular, the famous representative quality argument, which I examined in §9.2, is abandoned. I will argue that this is an improvement rather than a loss.

¹Pigden (2009). The references to Clarke are on pages 7 and 213, where his name is mentioned, but nothing more.

10.1. Malebranche against the passions

In §1.1, I introduced Hutcheson’s criticism of Clarke’s moral rationalism, but quickly set it to one side, the better to focus on the egoism debate. Now is the time to pick it up again and look at it more closely. To understand the view that Hume followed Hutcheson in opposing, however, the best place to start—according to Hume’s own account of the matter—is not with Clarke, but with Malebranche: “MALEBRANCHE, as far as I can learn, was the first that started this abstract theory of morals, which was afterwards adopted by CUDWORTH, CLARKE, and others” (M 3.34, n12, p. 197).

Malebranche’s two major influences were St. Augustine and René Descartes.² From the latter he inherited—among many other things—the doctrine of the real distinction between mind and body, and its attendant conception of the mind as a non-extended, thinking thing. From the former he inherited—again among other things—a neo-Platonic conception of the doctrine of Original Sin, according to which God is our only true good and is to be loved above all things, the corporeal world is to be shunned as leading us away from this sole good, and our close connection with our bodies (which draws us so strongly in the wrong direction) is God’s punishment to all mankind for Adam’s eating of the apple. It is the Augustinian strand of Malebranche’s thought that is of interest here, for it is this that dominates his treatment of the passions. Furthermore, it was precisely this sort of view that Hume was reacting against in his famous discussion of reason and passion.

Malebranche’s most extended examination of the passions is to be found in book 5 of his *De la Recherche de la Vérité* (1675), but for a slightly fuller picture we must begin with book 4. This preceding book discusses what Malebranche calls the “natural inclinations” of the mind, which are the purely mental aspects of the will, untainted by bodily interference. They stand to the passions as thought stands to sensation or imagination (p. 337).

²Nadler (2006, p. 1).

Because Malebranche's God loves himself above all else, and because he creates our wills in the image of his own, Malebranche concludes that our primary natural inclination must be towards God or his glory:

It is an unquestionable truth that God can have no other principal end for His operations than Himself... Since the mind's natural inclinations are undoubtedly the constant impressions of the Will of Him who has created and preserves them, it seems to me that these inclinations must be exactly like those of their Creator and Preserver. By their very nature, then, they can have no other principal end than His glory[.] (p. 226)

Since God's creation participates in his goodness, however, God may have as a secondary end the preservation of his creatures: "God wills His glory, then, as His principal end and the preservation of His creatures only for His glory" (p. 226). Again, our natural inclinations follow suit: "since God unceasingly imprints in us a love like His own... He also provides us with all those natural inclinations that... necessarily dispose us toward preserving our own being and the being of those with whom we live" (p. 267).

For reasons that are somewhat obscure (or which at any rate have so far escaped my understanding), God's self-love manifests itself in us as "the love of the good in general" (p. 267), rather than as the love of God (though for Malebranche, I take it, God is the only truly good thing, which perhaps goes some way towards explaining the switch). Whatever its basis, this love that we have of the good in general does not force us to love only what is *in fact* good; we cannot love what we do not *think* is good, but God has left us free to make mistakes in this regard: "the power of loving badly, or rather of loving well what we should not love at all, depends on us, because as free beings we can and do in effect determine toward particular, and consequently false, goods the good love that God unceasingly imprints in us" (p. 267).

The passions, now, are for Malebranche the chief causes of precisely these sorts of errors. He does allow that they have a useful purpose, namely the preservation of the body, by directing us towards things that are good for it:

The senses and the passions were given to us only for the good of the body. Sensible pleasure is the mark that nature has attached to the use of certain things in order that without having to bother with a rational examination we might use them for the preservation of the body[.] (p. 359)

On the whole, however, the passions do us more harm than good. The inclination that they give us towards the good of the body might have been fine by itself, if only we could remember that “we are not our body; it is a thing belonging to us” and that “[t]he good of our body is therefore not our good” (p. 359). But since the Fall of Adam, “God has withdrawn from us” (p. 360), weakening our connection with him and what is truly good, and strengthening our tie to our bodies.

As a result of God’s departure, we are now in constant peril of such sinful things as the love of strawberries (my example, not Malebranche’s, but this really does seem to be his position). For we mistakenly think that strawberries (or whatever else it might be) are the causes of our pleasure, not realising that they are merely the occasions on which God chooses to give us pleasure (a reference to Malebranche’s occasionalist view of causation). We mistakenly think that they are good for *us* and not just good for our bodies. And so we direct our love where it should not be directed (e.g. at strawberries), thereby exciting God’s righteous anger:

We can and must love what is capable of making us sense pleasure, granted. But it is for this reason that we must love only God, because only God can act in our soul and because sensible objects can do no more than move our sense organs. But what difference does it make, you will ask, where these pleasant sensations come from? I only want to enjoy them. Ingrate, take a look at the hand heaping goods upon you! You exact unjust rewards from a just God; you would have Him reward you for the crimes you commit against Him, and at the same time you commit them... But death will corrupt this body, and God, whom you have made serve your unjust desires, will mock you when His turn comes and then will make you serve His own just wrath. (pp. 359-60)

The mind, for Malebranche, lies “between God and body, between good and evil, between that which enlightens it and that which blinds it, that which sets it in order and that which disrupts it, between that which can make it perfect and happy and that which can make it unhappy and imperfect” (p. 363). The passions, as we have seen, are the emissaries of the body. They drag us away from God and true felicity; they “involve us in error with regard to the good” and “they must be resisted continuously” (p. 357). Reason, by contrast, is the emissary of God. The free consent of our wills must therefore shun the passions and follow instead our God-given reason:

This consent must be regulated and kept free in spite of all the efforts of the passions. Only to God should it subjugate its freedom; it should surrender only to the voice of the Author of nature, to inner certainty, to the secret reproaches of reason. (p. 357)

Hume’s famous discussion of reason and passion in the *Treatise (Of the influencing motives of the will*, T 2.3.3, pp. 413-8) begins with an outline of the view that he is opposing:

Nothing is more usual in philosophy, and even in common life, than to talk of the combat of passion and reason, to give the preference to reason, and assert that men are only so far virtuous as they conform themselves to its dictates... The eternity, invariableness, and divine origin of the former have been display’d to the best advantage: The blindness, unconstancy, and deceitfulness of the latter have been as strongly insisted on. (T 2.3.3.1, p. 413)

Samuel Clarke was certainly one of Hume’s targets in this argument, and I will turn to his view presently. But I trust it is clear from this summary that Malebranche was also firmly in his sights.³

³There is growing support these days for the extremely counterintuitive suggestion that Malebranche was a major *positive* influence on Hume, and that, notwithstanding their very great theological differences and “big picture” opposition, Hume got many of his particular ideas from this source. Thus Peter Kail, for example, sees Hume as “using Malebranchian

10.2. Clarke against the passions

Hume's other major target here, as I have said, was Samuel Clarke. Clarke's views on this matter are expounded at the start of his *Discourse on the Unchangeable Obligations of Natural Religion* (the second of his two sets of Boyle Lectures, delivered in 1705 and first published in 1706). The essence of his view, as stated in the contents, is this:

That from the Eternal and Necessary Differences of Things, there naturally and necessarily arise certain Moral Obligations, which are of themselves incumbent on all Rational Creatures, antecedent to all positive Institution, and to all Expectation of Reward and Punishment. (1706, p. 588)

That these obligations are incumbent on all rational creatures, antecedent to any expectation of reward and punishment, is of course an anti-egoist view; and Clarke devotes much of his discussion to a criticism of Hobbes. These arguments of his are not of any present interest, however, first because it is his rationalism rather than his anti-egoism that we are concerned with here, and secondly because Hume was also an anti-egoist (at least after the *Treatise*), as we have already seen at length.

The summary proposition just quoted refers to the necessary differences of things as the source of moral obligations, but when we get down to the details Clarke also speaks more simply of "the nature of things". One might safely assume that the differences between things simply follow from their individual natures, so that on one level this distinction does not really matter. But the distinction does seem to be playing a role in Clarke's account, since natures and differences appear to ground distinct moral obligations. Thus we are told

materials to arrive at conclusions that are squarely antithetical to those fundamental to the religious dimension of Malebranche's philosophy" (2008, p. 55). See also Wright (1983), Jones (1982), Lennon (1997), James (2005), Kail (2005, 2007a), Schmitter (2012). It seems to me that the similarities between these two thinkers, even at the level of detail, have been overstated (for a discussion of one such issue, see §12.2), and I would gladly see a return to the natural view of them as simply *opponents*, more or less all the way down. But I am content here merely to place Malebranche in the opposing camp on the central issue of the relationship between reason and passion.

first that “there is a *Fitness* or *Suitableness* of certain *Circumstances* to certain *Persons*, and an *Unsuitableness* of others; founded in the *nature of Things*”; and then that “from the different relations of *different Persons one to another*, there necessarily arises a fitness or unfitness of certain *manners of Behaviour* of some *Persons* towards others” (p. 608). The immediate example given to illustrate this latter kind of obligation is theological:

That *God* is infinitely superior to *Men*; is as clear, as that *Infinity* is larger than a *Point*, or *Eternity* longer than a *Moment*: And ’tis as certainly *Fit*, that *Men* should honour and worship, obey and imitate *God*, rather than on the contrary in all their *Actions* endeavour to dishonour and disobey him[.] (p. 608)

To illustrate the former kind of obligation, meanwhile, Clarke writes:

In like manner; in *Mens* dealing and conversing one with another; ’tis undeniably more *Fit*, absolutely and in the *Nature* of the thing itself, that all *Men* should endeavour to promote the *universal good and welfare of All*; than that all *Men* should be continually contriving the *ruin and destruction of All*. (p. 609)

However they are founded—whether on the differences or just the natures of things—these “fitness” facts are at the core of Clarke’s moral philosophy.

To begin with, Clarke offers no argument for this highly dubious deduction of the fitness of certain sorts of behaviour from the differences or natures of things. He simply asserts that “[t]hese things are so notoriously plain and self-evident, that nothing but the extremest stupidity of *Mind*, corruption of *Manners*, or perverseness of *Spirit*, can possibly make any *Man* entertain the least doubt concerning them” (p. 609). He goes on to accuse Hobbes of some such failing. Later, however, when he turns to examine some particular moral obligations, we do get more by way of argument. For example, he offers the following justification for the principle of equity, that every man must “deal always with another, as he would reasonably expect that *Others* should in like *Circumstances* deal with *Him*” (p. 619):

Whatever relation or proportion one Man in any Case bears to another; the same That Other, when put in like Circumstances, bears to Him. Whatever I judge reasonable or unreasonable for *another* to do for *Me*; That, by the same Judgment, I declare reasonable or unreasonable, that I in the like Case should do for *Him*. And to deny this either in Word or Action, is as if a Man should contend, that, though *two and three* are equal to *five*, yet *five* are not equal to *two and three*. (p. 619)

There are further arguments for some additional example obligations, but an examination of these would involve us in more detail than is presently required.

For what it's worth, this argument—which has a familiar rationalist appeal—strikes me as unsuccessful. All that it shows, if it shows anything at all, is that one judgment concerning what is reasonable or unreasonable follows from another such judgment. But this is beside the point: the question is where the *first* judgment came from, and whether, in particular, it can be deduced from the eternal differences or natures of things. Humeans of course deny that any such deduction is possible.

The final piece of Clarke's view is that the fitness of certain sorts of behaviour gives rise to an obligation on all rational agents to behave in that way. The argument at this point again strikes me as far from persuasive:

Wherefore all rational Creatures, whose Wills are not constantly and regularly determined, and their Actions governed, by right Reason... but suffer themselves to be swayed by unaccountable arbitrary Humours, and rash Passions; by Lusts, Vanity and Pride; by private Interest, or present sensual Pleasures; These, setting up their own unreasonable Self-will in opposition to the Nature and Reason of Things, endeavour (as much as in them lies) to make things be what they are not, and cannot be. Which is the highest Presumption and greatest Insolence, as well as the greatest Absurdity, imaginable. 'Tis acting contrary to that Understanding, Reason and Judgment, which God has implanted in their Natures on purpose to enable them to discern the difference between good and evil. (pp. 613-4)

First, it is unclear why agents who indulge their rash and lusty passions, however unreasonable this behaviour may be, are thereby endeavouring “to make things be what they are not, and cannot be”. That *might* be their intention, I suppose, but surely most of the time nothing is further from their (our) minds. Secondly, there is the merest assertion, without any supporting argument, that such an endeavour is immoral. This assertion is hard to credit, however, not least because the supposed endeavour is necessarily—and obviously—doomed to failure.

The contrast in the above quotation between “right Reason” and “rash Passions”, and the identification of the former as the source of all good behaviour and the latter of all evil, is a general theme throughout this section. God, in Clarke’s design, cannot but act reasonably; lesser intelligent beings, however, may “*wilfully and perversely* allow themselves to be over-ruled by *absurd Passions*, and *corrupt* or *partial Affections*” (p. 612). Such “*wilful Passions or Lusts*,” together with negligent misunderstanding, “are... the only Causes which can make a reasonable Creature act contrary to Reason” (p. 613). Clarke, like Malebranche, is one of Hume’s main targets, being a paradigm case of a philosopher “to talk of the combat of passion and reason, to give the preference to reason, and assert that men are only so far virtuous as they conform themselves to its dictates” (T 2.3.3.1, p. 413).

10.3. Hutcheson against the rationalists

As noted in §1.1, Hutcheson’s first work, *An Inquiry into the Original of Our Ideas of Beauty and Virtue*, prompted an objection from the rationalist Gilbert Burnet, who appealed to the views of Clarke and Wollaston. This objection, in turn, prompted Hutcheson’s subsequent *Illustrations on the Moral Sense*, with its sentimentalist attack on moral rationalism. This attack, as we will see in the next section, gave Hume much material for his own views on the matter.

Hutcheson’s first move, which we also briefly saw in §1.1, is to distinguish between moral *motivation* and moral *approbation*, and consequently between what he calls *exciting* and *justifying* reasons:

When we ask the Reason of an Action we sometimes mean, “*What Truth shews a Quality in the Action, exciting the Agent to do it?*” Thus, why does a *Luxurious* Man pursue *Wealth*? The Reason is given by this Truth, “Wealth is useful to purchase Pleasures.” Sometimes for a Reason of Actions we shew the *Truth expressing a Quality, engaging our Approbation*. Thus the *Reason* of hazarding Life in just War, is, that “it tends to preserve our honest Countrymen, or evidences publick Spirit:” The *Reason* for *Temperance*, and against *Luxury* is given thus, “Luxury evidences a selfish base Temper.” The former sort of Reasons we will call *exciting*, and the latter *justifying*. (1728, p. 138)

This would seem to be a variant of the distinction between reasons that explain why someone *does* something, and reasons that show why someone *should* do something. Notable modern variants include the popular distinction between *motivating* and *normative* reasons,⁴ and Bernard Williams’ distinction between *internal* and *external* reasons.⁵

Hutcheson’s criticism of moral rationalism proceeds on these two separate fronts. I will restrict my attention here to the first front, the case of exciting reasons. Although there is of course a close connection between the two, exciting reasons and motivation belong within the philosophy of emotion, while justifying reasons and approbation are more a topic for moral philosophy. This same division is also reflected very clearly in Hume’s own texts, where the issue of exciting reasons is taken up in the *Dissertation on the Passions* (section 5, pp. 24-5), and that of justifying reasons in the moral *Enquiry* (appendix 1, pp. 285-94). As space here is limited, and the latter would take me too far from my present concern, I will focus on the *Dissertation* topic. (When turning to Hume himself, I will be forced to glance in the direction of morals and approbation; but still my primary focus is on motivation.)

⁴Smith (1994, pp. 94-8), Parfit (1997, p. 99), Dancy (2000, pp. 1-5).

⁵Williams (1981). It seems to me, incidentally, that Williams set up the question in the wrong way, by asking whether there *are* any external reasons (and notoriously answering, of course, in the negative). At least in the eighteenth century, however, the question dividing sentimentalists and rationalists concerns not the existence but the *basis* of these things: are they grounded in reason, in perception of the moral facts, or in sentiment, in non-rational feelings of approval and disapproval? See Blackburn (2010).

Regarding exciting reasons or motivation, then, Hutcheson's core argument is very simple:

As to *exciting Reasons*, in every calm rational Action some *end* is desired or intended; no end can be intended or desired previously to some one of these Classes of Affections, *Self-Love*, *Self-Hatred*, or desire of private Misery, (if this be possible) *Benevolence* toward others, or *Malice*: All Affections are included under these; no *end* can be previous to them all; there can therefore be no *exciting Reason* previous to *Affection*. (1728, p. 139)

By a "calm rational Action", Hutcheson means one in which reason and reflection is involved, as opposed to one in which passion alone moves us suddenly and thoughtlessly (see below). By urging that the passions or affections are necessary even in these cases (to fix the ends of action), therefore, he is urging that they are necessary in *all* cases.

What Hutcheson is offering here is the first clear presentation of what is now widely known (rather unfairly) as the "Humean" theory of motivation. According to this account, there is no motivation without some end which the action is believed to further somehow (as a probable means to that end, a constitutive part, or whatever it might be); and there is no end without some affection or desire. As the formula goes, motive equals belief plus desire.⁶ On this view, reason and sentiment both play a necessary role, and accordingly Clarke's talk of the *combat* between reason and passion is dismissed as incoherent:

We have indeed many confused Harangues on this Subject, telling us, "We have two Principles of Action, *Reason*, and *Affection*, or *Passion* (*i.e.* strong Affection): the *former* in common with Angels, the *latter* with Brutes: No Action is wise, or good, or reasonable, to which we are not excited by *Reason*, as distinct from all *Affections*; or, if any such Actions as flow from *Affections* be good, 'tis only by *chance*, or *materially* and

⁶See e.g. Smith (1994, pp. 92-125), who offers in essence the same argument as Hutcheson did for the "Humean" theory.

not *formally*.” As if indeed *Reason*, or the Knowledge of the Relations of things, could excite to Action when we proposed no *End*, or as if *Ends* could be intended without *Desire or Affection*. (1728, p. 139)

Hutcheson accounts for this confusion by acknowledging that particular violent affections can sometimes prevent us from thinking clearly:

Perhaps what has brought the Epithet *Reasonable*, or *flowing from Reason*, in opposition to what flows from *Instinct, Affection, or Passion*, so much into use, is this, “That it is often observed, that the very best of our particular *Affections* or *Desires*, when they are grown violent and *passionate*, thro’ the *confused Sensations* and *Propensities* which attend them, do make us incapable of considering calmly the whole *Tendency* of our Actions, and lead us often into what is *absolutely pernicious*, under some Appearance of *relative* or *particular Good*.” This indeed may give some ground for distinguishing between *passionate Actions*, and *those* from *calm Desire* or *Affection* which employs our *Reason* freely: But can never set *rational Actions* in Opposition to those from *Instinct, Desire* or *Affection*. (1728, p. 175)

Here we see why Hutcheson restricted his attention earlier to cases of “calm rational Action”. Even in these cases, where reason is operating at its full, the agent must still have some end that is the object of their desire. Affection or sentiment must always play a role.

The debate between Hutcheson and the rationalists, it seems to me, is fundamentally about *how to divide up the mind*, and how to conceptualise human nature and agency at the most basic level. Malebranche and Clarke were operating with a *hierarchical* division of human psychology, with reason (mind, spirit) firmly on the top, and passion (body, matter) firmly on the bottom. Each of these two components, moreover, formed for them a complete and fully-functioning unit, capable of acting on its own account. For these thinkers, there is thus a constant *battle* in the soul between our higher self and our lower self, our angel and our devil.

Hutcheson, however, (and Hume after him, as we will see in the next section) operated with a fundamentally different view of the distinction between reason and passion, what we might call a *specialisation* view. On this understanding, there is no hierarchy: the rational and passionate components are on the same level. But more than this, the components are distinguished, not by their different position in the hierarchy, but by their different psychological *roles*, with each one having a necessary and indispensable part to play in motivation and agency: roughly, reason to gather the information, and sentiment to set the goals.

Hutcheson (and Hume after him) also thought, of course, that sentiment was responsible for *evaluating* the goals that it set. This is not to say that reason has no role to play here; but the final verdict is down to feeling. As Hume neatly puts it in the moral *Enquiry*:

The final sentence... which pronounces characters and actions amiable or odious, praise-worthy or blameable... depends on some internal sense or feeling, which nature has made universal in the whole species... But in order to pave the way for such a sentiment, and give a proper discernment of its object, it is often necessary, we find, that much reasoning should precede, that nice distinctions be made, just conclusions drawn, distant comparisons formed, complicated relations examined, and general facts fixed and ascertained. (M 1.9, pp. 172-3)

But though reason, when fully assisted and improved, be sufficient to instruct us in the pernicious or useful tendency of qualities and actions; it is not alone sufficient to produce any moral blame or approbation. Utility is only a tendency to a certain end; and were the end totally indifferent to us, we should feel the same indifference towards the means. It is requisite a *sentiment* should here display itself, in order to give a preference to the useful above the pernicious tendencies... Here, therefore, *reason* instructs us in the several tendencies of actions, and *humanity* makes a distinction in favour of those which are useful and beneficial. (M App1.3, p. 286)

It seems to me that this is a separate matter—although it is of course very closely related—and as I have said my focus here is on the issue of motivation more than

approbation. A proper examination of the latter is a large project in its own right, and well beyond the scope of this thesis.

To emphasise the distinctness of these two issues, notwithstanding their close relationship, note that it is possible to agree with Hutcheson and Hume in the motivational case, while disagreeing in the case of approbation.⁷ It is possible, that is to say, to hold that reason and sentiment must work together to generate motivation, in the way that Hutcheson proposed, but to think that reason alone is sufficient to produce moral approbation and disapprobation. The thought here, roughly, would be that there is information available about what is morally right and wrong, about what one should and should not pursue or avoid, and that reason can deliver this information.

For what it's worth, I have a suspicion that Hutcheson and Hume, though they both very clearly rejected this view, did so for rather different reasons. Hume rejected it because he believed that there was simply *no such information* to be had; but Hutcheson, or so I suspect, rejected this view because he thought that there was such information, but that we arrived at it through the direct perception of our moral sense, rather than through reason.

10.4. Hume against the rationalists

Hume's anti-rationalism with regard to motivation was, as far as I can see, not significantly different from Hutcheson's. It also remained the same throughout his life. Given what we have seen of Hutcheson in the previous section, therefore, making sense of Hume's remarks on this matter—both in section 2.3.3 of the *Treatise*, and section 5 of the *Dissertation*—is now an almost entirely straightforward matter. The reason why I say *almost* will emerge at the end of this section.

