The Blackwell Guide to Hume's *Treatise*

Blackwell Guides to Great Works

A proper understanding of philosophy requires engagement with the foundational texts that have shaped the development of the discipline and which have an abiding relevance to contemporary discussions. Each volume in this series provides guidance to those coming to the great works of the philosophical canon, whether for the first time or to gain new insight. Comprising specially commissioned contributions from the finest scholars, each book offers a clear and authoritative account of the context, arguments, and impact of the work at hand. Where possible the original text is reproduced alongside the essays.

Published

1   The Blackwell Guide to Plato's *Republic*   Gerasimos Santas
2   The Blackwell Guide to Descartes' *Meditations*   Stephen Gaukroger
3   The Blackwell Guide to Mill's *Utilitarianism*   Henry R. West
4   The Blackwell Guide to Aristotle's *Nicomachean Ethics*   Richard Kraut
5   The Blackwell Guide to Hume's *Treatise*   Saul Traiger

Forthcoming

The Blackwell Guide to Kant's Ethics   Thomas E. Hill Jr.
The Blackwell Guide to Hegel's *Phenomenology of Spirit*   Kenneth Westphal
The Blackwell Guide to Heidegger's *Being and Time*   Robert C. Scharff

THE BLACKWELL GUIDE TO

Hume's

Treatise

EDITED BY SAUL TRAIGER

Blackwell
Publishing

Contents

Notes on Contributors

Lilli Alanen is Professor of Philosophy at Uppsala University. She is the author of *Descartes's Concept of Mind* (2003) and numerous articles in Early Modern Philosophy. She is a co-editor of *Commonality and Particularity in Ethics* (1997) and *Feminist Reflections on the History of Philosophy* (2004), and is presently working on a book entitled *Passions and the Transformation of Reason: Descartes, Spinoza, and Hume*.

Donald L. M. Baxter is Professor of Philosophy at the University of Connecticut. He writes mostly in Metaphysics and Early Modern Philosophy. Recent work on Hume includes "Hume on steadfast objects and time" (*Hume Studies*, 2001), and "Hume's theory of space and time in its skeptical context," forthcoming in *The Cambridge Companion to Hume*, 2nd edition, edited by David Fate Norton.

Janet Broughton is Professor of Philosophy at the University of California, Berkeley. She is the author of *Descartes's Method of Doubt* (2002); with John Carriero, she is co-editing *The Blackwell Companion to Descartes*.

Rachel Cohon is Associate Professor of Philosophy at the University at Albany, State University of New York. She is the editor of *Hume: Moral and Political Philosophy* (2001) and author of several articles about Hume's moral philosophy and about the relation between morality and reasons, including "The common point of view in Hume's ethics" (*Philosophy and Phenomenological Research*, 1997) and "The roots of reasons" (*Philosophical Review*, 2000).

Lorne Falkenstein is Professor of Philosophy at the University of Western Ontario. He is the author of *Kant's Intuitionism* and of numerous articles on space and spatial representation in Hume, Reid, Kant, Berkeley, and Condillac.

Don Garrett is Professor of Philosophy at New York University. He is the author of *Cognition and Commitment in Hume's Philosophy* and the editor of *The Cambridge Companion to Spinoza*.

Mikael M. Karlsson is Professor of Philosophy at the University of Iceland, and Dean of the Faculty of Law and Social Sciences at the University of Akureyri. He is a long-standing member of the Hume Society and has served on its executive committee and as its Secretary-Treasurer. His recent publications include "Rational ends: Humean and non-Humean considerations" (*Sats – Nordic Journal of Philosophy*, 2000) and "Agency and patiency – back to nature?" (*Philosophical Explorations*, 2002).

Jane L. McIntyre is Professor of Philosophy at Cleveland State University and President of the Hume Society. She has published widely on British philosophy of the seventeenth and eighteenth centuries, and, in particular, on the work of David Hume. Her papers all argue for the centrality of Hume's account of the passions to his overall philosophical position.

William Edward Morris is Chair of the Department of Philosophy and Director of the Program in Cognitive Science at Illinois Wesleyan University. He was an editor of *Hume Studies* from 1993 to 2000, and has published extensively on various aspects of Hume's philosophy.