Section 2.3.3 of the *Treatise* begins with an introductory paragraph calling attention to the talk of the combat between reason and passion, and the ten-

⁷This, in fact, is precisely the view adopted in recent times by Smith (1994).

dency of many philosophers and moralists—Hume obviously had Malebranche and Clarke in mind, as we have seen—to give the preference to reason. Hume then announces his intention “to shew the fallacy of all this philosophy” (T 2.3.3.1, p. 413). Section 5 of the *Dissertation* omits this introduction, diving straight into the main point:

It seems evident, that reason, in a strict sense, as meaning the judgment of truth and falsehood, can never, of itself, be any motive to the will, and can have no influence but so far as it touches some passion or affection. *Abstract relations* of ideas are the object of curiosity, not of volition. And *matters of fact*, where they are neither good nor evil, where they neither excite desire nor aversion, are totally indifferent; and whether known or unknown, whether mistaken or rightly apprehended, cannot be regarded as any motive to action. (P 5.1, p. 24; cf. T 2.3.3.2-3, pp. 413-4)

Hume doesn't explicitly say here—as Hutcheson did—that passions or desires are necessary to set the ends or goals of action, but one can safely assume that this is what he thought.

Incidentally, the distinction appealed to here—between relations of ideas and matters of fact, which Hume had earlier drawn in the first *Enquiry* (E 4.1-2, pp. 25-6)—was also anticipated by Hutcheson in the context of the reason/passion debate:

Reason denotes either our Power of finding out *Truth*, or a collection of Propositions already known to be *True*. *Truths* are either *Speculative*, as “*When we discover, by comparing our Ideas, the Relations of Quantities, or of any other Object, among themselves;*” or *Practical*, as “*When we discover what Objects are naturally apt to give any Person the highest Gratifications or what means are most effectual to obtain such Objects.*” (Hutcheson & Burnet 1735, p. 18)

The truths on Hutcheson's practical side in this quotation are perhaps surprisingly narrow (as Hume's *matters of fact* are not), but that was presumably just because

of the motivational context of the argument, in which truths about means to given ends are particularly salient.

At this point in the *Treatise*, Hume says that, “[a]s this opinion may appear somewhat extraordinary, it may not be improper to confirm it by some other considerations” (T 2.3.3.4, p. 415). He then produces his famous representative quality argument (familiar from §9.2), and two follow-on paragraphs in the same vein (T 2.3.3.5-7, pp. 415-7). This material is not repeated in section 5 of the *Dissertation*, and I consider this fact to be significant. I will have more to say about it in the next section, but for now let us stay with the common aspects of the earlier and later discussions.

Hume’s next move, in both the *Treatise* and the *Dissertation*, is to offer an explanation or reinterpretation of the talk of the combat between reason and passion, one that is consistent with his specialisation view of the relationship between these two faculties. Here again, Hume apparently drew on Hutcheson for inspiration, for his account similarly turns on a distinction between calm and violent passions. The specifics of Hume’s suggestion, however, are slightly different. Where Hutcheson had claimed that the violent passions sometimes prevent us from thinking clearly, and so to that extent do genuinely *interfere* with reason, Hume says instead that his opponents confound reason with calm desire, because these two things feel the same (T 2.3.3.8, p. 417).

Hume’s explanation here calls to mind the earlier distinction, drawn at the start of *Treatise* Book 2, between calm and violent emotions (recall §3.4). As I said when discussing this earlier distinction, however, these two things should be kept apart. The first is a distinction between the passions (or violent emotions) and the calm emotions of moral and artistic approval and disapproval. This latter contrast, meanwhile, is not a distinction between *types* of passion. Rather, the point is that any passion can vary in its felt intensity, sometimes occurring more “violently” and sometimes more “calmly”.

Most familiar in this context is Hume's talk of the "force" or "violence" of a passion (e.g. T 2.1.4.4, 2.2.11.1, 2.3.4.1, pp. 284, 394, 418-9; P 6.3, 6.6, pp. 26, 27). But Hume also often speaks in terms of an increase of the passion itself, rather than an increase in its violence (e.g. P 1.12, 6.3, 6.6, 6.11, pp. 4, 26, 27, 28), about the "heightening" or "raising" of an emotion (e.g. P 2.9, 6.1, 6.3, pp. 8, 26; Tr 27, pp. 224-5), and about variations in "quantity" or "degree" (e.g. P 1.12, 2.21, 2.25, 2.28, 2.41, 2.42, 2.47, pp. 4, 11, 12, 15, 16). While it would be a mistake to suppose that Hume had different properties in mind for these different metaphors,⁸ it is important to be aware that he conceived of this singular property in (at least) two different ways: as akin to "violence" or "turbulence" (P 5.3, p. 24) on the one hand, and as akin to "degree" or "quantity" on the other. This point will be important in a later chapter (§12.3).

Section 5 of the *Dissertation* and section 2.3.3 of the *Treatise* then end with essentially the same paragraph, save for a few insignificant stylistic changes. In this paragraph Hume draws attention to one further point that Hutcheson didn't make: "The common error of metaphysicians has lain in ascribing the direction of the will entirely to one of these principles [i.e. calm general desires or violent passions], and supposing the other to have no influence"; but, says Hume, people in general are moved *both* by "the view of the greatest possible good" *and* by "present uneasiness" (P 5.4, p. 24; T 2.3.3.10, p. 418). This point puzzled me for some time, first because it just seems so obviously true, but also because neither Malebranche nor Clarke had ever denied it. Their position was not that we are moved by only one of the competing principles (a position that would after all make a nonsense of the idea that there is a constant combat between them). Rather, their claim was that we *should* be moved by only one of them. This puzzling paragraph was the reason for my qualification at the start of this section, when I said that understanding Hume's remarks concerning motivational anti-rationalism was *almost* an entirely straightforward matter.

⁸Aside from the fact that Hume deploys the different metaphors with considerable freedom and variety, with no hint of any underlying pattern, "violence" and "degree" appear to be fairly explicitly equated at T 2.2.2.22 (p. 343-4), in a passage that does not reappear in the *Dissertation on the Passions*.

So far as I am aware, no one has yet remarked on the curious nature of this concluding paragraph; attention, no doubt, has been drawn away by the other provocative things that Hume says in this context, especially in the *Treatise* presentation. Perhaps I thus have the honour of venturing the first explanation. In any case, the only way that I have been able to make sense of the comment is to suppose that it is aimed, not at Clarke or Malebranche, but at *Locke*. For Locke appears to have oscillated on precisely this point: in the first edition of his *Essay*, he wrote that “the greater Good is that alone which determines the Will” (1690, p. 251n); but in the second edition he said that “upon second thoughts I am apt to imagine [that what determines the Will] is not, as is generally supposed, the greater good in view: But some (and for the most part the most pressing) uneasiness a Man is at present under” (1694, pp. 250-1). Hume’s midway position on this matter seems obvious; but I suppose if Locke had swung from one extreme to the other, then the obvious point needed to be made.

This solution to the puzzle is not perfect, for Hume describes the view that he is rejecting as a *common* error of metaphysicians, and presents his rejection in the context of an attack that is obviously directed at Malebranche and Clarke. Thus we should expect it to be a point of agreement among all or most rationalists, rather than an error peculiar to Locke. Nevertheless, it is the most satisfying explanation that I have been able to come up with; I would gladly be persuaded by a better one.

10.5. The representative quality argument (again)

Section 5 of the *Dissertation* does not contain anything that was not already present in section 2.3.3 of the *Treatise*. However, when examining the former in the light of the latter, one cannot help but be struck by the very glaring omissions: all the best-known parts of the *Treatise* have been removed. This includes the representative quality argument already looked at in §9.2, but also the notorious claim that “[r]eason is, and ought only to be the slave of the passions” (T 2.3.3.4, p. 415), and the several provocative illustrations of the anti-rationalist position:

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter. (T 2.3.3.6, p. 416)

I suppose that the standard thought, insofar as anyone has given this curious fact any consideration at all, is that Hume made these striking cuts simply in order to keep things short. But a little reflection shows this explanation to be most unlikely. Section 2.3.3 of the *Treatise* was not long to begin with. If section 5 of the *Dissertation* had simply repeated it in its entirety, it would still have been the shortest section in this later work. As it is, section 5 is *radically* shorter than all of the other sections, and less than half the length of its *Treatise* ancestor. Considerations of space are thus not a plausible explanation, and we must look for some other reason for Hume's drastic editing of his earlier work.

The conclusion of Hume's first argument in this section of the *Treatise*, the argument that *is* repeated in the *Dissertation*, is that "reason alone can never produce any action, or give rise to volition" (T 2.3.3.4, p. 414); or that "reason... can never, of itself, be any motive to the will, and can have no influence but so far as it touches some passion or affection" (P 5.1, p. 24). This is very clearly a point about *motivation*, as I was of course assuming in the previous section. It is not, to be explicit, a point about *approbation*; and although Hume held an exactly parallel view in the case of approbation, and these two anti-rationalist theses are of course very closely related, they are nevertheless distinct claims (as I have already insisted).

In Hume's later work, the distinction between motivation and approbation is firmly in his grasp, and he tackles the two topics separately. The question of approbation first comes up, after the *Treatise*, in the essay *The Sceptic*, in which Hume makes his anti-rationalist stance very clear (at least if we take him to be in agreement with his Sceptic, which I do). It then comes up again, of course, in

the moral *Enquiry*. The question of motivation, meanwhile, is addressed in the *Dissertation on the Passions*, which is entirely silent on the matter of approbation. This is all very clear and straightforward.

In section 2.3.3 of the *Treatise*, however, Hume rather confusingly runs these two things together, attempting to take on his rationalist opponents on both fronts simultaneously. The representative quality argument is his case for denying that reason alone gives rise to *approbation*, its conclusion being that passions cannot be “contradictory to truth and reason” (T 2.3.3.5, p. 415), “contrary to reason” (T 2.3.3.6, p. 416), or “unreasonable” (T 2.3.3.7, p. 416), except in a trivial and uninteresting sense (i.e. when they are founded on some false belief). I suppose that one *might* be able to interpret this as a conclusion about motivation rather than approbation, but that would be a fairly serious strain on the words. The obvious interpretation of Hume’s thinking in these passages is that reason does have an evaluative role in human psychology—it can evaluate beliefs as either true or false—but that this evaluative role does not extend to classifying passions and desires as morally right or wrong.

One very good reason why Hume should have removed this argument from the *Dissertation on the Passions*, then, is that it doesn’t belong there: it belongs, if anywhere, in the moral *Enquiry*. But this is in fact only half of the explanation, for, most curiously, it isn’t in the moral *Enquiry* either. Hume didn’t just *move* the representative quality argument in his later work; he *abandoned* it. And he abandoned it, I believe, for a very good reason, namely that it isn’t a very good argument.

To see this, let us take a closer look at the argument again, in as clear a form as possible:

1. To be conformable or contrary to reason is to be either true or false respectively (because reason is the discovery of truth and falsehood).
2. Passions are not truth-apt.
3. Therefore, passions can be neither conformable nor contrary to reason.

The argument is obviously valid. It may even be sound. Nevertheless, it is a terrible argument for anti-rationalism, because the premises simply assume—without attempting to explain or justify—far too much of what is to be proved. As arguments go, it amounts to little more than a bare assertion of the claim to be established.

Given the obvious validity of the argument, rationalists wishing to escape its conclusion must deny at least one of the premises. But why *shouldn't* they do this? Hume has given us little or nothing in support of either. It is agreed on both sides of this debate that passions and desires can be morally right or wrong. The question concerns not the *reality* of this distinction, but its *foundation*.⁹ Clarke says that reason alone determines which side of the line things fall on, while Hume insists that sentiment has the final say. Suppose that you agree with Clarke, and that you agree with the second premise of Hume's argument (that the passions are not truth-apt). In this case you will *of course* deny that the only way for something to be contrary to reason is for it to be false. You will say that things can be contrary to reason either when they are false *or* when they are morally wrong. Or suppose instead that you agree with Hume's first premise; now you will obviously deny the second. You will say that passions can be true or false: they are true or false precisely when they are morally right or wrong. (I take it that the first line of defence is the more natural one, but this second is a logical possibility too.)

Hume and his fellow Humeans will not accept either of these rationalist responses to the argument, of course, and there is nothing in these considerations to persuade them that they should. But that is not the point. The point is that there is nothing in *Hume's* argument to persuade anyone that they *shouldn't*. The argument simply doesn't touch on the controversy at all. Hume helps himself to

⁹Following the uncharitable interpretation of the *Treatise* as “sapping the Foundations of Morality, by denying the natural and essential Difference betwixt Right and Wrong, Good and Evil” (L 19, p. 425), Hume was at pains to make this point very clear in the introduction to the moral *Enquiry* (M 1.2-3, pp. 169-70).

his preferred views of reason and truth in the premises, from which his anti-rationalist conclusion trivially follows. But the debate concerns whether these are indeed the right views of reason and truth. And so it was with very good reason, in my opinion, that Hume did not repeat this argument, either in the *Dissertation on the Passions* or in the moral *Enquiry*.

Chapter 11

The Chemistry of the Passions

In this chapter and the next, I look at section 6 of the *Dissertation on the Passions* together with the third of the four dissertations, *Of Tragedy*. As we will see, these two parts of the *Four Dissertations* very clearly belong together: in the former Hume introduces a general principle in his philosophy of emotion, which in the latter he then applies to a particularly intriguing case, namely the pleasure that we take in tragic art. This close connection between these two dissertations makes an important contribution to my Unity Thesis. The present topic also provides yet another point in support of my Difference Thesis, since in the *Treatise* Hume had offered a quite different account of this paradoxical pleasure.

11.1. Two overlapping theories

As we saw in the previous chapter (§10.4), Hume held that the passions can differ in their felt intensity, a difference that is conceived of both in terms of the *quantity* of the passion, and in terms of its *force* or *violence*. This idea is most familiar in the context of Hume's treatment of the relationship between reason and passion, and in particular his claim that philosophers have mistaken calm desires or passions for reason, giving rise to the mistaken belief that there is a combat between reason and passion (P 5.2-3, p. 24). Hume's interest in the intensity of the passions extends beyond this one anti-rationalist point, however, and he is also concerned in general with the causes of these felt differences. Thus, having made the anti-rationalist point in section 5 of the *Dissertation on the Passions*,

Hume then proceeds in section 6 to “enumerate some of those circumstances, which render a passion calm or violent, which heighten or diminish any emotion” (P 6.1, p. 26).

In Book 2 of the *Treatise*, the parallel discussion constitutes the core of part 3, taking up five of this final part’s ten sections (T 2.3.4-8, pp. 418-38). Section 5 discusses the effects of custom, section 6 looks at the influence of the imagination, while sections 7 and 8 examine the effects of distance in time and space. Hardly anything from these sections remains in the *Dissertation on the Passions*, but this later work does reproduce section 4 more or less in its entirety. In this section (which becomes section 6 of the *Dissertation*), Hume introduces and defends his *conversion principle*: “It is a property in human nature, that any emotion, which attends a passion, is easily converted into it... The predominant passion swallows up the inferior, and converts it into itself. The spirits, when once excited, easily receive a change in their direction” (P 6.2, p. 26; cf. T 2.3.4.2, pp. 419-20).¹

Section 1 of the *Dissertation on the Passions*, meanwhile, contains Hume’s account (examined in §7.3) of the generation of hope and fear from the mixture, in different proportions, of joy and sorrow. Having given this account, Hume then briefly expands it into a general theory concerning the mixing of “contrary” passions (i.e. one positive and one negative):

In contrary passions, if the objects be *totally different*, the passions are like two opposite liquors in different bottles, which have no influence on each other. If the objects be intimately *connected*, the passions are like an *alkali* and an *acid*, which, being mingled, destroy each other. If the relation be more imperfect, and consist in the *contradictory* views of the *same* object, the passions are like oil and vinegar, which, however mingled, never perfectly unite and incorporate. (P 1.24, p. 6)

¹This name for the principle seems to have been coined by Margaret Paton (1973, p. 121). It or something like it has since been used by most people writing on the topic: Hill (1982), Feagin (1983), Packer (1989), Schier (1989), Budd (1991), Yanal (1991, 1992), Neill (1992, 1998), Leibowitz (1993), Galgut (2001). I used to call it the “principle of the predominant passion” myself (Merivale 2011), taking my cue from Immerwahr (1994). I had a reason for this idiosyncrasy: I felt that the emphasis on *conversion*, at the expense of Hume’s other ways of describing the process (one passion swallowing up another, the spirits changing direction), was unfortunate. But I now think that the conversion idea probably is the most important after all, and in any case I have got tired of all those extra syllables.

The third of these principles is the one responsible for the production of hope and fear. Hope, on this chemical analogy, is like oil with a dash of vinegar, while fear is like vinegar with a dash of oil.

Hume does not say very much about the first two principles, in either the *Treatise* or the *Dissertation*. In the case of contrary passions concerning the same object cancelling each other out, he offers nothing at all beyond a bare statement of the principle (P 1.22, p. 6; cf. T 2.3.9.15, p. 442). In the case of contrary passions concerning different objects having no effect on each other, he merely offers one illustration: “when a man is afflicted for the loss of a lawsuit, and joyful for the birth of a son, the mind, running from the agreeable to the calamitous object; with whatever celerity it may perform this motion, can scarcely temper the one affection with the other, and remain between them in a state of indifference” (P 1.21, p. 6; cf. T 2.3.9.14, pp. 441-2).

Section 1 then ends with a hint of a fourth principle, concerning the mixture of *any* two passions, not just a positive and a negative: “The effect of a mixture of passions, when one of them is predominant, and swallows up the other, shall be explained afterwards” (P 1.25, p. 6). This is of course a reference ahead to section 6, and to the conversion principle that I have already introduced. Thus this principle belongs in the overlap between two of Hume’s general theories: concerning the mixture of passions, and concerning the causes of the violent passions. The two topics overlap, straightforwardly enough, because the effect of mixing two passions sometimes is precisely an increase in the force of one of them.

These two overlapping theories together form what I like to call Hume’s account of the *chemistry* of the passions, since his metaphors and analogies in this context are predominantly chemical and hydraulic. I have already looked at Hume’s account of hope and fear, and, as I said above, Hume says very little about his other two principles concerning the mixing of contrary emotions. That leaves the conversion principle on the mixture side. As for the causes of the violent passions, I will hereafter restrict my attention to this same principle. It will

become clear presently why this principle is of particular importance to me here. Furthermore, it is the only aspect of the general theory that Hume continued to discuss in detail after the *Treatise*, and so it makes sense that a study of Hume's later philosophy of emotion should exhibit the same bias.

The conversion principle is of considerable importance for both my Unity Thesis and my Difference Thesis. The principle itself was already present in Book 2 of the *Treatise* (T 2.3.4.2, pp. 419-20), and Hume's discussion of it in the *Dissertation* differs only in superficial points of style. It was only *after* writing the *Treatise*, however, that Hume came to realise that his principle might be applied to the curious phenomenon of the pleasure that we take in tragic art (see §§11.4 and 11.5 below). He develops this application, furthermore, not in the *Dissertation on the Passions*, but in the following dissertation *Of Tragedy*. One could hardly wish for a clearer case of unity among the *Four Dissertations*.

11.2. Du Bos

Hume's dissertation, *Of Tragedy*, is very familiar to philosophers of art for its presentation of one of the canonic solutions to the paradox of art and negative emotions. Put simply, the puzzle is this: why do we so much enjoy and appreciate things in art (tragedy, horror, blood and gore) that we find so unpleasant in real life? Hume motivates his own famous answer to this question with a brief look at the earlier solutions offered by Jean-Baptiste Du Bos and Bernard Le Bovier de Fontenelle. Probably because of Hume's criticisms, these solutions have since disappeared from the debate (though Hume's account continues to receive much attention). In the case of Fontenelle, I am not persuaded that this is any great loss. The case of Du Bos, however, is different.

Du Bos's principal work for our purposes (he also wrote some historical books) is the two-volume *Réflexions Critiques sur la Poésie et sur la Peinture* (1719).² This work contains a wealth of interesting psychological and art-

²An English translation by Thomas Nugent was published in 1748; I will give quotations and page numbers from volume 1 of this translation (I have no occasion to refer to volume 2).

historical observations, together with intriguing suggestions in the philosophy of art and art criticism. On another occasion, it would merit closer attention; for now, I will restrict my attention to just the introduction and the first three sections of volume 1.

The initial aim of this first volume is “to explain what the beauty of a picture or poem chiefly consists in” (1748, p. v). The nature of the pleasure that we receive from poetry and painting, Du Bos thinks, is on the face of it rather puzzling, because a great deal of the time it bears a striking resemblance to *sorrow* and other unhappy feelings: “The arts of poetry and painting are never more applauded, than when they are most successful in moving us to pity” (p. 1). This is of course the familiar paradox of art and negative emotions. Du Bos might even have been the first person to formulate this paradox in modern terms.³

Du Bos’s own solution, put forward in section 3, depends on an interesting view of human psychology developed in sections 1 and 2. The core of this view is that many of the emotions that we typically view as negative or painful in fact have a good deal to recommend them. The first point made in support of this is something that we have already seen Hume’s appropriation of (§4.4), namely the idea that “one of the greatest wants of man is to have his mind incessantly occupied” (p. 5). So strong is this desire, according to Du Bos, that we will often seek out great pain in order to satisfy it: “The heaviness which quickly attends the inactivity of the mind, is a situation so very disagreeable to man, that he frequently chuses to expose himself to the most painful exercises, rather than be troubled with it” (p. 5).

In section 2, Du Bos then turns more particularly to the emotions that we experience “upon seeing our fellow creatures in any great misfortune or danger” (p. 10). Although he begins by saying that these emotions have nothing to recommend them beyond the general relief from “ennui” introduced in section 1, his subsequent elucidation of the *sorts* of relief that they provide, and of what exactly it is that we enjoy about this emotional stimulation, could to my mind be just as well viewed as *additions* to the basic general point. I shall not dwell

³See Livingston (2013, pp. 401-4).

here on the details of Du Bos's examples. Some depend on more morally dubious sides of human nature, such as the pleasure that we take in public executions (p. 10), in bull-fights (p. 18), or in the Roman gladiatorial battles (p. 12). Others are more innocent, however, such as the thrill of tight-rope walking (pp. 11-2) or gambling (pp. 19-20).

Having argued that many negative experiences, for all their undoubted negativity, also have much to recommend them—emotional stimulation, the thrill of danger and high stakes, satisfaction of our bloodlust (whether we approve of this desire or not)—Du Bos then introduces his account of the pleasure of poetry and painting. What these art forms offer us, he says, are artificial experiences with all of these pleasant effects, but none of the usual drawbacks:

Since the most pleasing sensations that our real passions can afford us, are balanced by so many unhappy hours that succeed our enjoyments, would it not be a noble attempt of art to endeavour to separate the dismal consequences of our passions from the bewitching pleasure we receive in indulging them? Is it not in the power of art to create, as it were, beings of a new nature? Might not art contrive to produce objects that would excite artificial passions, sufficient to occupy us while we are actually affected by them, and incapable of giving us afterwards any real pain or affliction?
(p. 21)

The crucial idea here, which was hinted at in the previous sections, but now appears more clearly, is that in real life our emotional reactions to tragic, dangerous, and horrific events are *mixed*, containing both positive and negative aspects; but unfortunately the negative aspects are so strong that they typically drown out the positive. What poetry and painting do, thinks Du Bos, is present us with “imitations” of these events. And an imitation of this sort, because it is only fictional, “does not affect our reason, which is superior to the illusory attack of those sensations, ... [but] affects only the sensitive soul” (p. 23). As such, it is able to induce the pleasurable feelings in isolation, untainted by the pain that attends them in real life:

The pleasure we feel in contemplating the imitations made by painters and poets, of objects which would have raised in us passions attended with real pain, is a pleasure free from all impurity of mixture. It is never attended with those disagreeable consequences, which arise from the serious emotions caused by the object itself. (p. 24)

Du Bos is not entirely explicit in listing these “disagreeable consequences”, and a more careful and thorough interpretation would disentangle what he takes to be the positive and the negative aspects of our experience. But that is more than I have the space (or the need) for here, and I trust that the above has given a sufficient sense of the overall *structure*, at least, of Du Bos’s solution.

11.3. Fontenelle

Du Bos’s *Réflexions* is a substantial book, containing several astute psychological observations, and an intriguing and well thought-out solution to the paradox of art and negative emotions. By contrast, Fontenelle’s similarly titled *Réflexions sur la Poétique* (1742) is a medium-sized essay, containing just a single paragraph discussing this particular puzzle. This is the paragraph that Hume quotes, in his own (perfectly good) translation, in *Of Tragedy*:

“Pleasure and pain,” says he [Fontenelle], “which are two sentiments so different in themselves, differ not so much in their cause. From the instance of tickling, it appears, that the movement of pleasure, pushed a little too far, becomes pain; and that the movement of pain, a little moderated, becomes pleasure. Hence it proceeds, that there is such a thing as a sorrow, soft and agreeable: It is a pain weakened and diminished. The heart likes naturally to be moved and affected. Melancholy objects suit it, and even disastrous and sorrowful, provided they are softened by some circumstance. It is certain, that, on the theatre, the representation has almost the effect of reality; yet it has not altogether that effect. However we may be hurried away by the spectacle; whatever dominion the senses and imagination may usurp over the reason, there still lurks at the bottom

a certain idea of falsehood in the whole of what we see. This idea, though weak and disguised, suffices to diminish the pain which we suffer from the misfortunes of those whom we love, and to reduce that affliction to such a pitch as converts it into a pleasure. We weep for the misfortune of a hero, to whom we are attached. In the same instant we comfort ourselves, by reflecting, that it is nothing but a fiction: And it is precisely that mixture of sentiments, which composes an agreeable sorrow, and tears that delight us. But as that affliction, which is caused by exterior and sensible objects, is stronger than the consolation which arises from an internal reflection, they are the effects and symptoms of sorrow, that ought to predominate in the composition.” (Tr 7, pp. 218-9)

In this quite long paragraph, Fontenelle does not seem to me to offer a coherent explanation of the phenomenon, but rather presents a few different ideas (which are not obviously consistent with each other). The first is an echo of Du Bos’s idea that we have a strong desire for emotional stimulation of all kinds, and that ostensibly negative emotions may in truth be more desirable than is often thought; to repeat: “The heart likes naturally to be moved and affected. Melancholy objects suit it”.

A second idea in this paragraph also has some similarities with Du Bos’s account. This is the idea that the fictional nature of the events depicted weakens or softens the emotions, and that a passion that is typically painful, in fact becomes pleasurable when it is thus softened. Du Bos too, as we have seen, thought that the fictional nature of the events neutralised the unpleasant aspects of the experience. For Du Bos, however, our real-life experience combines pleasant with unpleasant emotional responses, and what the fictional case does is simply neutralise the unpleasant; for Fontenelle, at least according to this second idea, there is no pleasure until the falsehood of what we are seeing weakens the otherwise purely painful response. For Du Bos, the real-life reaction is mixed, and the artificial reaction is pure; for Fontenelle, both reactions are pure, but one is purely painful and the other purely pleasant.

There also seems to be yet a third idea towards the end of Fontenelle’s paragraph, where he suggests that our experience at the theatre is *not* after

all purely pleasant, but is rather a mixture of pleasure and pain, an “agreeable sorrow” that combines the sadness with the comforting thought that what we are seeing is not real. I leave to others the task of rendering this idea consistent with the rest.

11.4. Hume's change of mind

Though he does not name them explicitly in his early work, Hume had presumably read Du Bos and Fontenelle when he wrote the *Treatise*. For in a couple of passages in Book 1 he shows an awareness of the paradox of art and negative emotions, and gestures at a (somewhat vague) solution that contains elements of both of these earlier accounts:

We may add to this a remark; that in matters of religion men take a pleasure in being terrify'd, and that no preachers are so popular, as those who excite the most dismal and gloomy passions. In the common affairs of life, where we feel and are penetrated with the solidity of the subject, nothing can be more disagreeable than fear and terror; and 'tis only in dramatic performances and in religious discourses, that they ever give pleasure. In these latter cases the imagination reposes itself indolently on the idea; and the passion, being soften'd by the want of belief in the subject, has no more than the agreeable effect of enlivening the mind, and fixing the attention. (T 1.3.9.15, p. 115)

There is no passion of the human mind but what may arise from poetry; tho' at the same time the *feelings* of the passions are very different when excited by poetical fictions, from what they are when they arise from belief and reality. A passion, which is disagreeable in real life, may afford the highest entertainment in a tragedy, or epic poem. In the latter case it lies not with that weight upon us: It feels less firm and solid: And has no other than the agreeable effect of exciting the spirits, and rousing the attention. (T 1.3.10.10, p. 630-1)

At some point after writing the *Treatise*, however, Hume evidently gave the matter more thought, and came to be dissatisfied with this account. The result was the

third of the four dissertations, *Of Tragedy*, in which he abandons his earlier tacit allegiance to Du Bos and Fontenelle, and offers a very different account of his own.

Hume's later objection to Du Bos is disappointing. To begin with, his summary of Du Bos's solution mentions only the point about our strong desire for emotional stimulation of all kinds (Tr 3, p. 217). While his endorsement of this point (Tr 4, p. 217) should be taken entirely at face value (we may recall that Hume appealed to the same principle in his account of the love of family relations; §4.4), he makes no mention whatsoever of Du Bos's additional discussions of the inherent pleasures of gruesome spectacles, or of the thrill of risk and danger. More importantly, he completely ignores the core of Du Bos's actual solution, namely that the artificial passions aroused by art are pleasures *purified* from their usual negative aspects, by virtue of the fictional nature of the events depicted.

The extent of Hume's objection is simply this:

There is, however, a difficulty in applying to the present subject, in its full extent, this solution, however ingenious and satisfactory it may appear. It is certain, that the same object of distress, which pleases in a tragedy, were it really set before us, would give the most unfeigned uneasiness; though it be then the most effectual cure to languor and indolence. (Tr 6, p. 218)

As a criticism of Du Bos, this is simply beside the point.⁴ Du Bos after all explicitly *agrees*, as for example here:

The massacre of the innocents must have left most gloomy impressions in the imaginations of those, who were real spectators of the barbarity of the soldiers slaughtering the poor infants in the bosom of their mothers, all imbrued with blood. (1748, p. 24)

The point of Hume's objection seems to be that the desire for emotional stimulation *by itself* cannot explain the phenomenon; we also need to be told, at

⁴Cf. Livingston (2013, p. 405).

the very least, what the effective difference is between art and real life. Why are we so delighted in the former case, when we are so distressed in the latter? Whatever one thinks of the ultimate success of Du Bos's account, he does at least *have* an answer to this question: art works are imitations, and as such they affect only our sensation and imagination, not our reason; they excite only artificial emotions that have none of the negative consequences that go along with their real-life counterparts.

When Hume then introduces Fontenelle's account as offering an "addition" to Du Bos's that speaks to this objection (Tr 6, p. 218), by locating the key difference between art and real life in the *fictional* nature of the former, his unfairness to Du Bos becomes yet clearer. For Du Bos, as we have seen, the key difference between imitations and real life is *precisely* that the former are fictional. Fontenelle's supposed "addition" was thus already in Du Bos's account to begin with.

Be that as it may, Hume goes on to make short work of Fontenelle's novel idea that a pain weakened becomes pleasant: "You may by degrees weaken a real sorrow, till it totally disappears; yet in none of its gradations will it ever give pleasure" (Tr 10, p. 221). He is similarly dismissive of the suggestion that the fictional nature of art works is the key feature that distinguishes them from reality:

The pathetic description of the butchery, made by VERRES of the SICILIAN captains, is a masterpiece of this kind: But I believe none will affirm, that the being present at a melancholy scene of that nature would afford any entertainment. Neither is the sorrow here softened by fiction: For the audience were convinced of the reality of every circumstance. (Tr 8, p. 219)

With this remark Hume does after all give us (though apparently unintentionally) a genuine criticism of Du Bos, as well as of Fontenelle. The comment, something of an aside in Hume's dissertation, raises to my mind a substantial issue, to which I will return in the next section.

Given the general ineffectiveness of Hume's criticisms—this one point aside—it seems to me unlikely that Hume changed his mind on this topic because of these worries. They are not, on inspection, very compelling reasons for abandoning the ideas hinted at by Fontenelle, or the theory carefully developed by Du Bos. It seems possible, therefore, that Hume simply hit upon his own new solution, and then hunted around for objections to these earlier writers so as to motivate it. There is also a third and potentially more interesting possibility, however, namely that Hume had another reason for wanting a different explanation, a reason that is not brought out in his explicit and rather weak objections. I will tentatively suggest such a reason in the next section.

11.5. Hume's later account

What is most distinctive about Hume's later explanation of the pleasure that we take in tragic art, when compared to Du Bos's and Fontenelle's, is the introduction of a completely new set of emotional responses: responses to the artwork itself, rather than to the events that it depicts, such as our delight in its beauty and in the skill of the artist. With these (positive) emotions now in the picture, Hume is able to apply his conversion principle. The negative passions arising from the events themselves, Hume says, are converted into or swallowed up by the predominant feelings of delight evoked by the artistry:

The genius required to paint objects in a lively manner, the art employed in collecting all the pathetic circumstances, the judgment displayed in disposing them: the exercise, I say, of these noble talents, together with the force of expression, and beauty of oratorical numbers, diffuse the highest satisfaction on the audience, and excite the most delightful movements. By this means, the uneasiness of the melancholy passions is not only overpowered and effaced by something stronger of an opposite kind; but the whole impulse of those passions is converted into pleasure, and swells the delight which the eloquence raises in us... The impulse or vehemence, arising from sorrow, compassion, indignation, receives a new direction from the

sentiments of beauty. The latter, being the predominant emotion, seize the whole mind, and convert the former into themselves, at least tincture them so strongly as totally to alter their nature. (Tr 9, pp. 219-20)

If Hume had only engaged properly with Du Bos here, there would have been a nice contrast to be drawn between their solutions. As we have seen, Du Bos's position is that in the real life case there is a mixture of pleasure and pain, in which the pain predominates; while in the artistic case, the painful aspects are removed. Hume however holds that the real life case is purely painful, while in the artistic case there is a mixture of pleasure and pain in which the pleasure predominates (before ultimately "swallowing up" the pain).

In the *Treatise*, when Hume introduces Du Bos's idea of the strong desire for emotional stimulation of all kinds, he attributes it (without mentioning Du Bos by name) to those "who take a pleasure in declaiming against human nature" (T 2.2.4.4, p. 352). He is in this way implicitly placing Du Bos in the wider group (which includes the egoists) of those who "give us a mean opinion of our nature" (DM 2, p. 81; recall §2.3). This points to the important Hobbesian context of Du Bos's solution. As Hume was no doubt aware, Hobbes had earlier expressed some closely related ideas:

[F]rom what Passion proceedeth it, that men take *pleasure* to *behold* from the shore the *danger* of them that are at Sea in a tempest; or in Fight, from a safe Castle to behold two armies charge one to another in the field? It is certainly, in the whole sum, *Joy*; else men would never flock to such a spectacle. Nevertheless there is in it both *Joy* and *Grief*: for as there is novelty and remembrance of our own Security present, which is *delight*: so there is also *Pity*, which is Grief: But the Delight is so far predominant, that men usually are content in such a case to be spectators of the misery of their friends. (1651, p. 114)

Whether Du Bos was directly inspired by Hobbes, I do not know, but there is a striking similarity in that both claim our experience in such cases is a mixture of pleasure and pain. Du Bos also discusses the same two examples of watching a

shipwreck from the shore, and a battle from a castle (1733, p. 13; 1748, p. 11). There is nothing in that to tie Hobbes *directly* to Du Bos—the examples are from Lucretius’s poem *De Rerum Natura*⁵—but the common source is significant in itself. Lucretius’s poem is an explication of Epicurean philosophy (indeed, it is one of our major sources of the ideas of this school), and both Hobbes and Du Bos are recognisably influenced by the Epicurean tradition in their psychological theories.

This context makes Hume’s rejection of Du Bos’s solution—indeed, his complete failure to engage with it properly—potentially more intelligible. As I argued extensively in Part 1 of this thesis, Hume’s philosophy of emotion underwent a major change after the *Treatise* away from egoism and the Epicurean tradition, and towards sentimentalism and the rival Stoical tradition. I suggested at the end of the previous section that Hume may have abandoned his earlier solution to the paradox of art and negative emotions simply because he came up with a novel one that seemed better. But I also suggested a more interesting possibility: that he later had other reasons for disliking Du Bos’s account that he did not make explicit. One possible such reason is his general attempt to disassociate himself from Hobbes and Epicurus, and to offer more morally palatable psychological explanations of the phenomena that he was investigating.

Returning to the other unanswered question from the previous section—of what the crucial distinguishing feature of artworks is, that makes our response different from what it would be in real life—the obvious first thought, given what we have now seen of his solution, is that Hume’s answer ought to be simply *that it is an artwork*. The difference in reactions, thinks Hume, is down to the addition of some positive emotions, but these are not in any way a response to the fictional nature of the events depicted (the events need not even be fictional). Rather, they arise because the events are depicted so beautifully, with such great skill and artistry.

⁵“Tis sweet, when on the might sea the storm winds rouse the main, / To watch from shore another toil with all his might in vain: / Not that the hurt of others can to us delightful be, / But that we like to look on ills from which ourselves are free. / Sweet is it too to view in line the mighty strife of war / Arrayed across the plains when we from danger stand afar.” (Baring 1884, p. 55)

This is certainly right as far as it goes. On closer inspection, however, it cannot be the whole of Hume's answer, for the following reason. Supposing that some work of art evokes pleasant emotions because of its beauty, and negative emotions because of its content, it is still an open question which of these two sets of emotions will be *predominant* in the mixture. In some cases, Hume says, the negative feelings dominate, and all the eloquence in the world will only serve to make us feel worse:

Who could ever think of it as a good expedient for comforting an afflicted parent, to exaggerate, with all the force of elocution, the irreparable loss, which he has met with by the death of a favourite child? The more power of imagination and expression you here employ, the more you encrease his despair and affliction. (Tr 21, p. 223)

More examples are given in the following paragraphs: Cicero's rhetorical skill in describing Verres' massacre delights the audience, but only distresses Verres himself all the more (Tr 22, p. 223); Clarendon overlooks the death of King Charles I in his history of the civil war, because "[h]e himself, as well as the readers of that age, were too deeply concerned in the events" (Tr 23, p. 223); the English theatre depicts too much blood and gore, creating a sense of horror that no amount of artistry can overcome (Tr 24, p. 224); even our sorrow at seeing virtuous characters suffer at the hands of the vicious predominates over any feelings of beauty, unless softened by having the virtue turn into a courageous despair, or the villains receive their proper punishment (Tr 25, p. 224).

Though this shows that a positive reaction to the beauty and brilliance of the artwork cannot be the sole difference, however, it does not show us what the crucial missing factor is. What is different about the cases in which this positive reaction predominates, as opposed to those in which it only serves to increase our pain? Having thus excited the reader's curiosity, I will follow Hume's advice (Tr 13, p. 221) and refrain from revealing the answer to this question until the next chapter (§12.4).

Chapter 12

The Conversion Principle

In the previous chapter, I placed Hume's conversion principle in its context, and looked at how Hume came to apply it to the paradox of art and negative emotions. I ended with an unanswered question: what is it that makes one passion predominant over another? Why is it, for example, that for most of us the pleasure of Cicero's rhetoric is heightened by the horrific subject matter of Verres' massacre, while for Verres himself the shame and terror is heightened by the rhetoric? This is one of three main interpretative questions regarding Hume's principle. The other two concern the nature of the process itself (how the predominant and subordinate passions interact), and its end result (what becomes of these two passions after their interaction).

In this chapter, I offer answers to each of these questions. It will be convenient to tackle the middle question first, concerning the nature of the process itself, since this question is both easier to answer, and has an impact on the answers to the other two questions. From the outset I should say that we cannot assume, simply because we want answers to these other two questions, that Hume himself had precise answers in his own mind. I will initially proceed on the basis of this assumption, and see which answers are best supported by the text. In the final section, however, I will argue that Hume was not in fact quite so clear on these matters as we might have wished.¹

¹I have presented much the same view before in answer to the third question (*what becomes of these two passions after their interaction*); see Merivale (2011). Here I extend my answer to include the first question as well (*what makes one passion predominant over the other*).

12.1. Swallowing up: passions or emotions?

Hume's statement of his conversion principle includes three images or metaphors:

It is a property in human nature, that any emotion, which attends a passion, is easily converted into it; though in their natures they be originally different from, and even contrary to each other... The predominant passion swallows up the inferior, and converts it into itself. The spirits, when once excited, easily receive a change in their direction; and it is natural to imagine, that this change will come from the prevailing affection. (P 6.2, p. 26; cf. T 2.3.4.2, pp. 419-20)

The first image invoked is that of the *conversion* of one passion into another. This is perhaps the most important idea; certainly it is the one that has dominated discussions of Hume's view, and indeed given the principle its familiar name. I will say more about this idea in §12.3 below. The second image is a meal metaphor: Hume talks of the predominant passion "swallowing up" the subordinate. The third is a gesture at the underlying physiology: the spirits are said to receive a change in their direction. I will say more about this last in §12.2.

The meal metaphor does not seem especially puzzling or problematic, and I have nothing to say about it directly. It is well to be aware of it, at least, for it will prove helpful when turning to consider what happens to the passions following their interaction (see §12.4). But in itself it does not seem to raise any interpretative questions. This metaphor does however provide the occasion for settling an important preliminary issue.

The issue in question arises from an influential suggestion by Alex Neill.² Neill's idea is that it is not the subordinate *passion* itself that is swallowed up by the predominant, in Hume's process, but rather a distinct item called an "emotion" that "attends" the subordinate passion. The textual springboard for this idea is Hume's very first statement of the principle: "It is a property in

²Neill (1998).

human nature, that any emotion, which attends a passion, is easily converted into it” (P 6.2, p. 26).

Interesting though this suggestion is, it does not seem to be supported by the text. As we saw in chapter 3, Hume does sometimes use “emotion” as a more general term than “passion”, to refer to *all* secondary impressions (including the moral and critical sentiments). But as this distinction is seldom Hume’s focus, for the most part the words are effectively just stylistic variants. Many passages are suggestive of this straightforward interpretation, but perhaps the clearest piece of evidence comes from a revision that Hume made to a paragraph from *Treatise* Book 2:

Thus a man, who, by any injury from another, is very much discompos’d and ruffled in his temper, is apt to find a hundred subjects of discontent, impatience, fear, and other uneasy passions; especially if he can discover these subjects in or near the person, who was the object of his first passion.
(T 2.1.4.4, p. 284)

The repetition of “passion” at the end here is somewhat inelegant, as Hume himself no doubt realised later. For in the corresponding passage in the *Dissertation on the Passions* it is avoided by using “emotion” instead:

Thus, a man, who, by an injury received from another, is very much discomposed and ruffled in his temper, is apt to find a hundred subjects of hatred, discontent, impatience, fear, and other uneasy passions; especially, if he can discover these subjects in or near the person, who was the object of his first emotion. (P 2.8, p. 8)

This provides clear evidence that these terms are just stylistic variants.³

³There is one other change in this paragraph that I have hidden from the above quotations: in the *Treatise*, Hume spoke of the person in question as the “*cause* of the first passion” (my emphasis), rather than the *object* of this passion. This was clearly an error, and one that Hume corrected in the *Dissertation on the Passions* (from its very first edition): recall that the passion of hatred, like all the double-relation passions, has a person for its *object*, not its cause (§9.4). Since this is a potentially misleading matter of substance, Millican and I have corrected it in our on-line edition of the *Treatise*, from which I am quoting throughout this thesis.

Since Hume never explicitly draws a distinction between “passions” and their attendant “emotions”, and since there was no such standard distinction that would have been familiar to Hume’s readers, particularly strong evidence is required to suggest that Hume really was operating with some such contrast in mind. Neill’s evidence, however, is all perfectly consistent with the more straightforward synonymy interpretation. For example, is it unsurprising that Hume should say “any emotion, which attends a passion, is easily converted into it” (P 6.2, p. 26), rather than use the more clumsy expression “any passion, which attends a passion”. Moreover, Neill seems to have misread this passage: the emotion in question is said to attend the *predominant* passion, but Neill’s interpretation requires that it attends the *subordinate*.

The effective synonymy of these two words in the present context is further established by the fact that Hume refers to the subordinate item, the thing that is converted or swallowed up, as *both* a “passion” *and* an “emotion”. Neill acknowledges one place in which Hume refers to the subordinate item as a passion (T 2.3.4.2, p. 420; cf. P 6.2, p. 26), but goes on to say that this “is in fact very much an exception to the rest of his talk about affective conversion”, in which “emotion” is the more common term.⁴ But Neill’s selective quotations in support of this claim are not representative. In the *Dissertation on the Passions and Of Tragedy* combined, the tally is in fact exactly equal, with the subordinate item being referred to as a “passion” five times and an “emotion” five times.⁵ The relevant section of the *Treatise* (T 2.3.4, pp. 418-22), which includes most of the *Dissertation*’s references, does not alter this picture.⁶ Thus Neill’s proposal is not borne out by the text.