Tony Pitson is Honorary Research Fellow in Philosophy at the University of Stirling. He has published widely on the philosophy of the Scottish Enlightenment and is the author of *Hume's Philosophy of the Self* (2002).

Wade L. Robison is the Ezra A. Hale Professor of Applied Ethics at the Rochester Institute of Technology. He is a past president of the Hume Society. His books include *Decisions in Doubt: the Environment and Public Policy* (1994) and *Ethical Decision Making in Social Work* (2000), with L. Reeser.

Abraham Sesshu Roth is Associate Professor of Philosophy at the University of Illinois at Chicago. He is author of several papers on Hume, including "What was Hume's problem with personal identity?" (*Philosophy and Phenomenological Research*, 2000), as well as papers in the philosophy of action, including "Shared agency and contralateral commitments" (forthcoming in *Philosophical Review*).

Corliss Gayda Swain is Associate Professor of Philosophy at St Olaf College and the Executive Secretary-Treasurer of the Hume Society. She has written numerous articles on Hume's metaphysics and moral theory.

Jacqueline Taylor is Assistant Professor of Philosophy at the University of San Francisco. She has written numerous articles on Hume's moral philosophy and on

contemporary Humean ethics, including "Hume on the standard of virtue" (*Journal of Ethics*, 2002) and "Humean ethics and the politics of sentiment" (*Topoi*, 2002).

John P. Wright is Professor of Philosophy at Central Michigan University and Visiting Professor at the University of Edinburgh. He is author of *The Sceptical Realism of David Hume* (1983), and co-editor of *Hume and Hume's Connexions* (1994), *Psyche and Soma* (2000), *John Locke:* An Essay Concerning Human Understanding *in Focus* (2000), and *David Hume's* A Treatise of Human Nature (2003).

References to the *Treatise, Abstract*, and *Enquiries*

References to *A Treatise of Human Nature* (T), *An Abstract of A Treatise of Human Nature* (A), *An Enquiry Concerning Human Understanding* (EHU), and *An Enquiry Concerning the Principles of Morals* (EPM) are to the Oxford Philosophical Texts editions:

A Treatise of Human Nature, edited by David Fate Norton and Mary Norton, Oxford: Oxford University Press, 2000.

An Abstract of A Treatise of Human Nature, in *A Treatise of Human Nature*, edited by David Fate Norton and Mary Norton, Oxford: Oxford University Press, 2000.

An Enquiry Concerning Human Understanding, edited by Thomas Beauchamp, Oxford: Clarendon Press, 2000.

An Enquiry Concerning the Principles of Morals, edited by Thomas Beauchamp, Oxford: Clarendon Press, 1998.

For the *Treatise*, citations are by Book, Part, section, and paragraph number in arabic numerals (the style of the new editions) and *also* by page number from the Selby-Bigge/Nidditch edition (SBN). A citation to the *Treatise* will look like:

> (T 1.4.2.3; SBN 188–9)

For the *Abstract*, citations are by paragraph number and by corresponding page number in the Selby-Bigge/Nidditch edition. For the Preface to the *Abstract*, "A Pref." is used. Sample citations are:

> (A 3; SBN 646)
> (A Pref. 1; SBN 643)

For the *Enquiries*, citations are by section and paragraph number, and by page number from the Selby-Bigge/Nidditch edition. Sample citations are:

(EHU 7.6; SBN 63)
(EPM 2.12; SBN 180)

The Selby-Bigge/Nidditch editions are:

A Treatise of Human Nature, edited by L. A. Selby-Bigge, 2nd edition edited by P. H. Nidditch, Oxford: Clarendon Press, 1978.

An Abstract of A Treatise of Human Nature, in *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd edition edited by P. H. Nidditch, Oxford: Clarendon Press, 1978.

Enquiries Concerning Human Understanding and Concerning the Principles of Morals, edited by L. A. Selby-Bigge, 3rd edition edited by P. H. Nidditch, Oxford: Clarendon Press, 1975.