⁴Neill (1998, p. 345).

⁵The “subordinate” or “inferior” item is termed a “passion” three times (P 6.2, p. 26; Tr 13, 27, pp. 221, 224). The item that stands in the relevant relation to the predominant passion, meanwhile, when it is not explicitly labelled as “subordinate”, is termed a “passion” twice (P 1.25, 6.4; 6, 27) and an “emotion” five times (P 6.2, 6.3, 6.6, 6.8, pp. 26, 27; Tr 14, p. 221), bringing the total to five a piece.

⁶The item said to be “converted” or “transfused” into the predominant passion is called an “emotion” twice (T 2.3.4.2, 2.3.4.5, pp. 419, 421) and a “passion” twice (T 2.3.4.2, 2.3.4.7, pp. 420, 421). There is a fifth reference to the “inferior emotion” (T 2.3.4.3, p. 420).

12.2. Movements in the spirits

Hume's reference to "spirits" in connection with his conversion principle is to be understood in the context of the physiological theory of the time. In the second century, Galen had hypothesised that food was converted into *natural spirit* by digestion, which latter was then carried (by the veins) to nourish the body's tissues. It was also sent to the heart, where—mixed with air breathed into the lungs—it was converted into *vital spirit*, which was transported throughout the body by the arteries. Some of this vital spirit, when it reached the brain, was transformed into *animal spirits*, which were sent through the nervous system. These animal spirits served a two-way communicative purpose: to send information from the sense organs to the brain, and to send instructions from the brain to the muscles.⁷

Aside from differences at the level of detail that needn't concern us here (chiefly concerning *how* the animal spirits effected these two communicative tasks), Galen's view of the nervous system was the standard one until Luigi Galvani produced compelling evidence (in 1780, shortly after Hume's death) that electricity rather than the hypothesised animal spirits performed the communicative function. (Other aspects of Galen's view were most notably challenged by William Harvey in 1628, who argued for the circulation of the blood, with the heart as its pump.) It is thus no surprise to find Descartes and Malebranche positing movements in the animal spirits as the immediate causes of the passions (or the occasions for God to bring about the passions, for Malebranche).⁸ The two had their own particular take on how the spirits managed to do this, but otherwise this claim was simply received wisdom: all the perceptions of the mind were presumed to be brought about in this way.

⁷Though Galen made some notable progress in our understanding of these matters, not all of these ideas were original, and the three sorts of spirits were not first hypothesised by him. For a thorough, book-length treatment of the history of this subject—from a physiologist, who therefore may be presumed to understand it all much better than I do—see Ochs (2004).

⁸Descartes, moreover, was an influential early critic of Harvey, agreeing with the principle of circulation, but rejecting the idea of the heart as a pump. See Gorham (1994).

Hume, like any learned man of his age, was familiar with the gist of the physiological theory of the day, as we may see particularly clearly in his essay *The Epicurean*:

The stomach digests the aliments: The heart circulates the blood: The brain separates and refines the spirits[.] (Ep 5, p. 140)

It should go without saying, however, that Hume was no physiologist. Though his work on the passions is peppered—very lightly—with mentions of the spirits and of movements in the spirits, in every case he is simply taking the received wisdom for granted. With just one exception, these claims never play any indispensable role in his psychological theories, and in the exceptional case Hume explicitly apologises for the anomaly:

’Twou’d have been easy to have made an imaginary dissection of the brain, and have shewn, why upon our conception of any idea, the animal spirits run into all the contiguous traces, and rouze up the other ideas, that are related to it. But tho’ I have neglected any advantage, which I might have drawn from this topic in explaining the relations of ideas, I am afraid I must here have recourse to it, in order to account for the mistakes that arise from these relations. (T 1.2.5.20, p. 60)

It may also be noted that this is in the context of Hume’s discussion of the association of ideas; in the case of the passions, remarks about the spirits add some nice visual imagery, but never serve any significant theoretical purpose.

In the context of the conversion principle, Hume’s talk of the spirits receiving a change in their direction is obviously to be understood as a gesture at the underlying physiology. But the principle itself is a *psychological* one, as far as Hume is concerned, and he supports it exclusively with *psychological* examples; the physiological descriptions of it do not seem to play any essential role in Hume’s argument, as far as I can see, and might just as well be deleted.

John Wright, as part of his insistence that Malebranche was an important positive influence on Hume, is keen on emphasising Hume’s occasional references to animal spirits. He claims that Hume “seems to have thought that the search

for something like the physical causes described in Cartesian psychophysiology was directly relevant to his own activities as a *moral* philosopher”.⁹ But apart from the isolated and apologetic example regarding the association of ideas that we have already seen, I am aware of no textual evidence to support this surprising claim. With regard to his Malebranchean interpretation of Hume’s associationism, Wright does acknowledge that “[t]his part of his account may not be entirely explicit”.¹⁰ And again with regard to Hume’s account of personal identity: “there is no explicit reference back to the neurophysiological account when Hume explains how we develop our idea of the continuous identical self”.¹¹ A less ingenious reader might rather contend that these things are entirely absent.¹²

Hume’s allusions to the physiology underlying his conversion principle may provide us with some clues pertaining to the interpretation of that principle (see §12.4 below). But it would be quite inappropriate to read them as serious attempts to uncover the physical processes of the brain. Hume seems to have been broadly familiar with the theory of the day, and that familiarity shows through in some of his writing, but we should not read too much into this unsurprising fact.

12.3. Conversion

This brings us to the aspect of Hume’s general description of the conversion principle that has dominated recent commentaries: the talk of one passion “converting” another into itself (P 6.2, 6.6, pp. 26, 27; Tr 9-12, 19-20, pp. 219-21, 222-3). There has been much speculation as to what Hume might have meant by this conversion in his discussion of tragedy. The most likely candidate, I suggest, is not quite the natural reading of the term. It is difficult to make sense of a way in which, for Hume, a particular perception of pride (for example) might

⁹Wright (1983, p. 15).

¹⁰Wright (1983, p. 68).

¹¹Wright (1983, p. 73).

¹²Recall §10.1, note 3.

have previously been a different passion. It seems that all Hume really *can* mean by conversion in this context is *causation*: when one passion is converted into another, this amounts to the former passion causing the latter to appear (and presumably vanishing itself in the process; see §12.4 below).

This suggestion is supported by Hume's use of the notion in other contexts. In section 2 of the *Dissertation on the Passions*, in his discussion of the causal genesis of pride and humility, Hume writes:

Beauty of all kinds gives us a peculiar delight and satisfaction; as deformity produces pain, upon whatever subject it may be placed, and whether surveyed in an animate or inanimate object. If the beauty or deformity belong to our own face, shape, or person, this pleasure or uneasiness is *converted into* pride or humility; as having in this case all the circumstances requisite to produce a perfect transition, according to the present theory. (P 2.17, p. 10; my emphasis)

The “present theory” here referred to is that of the double relation of sentiments and ideas, according to which the pleasure or uneasiness that Hume says is *converted into* pride or humility is a key step in the causal chain that *produces* these passions (recall §§4.1 and 9.5). Although this is the only one to remain in the *Dissertation*, there are more examples of this talk of conversion to be found in Book 2 of the *Treatise* (e.g. T 2.1.5.5, 2.1.6.2, 2.1.12.8, pp. 286-7, 290, 327).

It doesn't immediately follow, of course, that Hume meant the same thing by this word in his discussion of the conversion principle; but there is a default assumption in favour of this simple hypothesis. Moreover, there is direct evidence of synonymy when this principle is first described, in two sentences that I skipped over when quoting this passage at the start of the chapter (page 228):

It is a property in human nature, that any emotion, which attends a passion, is easily converted into it; though in their natures they be originally different from, and even contrary to each other. It is true, in order to cause a perfect union amongst passions, and make one produce the other, there is always required a double relation, according to the theory above delivered. But when two passions are already produced by their separate causes,

and are both present in the mind, they readily mingle and unite; though they have but one relation, and sometimes without any. The predominant passion swallows up the inferior, and converts it into itself. (P 6.2, p. 26)

Notice how Hume relates the conversion mechanism here to his theory of the double relation of sentiments and ideas. In order to make one passion “produce” another (cause and effect), a double relation is required between both the sentiment components and the idea components. But when both passions are already present in the mind, one will be converted into the other with just a single relation (meaning a relation of ideas).¹³ I suggest that this is not just an artful juxtaposition: production and conversion boil down to one and the same thing, namely, the causing of one passion by another.¹⁴

In the case of the conversion principle, the predominant passion in question is already present in the mind. And what good, one might reasonably ask, is a cause that only ever operates when its effect has already been produced by independent means? To make sense of this, we need to remember Hume’s two ways of describing the difference between calm and violent passions (§10.4). On the one hand, it is thought of in terms of a difference in violence or turbulence; the talk of conversion does not seem to make sense on this model, and this is the reason for the present worry. On the other hand, however, it is also thought of in terms of an increase in *quantity*, and on this model the idea of conversion (as causation) makes perfect sense. When two passions are both present to the mind, and one is converted into the other, this amounts to the one causing *more* of the other to appear.

¹³To be precise, Hume says with “but one relation, *and sometimes without any*” (my emphasis). It is unclear what this further qualification is doing, however. Hume never gives us any examples of one passion being converted into another even when their objects are unrelated. In fact, he told us earlier that “[i]n contrary passions, if the objects be totally different, the passions are like two opposite liquors in different bottles, which have no influence on each other” (P 1.24, p. 6). It seems, therefore, that this later qualification was a slip.

¹⁴The notion of conversion, we may recall, is also present in Hume’s account of the sympathetic communication of passions (§4.3). Here, however, it is not one sentiment being converted into another, but the *idea* of a sentiment being converted into that very sentiment. In *this* case, I take it that there is a genuine conversion, i.e. the very same thing changes from an idea into an impression. Since impressions and ideas differ only in their felt intensity, all that is required for this conversion is an injection of force and vivacity.

12.4. Before and after

Now that we have a sense of how Hume pictured the process itself, we may tackle the other two questions that I posed at the start of this chapter: what determines which of the two passions will predominate, and what happens to the passions at the end of the process? What happens to the *predominant* passion at the end of the process, at least, is obvious: the main result of the interaction is “to give additional force to the prevailing passion” (P 6.3, p. 26). Hume states this unambiguously, and we could have deduced it easily enough anyway from the context that we saw in the previous chapter. This principle is one of those that Hume offers concerning the causes of the violent emotions, and it wouldn’t belong in this context if its effect was not an increase in force or violence. This is also presumably the point of Hume’s talk of one passion swallowing up another: the predominant passion is *strengthened* by its metaphorical meal.

The question of what happens to the *subordinate* passion at the end of the process, however, is not so easily settled. Furthermore, this question has often been thought to be particularly important for the evaluation of Hume’s account, since there is some evidence to suggest that he supposes the mechanism utterly annihilates these melancholy passions. We might call this the *destruction* interpretation of the principle. This destruction interpretation, furthermore, is sometimes thought to entail that our experience of tragic drama is entirely pain-free. This, in turn, is believed to be incorrect as a description of the experience, and the inevitable conclusion is that Hume’s account fails.¹⁵

Taken at face value, the talk of conversion plainly implies the death of the passion that gets converted. This is certainly so if we give the word “conversion” its usual interpretation, and understand Hume as saying that the subordinate passion itself somehow *turns into* the predominant. But even if we understand conversion as essentially the causing of one passion by another, as I suggest we

¹⁵See Hill (1982, pp. 323-4), Packer (1989, p. 212), and Budd (1991, pp. 103-4). Those who don’t accuse Hume of this error still typically accept that it would be an error, and develop interpretations explicitly designed to overcome it; Yanal (1991, 1992), Neill (1992, 1998).

should, the natural reading is still that of a causal process involving the destruction of the cause. Nor can we dismiss this as a possibly unintended connotation of the word, for Hume could perfectly well have chosen a causal term with no such connotation (e.g. “produces”, “creates”). It is chiefly with a view to sidestepping this evidence in favour of the destruction interpretation that Neill introduces his distinction between “passions” and “emotions” (recall §12.1), urging that the emotion is converted rather than the passion, and therefore that the subordinate passion itself does not vanish at the end of the process. We have seen, however, that this distinction has no convincing basis in the text.

Hume’s gesture at the underlying physiology points in the same direction. If the stream of the subordinate passion is diverted, the spirits will no longer “rummage” the appropriate cell and actuate the felt emotion (T 1.2.5.20, pp. 60–1), and so it seems from this picture too that the subordinate passion is fated to disappear. As for the meal metaphor, once again the evidence clearly points in the same direction; for it is difficult to suppose that the subordinate passion might live on in the belly of the predominant, so to speak.¹⁶

This evidence is all in principle defeasible. It is open to Neill, for example, to maintain that Hume chose his general descriptions carelessly or badly, and that he never intended these implications. This hypothesis might then be backed up by reference to what Hume says about *particular instances* of his principle. And indeed it is Neill’s position that the particular instances—with one notable exception—point to the survival of the subordinate emotion.¹⁷ The exception is at Tr 27 (p. 225), where Hume says that “too much jealousy *extinguishes* love” (my emphasis). Neill acknowledges the exception,¹⁸ but thinks that the other examples outweigh the evidence of this one. On inspection, however, his arguments at this stage all rest on the supposed implausibility of applying the destruction interpretation to the particular cases (for an example, see page 241

¹⁶Neill also reads the meal metaphor as implying the destruction of the subordinate passion (Neill 1998, p. 344).

¹⁷Neill (1998, pp. 348ff).

¹⁸Neill (1998, p. 350).

below), and not on any direct textual evidence. At no point does Hume in fact say or even imply that the subordinate passion does survive the process.

It seems to me that many of Neill's criticisms of the destruction idea are not in fact decisive. I have argued this in detail elsewhere, and will not repeat these arguments here.¹⁹ The present point is simply that they are *philosophical* objections, not *textual*. Ultimately, however, I do think that Neill came very close to putting his finger on a difficulty here that is both philosophical *and* textual, and in the next section I will suggest a subtler refinement of the destruction interpretation. For now, however, I submit that interpretation as provisionally the best of the available options.

So much for the fate of the passions *after* the process. I turn now to the question of what makes one passion predominant over the other in the first place. To the best of my knowledge, only one answer has been given to this question in the secondary literature, namely that the predominant passion is the more forceful of the two.²⁰ I agree that this is the best available answer, but it is worth pointing to the reasons in support of this interpretative claim, since those who have made it before seem to have done so without offering any such support (and the quality of this evidence will also be important in §12.5 below). The general argument runs as follows: there is only one way in which a passion can vary (other than in its causes or objects), and that is in its force or violence; this is therefore the only salient way in which the predominant passion might differ from the subordinate one; and since it is absurd to suppose that the predominant passion is the *less* forceful of the two, it must be the more forceful.

To this general argument we can add two specific passages from *Of Tragedy*. Towards the end of this dissertation, Hume offers a handful of examples in support of the "inversion" of the conversion principle (Tr 27, p. 224), which are in fact just examples of the principle itself, but with the predominant passion being

¹⁹See Merivale (2011).

²⁰This answer is first given clearly and explicitly by Hill (1982, p. 322). Hill attributes it to Paton (1973), who does seem to hint at it (pp. 127-8), but I do not myself find it obvious that this is what she is saying. The same answer is later given by Budd (1991, p. 95).

a passion that in other (more common) situations is subordinate. The shame of Verres on hearing Cicero's speech in prosecution of him, we are told, would have been "*too strong* for the pleasure arising from the beauties of elocution" (Tr 22, p. 223; my emphasis). This is what makes that painful passion predominate in Verres' mind; and "strength" here is presumably to be equated with "force" and "violence".²¹ Furthermore, Hume later adds that "[*t*]oo much jealousy extinguishes love" (Tr 27, p. 225; my emphasis again). Here the metaphor is that of volume rather than violence, but, as we have seen, the underlying property is the same.

There is a difficulty with this answer to our first question, however, one that has to some extent been noted already by Malcolm Budd, who criticises this aspect of Hume's proposed solution to the problem of tragic pleasure.²² His objection, in a nutshell, is that it just isn't plausible to suppose that the conversion principle, thus understood, applies in the case of tragedy, because one would have to suppose that the pleasurable passions roused by the eloquence with which the melancholy scene is depicted are always more forceful, to begin with, than the painful passions caused by the scene itself.

The example of Verres gives us a hint of why Hume might have thought that the pleasurable passions are (in the appropriate cases) always more forceful than the painful. All Cicero's rhetorical skill supposedly could not induce feelings strong enough to efface Verres' shame and guilt, which painful passions instead predominated and thereby gained greater force as a result of the orator's eloquence (Tr 22, p. 223). Why is it, then, that for *us* Cicero's rhetoric wins out over the horror? The salient difference for Hume would seem to be that we are not so intimately connected with the tragic facts. Similarly with the case of

²¹In one place in the *Treatise* Hume does distinguish between "violent" and "strong" passions, the latter being those that have a greater influence on the will, regardless of their felt intensity (T 2.3.4.1, pp. 418-9). But elsewhere "strength" is clearly just another term for "violence" (e.g. T 2.2.2.23, p. 344), and the present passage makes much more sense read in this way than in the other. It is not likely that Hume means to say here that Verres' shame has a greater influence on his actions or volitions than the sentiment of beauty. That may be true, but hardly seems relevant.

²²Budd (1991, pp. 96ff).

Lord Clarendon overlooking King Charles I's death in his history of the civil war: "He himself, as well as the readers of that age, were *too deeply concerned in the events*, and felt a pain from subjects, which an historian and a reader of another age would regard as the most pathetic and most interesting, and, by consequence, the most agreeable" (Tr 23, pp. 223-4; my emphasis). The difference, for Hume, is in the distance.

This at least is how Hume might attempt to meet Budd's objection, though how successful such an attempt could be I leave for others to judge. For even if the objection can be met, a broader worry would still remain regarding this aspect of the conversion principle. This broader worry, meanwhile, is closely analogous to a similar worry with the other aspect of this principle discussed here (the fate of the subordinate passion), and I will therefore examine both matters in the next section. For now, suffice it to conclude that the best reconstruction of Hume's principle, on the basis of the text, is as I have been arguing: the predominant passion (i.e. the antecedently more forceful of the two) destroys the subordinate passion, and acquires yet greater force as a result.

12.5. A subtler interpretation

The textual evidence that I offered in the previous section is far from decisive. Hume's metaphors certainly *indicate* that the fate of the subordinate passion is to disappear, but there is only one comment that backs this up in a particular case. It seems that the only plausible answer to the question of which passion predominates, meanwhile, is that the predominant passion is the more forceful of the two; but again only two particular examples back this up explicitly. I have argued that these answers are more defensible than any alternative, but the weakness of the evidence in *any* direction here suggests a subtler interpretation: that although this might be (or should be) Hume's *considered* position, he never actually considered either of these questions very closely, so as to frame clear and definite answers in his own mind. I shall now give some reasons for thinking that this subtler interpretation is in fact closer to the mark.

The main textual difficulty with the cruder interpretation that I have been developing so far is also, it seems to me, the main philosophical difficulty with Hume's account. Hume wants his conversion principle to be a general psychological principle, accounting for a range of phenomena. The particular examples that he appeals to must therefore all have a sufficient amount in common, with these common elements comprising the essence of the principle. It is not enough, clearly, that the various cases all involve two passions interacting somehow. Rather, the two passions must be distinguished in the same way every time, and the end results for both must be the same every time.

The difficulty here is that Hume himself seems to have been somewhat lax in meeting these requirements. Nowhere does he take the time to argue, for *any* of his examples, that the subordinate passion is destroyed at the end of the process. Nor does he ever take the time to argue, for any of his examples, that the passion that predominates is, at the start of the process, the more forceful of the two. In the absence of this detail, Hume's justification of his principle falls far short. His several examples, we can allow, are all *similar* at some broad level of description. But are they similar *enough*? Are they really instances of *the very same* phenomenon?

We have seen hints of this problem before that can now be brought out more clearly. Neill, we may recall, thinks that in several of Hume's particular examples of the conversion principle it is highly implausible to suppose that the subordinate passion is destroyed at the end of the process. (This is in the context of arguing that Hume *didn't* in fact suppose this; if I am right in thinking this argument unsuccessful, the point now becomes an objection to Hume himself.) Take the case of grief at a friend's death increasing one's love for him, for example (Tr 16, p. 222). Neill thinks that to suppose one's love destroys the grief in this case is "grotesquely implausible",²³ and I am inclined to agree. But that Hume (if he was committed to the destruction of the subordinate passion) might be *wrong* about some of his examples is not so much the problem; more immediately problematic

²³Neill (1998, p. 351).

is the fact that he *systematically fails to give any argument* for this commitment. I have said that the destruction interpretation is the most textually plausible of the available options here. But if we take a step back, ultimately the most plausible interpretation would seem to be that Hume never quite realised the need to give a definite answer to the question of what happens to the subordinate passion at the end of the process.

Much the same is true concerning the question of what makes one passion predominant over the other. Budd, we have seen, comes close to appreciating the difficulty here when he challenges the claim that, in the case of tragic pleasure, the joys of the eloquence are always stronger than the pains of the drama. I went some way, in the previous section, in defending Hume against this charge, but really this is just the tip of the iceberg. For it is not enough for this to be true in the case of tragedy; it must also be true in *every* case that Hume produces as support for his principle. Hume makes no attempt, however, to argue for this in *any* of the cases, and it is far from obvious that it is indeed true for them all.

Consider, for example, Hume's claim that "when good or evil is placed in such a situation as to cause *any* particular emotion, besides its direct passion of desire or aversion, this latter passion must acquire new force and violence" (P 6.5, p. 27; my emphasis). For this to be true, it must also be true that no passion is ever more violent than the passions of desire and aversion; for if there were such a passion, it would convert the desire or aversion into itself, rather than the other way round. Or again, consider Hume's account of why "every thing that is new, is most affecting" (P 6.12, p. 28), namely that the passion of surprise or wonder nourishes any emotion attending the novel object. For this to be true, it must be the case that surprise or wonder is the weakest of all the passions; for if something new were the object of a weaker passion, this weaker passion would nourish our surprise, rather than be nourished by it.