Introduction

Saul Traiger

This book introduces the reader to one of the most important philosophical works of modern philosophy, David Hume's *A Treatise of Human Nature*. Although its initial reception led Hume to describe it as having fallen "*dead-born from the Press*," the influence of *Treatise* on the philosophical world is incalculable. Immanuel Kant, for example, admitted that Hume's ideas caused him to awake from his dogmatic slumbers and begin work on *The Critique of Pure Reason*. While the appreciation of the *Treatise* in the nineteenth century was muted, the twentieth century witnessed an explosion of Hume scholarship. Today the extraordinary interest in the *Treatise* continues to grow. The purpose of this book is to introduce the student of modern philosophy to the main themes, arguments, interpretive issues in Hume's *Treatise*, including key results and trends in Hume scholarship.

Our understanding of Hume's *Treatise* has advanced on four major fronts. First, we now know much more about the philosophical and literary context of the writing of the *Treatise* than we did in the 1960s and 1970s. Careful historical and textual scholarship has uncovered many of the influences on Hume. Much has been discovered about the writing and publishing of the *Treatise*, and its early and long-term reception. Second, Hume's metaphysics and epistemology have received intense scrutiny. Work done in the late twentieth century in this area explores ideas about the naturalistic aspects of Hume's treatment of causation first advanced by Norman Kemp Smith. New debates have arisen about whether Hume was a realist about causation, about the nature and scope of his skepticism, and about the interpretation of Hume's conclusion of Book I. Third, beginning with seminal papers by Annette C. Baier and Donald Davidson in the 1970s, Hume's theory of the passions began to receive the attention it deserves, and a growing body of Hume interpretation concentrates on this aspect of his theory of human nature. Finally, Hume's moral theory has come to be seen as a central item of the vocabulary of contemporary theorizing about the role of reason in morals. As a result, many of the central contemporary moral theorists have couched their own views

in relation to Hume's moral theory, and in terms of the relation of Hume's moral theory to that of Kant. This has led to wide-ranging discussions of the proper interpretation of Hume's moral theory.

The organization and content of this book reflects these four areas of developments in Hume scholarship. Each chapter is a new, self-contained, independent essay which provides the reader with the scholarly and interpretive tools needed to begin mining this rich text for philosophical insight. Readers can go directly to the essay or essays that address the topic he or she wants to investigate. The chapters are, however, ordered and coordinated to correspond to the main topics of the *Treatise*, and careful attention has been paid to allowing the reader to move easily between essays.

The *Guide* has four parts. Part I includes a discussion of the formulation, reception, and scope of the *Treatise*: the writing of the *Treatise*, including the philosophical, literary and cultural influences on its young author. It covers the reception of the *Treatise* from the time of Hume's contemporaries, through the nineteenth century, and up to Kemp Smith and the early twentieth century. Part I also covers the relation of the *Treatise* to Hume's other works; this and other chapters encourage students to read the *Treatise* in relation not only to the *Enquiries*, but also to Hume's *History of England* and his *Essays*. The essays of Part II address Book 1 of the *Treatise*. They emphasize the current issues in Hume interpretation, including such questions as whether Hume was a realist about causal connection, the relation of his skepticism to his normative epistemological program, and the proper interpretation of his review of his metaphysics and epistemology at the end of Book 1. Parts III and IV, "The Passions" and "Morals" respectively, contain essays on material from Books 2 and 3 of the *Treatise*. Part III treats the passions and the will, topics to which Hume devoted all of Book 2 of the *Treatise*, but which are rarely treated with the careful attention found in the essays of this volume. Part IV treats the moral sentiments and the relation of moral sentiment to reason, Hume's theory of justice as an artificial virtue, the role of sympathy, and the evaluation of moral character. In all the essays in this volume the reader will find carefully crafted and accessible explanations of the positions Hume took, close readings of the relevant passages from the *Treatise*, and illuminating discussions both of alternative interpretations and analyses of Hume's contribution to ongoing contemporary philosophical debates.