None of these commitments seems particularly plausible. But again, the immediate problem is not so much that they might be false, but rather that Hume systematically makes no attempt to argue for them. I have suggested that

the best answer to the question of what makes one passion predominant over another, for Hume, is that it is the antecedently more forceful of the two. He does more or less say as much in two places in *Of Tragedy*, and in any case this seems like the only answer that Hume has available to him. But if we take a step back, ultimately the most plausible interpretation seems to be that Hume never really thought through this commitment fully.

Regarding both of these questions, then, concerning the status of the passions before and after the conversion process, I submit that no definite answer can be given. If we must give answers, then the view best supported by the text seems to be that the predominant passion is the more forceful of the two, and that it destroys the subordinate passion. But it would be overstating the case to say that this is definitely what Hume thought. I suspect that he had no definite thoughts on the matter one way or the other. And after all, this is perhaps only to be expected: Hume was pioneering a new science of experimental psychology. It is unreasonable to suppose that his principles must be as clearly and precisely formulated as we would expect today.

Chapter 13

Passion and Value

In chapter 10, we saw that, as a consequence of the specialisation view of reason and passion, the passions for Hume cannot be contrary to reason (except in a trivial sense, when they are founded on some false belief). This naturally prompts the question of how passions can be evaluated beyond the minimal constraint that they be based only on true beliefs. In this final pair of chapters, I turn to the fourth dissertation, *Of the Standard of Taste*, and to a part of Hume's reply to this question, namely the part concerning our emotional responses to works of art. Since the appreciation of artistic beauty, for Hume, just *is* an emotional response, this comes to the question of the value of art quite generally. In reconstructing Hume's position, more points will emerge in support of my Unity Thesis. The differences between this fourth dissertation and Hume's remarks on beauty in Book 2 of the *Treatise*, meanwhile, have already been examined in §5.5.

13.1. The sceptical principle

At the heart of Hume's anti-rationalism, as we saw in chapter 10, is what I have called the specialisation view of reason and passion (as opposed to Malebranche's and Clarke's hierarchical view). On this picture, both reason and sentiment have a role to play in motivation; roughly, the latter to set the goals, and the former to tell us how to achieve those goals. My focus then was on motivation, but as I noted at the time Hume held a similar view with regard to approbation, in the case of morals as well as art:

Thus the distinct boundaries and offices of *reason* and of *taste* are easily ascertained. The former conveys the knowledge of truth and falsehood: The latter gives the sentiment of beauty and deformity, vice and virtue. (M App1.21, p. 294)

In his essay *The Sceptic*, Hume has the title character endorse something that I will refer to as Hume's *sceptical principle*:

If we can depend upon any principle, which we learn from philosophy, this, I think, may be considered as certain and undoubted, *that there is nothing, in itself, valuable or despicable, desirable or hateful, beautiful or deformed; but that these attributes arise from the particular constitution and fabric of human sentiment and affection.* What seems the most delicious food to one animal, appears loathsome to another: What affects the feeling of one with delight, produces uneasiness in another. This is confessedly the case with regard to all the bodily senses: But if we examine the matter more accurately, we shall find, that the same observation holds even where the mind concurs with the body, and mingles its sentiment with the exterior appetite. (Sc 8, p. 162; my emphasis)

There is no guarantee, from this passage alone, that Hume himself endorsed this sceptical principle (for he is not speaking here in his own voice). But the principle is closely related to the specialisation view of reason and passion, being arguably just another way of saying the same thing. And in any case, Hume explicitly endorses the principle himself, at least in the case of art (my present focus), in *Of the Standard of Taste*: “it [is] certain, that beauty and deformity, more than sweet and bitter, are not qualities in objects, but belong entirely to the sentiment, internal or external” (ST 16, p. 235).

The obvious worry, for those who are disinclined to agree with this sceptical principle or the specialisation view of reason and passion, is that it leads to relativism. When different people find different things desirable, valuable, or beautiful, on account of their different constitutions (and not because of any differences of opinion concerning the facts), reason—on this view—is not in a

position to adjudicate between the rival sets of sentiments. But if not reason, then what? On what basis can we pronounce one constitution or resulting sentiment better than another?

Hume himself was keenly aware of this difficulty, and the dissertation *Of the Standard of Taste* constitutes his attempt to respond to it in the case of beauty in works of art. (His parallel attempt in the case of morals and motivation is of course to be found in the moral *Enquiry*.) Putting the case forcefully on the other side, he begins this fourth dissertation as follows:

The difference, it is said, is very wide between judgment and sentiment. All sentiment is right; because sentiment has a reference to nothing beyond itself, and is always real, wherever a man is conscious of it.^[1] But all determinations of the understanding are not right; because they have a reference to something beyond themselves, to wit, real matter of fact; and are not always conformable to that standard. Among a thousand different opinions which different men may entertain of the same subject, there is one, and but one, that is just and true... On the contrary, a thousand different sentiments, excited by the same object, are all right: Because no sentiment represents what is really in the object... Beauty is no quality in things themselves: It exists merely in the mind which contemplates them; and each mind perceives a different beauty... [E]very individual ought to acquiesce in his own sentiment, without pretending to regulate those of others. To seek the real beauty, or real deformity, is as fruitless an enquiry, as to pretend to ascertain the real sweet or real bitter. (ST 7, p. 229-30)

The starting point for this relativistic argument is precisely the sceptical principle, which as I have said Hume himself accepts. But he is anxious to resist the

¹Hume's language here is reminiscent of the problematic premise in his representative quality argument, which states that "[w]hen I am angry, I am actually possest with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high" (T 2.3.3.5, p. 415). As I argued in §9.2, Hume's point there is not that the passions have no intentional objects, but rather that they are not truth-apt, that they do not purport to correspond to reality in the way that beliefs do. It should be clear, I trust, that Hume means exactly the same thing here. This passage may indeed be considered further confirmation of my earlier claim.

relativist's inference from this principle to "the natural equality of tastes" (ST 8, p. 231). He is anxious, that is to say, to erect a *standard of taste*, according to which sentiments can be assessed as either right or wrong.

The essence of Hume's view, put very simply, is that the right sentiments are those felt by the *right* sort of people, namely the "true judges": "the joint verdict of such, wherever they are to be found, is the true standard of taste and beauty" (ST 23, p. 241). This move seems structurally sound, and a coherent way of resisting the relativist's conclusion while nevertheless maintaining the sceptical premise. But of course the two questions now are, who are these true judges, and—more importantly—why are *they* the right sort of people? Why is it that *their* sentiments are fit to provide the standard for all of us, rather than the sentiments of some other group of people (or none at all, as the relativist maintains)? The remainder of this chapter will focus on the first question; in the next chapter I will attempt to reconstruct Hume's answer to the second.

13.2. The true judges

The first quality that Hume insists the true judge must possess is a delicacy of taste or imagination, which is a heightened sensitivity to the causes of the sentiment of beauty:

Though it be certain, that beauty and deformity, more than sweet and bitter, are not qualities in objects, but belong entirely to the sentiment, internal or external; it must be allowed, that there are certain qualities in objects, which are fitted by nature to produce those particular feelings. Now as these qualities may be found in a small degree, or may be mixed and confounded with each other, it often happens, that the taste is not affected with such minute qualities, or is not able to distinguish all the particular flavours, amidst the disorder, in which they are presented. Where the organs are so fine, as to allow nothing to escape them; and at the same time so exact as to perceive every ingredient in the composition: This we call delicacy of taste, whether we employ these terms in the literal or metaphorical sense. (ST 16, p. 235)

Hume illustrates what he has in mind here with a story from *Don Quixote*, in which Sancho Panza reports that two of his ancestors were able to discern a slight taste of iron and leather respectively in an otherwise fine wine. Sure enough, when the barrel was emptied, an iron key with a leather cord attached was found at the bottom. Hume's delicacy of taste is intended to be the mental equivalent of this refined palate.

Some people, Hume seems to think, will naturally have a more delicate taste than others. But the skill can be honed with sufficient practice:

But though there be naturally a wide difference in point of delicacy between one person and another, nothing tends further to encrease and improve this talent, than practice in a particular art, and the frequent survey or contemplation of a particular species of beauty. (ST 18, p. 237)

So as to fix the proper degree of beauty or deformity, furthermore, true judges must also be experienced in a range of works. Otherwise they will be in danger of rating too highly something that is merely quite good:

A man, who has had no opportunity of comparing the different kinds of beauty, is indeed totally unqualified to pronounce an opinion with regard to any object presented to him. By comparison alone we fix the epithets of praise or blame, and learn how to assign the due degree of each... The most vulgar ballads are not entirely destitute of harmony or nature; and none but a person, familiarized to superior beauties, would pronounce their numbers harsh, or narration uninteresting... One accustomed to see, and examine, and weigh the several performances, admired in different ages and nations, can alone rate the merits of a work exhibited to his view, and assign its proper rank among the productions of genius. (ST 20, p. 238)

It is common to count five characteristics of Hume's true judge, with those just mentioned—delicacy of taste, practice, and comparison—being the first three. (I will come to the last two shortly.) To be precise, however, practice and comparison are not additional characteristics in their own right; rather, they are ways of

honing one's delicacy of taste. As Hume says, the true judge has a "delicate sentiment, *improved* by practice, *perfected* by comparison" (ST 23, p. 241; my emphases).² But this is a point of no great significance.

The need for practice and comparison is not, for Hume, something peculiar to the sentiments of beauty and deformity, or to the appreciation of art. On the contrary, he maintains that it is a quite general feature of human psychology that our sentiments require experience and comparison to settle into something suitably ordered:

[I]f a person full grown, and of the same nature with ourselves, were on a sudden transported into our world, he would be much embarrassed with every object, and would not readily determine what degree of love or hatred, of pride or humility, or of any other passion should be excited by it. The passions are often varied by very inconsiderable principles; and these do not always play with perfect regularity, especially on the first trial. But as custom or practice has brought to light all these principles, and has settled the just value of every thing; this must certainly contribute to the easy production of the passions, and guide us, by means of general established rules, in the proportions, which we ought to observe in preferring one object to another. (P 2.47, pp. 16-7; cf. T 2.1.6.9, pp. 293-4)

My main case for placing *Of the Standard of Taste* in the context of Hume's philosophy of emotion will be given in the next chapter. But the present observation may serve as a hint of the fuller argument to come. The sentiments of the true judges determine the standard of taste; but it is the science of the passions that determines what the true judges must be like.

The next feature of the true judge—either the second or the fourth, depending on whether practice and comparison count—is a suitable lack of prejudice (ST 21, pp. 239-40). This impartiality takes two forms. First, one must consider oneself "as a man in general" (ST 21, p. 239), forgetting any particular relationship that one may have with any of the people involved (notably the author of the

²Cf. Jones (1976, pp. 333-4), Weiland (1984, pp. 135-6).

artwork). Secondly, one must take account of changing times and customs. Art is not created in a historical vacuum, but is designed with a particular audience in mind. If we are not ourselves part of that intended audience, and we wish to assess the merit of the piece, we must imaginatively enter into their point of view. For example:

An orator addresses himself to a particular audience, and must have a regard to their particular genius, interests, opinions, passions, and prejudices; otherwise he hopes in vain to govern their resolutions, and inflame their affections... A critic of a different age or nation, who should peruse this discourse, must have all these circumstances in his eye, and must place himself in the same situation as the audience, in order to form a true judgment of the oration. (ST 21, p. 239)

This second form of impartiality is not so much a lack of prejudice, as an ability to take on the appropriate prejudices. As Michelle Mason nicely puts it, “Hume’s judge [is] less an impartial observer than a cultural chameleon”.³

Finally, the true judge must also be suitably intelligent: “reason, if not an essential part of taste, is at least requisite to the operations of this latter faculty” (ST 22, p. 240). This parallels Hume’s view in the moral case that “in order to pave the way for such a sentiment [i.e. moral approbation or disapprobation], and give a proper discernment of its object, it is often necessary, we find, that much reasoning should precede, that nice distinctions be made, just conclusions drawn, distant comparisons formed, complicated relations examined, and general facts fixed and ascertained” (M 1.9, p. 173). Reason’s role in the appreciation of art includes such things as determining whether a work succeeds in what it is trying to do, and whether the characters depicted behave in credible ways. Though Hume doesn’t say this himself, we should surely add that certain background knowledge is generally required: of the history of the genre, say, of any symbolic

³Mason (2001, p. 60). Mason sees this interpretation of freedom from prejudice as an *alternative* to the more literal interpretation according to which the true judge must have no prejudices. But it seems to me clear that there are *two* aspects to the impartiality that Hume has in mind, as described above; and the first aspect is indeed a literal lack of prejudice.

conventions that the artist might use, or of any recent or current events that might form the context of the work.

Such, then, is the character of Hume's true judge. And it is the sentiments of these people that set the standard for all of us. The pressing question, of course, is why these people or their sentiments are so special. As I have said, however, I will save tackling this question for the next chapter. In the meantime, there is a preliminary matter to set straight.

13.3. The threat of circularity

This preliminary matter is the well-known threat of circularity in Hume's account: "whereby"—as Peter Kivy puts it—"good art is defined in terms of the good critic and the good critic in terms of good art".⁴ It may help to illustrate the worry here with a caricature of Hume's view. Suppose, for example, that we want to know whether Scottish poetry is better than English poetry. The matter is particularly puzzling, because there seems to be no general agreement: the Scottish prefer Scottish poetry, but the English are convinced that theirs is greatly superior. Our Humean caricature has a quick way to settle the matter. Scottish poetry is better, he says, because the Scottish are much better judges. But why, we now ask, are the Scottish better judges? The answer is surely obvious: because they prefer the better poetry.

This account is clearly hopeless; such circular reasoning can be applied just as well—which is to say, just as badly—in support of precisely the opposite conclusion. Now this is certainly a caricature of Hume's position, and it is easy to see why. Hume is indeed committed to the view that some works are better than others because they are preferred by the better judges. But he does not say, at least not explicitly, that these judges are better because they prefer the better works. He says that they are better because they have a more delicate taste, more experience, more intelligence, and are free from prejudice.

⁴Kivy (1967, p. 60).

The threat of circularity is not so easily avoided, however. For the worry is that, once we look *in detail* at the characteristics of the true judge, we will find some ineliminable reference to the *quality* of works of art. The difficulty here is not with intelligence or freedom from prejudice, but with delicacy of taste and the practice and comparison that is supposed to hone this skill. Roughly, the worry is that a delicate taste ultimately boils down to an ability to appreciate *good* works of art. In which case, the circle is slightly bigger than in the caricature above, but it is still a circle.

According to Kivy, Hume's view of practice and comparison falls afoul of precisely this problem:

Practice Hume thinks of as 'the frequent survey of a particular species of beauty'. *Use of comparisons* requires juxtaposing 'the several species and degrees of excellence'. But we must be able to recognize the beautiful before we are able to determine whether a critic has or has not been engaged in 'the frequent survey or contemplation of a particular species of beauty'. We must know what is excellent before we are able to determine whether or not a critic has compared 'the several species and degrees of excellence'.⁵

To put the point another way, there is a problem of familiarity here. Those who have more experience of Scottish poetry may prefer it to English verse, while those who have more experience of the latter may have the contrary preference. Whom then do we trust? On Kivy's interpretation, Hume answers that we must trust those who have more experience of the *better* poetry. But the point is that we are not yet in a position to know which poetry is better; that is the very question to be settled.

This particular criticism, it seems to me, can be easily met, for Kivy's interpretations of practice and comparison are mistaken. The passages that he quotes do suggest, out of context, that Hume thought the true judges must have more experience of *better* works of art. But the more plausible interpretation overall is that he held, more simply, that the true judges must have more experience of

⁵Kivy (1967, pp. 60-1).

works of art *in general*, both good and bad. In the sentences immediately following the first passage that Kivy quotes, Hume talks of “beauties *and defects*” (ST 18, p. 237; my emphasis), indicating that the “species of beauty” in question is rather a species of *attempted* or *intended* beauty, a set that will include bad instances of the type alongside the good. We must give a similar interpretation to the “several species and degrees of excellence”, which are said to include “[t]he coarsest daubing” and “[t]he most vulgar ballads” (ST 20, p. 238) alongside the genuinely excellent works.

In order to establish who has the better experience *in general*, it is of course not necessary to know in advance which works are good and which are bad. And to return to the question of whom we are to trust—those who have more experience of Scottish poetry, or those who have more experience of English poetry—Hume’s answer is *neither*. Rather, we should seek out someone (intelligent and unprejudiced) who has a decent experience of *both*, and erect our standard on the strength of their opinion.

The main threat of circularity, however, comes not from practice and comparison, but from the delicacy of taste that these things are supposed to refine and perfect. There is a genuine worry here as to whether this characteristic can be fleshed out independently of any reference to good art. The worry emerges from Hume’s own claim that “the best way of ascertaining” whether someone has delicacy of taste, “is to appeal to those models and principles, which have been established by the uniform consent and experience of nations and ages” (ST 17, p. 237). Here Hume is explicitly appealing to *good* works of art, or principles linking works with the *positive* sentiments such as beauty, and the threat of circularity looms large. It was this passage, indeed, that caused S. G. Brown to pose the original problem, in the earliest statement of it that I have come across.⁶

13.4. Getting into the circle

Hume also discusses delicacy of taste in his essay *Of the Delicacy of Taste and Passion*, in which—as one may guess from the title—he contrasts it with delicacy

⁶Brown (1938).

of passion. Where delicacy of taste, as we have seen, is a heightened sensitivity to (the causes of) the sentiments of beauty and deformity, delicacy of passion is a corresponding sensitivity to the general ups and downs of life: it makes those suffering from it “extremely sensible to all the accidents of life, and gives them a lively joy upon every prosperous event, as well as a piercing grief, when they meet with misfortunes and adversity” (DT 1, pp. 3-4).

Kivy, although he thinks that practice and comparison do give rise to a vicious circle in Hume’s account, argues that delicacy of taste itself is not thus problematic. His proposed solution in this case, however, is both philosophically and textually dubious. His claim is that delicacy of taste is an invariable concomitant of delicacy of *passion*, and that delicacy of passion can be identified without reference to art works at all (let alone exclusively good ones). Thus to spot someone with a delicate taste we need only look out for someone with delicacy of passion, which we can do without any knowledge of art at all.

Kivy’s proposal is certainly structurally sound, and it is also true that delicacy of passion is identifiable without reference to works of art. But his first claim—that delicacy of passion invariably goes along with delicacy of taste—seems in itself highly implausible.⁷ In any case, this is not a way out of the circle that is available to *Hume*, for he explicitly rejects the claim. The point of his essay *Of the Delicacy of Taste and Passion* is precisely to recommend cultivating the former but steering clear of the latter. This advice would hardly count for much if Hume thought that, as a matter of fact, you couldn’t have one without the other. Thus the threat of circularity still remains.

I believe, however, that the criticism is ultimately unfounded. The first point to make in responding to it is that there are potentially two different kinds of circle here, a *metaphysical* one and an *epistemic* one. On the metaphysical circle, being approved of by good critics is what *makes* something a good work of art, while approving of good works of art is (at least a part of) what *makes* someone a good critic. On the epistemic circle, by contrast, being approved of by good critics is *evidence that* something is a good work of art, while approving of good

⁷For a criticism, see Carroll (1984).

works of art is *evidence that* someone is a good critic. The metaphysical circle would certainly be a problem; but, I will argue, Hume is not committed to it. Hume may well be committed to the epistemic circle, meanwhile. But that, I will argue, is not a problem.

Being approved of by the good critics is indeed what makes something a good work of art for Hume. But approving of good works of art is not what makes someone a good critic. What makes someone a good critic are the qualities examined in §13.2: good sense, freedom from prejudice, and delicacy of taste. And what constitutes delicacy of taste, in particular, is that “the organs are so fine, as to allow nothing to escape them; and at the same time so exact as to perceive every ingredient in the composition” (ST 16, p. 235). There is nothing circular here.

The passage that troubled Brown, and was among those that started the original worry, merely states that established models and principles are the best way of *ascertaining* whether someone has delicacy of taste. This is an evidential or epistemic claim, and not one that can be in any danger of landing Hume in a metaphysical circle. And indeed, the problem seems to have persisted in an epistemic form, albeit not always explicitly so. Kivy, for example, initially sets up the difficulty in terms of two things being *defined* with circular reference to each other, suggesting a metaphysical circle; but when he gets into the details, his discussion is clearly epistemic: “we *know* a good critic to possess these qualities because he approves of good art”; “*How are we to determine* whether or not a critic possesses delicacy of taste in the aesthetic sense?”; “Here again is a quality of good critics that *can be recognised* prior to any knowledge of the critic’s aesthetic performance”.⁸ Nor is Kivy by any means unusual in this regard; for Jeffrey Weiland, for example, the key question is: “how do we *know* that someone has the five characteristics?”⁹

We can take it as read that approving of good art is evidence that someone is a good critic, and, in particular, evidence that they have delicacy of taste. We

⁸Kivy (1967, pp. 60, 61, 63; my emphases).

⁹Weiland (1984, p. 138; my emphasis).

can also take it as read that being approved of by good critics is evidence that something is a good work of art. It would be absurd for Hume to deny either of these claims. The existence of an evidential circle here is not problematic in itself. The trouble would be if this was the *only* evidence that we had on either side, so that there was no way into the circle from the outside.

But why suppose that this is the only evidence that we have on either side? For Hume, at least, it is not. As Kivy himself points out, there are various ways of establishing whether someone has good sense and is free from prejudice, independently of their artistic judgements.¹⁰ But to go straight to the heart of the matter: we don't necessarily need good critics even to establish the established models. These can be established, for example, by their ability to stand the test of time (as with Homer; ST 11, p. 233). Alternatively, find someone with good sense, who is free from prejudice (you may be such a person yourself), and then see how they (you) feel about works that don't require a great deal of delicacy to appreciate; the ones whose features are presented "singly and in a high degree" (ST 16, p. 235). However you do it, gather some evidence that such-and-such is a good feature of works of art. If you then find yourself enjoying works that possess that quality "in a smaller degree" (*ibid.*), you have grounds for thinking that your taste is delicate.

13.5. Passing the buck

Kivy follows up his treatment of the circle with the challenge of an infinite regress. All Hume's account does, he maintains, is pass the evaluative buck:

The phrase *good sense describes*; it also *approves*. What has happened is that in his attempt to reduce disagreements about aesthetic values to disagreements about facts, Hume has simply pushed the value judgement a step back: the question *Is x a good poem?* has become: *Does y have good sense?* And both are evaluative questions, questions of 'sentiment', not (solely) questions of fact. (p. 64)

¹⁰Kivy (1967, pp. 62-3).