Part I

Formulation, Reception, and Scope of the *Treatise*

1

The *Treatise*: Composition, Reception, and Response

John P. Wright

In *My Own Life*, the short autobiography that Hume wrote a few months before he died, he stated that he "composed [his] *Treatise of human nature*" during his "Retreat in France, first at Reims, but chiefly at La fleche in Anjou" (Greig 1932, I: 2). Hume had gone to Paris in the summer of 1734 at age 23 and, after a few weeks there, went to Reims, a "famous & antient Town & University" about 130 miles from Paris (p. 22). He spent about eight months in Reims and then moved to La Flèche, where he spent over two years writing the *Treatise*. The latter town had much natural beauty, and offered Hume a serene place to study and write. But clearly its major attraction was the famous Jesuit College, where Descartes had gone to school a century earlier. Hume described it as "a College of a hundred Jesuits, which is esteem'd the most magnificent both for Buildings & Gardens of any belonging to that Order in France or even in Europe" (Mossner 1980: 32). It is likely that Hume would have had access to the substantial library at the College, and he certainly had conversations with the learned men who taught there. We know that he thought up his famous argument concerning the lack of credibility of testimony about miracles when having a discussion with a learned Jesuit in the cloisters of the College.

When Hume arrived in London in early September 1737, he began immediately to approach printers; however he did not sign a contract for publication of the first two books of the *Treatise* for another year. He probably spent much of that time editing his work. When he wrote to his friend Henry Home in December 1737, he told him that he had spent the three months since he arrived in London improving his "style and diction" and, more significantly, "castrating [his] work, that is, cutting off its nobler parts" in order that "it shall give as little offence [to religious people] as possible" (Greig 1932, I: 24–5). Hume mentions in the letter that he had once thought of publishing his argument about miracles as part of the *Treatise*, but had thought better of it. He removed other sections which could offend religious sensibilities. These may have included discussions in which he

showed the unsoundness of philosophical arguments for the immortality of the soul, for the belief in an afterlife, and for the existence of God.

The first two books of *A Treatise of Human Nature* – those concerning the understanding and the passions – were not published until late January 1739, when Hume was 27 years old. Shortly afterwards, he left London for his home in Scotland where he eagerly awaited news of the reception of the *Treatise*. By the summer, when he learned that few copies of his book had been sold, Hume began to doubt not only the success of the book, but also whether he had really made the significant discoveries which he had thought he had made. Nevertheless, he continued work on the third book of the *Treatise*, on morals. Also, while he was waiting for reviews of the first two books of the *Treatise*, he wrote An *Abstract . . . of A Book lately Published; Entitled A Treatise of Human Nature*, a "puff" of his own book, outlining the main argument of Book 1.

Reception of the *Treatise* by Francis Hutcheson and Hume's Revisions to Book 3

Shortly after the first two books of the *Treatise* were published, Henry Home sent a copy to Francis Hutcheson, Professor of Moral Philosophy at the University of Glasgow, among the most famous moral philosophers of the day. Hutcheson's reception of these two books was very positive. He had read through them by April and wrote back to Home that he "was every where surprised with a great acuteness of thought and reasoning in a mind wholly disengaged from the preju- dices of the Learned as well as those of the Vulgar" (Ross 1966: 69–72). He said that he appreciated the skepticism of the anonymous author, and that he himself in recent years had begun to despair "of Certainty in the most important subjects" and thought that one must be "satisfied with a sort of Probable knowledge" which would be "sufficient for the Conduct of Life." He also expressed a desire to meet the author of the book.

Encouraged by Hutcheson's letter to Home, Hume sent Hutcheson a draft of Book 3 and probably went to Glasgow to meet him. From the letters from Hume to Hutcheson which have survived it is clear that Hutcheson's reception of the book on morals was far more critical than his reception of the first two books. Among a number of other criticisms, Hutcheson told Hume that his writing lacked a "certain Warmth in the Cause of Virtue" – that is, it did not encourage its readers to become virtuous (Greig 1932, I: 32–5). Hume responded in a letter of September 1739 that he was writing a scientific (metaphysical) work explaining how and why people make the moral evaluations that they do make; he was not writing as a moralist, who tries to change people's behavior. He used the analogy of the "Anatomist" and "Painter" to explain the necessity of separating these two tasks: he "cannot easily conceive" the two tasks of anatomy and painting "united in the same Work." Nevertheless, Hume told Hutcheson that he would make a

new effort to see if he could "make the Moralist & Metaphysician agree a little better."

Hutcheson was also concerned about certain passages that he thought "defective in Point of Prudence" because they would offend religious people (Greig 1932, I: 34). Hume agreed to follow Hutcheson's advice and remove these passages; so once again he set about "castrating" his writing. But he also told Hutcheson that he thought him "a little too delicate" and that a person's philosophical speculations are not really relevant to his character unless he is a clergyman or directly involved in "the Instruction of Youth." Since Hume was neither he did not think he needed to be overly concerned about the orthodoxy of his philosophical views. This response came back to haunt Hume five years later when he sought a teaching position, and Hutcheson opposed his candidacy for the Edinburgh chair of moral philosophy.

Other criticisms from Hutcheson concerned Hume's views that justice was not a "natural virtue" and that natural abilities are virtues. After completing his revisions to Book 3, Hume wrote to Hutcheson on March 4, 1740 asking for his help finding a publisher. He assured Hutcheson that the revisions he had made left nothing to which religious people should take offence, though he added that "the Clergy . . . [are] always Enemys to Innovations in Philosophy" (Greig 1932, I: 36–8). Hume sent Hutcheson his altered Conclusion which, he claimed, would show even "the strictest most rigid" persons the "Soundness of [his] Morals." In it he argued that his own system of "sympathy" shows how our happiness depends upon the approval of others of our virtuous characters. He also claimed that his system is superior to that of philosophers who depend on "original instincts of the human mind" – an apparent reference to Hutcheson's own theory of moral sense (T 3.3.6.3; SBN 619). Moreover, while reaffirming his original claim to Hutcheson that "anatomy" and "painting" cannot be combined in one work, Hume argued that the anatomy of human nature that *he* has provided will be useful to the painter. The latter must base his exhortations to virtue on facts about human nature (T 3.3.6.6; SBN 620–1).

The Early Reviews of the *Treatise* and Hume's Response

The first review of the *Treatise* was published in *The History of the Works of the Learned* in late 1739 (Fieser 2000: 3–40). It began by objecting to the fact that Hume had denied the freedom of the will. The reviewer went on to complain of the author's pretentiousness in refuting famous authors such as John Locke and Samuel Clarke, his excessive use of the word "I," and the obscurity of his arguments. The final paragraph is complimentary in a rather backhanded way: the book "bears . . . incontestable Marks of a great Capacity" and "a soaring Genius, but young, and not yet thoroughly practised" (pp. 39–40). The reviewer says that after

the author has matured, the *Treatise* may be regarded like "the *Juvenile* Works of *Milton*" or the first paintings of a genius like Raphael.

The second review (apart from Hume's own *Abstract*, which had been published in February or early March) appeared in the summer of 1740 in the *Bibliothèque raisonée* (Fieser 2000: 48–63). It consisted mostly of a French translation of the *Abstract*, but there were also three paragraphs which repeated some of the criticisms in the first review in *The History of the Works of the Learned*. The reviewer noted the extreme Pyrrhonian skepticism that marks Book 1 of the *Treatise*, the inconsistencies of many propositions in the book with each other, the dangerous results for religion that can be inferred from its principles, and the dogmatism with which the author "dares to substitute his speculations for the opinions of the greatest philosophers" (p. 63). The reviewer says he had never encountered a more dogmatic skepticism. He ends the review by saying that far better work can be expected once age "has matured" the author's "taste and given him time to think anew" on the subjects he has dealt with in the *Treatise*.

Other reviews appeared in the *Gottingische Zeitungen* (January 1740) and the *Nouvelle Bibliothèque* (July and September 1740). They consist mostly of quotations or summaries of Book 1 of the *Treatise*. The one interesting critical remark in the second of these was that Hume "confused belief with an act that follows immediately from it, namely, judgment" (Fieser 2000: 77). This is a criticism that is still made of his philosophy (see Stroud 1977: 232).

The first philosophical consideration of Hume's arguments in the *Treatise* appeared in July, in an anonymous letter to the editor of *Common Sense: or the Englishman's Journal* (Fieser 2000: 86–91). The critic took issue with Hume's discussion "Of liberty and necessity" in Part 3 of Book 2. He began by complaining of the obscurity of Hume's view, contrasting it with the clarity of the argument against free will of Anthony Collins (whom he had already criticized in his *An Essay towards demonstrating the Immateriality and Free-Agency of the Soul*, 1740). But then the critic goes on to engage Hume's arguments in a significant way.