Kivy focuses his complaint here on the characteristic of good sense (i.e. intelligence), and he was perhaps misled by the mere presence of the word “good”. Reason has a definite purpose for Hume, namely to discern truth from falsehood, and it is an objective matter whether or not it is successful in this task. Thus the question of whether or not someone has good sense is indeed a factual one, not a matter of sentiment. Good sense is not the ability to reason in a *pleasing* way, but in an *accurate* way.

That being said, there is nevertheless a legitimate question in this area, and it is the one that I have already raised and postponed. Let us grant that the question of whether or not someone possesses the qualities of the true judge is a *factual* question, rather than a matter of sentiment. It is not dependent on a certain sentimental response to works of art (as in the problem of the circle), and nor is it dependent on a certain sentimental response to the judges’ qualities themselves (as in this new problem of Kivy’s). Even so, the judgment that these people—however impartially characterised—are the *right* sort of people, the people whose opinions we *ought* to erect as our standard, would still seem to be a matter of sentiment. In any case, whether it is a matter of sentiment or not, the pressing question for Hume’s account is *why* the people thus identified set the standard for the rest of us.

Hume has recently been charged, by Brian Ribeiro, of failing to have an answer to this question.¹¹ Ribeiro characterises Hume’s true judges, not as those people with the specific qualities that Hume singles out, but rather as the people who emerge after a certain “programme of aesthetic education”.¹² As far as I can tell nothing hangs on this characterisation; I mention it simply so that the following quotation will make sense out of context. For here is the objection forcefully stated:

I am not simply raising the complaint that Hume is trapped in circularity:
we identify the good art by identifying the good judges and identify the

¹¹Ribeiro (2007).

¹²Ribeiro (2007, p. 21).

good judges by their ability to identify the good art. Hume could break that circle on either side if doing so would not leave him sitting on top of his educative programme... My argument is that whether he plumps for the programme directly, or indirectly by plumping for judges that exemplify that programme's details, he ultimately plumps for his programme. And he has no argument—*no argument*—for his programme over against other competing programmes. So the problem is not circularity (or regress). The problem is that Hume is committed to the arbitrary, foot-stomping horn of Agrippa's Trilemma. The question that expresses this horn is this: why should we accept Hume's method... of producing consensus about matters of taste and not some other method, given that each would produce comparable degrees of consensus? Hume has no answer to that question as far as I can see. (Because there is no answer to that question as far as I can see.)¹³

The objection is certainly worrying, but I believe that it can be met. In order to meet it fully, however, it is necessary to place Hume's account within his philosophy of emotion, by examining *Of the Standard of Taste* in the context of the *Four Dissertations*. It is to this task that I turn in the next and final chapter.

¹³Ribeiro (2007, p. 25).

Chapter 14

The Science of Criticism

In the previous chapter, we stated the problem: how can there be a standard of taste, which pronounces some sentiments right and others wrong, given the sceptical principle that beauty (value, desirability) is not a quality of objects themselves? We saw the basic structure of Hume’s reply. The right sentiments, he says, are those felt by the true judges; and the true judges are those with a delicate taste, sufficient experience, requisite knowledge and intelligence, and freedom from prejudice. I argued that there was no threat of circularity here, and that Hume’s account was structurally sound. But from what we have seen so far, it remains incomplete. Some answer must be given to the crucial question of why the true judges thus defined set the standard for the rest of us: why are *their* sentiments the *right* ones? To answer this question, as I have already said, we must place Hume’s account in the broader context of his philosophy of emotion. This is the final piece of my Unity Thesis, to be defended in this chapter.

14.1. Relevant virtues

Jerrold Levinson claims to have uncovered “the real problem” with Hume’s proposed account of the standard of taste. “Why,” he asks, “are the works enjoyed and preferred by *ideal* critics¹ characterized as Hume characterizes them ones that

¹I prefer to call them “true” rather than “ideal” myself, both because that is the term that Hume himself uses, and because it is controversial whether or to what extent Hume’s true judges are ideal. See Ross (2008).

I should, all things being equal, aesthetically pursue? Why not, say, the objects enjoyed and preferred by *izeal* critics—who are introverted, zany, endomorphic, arrogant, and left-handed?”² I am doubtful that zaniness is an objectively identifiable property, and that left-handedness has any systematic bearing on one’s artistic sentiments; but setting these issues aside, the structural problem remains. Levinson intimates that Hume failed to give any answer to this question.

The way in which Levinson sets up the problem may allow for an easy response. Hume is concerned to establish that the sentiments of the true judges are *right* (thereby silencing the relativist). It is a further question whether anyone wants to acquire the character of the true judge so as to experience the right sentiments: perhaps some of us quite legitimately have better things to do; life is short. Levinson is unmoved by this line of objection (it was already raised by Jeffrey Weiland); in a subsequent paper he emphasised that he did not mean to raise a problem specifically *for Hume*, but merely a difficulty that strikes people of a certain turn of mind upon encountering Hume’s essay.³

This response is somewhat puzzling, however. For once the question is clarified in this way, it becomes apparent, not only that this isn’t a problem *for Hume*, but that it isn’t even a problem that has anything in particular to do with the sceptical principle or the Humean attempt to erect a standard of taste that is consistent with that principle. It is a perfectly general problem that might arise in the philosophy of art, *regardless* of one’s views on the nature of value. For suppose that beauty is in fact an objective quality of things in themselves; it would still be a further question why anyone should want to learn how to appreciate it. They might be quite happy enjoying objectively ugly things, or at least sufficiently content not to want to put in the requisite hours to correct their critical responses. If we are wrong about what sort of food is good for us, we have an obvious pragmatic reason for correcting the mistake. But if we are wrong about what sort of art is beautiful, it is not thus immediately clear why this should trouble us.

²Levinson (2002, p. 229).

³Weiland (2003), Levinson (2003).

If anything, indeed, it seems to me that Hume is *better* placed to answer this question than the objectivist. For if the question is why we should acquire the character of the true judge, rather than why we should enjoy the objectively good things, it is more readily apparent what one might say. Good sense, first of all, is obviously a useful quality to possess, for all sorts of reasons (by no means limited to the appreciation of art). And Hume makes a point of saying, secondly, that delicacy of taste is agreeable:

A very delicate palate, on many occasions, may be a great inconvenience both to a man himself and to his friends: But a delicate taste of wit or beauty must always be a desirable quality; because it is the source of all the finest and most innocent enjoyments, of which human nature is susceptible.
(ST 17, p. 236)

In *Of the Delicacy of Taste and Passion*, Hume argues at length for the agreeableness of this quality (and disagreeableness of the parallel delicacy of passion, a heightened sensitivity to the general ups and downs of life).

Perhaps scepticism is appropriate, finally, about the extent to which we should free ourselves from prejudice (at least of the morally blameless and cognitively harmless kind). Hume after all considers some degree of prejudice to be quite acceptable in the appreciation of art: “it is almost impossible not to feel a predilection for that which suits our particular turn and disposition. Such preferences are innocent and unavoidable, and can never reasonably be the object of dispute” (ST 30, p. 244). But in general, the character of the true judge is surely a desirable one, quite independently of that character’s emotional responses to works of art.

However that may be, there is a closely related question here that needs to be addressed, namely, why Hume supposes that the three characteristics he chooses will ensure that someone has the *right* sentiments. Why should these characteristics ensure consistency in sentiments, and why should they be preferred to other (possibly zany) characteristics? We can go some way to answering this question by noting that the three characteristics are *relevant virtues* in the appreciation of

art. That they are virtues ought to be uncontroversial. At any rate Hume takes it to be so: “that such a character is valuable and estimable will be agreed in by all mankind” (ST 25, p. 242). As to their relevance, meanwhile, Hume has a few things to say. For example, prejudice may cause us to overestimate the work of a friend, or to underestimate that of an enemy (ST 21, p. 239); good sense is necessary to judge whether characters in a drama behave in accordance with their nature and circumstances (ST 22, p. 240); and without delicacy of sentiment, various features of a work of art will simply pass us by (ST 17, p. 236).

From what we have seen so far, it is plausible to read Hume’s answer to the relativist about taste as what we might call a *buck-passing* response. The question of the value of art is shifted onto the question of the value of certain character traits: good art is the art appreciated by (relevantly) virtuous people. This may be a satisfactory answer to those subjectivists or sceptics who are predisposed to reject relativism. But hard-line relativists—who maintain that the sentiments of insensitive, partial, and unintelligent people, with comparatively little experience or knowledge of art, are just as valid as the sentiments of the contrary sort of people—will find nothing to trouble themselves in this interpretation of Hume’s argument.

It seems to me that Hume does in fact have something to say even against these hard-line relativists. I do not quite want to say that the buck-passing version of Hume’s answer is *mistaken*. It is *faulty*, but more because of what it leaves out than because of what it says. Once these omissions are made plain, we will have more Humean tools at our disposal with which to answer the present question.

14.2. A textual problem

What the buck-passing interpretation is missing, from a textual point of view, is an account of what Hume variously terms the “rules of composition”, “rules of art”, “rules of criticism”, and “rules of beauty” (ST 9, 10, 16; pp. 231, 232, 235). The textual problem for this interpretation, then, is just this: why does Hume

spend so much time talking about these things, if they play no role in his answer to the relativist?

Before filling in this blank, a slight digression may be useful, if only to bring the issue into clearer focus. Consider the following passages from Hume's essay:

It is natural for us to seek a *Standard of Taste*; a rule, by which the various sentiments of men may be reconciled; at least, a decision afforded, confirming one sentiment, and condemning another. (ST 6, p. 229)

In each creature, there is a sound and a defective state [of the organ]; and the former alone can be supposed to afford us a true standard of taste and sentiment. (ST 12, pp. 233-4)

Thus, though the principles of taste be universal, and nearly, if not entirely the same in all men; yet few are qualified to give judgment on any work of art, or establish their own sentiment as the standard of beauty. (ST 23, p. 241)

[T]he joint verdict of such [true judges], wherever they are to be found, is the true standard of taste and beauty. (ST 23, p. 241)

In these four passages, Hume is seen to describe the standard of taste or beauty as, variously, (i) a rule for reconciling differing sentiments, (ii) a decision confirming one sentiment and condemning another, (iii) a sound state of mind (at least, such a thing "affords us" a standard), (iv) a sentiment of someone suitably qualified, and (v) a joint verdict of suitably qualified people.

This apparent variety poses an interpretative puzzle. Is there only a superficial difference here, so that all of Hume's characterisations of the standard can be seen as getting at essentially the same thing, just in a slightly different way or from a slightly different angle? In which case, what is this common core? Or is there some substantial difference here? In which case, which of the varieties should we take to be Hume's official or considered position?

This puzzle, or something like it, was first drawn to our attention by Weiland, and later emphasised by James Shelley.⁴ Both Shelley and Weiland somewhat misrepresent the issue, however. Most importantly, they introduce a new contender for the prize of Hume's standard, namely the rules of art or criticism mentioned a moment ago: "Shortly after giving his definition [of the standard of taste], Hume identifies the standard with what he variously calls 'rules of composition', 'laws of criticism,' and 'rules of art'".⁵ Interesting though this suggestion is, I can find no support for it in the text; at no point in the essay, this is to say, does Hume identify the standard of taste with these rules of art. That standard may be many things, but I can discover no evidence that this is one of the candidates.

As well as adding an unwarranted candidate, both Shelley and Weiland ignore several of the others, by presenting the joint verdict of the true judges—number (v) above—as the chief or only rival to the rules of art. Weiland apparently equates the rules of art with the rule for reconciling sentiments (i), and the joint verdict of the judges (v) with the decision confirming one sentiment and condemning another (ii). Shelley explicitly asks the question whether these identifications are right.⁶ What concerns him, however, is the switch from the singular to the plural (*a rule* versus *the rules*; confirmation of *one* sentiment versus a *joint* verdict);⁷ the overall framework seems not to be in doubt. The sound state of the internal organs (iii) and the sentiments themselves (iv), meanwhile, are never even considered.

I will return to the problem of Hume's possibly different accounts of the standard when I have said more about the rules of art, after which the issue will be easier to address, since we will see that there is an account of the standard that renders all of these definitions consistent. For now, the point to note is that the textual problem for the buck-passing interpretation of Hume's answer—that

⁴Weiland (1984), Shelley (1994).

⁵Shelley (1994, p. 437).

⁶Shelley (1994, p. 437).

⁷Shelley (1994, p. 443).

it fails to say anything about the rules of art—is not that it fails to say anything about Hume’s other definitions of the standard of taste. Certainly it does focus on the joint verdict of the true judges, at the expense of the other definitions, and perhaps that could turn out to be a significant lack. But the more immediate problem, or at least the one that I wish to tackle first, is that it focuses on the standard of taste (however defined) at the expense of the rules of art.

To understand these rules of art and their place in Hume’s response to the relativist, we must situate *Of the Standard of Taste* in its proper context, as part of Hume’s philosophy of emotion. If, in doing so, we are able to make better sense of this fourth dissertation, I will take this final aspect of my Unity Thesis to be well established. This will be the task of the next two sections.

14.3. Art and the passions

The immediate purpose of Hume’s fourth dissertation is, obviously enough, to justify a standard of taste with which to silence the relativist. But Hume also has a wider purpose in this work, one that emerges more clearly when we see it in the context of his philosophy of emotion. *Of the Standard of Taste* is a manifesto for an experimental science of criticism, a science that Hume barely began, but which he considered to be both possible and desirable.

Work on the science of criticism was first promised in the advertisement to the *Treatise*:

The subjects of the understanding and passions make a compleat chain of reasoning by themselves; and I was willing to take advantage of this natural division, in order to try the taste of the public. If I have the good fortune to meet with success, I shall proceed to the examination of morals, politics, and criticism; which will compleat this Treatise of human nature.

(Ad 1739, p. xii)

As is well known, Hume did not have the good fortune to meet with success. Nevertheless, a third book on morals was published, as well as the two *Enquiries*

and a number of essays on political subjects. Regarding criticism, however, we have only scraps; though Hume continued to be conscious of the need for more, as we learn from this comment in the essay *Of Civil Liberty*:

Men, in this country, have been so much occupied in the great disputes of *Religion, Politics, and Philosophy*, that they had no relish for the seemingly minute observations of grammar and criticism. And though this turn of thinking must have considerably improved our sense and our talent of reasoning; it must be confessed, that, even in those sciences above-mentioned, we have not any standard-book, which we can transmit to posterity: And the utmost we have to boast of, are a few essays towards a more just philosophy; which, indeed, promise well, but have not, as yet, reached any degree of perfection. (CL 8, p. 92)

This is from an essay first published in 1741, sixteen years before *Of Tragedy* and *Of the Standard of Taste* appeared. Which essays Hume had in mind as moving “towards a more just philosophy” I do not know, but presumably Addison and Shaftesbury (to whom he had referred four paragraphs earlier) cannot have been far from his mind.⁸

In understanding Hume’s conception of the science of criticism, it is important to appreciate how closely related it is to the science of the passions. One of the main purposes of art, for Hume, is to excite our passions; thus those familiar with the laws governing the interplay of the passions will be well placed to create good art (or to theorise about the merits of existing art). Consider Hume’s praise of his friend John Home’s tragedy *Douglas*, given in the dedication of the *Four Dissertations*:

⁸Hugh Blair, in his *Lectures on Rhetoric and Belles Lettres* (first published in 1783, but given to students at Edinburgh at least fifteen years earlier), echoes Hume’s lament in *Of Civil Liberty* that “[w]e are far from having yet attained to any system concerning this subject” (volume 1, p. 52), but mentions Addison’s *Essay on the Pleasures of the Imagination* (*Spectator*, volume 6) as the first attempt at systematisation (*ibid.*). The lectures also refer to Hume’s *Of the Standard of Taste* (volume 1, p. 20), and Blair’s response to relativism is very much in the same spirit as Hume’s.

[T]he unfeigned tears which flowed from every eye, in the numerous representations which were made of it on this theatre; the unparalleled command, which you appeared to have over every affection of the human breast: These are incontestible proofs, that you possess the true theatric genius of Shakespear and Otway, refined from the unhappy barbarism of the one, and licentiousness of the other. (Ded 5)

Thus, though Hume never pursued the science of criticism fully or in detail, his work on the passions may be taken to provide much of its foundations. The dissertation *Of Tragedy* is of course clearly in the overlap between these two sciences.

The view that command of the passions was high among the artist's skills was by no means uncommon in the eighteenth century. Alexander Gerard, for example, in his Edinburgh Society prize-winning *Essay on Taste*, states that “[a] very great part of the merit of most works of genius arises from their fitness to agitate the heart with a variety of passions” (1759, p. 87). Furthermore, in a passage that strikes an obvious note of agreement with Hume's essay, Gerard writes:

Genuine criticism... investigates those qualities in its objects which, from the invariable principles of human nature, must always please or displease; describes and distinguishes the sentiments which they in fact produce; and impartially regulates its most general conclusions according to real phaenomena. (p. 181)

Or consider also this passage from chapter 2 of Henry Home's (Lord Kames') *Elements of Criticism* (the second longest chapter in the work by a long way, and devoted exclusively to the passions):

The design... of this chapter is to delineate that connection [between passions and the fine arts], with the view chiefly to ascertain what power the fine arts have to raise emotions and passions. To those who would excel in the fine arts, that branch of knowledge is indispensable; for without it the critic, as well as the undertaker, ignorant of any rule, have nothing left but

to abandon themselves to chance. Destitute of that branch of knowledge, in vain will either pretend to foretell what effect his work will have upon the heart. (1762, p. 32)

It should thus come as no surprise that Hume first published *Of the Standard of Taste* alongside his *Dissertation on the Passions*. This juxtaposition would not have seemed remotely unusual to his readers at the time.

Another thing to note about Hume's conception of the science of criticism is that it primarily concerns, not art in general, but only those arts involving words, such as poetry and theatre. (Notice how, in the quotation given above from *Of Civil Liberty*, criticism and grammar go so closely together.) It is not, I am sure, that Hume explicitly discounted music and the visual arts from its remit—certainly he never did so in print, and I doubt that he did in his own mind either—but more that he simply showed no interest in them, and tended to ignore them or unconsciously set them aside. This focus on poetry adds weight to the claim that understanding the passions was crucial to Hume's science of criticism, especially when we consider that, for Hume, “[t]he object... of poetry [is] to please by means of the passions and the imagination” (ST 22, p. 240).

Thus the hunt was on for the “rules of composition”, “rules of art”, “rules of criticism”, or “rules of beauty” (ST 9, 10, 16; pp. 231, 232, 235): general causal laws governing what features give rise to which passions, and in particular which properties give rise to the positive sentimental responses, thus rendering the works that possess these features good or beautiful.⁹ Hume does not venture any of these rules in *Of the Standard of Taste*, but he does not doubt their existence. Our question now is, how do these rules of art relate to the standard of taste, and to Hume's attempt to refute the relativist?

⁹Sometimes the prompting of negative sentiments, for Hume, can be valuable in a work (as in the case of tragedy), but only because they increase the predominant pleasurable sentiments. This hedonistic view of the value of art is presumably too simplistic to be very convincing, but it is Hume's view (as it was many other people's at the time). I suspect that it can, however, be divorced from his general framework, while still leaving that framework intact. Perhaps for example we could say that the sentiments invoked in us by good works of art are ones that we somehow value or want to experience; even when they are painful, or not even best characterised in terms of pleasure and pain at all.

14.4. The rules of art

There is no *a priori* guarantee that a science of criticism, of the sort that Hume envisaged, will be possible. A science of this sort is concerned with uncovering general principles, but what if the relevant phenomena are simply too diverse to admit of any substantial generalisations? What if individual tastes are simply too different to be accounted for by any general rules of art or criticism? This is the thesis of (let us say) *particularism*, and Hume's science of criticism is premised on its falsehood.

Particularism about art these days typically means a particularism about *properties*, the view that no general rules exist conferring value on properties of works of art: the most we can say is that a property makes *this particular* work of art valuable, not that it renders valuable, to a degree, *any* work possessing that feature. This is not the view that I have in mind here (though I take it that Hume was opposed to this view as well).¹⁰ The view that I have in mind here is rather a particularism about *people*; particularists of this sort may allow that a certain property is always valuable in any work of art *relative to a particular person or group of people*. What they deny is that there are any general rules of this nature governing *all* people (or all groups of people). Some people like this sort of thing, and other people like that sort of thing, and that is all we can say at the general level.

It will be apparent, I trust, how close the relationship is between particularism (in my sense) and relativism. But the two are not the same. Crucially, particularism is a purely descriptive doctrine. It claims that different people are, in matters of taste, too different to allow for any general science of taste or criticism that makes substantial claims about all of them. Perhaps we can do a general science of *these* people's taste, or *those* people's taste; but a general science of *human* taste will not be possible. The relativist, by contrast, makes an

¹⁰See the exchange between Dickie (2003), a generalist, and Shelley (2002, 2004), a particularist (in the modern sense). Shelley argues, convincingly to my mind, that Hume was a generalist in the sense that he himself rejects.

evaluative claim: that different people have different tastes, and that we cannot legitimately criticise others for having a taste that is different from our own, or pronounce any one taste superior to any other.

Hume is firm in his conviction of the possibility of uncovering general rules of art. He acknowledges, of course, the great variety of taste in the world, but he urges that these differences are not simply brute, inexplicable facts, as the particularist maintains. Rather, there are general principles of human nature governing *everyone's* sentimental responses, with the divergent sentiments that we actually experience being the result of causal “noise” from other principles which operate in different people to different degrees:

But though all the general rules of art are founded only on experience and on the observation of the common sentiments of human nature, we must not imagine, that, on every occasion, the feelings of men will be conformable to these rules. Those finer emotions of the mind are of a very tender and delicate nature, and require the concurrence of many favourable circumstances to make them play with facility and exactness, according to their general and established principles. The least exterior hindrance to such small springs, or the least internal disorder, disturbs their motion, and confounds the operation of the whole machine. (ST 10, p. 232)

As I said in §14.1 above, Hume holds that the defining characteristics of his true judges are *virtues*, i.e. qualities that we all agree are valuable. They are moreover *relevant* virtues in the appreciation of art. I have given some motivation for this claim already, but we are now in a position to see Hume's main justification for it, or at least the general backdrop of his particular justifications. The central idea is that these virtues enable us to remove the causal “noise” that prevents us from experiencing the “catholic and universal beauty” (ST 10, p. 233), i.e. from experiencing the sentiments that arise in accordance with the general principles of taste that (Hume believes) lie in each and every one of us.

We must of course ask the very important question of why Hume thinks he is entitled to assert that there are these general principles in all of us, or

on what basis he concludes that a general experimental science of criticism is indeed possible. This question is all the more urgent given his failure to pursue this science in very much detail (and indeed his acknowledgement, in *Of Civil Liberty*, that he and his predecessors had yet to get very far in this task). And the answer, I fear, is somewhat disappointing. Hume's discussion of the virtuous features of his true judges, and of the ways in which the lack of these virtues can interfere with the sentiments that we feel, certainly provides the *basis* for a reply to the particularist, by offering a general method for accounting for the diversity, consistently with the view that there are general rules. But much more would need to be done to establish, not only that it is *possible* there are such general rules, but that there in fact *are*. The only clear piece of evidence that I can find Hume offering in this regard is the example of Homer, for whom "[a]ll the changes of climate, government, religion, and language, have not been able to obscure his glory" (ST 11, p. 233). This, I allow, is evidence in the right direction; but if this is all that Hume can muster, it seems a fairly unconvincing case.¹¹

But however weak Hume's argument may be at this point, we must return to the issue of his reply to the relativist. Suppose, then, that there *are* general rules of art or criticism, governing the sentiments of all human beings. If this is indeed true, it does seem quite legitimate to erect a standard of taste on the back of these rules. There can be no suspicion of arbitrariness in choosing to fix on the sentiments of the true judges rather than those of any other group, since there will be an important sense in which these are the sentiments of *all* of us (although in many cases various external factors prevent us from feeling them when and where we otherwise would). It might be, I hasten to add, that the particular characteristics Hume picks out as defining the true judges are the wrong ones, or at least form an incomplete set; but any complete science of criticism will include, not only the general rules, but also all the various factors that prevent them from operating. Assuming that we have such a complete science, we will then simply be able to read off the relevant virtues.

¹¹The poor quality of Hume's argument here should not be particularly surprising. His belief in determinism quite generally was more optimistic than well-reasoned; see Millican (2010, pp. 635-7).

14.5. Many standards

One loose end remains to be tied up: the problem of Hume's several characterisations of the standard of taste (introduced in §14.2 above). With a little care, we can see that Hume's various statements are not really in conflict, and that we can find in this essay a clear definition of the standard of taste. Let us begin with the first potentially problematic statement, the one that Shelley homes in on as the main source of the trouble:

It is natural for us to seek a *Standard of Taste*; a rule, by which the various sentiments of men may be reconciled; at least, a decision afforded, confirming one sentiment, and condemning another. (ST 6, p. 229)

The simple point to make here is that the main *purpose* of a standard of taste is to settle the question of which sentiments are right and which are wrong. If we had some such standard, therefore, it would by definition enable us to reconcile conflicting sentiments; if, that is, all parties to the dispute were aware of the standard, were unbiased, and were determined to find the right sentiment. Failing that, it would at least enable those familiar with the standard to decide who is right and who is wrong in any such dispute.

This is how I read Hume's supposed "two definitions" of the standard of taste; which is to say, I do not read them as *definitions* at all, but simply as a (pretty trivial) statement of why we might *want* such a standard (whatever it is), and of what *use* we might put it to. On this innocuous reading of the passage, there is no danger that Hume might be offering two potentially conflicting definitions of the standard. Nor is there any danger of this passage conflicting with any subsequent definitions that Hume might give.

Shelley picks out this paragraph as "perhaps... the most important and least understood" in Hume's essay. He laments that, "in all that has been written on Hume's aesthetics, almost nothing has been said" about it, mentioning Weiland as the only exception to this rule.¹² If I am right, however, the lack of attention

¹²Shelley (1994, p. 437).

that this paragraph has received is entirely understandable, indeed *appropriate*. Obviously this is what we want a standard of taste for: to reconcile conflicting sentiments, or at least (if the relevant party won't budge) to decide who is right and who is wrong. Hume is here saying nothing substantial or controversial, and the paragraph in question is perhaps one of the least important and (Shelley and Weiland excepted) the best understood.

That still leaves three potentially conflicting statements in Hume's essay about the nature or essence of the standard of taste, which for ease I will repeat here:

In each creature, there is a sound and a defective state [of the organ]; and the former alone can be supposed to afford us a true standard of taste and sentiment. (ST 12, pp. 233-4)

Thus, though the principles of taste be universal, and nearly, if not entirely the same in all men; yet few are qualified to give judgment on any work of art, or establish their own sentiment as the standard of beauty. (ST 23, p. 241)

[T]he joint verdict of such [true judges], wherever they are to be found, is the true standard of taste and beauty. (ST 23, p. 241)

There is a consistent reading of all of these passages available, however, a reading that also closely parallels Hume's account of the standard of judgment. The standard of taste, on this reading, is a sentiment (or the sentiments) felt by someone suitably qualified, i.e. someone with the defining characteristics of the true judge, i.e. someone whose mental "organs" are in a sound state. Judgments, recall, "have a reference to something beyond themselves, to wit, real matter of fact" (ST 7, p. 230), and real matter of fact affords us the standard of judgment. Sentiments, by contrast, do not have a reference to anything beyond themselves (in that same sense), and so the only things that can afford us a standard of taste are other sentiments. Which sentiments, then? Hume answers: the *natural* ones, the ones felt in the *sound* state of the organ, or equivalently those felt by the *true* judges.

This still leaves one thing that Hume says unaccounted for, namely the sudden and unexplained introduction of a *joint* verdict of several (all?) true judges. Perhaps Hume intended, with this move, to deal with the limited degree of relativism that he later mentions, towards the end of the essay: true judges will all still have their own favourite authors and genres, and be subject to a handful of blameless differences of temperament or even prejudices, resulting in somewhat different sentiments (ST 28-31, pp. 243-5). Perhaps, then, Hume thought that a joint verdict of enough of these judges would help to settle these disputes. Difficult questions arise here, however, about how exactly we are to “join” the conflicting sentiments. Do we take some sort of average? Do we combine all the positive ones, and delete the conflicting negative ones? Or what?

It seems to me quite unlikely that this is what Hume had in mind. If it was, we would expect some answers to these questions, which he nowhere attempts to give. Furthermore, such answers would conflict with Hume’s acknowledgement that the appropriate response in these cases of disagreement is in fact relativism: “we seek in vain for a standard, by which we can reconcile the contrary sentiments” (ST 28, p. 224). Much more likely, then, is that, when Hume spoke of a joint verdict of several judges, his point was merely to emphasise the *agreement* between all such people; for without this agreement, his proposed standard of taste would not work: “If, in the sound state of the organ, there be an entire or a considerable uniformity of sentiment among men, we may thence derive an idea of the perfect beauty” (ST 12, p. 234). If not, however, there is nothing more that we can say against the relativist. The joint verdict, then, is precisely the degree to which the true judges agree. Whatever differences there are in the sentiments of the true judges—and Hume thinks there certainly will be some—must remain blameless and irreconcilable.

Conclusion

The philosophy of emotion naturally commands less mainstream attention than epistemology, metaphysics, or moral philosophy. This was as true in the eighteenth century as it is today, and it is no surprise that Book 2 was the last of the *Treatise* volumes that Hume revisited in his mature work. This topic nevertheless continued to be very important to him, with clear connections to his philosophy of religion and philosophy of art, not to mention his moral philosophy. Though the *Four Dissertations* will probably always be overshadowed by the two *Enquiries*, I hope to have shown that they hang together very nicely as a set, and contain much of interest to the philosopher and the student of Hume.

The current scholarly emphasis on the *Treatise*, at the expense of Hume's later works, has become so familiar that it now scarcely raises an eyebrow. In truth, however, it ought to be very puzzling. Not only does common sense suggest that Hume's thought would have got better over time, but we also know that Hume himself deeply regretted publishing the *Treatise*, and in later life had a clear preference for the *Enquiries* and *Dissertations*.

In October 1775, most notably, he wrote to William Strahan asking that an advertisement be prefixed to any remaining copies of volume 2 of his *Essays and Treatises* (in the event it was prefixed to volume 2 of the posthumous 1777 edition). I already quoted some of this advertisement at the end of Part 1, but here it is in full:

Most of the principles, and reasonings, contained in this volume, were published in a work in three volumes, called *A Treatise of Human Nature: A work which the Author had projected before he left College, and which*

he wrote and published not long after. But not finding it successful, he was sensible of his error in going to the press too early, and he cast the whole anew in the following pieces, where some negligences in his former reasoning and more in the expression, are, he hopes, corrected. Yet several writers, who have honoured the Author's Philosophy with answers, have taken care to direct all their batteries against that juvenile work, which the Author never acknowledged, and have affected to triumph in any advantages, which, they imagined, they had obtained over it: A practice very contrary to all rules of candour and fair-dealing, and a strong instance of those polemical artifices, which a bigoted zeal thinks itself authorised to employ. Henceforth, the Author desires, that the following Pieces [the two *Enquiries*, the *Dissertation on the Passions*, and the *Natural History of Religion*] may alone be regarded as containing his philosophical sentiments and principles. (Ad 1777, p. 2)

Hume had in mind here the criticisms of him by Thomas Reid (1764) and James Beattie (1770); in the letter he writes that the advertisement "is a compleat Answer to Dr Reid and to that bigoted silly Fellow, Beattie" (HL 2, p. 301). But his dissatisfaction with the *Treatise* goes back much further. In the published *Letter from a Gentleman to his Friend in Edinburgh* (1745), for example, written in response to a pamphlet accusing the *Treatise* of endorsing "Universal Scepticism" (L 14, p. 425), "Principles leading to downright Atheism" (L 15, p. 425), and "With sapping the Foundations of Morality" (L 19, p. 425), Hume admitted, after denying the charges, "that the Author had better delayed the publishing of that Book; not on account of any dangerous Principles contained in it, but because on more mature Consideration he might have rendered it much less imperfect by further Corrections and Revisals" (L 41, p. 431).

The same attitude emerges from his private correspondence. In March 1740, even before Book 3 had been published, Hume was already writing thus to Hutcheson: "I wait with some impatience for a second Edition principally on Account of Alterations I intend to make in my Performance" (HL 1, p. 38). Of course he never did produce a second edition, but wrote instead the *Enquiries* and *Disser-*

tations. Indeed, in a letter to John Stewart some time later, he wrote that the “positive Air” of the *Treatise* “so much displeases me, that I have not Patience to review it” (HL 1, p. 187). Once the first *Enquiry* was in print, moreover, he expressed a clear preference for this work in a letter to Gilbert Elliot:

I believe the philosophical Essays contain every thing of consequence relating to the Understanding, which you would meet with in the *Treatise*; & I give you my Advice against reading the latter. ... The Philosophical Principles are the same in both: But I was carry'd away by the Heat of Youth & Invention to publish too precipitately. So vast an Undertaking, plan'd before I was one and twenty, & compos'd before twenty five, must necessarily be very defective. I have repented my Haste a hundred, & a hundred times. (HL 1, p. 158)

The evidence is unambiguous, and there is no doubt that Hume came to dislike his own *Treatise* almost immediately after publishing it.

While this much is uncontroversial, however, the more difficult question is *why* Hume was so unhappy with his first performance. There are two possible answers. The less interesting one holds that the basis of Hume's attitude was solely or at least predominantly stylistic. There is certainly some *prima facie* evidence in support of this view. In the *Letter from a Gentleman*, Hume admits that he had better delayed the publishing of the *Treatise*, but not because of its “dangerous Principles”. In the 1777 advertisement, he tells us that “[m]ost of the principles, and reasonings” from his later work were already to be found in the *Treatise*, and while he acknowledges “some negligences in his former reasoning”, it appears that he is more unhappy with the “expression”. Again, the letter to Elliot states, at least of Book 1 and the first *Enquiry*, that “[t]he Philosophical Principles are the same in both”. And in his autobiographical *My Own Life*, Hume said he felt that the failure of the *Treatise* “proceeded more from the manner than the matter” (MOL 8, p. xxxv) of the work.

This answer, I take it, is the one that most scholars today favour; a view that goes some way to explaining the general preference for the *Treatise* over

Hume's later works. But the evidence of Hume's explicit remarks on the matter is also consistent with a more intriguing possibility: that his dissatisfaction with the *Treatise* was not merely stylistic, but also grounded in some of its substantial *content*. *Most* of the principles are not *all* of the principles, after all; negligences in reasoning, however reticently acknowledged, are certainly matters of substance; and Hume can hardly have thought that he answered Reid's and Beattie's philosophical criticisms merely by improving his writing style.¹³ Needless to say, it is this latter answer that I favour: as I have argued throughout this thesis, there are several *substantial* differences between *Treatise* Book 2 and the *Four Dissertations*. I have also noted along the way some differences between Book 3 and the moral *Enquiry*.

The official statement of my Difference Thesis in the introduction was given in neutral terms, but I needn't be coy. It has no doubt been apparent from very early on that I believe the philosophy of emotion that we find in the *Four Dissertations* to be not only different from that of *Treatise* Book 2, but also significantly *better*. It used to be thought that the *Treatise* was Hume's principal philosophical work, and that, prompted by its lack of immediate success, Hume dumbed down his ideas in the later *Enquiries* so as to reach a wider audience (while throwing in some provocative irreligious arguments for good measure). No one today, I take it, holds this uncharitable interpretation of Hume's literary and intellectual motives.¹⁴ Old habits die hard, however, and the *Treatise* is still considered by most scholars to be the "heavyweight" of Hume's philosophical texts. Success for my argument would entail nothing less than the relegation of Book 2 from its current exalted position as the pinnacle of Hume's thought on the passions, to the more humble but philosophically and historically more plausible status of a first draft.

¹³This last point is made by John Nelson (1972), who may also have been the first person (I haven't done a thorough search) to ask the present question clearly in these terms. For a detailed attempt to explain how Hume's first *Enquiry* "answers" Reid's criticisms, see Millican (2006).

¹⁴The classic refutations of it are Kemp Smith (1941/2005, pp. 526-30) and Mossner (1950).

Bibliography of Primary Literature

Page references to Hume's texts and letters, unless otherwise stated, are from the following editions (page references to the *Treatise* are from Selby-Bigge's edition; page references to the *Letter from a Gentleman* are from the Norton's):

Greig, J. Y. T. (1932). *The Letters of David Hume*. Oxford: Oxford University Press.

Selby-Bigge, L. A. (1975). *Enquiries concerning Human Understanding and concerning the Principles of Morals*. 2nd edition, revised by P. H. Nidditch. Oxford: Oxford University Press.

Selby-Bigge, L. A. (1978). *A Treatise of Human Nature*. 3rd edition, revised by P. H. Nidditch. Oxford: Oxford University Press.

Miller, Eugene F. (1987). *Essays, Moral, Political, and Literary*. Indianapolis: Liberty Fund.

Beauchamp, Tom L. (2007). *A Dissertation on the Passions and the Natural History of Religion*. Oxford: Clarendon Press.

Norton, David Fate and Mary J. (2007). *A Treatise of Human Nature and A Letter from a Gentleman*. Oxford: Clarendon Press.

References to other texts are as follows. Quotations and page numbers, unless otherwise stated, are from the most recent edition shown here (a modern edition, where one was easily available).

Beattie, James (1770). *An Essay on the Nature and Immutability of Truth; in Opposition to Sophistry and Scepticism*. Edinburgh.

Berkeley, George (1732). *Alciphron: Or, the Minute Philosopher*. Dublin.

— (1993). *Alciphron in Focus*. Ed. by David Berman. London & New York: Routledge.

- Blair, Hugh (1783). *Lectures on Rhetoric and Belles Lettres*, in 3 volumes. Dublin.
- Du Bos, Jean-Baptiste (1719). *Réflexions Critiques sur la Poesie et sur la Peinture*, in 2 volumes. Paris.
- (1748). *Critical Reflections on Poetry, Painting, and Music*, in 3 volumes (all page references are to volume 1). Translated by Thomas Nugent. London.
- Burnet, Gilbert and Francis Hutcheson (1735). *Letters concerning the True Foundation of Virtue or Moral Goodness*. London.
- Butler, Joseph (1726). *Fifteen Sermons preached at the Rolls Chapel*. London.
- (1729). *Fifteen Sermons preached at the Rolls Chapel*, 2nd edition. London.
- (1736). *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature*. London.
- (2006). *The Works of Bishop Butler*. Ed. by David E. White. Rochester: University of Rochester Press. (Page references are to this edition, but quotations are from the originals, since White's modernisation of the text is not quite to my taste.)
- Clarke, John (1726). *An Examination of the Notion of Moral Good and Evil*. London.
- Clarke, Samuel (1706). *A Discourse Concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*. London.
- (1738). *The Works of Samuel Clarke*, volume 2. London. (Reprinted in 2002, New York: Garland Publishing Co.)
- Clayton, Robert (1753). *Some Thoughts on Self-Love, Innate-Ideas, Free-Will, Taste, Sentiment, Liberty and Necessity, &c.* Dublin.
- Cumberland, Richard (1672). *De Legibus Naturae*. London.
- Fontenelle, Bernard (1742). "Réflexions sur la Poétique". In *Oeuvres de Fontenelle*. Paris.
- Gerard, Alexander (1759). *An Essay on Taste*. London.
- Hobbes, Thomas (1650a). *Humane Nature: Or, the Fundamental Elements of Policy*. London.
- (1650b). *De Corpore Politico: Or, the Elements of Law*. London.
- (1840). *The English Works of Thomas Hobbes*, volume 4. Ed. by William Molesworth. London.
- Hobbes, Thomas (1651). *Leviathan*. London.

- (1996). *Leviathan*. Ed. by Richard Tuck. Cambridge: Cambridge University Press.
- Hutcheson, Francis (1725). *An Inquiry into the Original of our Ideas of Beauty and Virtue*. London.
- (1738). *An Inquiry into the Original of our Ideas of Beauty and Virtue*, 4th edition. London.
- (2004). *An Inquiry into the Original of our Ideas of Beauty and Virtue*. Ed. by Wolfgang Leidhold. Indianapolis: Liberty Fund.
- (1728). *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*. London.
- (2002). *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*. Ed. by Aaron Garrett. Indianapolis: Liberty Fund.
- Johnson, Samuel (1755). *A Dictionary of the English Language*, in 2 volumes. London.
- Kames, Lord [Henry Home] (1762). *Elements of Criticism*, in three volumes. Edinburgh.
- (2005). *Elements of Criticism*, in two volumes. Ed. by Peter Jones. Indianapolis: Liberty Fund.
- Locke, John (c. 1686-8). “Of Ethic in General”. MS, reprinted in the below.
- (1997). *Locke: Political Essays*. Ed. by Mark Goldie. Cambridge: Cambridge University Press.
- (1690). *An Essay concerning Humane Understanding*. London.
- (1694). *An Essay concerning Humane Understanding*, 2nd edition. London.
- (1975). *An Essay concerning Human Understanding*. Ed. by P. H. Nidditch. Oxford: Clarendon Press.
- Malebranche, Nicolas (1675). *De la Recherche de la Vérité*. Paris.
- (1997). *The Search After Truth*. Translated by Thomas M. Lennon and Paul J. Olscamp. Cambridge: Cambridge University Press.
- Mandeville, Bernard (1705). *The Grumbling Hive: Or, Knaves Turned Honest*. London.
- (1714). *The Fable of the Bees: Or, Private Vices, Publick Benefits*. London.
- (1723). *The Fable of the Bees: Or, Private Vices, Publick Benefits*, 2nd edition.

London.

— (1988). *The Fable of the Bees*, volume 1. Ed. by F. B. Kaye. Indianapolis: Liberty Fund.

— (1729). *The Fable of the Bees, Part II*. London.

— (1988). *The Fable of the Bees*, volume 2. Ed. by F. B. Kaye. Indianapolis: Liberty Fund.

Reid, Thomas (1764). *An Inquiry into the Human Mind, on the Principles of Common Sense*. Edinburgh.

Shaftesbury, Earl of [Anthony Ashley Cooper] (1711). *Characteristicks of Men, Manners, Opinions, Times*.

— (2000). *Characteristics of Men, Manners, Opinions, Times*. Ed. by Lawrence E. Klein. Cambridge: Cambridge University Press.

— (1716). *Several Letters Written by a Noble Lord to a Young Man at the University*. London.

Wollaston, William (1722). *The Religion of Nature Delineated*.

Bibliography of Secondary Literature

- Aaron, Richard I. (1971). *John Locke*. Oxford: Clarendon Press.
- Alanen, Lilli (2006). "The Powers and Mechanisms of the Passions". In: *The Blackwell Guide to Hume's Treatise*. Ed. by Saul Traiger. Oxford: Blackwell, pp. 179–98.
- Árdal, Páll S. (1966). *Passion and Value in Hume's Treatise*. Edinburgh: Edinburgh University Press.
- Baier, Annette C. (1991). *A Progress of Sentiments: Reflections on Hume's Treatise*. Cambridge, MA: Harvard University Press.
- Baring, Thomas Charles (1884). *The Scheme of Epicurus: A rendering into English verse of the unfinished poem of Lucretius, entitled 'De Rerun Natura'*. London: Kegan Paul, Trench, & Co.
- Beauchamp, Tom L., ed. (2000). *An Enquiry concerning Human Understanding, by David Hume*. Oxford: Oxford University Press.
- ed. (2007). *A Dissertation on the Passions and the Natural History of Religion, by David Hume*. Oxford: Oxford University Press.
- Beebee, Helen (2006). *Hume on Causation*. London: Routledge.
- Bell, Martin (1999). "Hume on Superstition". In: *Religion and Hume's Legacy*. Ed. by D. Z. Phillips and Timothy Tessin. Basingstoke: Macmillan, pp. 159–70.
- Blackburn, Simon (1998). *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.
- (2010). "The Majesty of Reason". *Philosophy* 85.1, pp. 5–27.
- Bricke, John (1984). "Locke, Hume and the Nature of Volitions". *Hume Studies* 10th Anniversary Issue, pp. 15–51.
- Broackes, Justin (2002). "Hume, Belief and Personal Identity". In: *Reading Hume on Human Understanding*. Ed. by Peter Millican. Oxford: Oxford University Press, pp. 187–210.

- Broad, C. D. (1965). *Five Types of Ethical Theory*. Totowa, New Jersey: Littlefield, Adams.
- Brooke, Christopher (2012). *Philosophic Pride: Stoicism and Political Thought from Lipsius to Rousseau*. Princeton: Princeton University Press.
- Brown, S. G. (1938). "Observations on Hume's Theory of Taste". *English Studies* 20.5, pp. 193–8.
- Buckle, Stephen (2008). "Hume in the Enlightenment Tradition". In: *A Companion to Hume*. Ed. by Elizabeth S. Radcliffe. Oxford: Blackwell, pp. 21–37.
- Budd, Malcolm (1991). "Hume's Tragic Emotions". *Hume Studies* 17.2, pp. 93–106.
- Carroll, Noel (1984). "Hume's Standard of Taste". *The Journal of Aesthetics and Art Criticism* 43.2, pp. 181–94.
- Chapman Sharp, Frank (1921). "Hume's Ethical Theory and Its Critics". *Mind, New Series* 30.117, pp. 40–56.
- Cohon, Rachel (1994). "On an Unorthodox Account of Hume's Moral Psychology". *Hume Studies* 20.2, pp. 179–94.
- (2008a). "Hume's Indirect Passions". In: *A Companion to Hume*. Ed. by Elizabeth S. Radcliffe. Oxford: Blackwell, pp. 159–84.
- (2008b). *Hume's Morality: Feeling and Fabrication*. Oxford: Oxford University Press.
- (2010). "Hume's Moral Sentiments as Motives". *Hume Studies* 36.2, pp. 193–213.
- Coleman, Dorothy, ed. (2007). *Dialogues concerning Natural Religion, and other writings, by David Hume*. Cambridge: Cambridge University Press.
- Colman, John (1983). *John Locke's Moral Philosophy*. Edinburgh: Edinburgh University Press.
- Connolly, John (1987). "David Hume and the Concept of Volition: The Will as Impression". *Hume Studies* 13.2, pp. 276–305.
- Cunningham, Andrew S. (2005). "The Strength of Hume's 'Weak' Sympathy". *Hume Studies* 30.2, pp. 237–56.
- Dancy, Jonathan (2000). *Practical Reality*. Oxford: Oxford University Press.
- Darwall, Stephen (1993). "Motive and Obligation in Hume's Ethics". *Noûs* 27.4, pp. 415–48.

- (1995). *The British Moralists and the Internal 'Ought': 1640-1740*. Cambridge: Cambridge University Press.
- Debes, Remy (2007a). "Has Anything Changed? Hume's Theory of Association and Sympathy after the *Treatise*". *The British Journal for the History of Philosophy* 15.2, pp. 313–38.
- (2007b). "Humanity, Sympathy and the Puzzle of Hume's Second *Enquiry*". *The British Journal for the History of Philosophy* 15.1, pp. 27–57.
- Deigh, John (1994). "Cognitivism in the Theory of Emotions". *Ethics* 104.4, pp. 824–54.
- Dickie, George (2003). "James Shelley on Critical Principles". *The British Journal of Aesthetics* 43.1, pp. 57–64.
- Dietl, Paul (1968). "Hume on the Passions". *Philosophy and Phenomenological Research* 28.4, pp. 554–66.
- Falkenstein, Lorne (2003). "Hume's Project in 'The Natural History of Religion'". *Religious Studies* 39.1, pp. 1–21.
- Feagin, Susan L. (1983). "The Pleasures of Tragedy". *American Philosophical Quarterly* 20.1, pp. 95–104.
- Ferreira, M. Jamie (1995). "Hume's *Natural History*: Religion and 'Explanation'". *The Journal of the History of Philosophy* 33.4, pp. 593–611.
- (1999). "Hume's 'Mitigated Scepticism': Some Implications for Religious Belief". In: *Religion and Hume's Legacy*. Ed. by D. Z. Phillips and Timothy Tessin. Basingstoke: Macmillan, pp. 47–67.
- Fieser, James (1992). "Hume's Classification of the Passions and Its Precursors". *Hume Studies* 18.1, pp. 1–17.
- Frasca-Spada, Marina and P. J. E. Kail, eds. (2005). *Impressions of Hume*. Oxford: Clarendon Press.
- Galgut, Eliza (2001). "The Poetry and the Pity: Hume's Account of Tragic Pleasure". *The British Journal of Aesthetics* 41.4, pp. 411–24.
- Garrett, Don (1997). *Cognition and Commitment in Hume's Philosophy*. New York & Oxford: Oxford University Press.
- (2005). "Introduction". In: *The Philosophy of David Hume, by Norman Kemp Smith*. Basingstoke: Palgrave Macmillan, pp. xxv–xl.

- Garrett, Don (2006). "Hume's Naturalistic Theory of Representation". *Synthese* 152.3, pp. 301–19.
- (2007). "The First Motive to Justice: Hume's Circle Argument Squared". *Hume Studies* 33.2, pp. 257–88.
- Gaskin, J. C. A. (1988). *Hume's Philosophy of Religion*. Basingstoke: Macmillan.
- (2009). "Hume on Religion". In: *The Cambridge Companion to Hume*. Ed. by David Fate Norton and Jacqueline Taylor. Cambridge: Cambridge University Press, pp. 480–513.
- Gert, Bernard (1967). "Hobbes and Psychological Egoism". *The Journal of the History of Ideas* 28.4, pp. 503–20.
- (2006). "Hobbes's Psychology". In: *The Cambridge Companion to Hobbes*. Ed. by Tom Sorell. Cambridge: Cambridge University Press, pp. 157–74.
- Gill, Michael B. (2006). *The British Moralists on Human Nature and the Birth of Secular Ethics*. Cambridge: Cambridge University Press.
- Glathe, A. B. (1950). *Hume's Theory of the Passions and of Morals*. Berkeley: University of California Press.
- Gorham, Geoffrey (1994). "Mind-Body Dualism and the Harvey-Descartes Controversy". *The Journal of the History of Ideas* 55.2, pp. 211–34.
- Green, T. H. (1882a). "General Introduction". In: *A Treatise of Human Nature and Dialogues concerning Natural Religion, by David Hume*. Ed. by T. H. Green and T. H. Grose. Vol. 1. London: Longmans, Green, and Co., pp. 1–299.
- (1882b). "Introduction to the Moral Part of the *Treatise*". In: *A Treatise of Human Nature and Dialogues concerning Natural Religion, by David Hume*. Ed. by T. H. Green and T. H. Grose. Vol. 2. London: Longmans, Green, and Co., pp. 1–71.
- Guyer, Paul (2002). "Beauty and Utility in Eighteenth-Century Aesthetics". *Eighteenth-Century Studies* 35.3, pp. 439–53.
- Harris, James (2009). "Of Hobbes and Hume: A Review of Paul Russell, *The Riddle of Hume's Treatise: Skepticism, Naturalism, and Irreligion*". *Philosophical Books* 50.1, pp. 38–46.
- Hearn, Thomas K. (1973). "Árdal on the Moral Sentiments in Hume's 'Treatise'". *Philosophy* 48.185, pp. 288–92.

- Hedley, Douglas and Sarah Hutton, eds. (2008). *Platonism at the Origins of Modernity*. Dordrecht: Springer.
- Hill, Eric (1982). "Hume and the Delightful Tragedy Problem". *Philosophy* 57.221, pp. 319–26.
- Immerwahr, John (1994). "Hume's *Dissertation on the Passions*". *The Journal of the History of Philosophy* 32.2, pp. 225–40.
- Inoue, Haruko (2003). "The Origin of the Indirect Passions in the *Treatise*: An Analogy Between Books 1 and 2". *Hume Studies* 29.2, pp. 205–22.
- James, Susan (2005). "Sympathy and Comparison: Two Principles of Human Nature". In: *Impressions of Hume*. Ed. by Marina Frasca-Spada and P. J. E. Kail. Oxford: Clarendon Press, pp. 107–24.
- Jenkins, John J. (1992). *Understanding Hume*. Edinburgh: Edinburgh University Press.
- Jones, Peter (1976). "Cause, Reason, and Objectivity in Hume's Aesthetics". In: *Hume: A Re-Evaluation*. Ed. by D. W. Livingston and J. T. King. New York: Fordham University Press, pp. 323–42.
- (1982). *Hume's Sentiments: Their Ciceronian and French Context*. Edinburgh: Edinburgh University Press.
- Kail, P. J. E. (2005). "Hume's Ethical Conclusion". In: *Impressions of Hume*. Ed. by Marina Frasca-Spada and P. J. E. Kail. Oxford: Clarendon Press, pp. 125–39.
- (2007a). *Projection and Realism in Hume's Philosophy*. Oxford: Oxford University Press.
- (2007b). "Understanding Hume's Natural History of Religion". *The Philosophical Quarterly* 57.227, pp. 190–211.
- (2008). "On Hume's Appropriation of Malebranche: Causation and Self". *European Journal of Philosophy* 16.1, pp. 55–80.
- Karlsson, Mikael M. (2006). "Reason, Passion, and the Influencing Motives of the Will". In: *The Blackwell Guide to Hume's Treatise*. Ed. by Saul Traiger. Oxford: Blackwell, pp. 235–55.
- Kemp Smith, Norman (1905). "The Naturalism of David Hume (II)". *Mind, New Series* 14.55, pp. 335–47.
- (1941/2005). *The Philosophy of David Hume*. Basingstoke: Palgrave Macmillan.

- Kenny, Anthony (1963/2003). *Action, Emotion and Will*. London: Routledge & Kegan Paul.
- Kivy, Peter (1967). "Hume's Standard of Taste: Breaking the Circle". *The British Journal of Aesthetics* 7.1, pp. 57–66.
- Leibowitz, Flo (1993). "The Logic of Hume's Conversion Theory". *The Journal of Aesthetics and Art Criticism* 51.4, pp. 625–6.
- Lennon, Thomas M. (1997). "Introduction". In: *The Search After Truth, by Nicolas Malebranche*. Ed. by Thomas M. Lennon and Paul J. Olscamp. Cambridge: Cambridge University Press, pp. vii–xxiii.
- Levinson, Jerrold (2002). "Hume's Standard of Taste: The Real Problem". *The Journal of Aesthetics and Art Criticism* 60.3, pp. 227–38.
- (2003). "The Real Problem Sustained: Reply to Weiland". *The Journal of Aesthetics and Art Criticism* 61.4, pp. 398–9.
- Livingston, Paisley (2013). "Du Bos' Paradox". *The British Journal of Aesthetics* 53.4, pp. 393–406.
- Loeb, Louis E. (1977). "Hume's Moral Sentiments and the Structure of the *Treatise*". *The Journal of the History of Philosophy* 15.4, pp. 395–403.
- Magri, Tito (2008). "Hume on the Direct Passions and Motivation". In: *A Companion to Hume*. Ed. by Elizabeth S. Radcliffe. Oxford: Blackwell, pp. 185–200.
- Mason, Michelle (2001). "Moral Prejudice and Aesthetic Deformity: Rereading Hume's 'Of the Standard of Taste'". *The Journal of Aesthetics and Art Criticism* 59.1, pp. 59–71.
- McGilvary, E. B. (1903). "Altruism in Hume's *Treatise*". *The Philosophical Review* 12.3, pp. 272–98.
- Merivale, Amyas (2009). "Hume's Mature Account of the Indirect Passions". *Hume Studies* 35.1&2, pp. 185–210.
- (2011). "Mixed Feelings, Mixed Metaphors: Hume on Tragic Pleasure". *The British Journal of Aesthetics* 51.3, pp. 259–69.
- Millgram, Elijah (1995). "Was Hume a Humean?" *Hume Studies* 21.1, pp. 75–94.
- Millican, Peter, ed. (2002a). *Reading Hume on Human Understanding*. Oxford: Oxford University Press.

- (2002b). “The Context, Aims, and Structure of Hume’s First *Enquiry*”. In: *Reading Hume on Human Understanding*. Ed. by Peter Millican. Oxford: Oxford University Press, pp. 27–65.
- (2006). “Hume’s ‘Compleat Answer to Dr Reid’”. Hume Conference, University of Koblenz, Germany. URL: <http://www.davidhume.org/papers/millican.html>.
- (2007a). “Against the ‘New Hume’”. In: *The New Hume Debate*. Ed. by Rupert Read and Richman. 2nd edition. London: Routledge, pp. 211–77.
- ed. (2007b). *An Enquiry concerning Human Understanding, by David Hume*. Oxford: Oxford University Press.
- (2007c). “Hume’s Old and New: Four Fashionable Falsehoods, and One Unfashionable Truth”. *Proceedings of the Aristotelian Society, Supplementary Volumes* 81, pp. 163–99.
- (2009a). “Hume, Causal Realism, and Causal Science”. *Mind, New Series* 118.471, pp. 647–712.
- (2009b). “Hume on Induction and the Faculties”. Draft Article. URL: <http://www.davidhume.org/papers/millican.html>.
- (2010). “Hume’s Determinism”. *The Canadian Journal of Philosophy* 40.4, pp. 611–42.
- (2011). “Hume, Causal Realism, and Free Will”. In: *Causation and Modern Philosophy*. New York & Oxford: Routledge, pp. 123–65.
- (2012). “Hume’s ‘Scepticism’ About Induction”. In: *The Continuum Companion to Hume*. Ed. by Alan Bailey and Dan O’Brien. London: Continuum, pp. 57–103.
- (forthcoming). “Hume’s Chief Argument”. In: *The Oxford Handbook of David Hume*. Ed. by Paul Russell. Oxford: Oxford University Press.
- Moore, James (1995). “Hume and Hutcheson”. In: *Hume and Hume’s Connexions*. Ed. by M. A. Stewart and John P. Wright. Pennsylvania: Penn State University Press, pp. 23–57.
- Mossner, E. C. (1936). “The Enigma of Hume”. *Mind, New Series* 45.179, pp. 334–49.
- (1950). “Philosophy and Biography: The Case of David Hume”. *The Philosophical Review* 59.2, pp. 184–201.

- Nadler, Steven, ed. (2006). *The Cambridge Companion to Malebranche*. Cambridge: Cambridge University Press.
- Nagel, Thomas, ed. (1970). *The Possibility of Altruism*. Princeton, New Jersey: Princeton University Press.
- Neill, Alex (1992). "Yanal and Others on Hume on Tragedy". *The Journal of Aesthetics and Art Criticism* 50.2, pp. 151–4.
- (1998). "'An Unaccountable Pleasure': Hume on Tragedy and the Passions". *Hume Studies* 24.2, pp. 335–54.
- Nelson, John O. (1972). "Two Main Questions Concerning Hume's *Treatise* and *Enquiry*". *The Philosophical Review* 81.3, pp. 333–50.
- Noonan, Harold (1999). *Hume on Knowledge*. London: Routledge.
- Norton, David Fate (2009). "The Foundations of Morality in Hume's *Treatise*". In: *The Cambridge Companion to Hume*. Ed. by David Fate Norton and Jacqueline Taylor. Cambridge: Cambridge University Press, pp. 270–310.
- Norton, David Fate and Mary J. Norton (1996). *The David Hume Library*. Edinburgh: Edinburgh Bibliographical Society with the National Library of Scotland.
- (2007). "Editor's Annotations". In: *A Treatise of Human Nature, by David Hume*. Vol. 2. Oxford: Clarendon Press, pp. 685–979.
- Norton, David Fate and Jacqueline Taylor, eds. (2009). *The Cambridge Companion to Hume*. Cambridge: Cambridge University Press.
- Nussbaum, Martha C. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.
- Ochs, Sidney (2004). *A History of Nerve Functions: From Animal Spirits to Molecular Mechanisms*. Cambridge: Cambridge University Press.
- Owen, David (1999). *Hume's Reason*. New York: Oxford University Press.
- Packer, Mark (1989). "Dissolving the Paradox of Tragedy". *The Journal of Aesthetics and Art Criticism* 47.3, pp. 211–19.
- Parfit, Derek (1997). "Reasons and Motivation". *Proceedings of the Aristotelian Society, Supplementary Volumes* 71, pp. 99–130.
- Paton, Margaret (1973). "Hume on Tragedy". *The British Journal of Aesthetics* 13.2, pp. 121–32.

- Peach, Bernard (1970). "The Correspondence Between Francis Hutcheson and Gilbert Burnet: The Problem of the Date". *The Journal of the History of Philosophy* 8.1, pp. 87–91.
- Penelhum, Terence (1975). *Hume*. London and Basingstoke: Macmillan.
- (2009). "Hume's Moral Psychology". In: *The Cambridge Companion to Hume*. Ed. by David Fate Norton and Jacqueline Taylor. Cambridge: Cambridge University Press, pp. 238–69.
- Phillips, D. Z. and Timothy Tessin, eds. (1999). *Religion and Hume's Legacy*. Basingstoke: Macmillan.
- Phillips, David (2005). "Hume on Practical Reason: Normativity and Psychology in *Treatise* 2.3.3". *Hume Studies* 31.2, pp. 299–316.
- Pigden, Charles R., ed. (2009). *Hume on Motivation and Virtue*. Basingstoke: Palgrave Macmillan.
- Pitcher, George (1965). "Emotion". *Mind, New Series* 74.295, pp. 326–46.
- Pitson, Tony (1996). "Sympathy and Other Selves". *Hume Studies* 22.2, pp. 255–72.
- Popkin, Richard H. (2003). *The History of Scepticism: From Savoranola to Bayle*. New York: Oxford University Press.
- Postema, Gerald J. (2005). "'Cemented with Diseased Qualities': Sympathy and Comparison in Hume's Moral Psychology". *Hume Studies* 31.2, pp. 249–98.
- Prinz, Jesse J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.
- Radcliffe, Elizabeth S., ed. (2008). *A Companion to Hume*. Oxford: Blackwell.
- Ribeiro, Brian (2007). "Hume's Standard of Taste and the *De Gustibus* Sceptic". *The British Journal of Aesthetics* 47.1, pp. 16–28.
- Rickless, Samuel C. (2013). "Hume's Theory of Pity and Malice". *The British Journal for the History of Philosophy* 21.2, pp. 324–44.
- Ross, Stephanie (2008). "Humean Critics: Real or Ideal?" *The British Journal of Aesthetics* 48.1, pp. 20–8.
- Russell, Paul (2008). *The Riddle of Hume's Treatise: Skepticism, Naturalism, and Irreligion*. Oxford: Oxford University Press.
- Schier, Flint (1989). "The Claims of Tragedy: An Essay in Moral Psychology and Aesthetic Theory". *Philosophical Papers* 18.1, pp. 7–26.

- Schmitter, Amy M. (2012). "Family Trees: Sympathy, Comparison, and the Proliferation of the Passions in Hume and his Predecessors". In: *Emotion and Cognitive Life in Medieval and Early Modern Philosophy*. Ed. by Martin Pickavé and Lisa Shapiro. Oxford: Oxford University Press, pp. 255–78.
- Selby-Bigge, L. A. (1893/1975). "Introduction". In: *Enquiries concerning Human Understanding and concerning the Principles of Morals, by David Hume*. Ed. by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press, pp. vii–xxxii.
- Shelley, James (1994). "Hume's Double Standard of Taste". *The Journal of Aesthetics and Art Criticism* 52.4, pp. 437–45.
- (2002). "The Character and Role of Principles in the Evaluation of Art". *The British Journal of Aesthetics* 42.1, pp. 37–51.
- (2004). "Hume's Principles of Taste: A Reply to Dickie". *The British Journal of Aesthetics* 44.1, pp. 84–9.
- Sheridan, Patricia (2007). "Pirates, Kings and Reasons to Act: Moral Motivation and the Role of Sanctions in Locke's Moral Theory". *The Canadian Journal of Philosophy* 37.1, pp. 35–48.
- Smith, Michael (1994). *The Moral Problem*. Oxford: Blackwell.
- Sober, Elliot (2000). "Psychological Egoism". In: *The Blackwell Guide to Ethical Theory*. Ed. by Hugh LaFollette. Oxford: Blackwell, pp. 129–48.
- Solomon, Robert C. (1977). "The Logic of Emotion". *Noûs* 11.1, pp. 41–9.
- (2004a). "Emotions, Thoughts, and Feelings: Emotions as Engagements with the World". In: *Thinking About Feeling: Contemporary Philosophers on Emotions*. Ed. by Robert C. Solomon. Oxford: Oxford University Press, pp. 76–88.
- ed. (2004b). *Thinking About Feeling: Contemporary Philosophers on Emotions*. Oxford: Oxford University Press.
- Stalley, R. F. (1986). "The Will in Hume's *Treatise*". *The Journal of the History of Philosophy* 24.1, pp. 41–53.
- Stewart, M. A. (2005). "Hume's Intellectual Development, 1711-1752". In: *Impressions of Hume*. Ed. by Marina Frasca-Spada and P. J. E. Kail. Oxford: Clarendon Press, pp. 11–58.
- Sutherland, Stewart R. (1976). "Hume on Morality and the Emotions". *The Philosophical Quarterly* 26.102, pp. 14–23.

- Taylor, Jacqueline (2002). "Hume on the Standard of Virtue". *The Journal of Ethics* 6.1, pp. 43–62.
- (2007). "Hume and the Nortons on the Passions and Morality in Hume's *Treatise*". *Hume Studies* 33.2, pp. 305–12.
- (2008). "Hume on Beauty and Virtue". In: *A Companion to Hume*. Ed. by Elizabeth S. Radcliffe. Oxford: Blackwell, pp. 273–92.
- (2009). "Hume's Later Moral Philosophy". In: *The Cambridge Companion to Hume*. Ed. by David Fate Norton and Jacqueline Taylor. Cambridge: Cambridge University Press, pp. 311–40.
- ed. (forthcoming). *Reading Hume on the Principles of Morals*. Oxford: Oxford University Press.
- Traiger, Saul, ed. (2006). *The Blackwell Guide to Hume's Treatise*. Oxford: Blackwell.
- Von Leyden, W. (1954/2002). "Introduction". In: *Essays on the Law of Nature, by John Locke*. Ed. by W. Von Leyden. Oxford: Clarendon Press.
- Webb, Mark (1991). "The Argument of the *Natural History*". *Hume Studies* 17.2, pp. 141–60.
- Weiland, Jeffrey (1984). "Hume's Two Standards of Taste". *The Philosophical Quarterly* 34.135, pp. 129–42.
- (2003). "Hume's Real Problem". *The Journal of Aesthetics and Art Criticism* 61.4, pp. 395–8.
- Weller, Cass (2002). "The Myth of Original Existence". *Hume Studies* 28.2, pp. 195–230.
- Williams, Bernard (1981). "Internal and External Reasons". In: *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press, pp. 101–13.
- Wilson, Catherine, ed. (2008). *Epicureanism at the Origins of Modernity*. Oxford: Clarendon Press.
- Wright, John P. (1983). *The Sceptical Realism of David Hume*. Manchester: Manchester University Press.
- Yanal, Robert J. (1991). "Hume and Others on the Paradox of Tragedy". *The Journal of Aesthetics and Art Criticism* 49.1, pp. 75–6.
- (1992). "Still Unconverted: A Reply to Neill". *The Journal of Aesthetics and Art Criticism* 50.4, pp. 324–6.

Yandell, Keith E. (1990). *Hume's Inexplicable Mystery*. Philadelphia: Temple University Press.

Zemach, Eddy M. (2001). "What is Emotion?" *American Philosophical Quarterly* 38.2, pp. 197–207.