First, against Hume's claim that the ultimate causal connection of external bodies is not "discoverable either by our senses or reason" (T 2.3.1.4; SBN 400), he replies that Newton discovered this causal connection in "so far as it depends upon, or proceeds from any *Quality* in *Matter*" (Fieser 2000: 88). But, he continues, "*Matter* is in itself absolutely *passive*" and "can never *of itself* produce any Effect." Physical changes must ultimately derive from an "*active Being*, supreme or subordinate, which we call *Spirit*." This claim, that the ultimate unknown power from which natural effects result must be spiritual, underlay criticisms of Hume's philosophy until well into the nineteenth century.

Then the critic objects to Hume's account of the origin of the idea of necessity, arguing that we would never have that idea unless we "perceived in ourselves a Motion *that is not necessary*" (Fieser 2000: 88). Since all actions, even those we normally call free, are on Hume's view necessitated, we can have no contrast by which we can understand necessity. (If all objects in the world were blue, we would

have no concept of blue objects; similarly, if everything were necessitated then we would lack the contrast needed to form a meaningful concept of necessity.)

Finally, this critic engages Hume's argument that from observation of behavior we discover "that our actions have a constant union with our motives, tempers, and circumstances," and because all we observe is constant succession of cause and effect both in the case of external bodies and human actions, there is as much reason to ascribe necessity in the latter case as in the former (Fieser 2000: 89; cf. T 2.3.1.4; SBN 400). He acknowledges that the evidence Hume gives from external observation establishes that our motives, character and circumstances have a great deal of *influence* on our actions; nevertheless, he claims, we learn from introspection that "this influence is not *absolute* and *necessary*." And the critic adds that "*Self Conviction* is a much stronger Proof than any we can have from our *Observation* of *external Objects*." Thus he recognizes that a central issue between his own view and that of Hume is epistemological: it depends on whether introspection or observation of external behavior is accepted as more trustworthy.

The sole review of Book 3, "Of morals" (which had been published in November 1740), appeared in the *Bibliothèque raisonée* in the summer of 1741 (Fieser 1999: 2–10). The tone of the review, like the earlier one of Books 1 and 2 in the same journal, is negative. The reviewer begins by complaining of the obscurity of Hume's metaphysics, though he recognizes that Hume has tried to be clearer in this book than in the earlier ones. He commends Hume's use of examples in discussing moral philosophy, but he complains of the book's organization. The reviewer summarizes the three main arguments of section 1 of Part 1, where Hume argues against the view that moral judgments are based on reason. He quotes the passage toward the end of section 1 where Hume says that moral judgments are based on feeling or sentiment arising from the constitution of our nature (T 3.3.1.26; SBN 469), and then proceeds to complain that Hume has not looked at the criticisms put forward against this view by Gilbert Burnet and George Berkeley when it was first proposed by Francis Hutcheson. Finally, the author notes that Hume has diverged from Hutcheson's philosophy by calling justice an artificial virtue, and claims that this second part of Hume's theory is only that of "HOBBES'S system presented in a new form" (p. 10). But here the review breaks off without any attempt to analyse Hume's moral philosophy any further.

In *My Own Life* Hume stated that there was never a "literary Attempt . . . more unfortunate than my Treatise of human Nature. It fell *dead-born from the Press*; without reaching such distinction as even to excite a Murmur among the Zealots" (Greig 1932, I: 2). This latter claim was not entirely accurate for, as we have just seen, the early reviews constitute at least a "murmur" from reviewers who suspected Hume of religious impiety. After the first review appeared in the *History of the Works of the Learned*, Hume wrote to Hutcheson that it was "somewhat abusive" (Greig 1932, I: 38). He also stated that he now thought that the *Treatise* would "be more useful by furnishing Hints & exciting People's Curiosity than as containing any Principles that will augment the Stock of Knowledge that must

pass to future Ages" (p. 39). He also told Hutcheson that he suspected that he had been foolish to try to publish so many new ideas at his age, and that he could at least "plead as [his] Excuse that very Circumstance of Youth, which may be urg'd against [him]." Still, at this time he was planning a second edition of Books 1 and 2 of the *Treatise*, a project which he eventually abandoned. Hume did write a short Appendix, which he added to Book 3 and in which he acknowledged three philosophical errors, one major (in his account of personal identity) and two minor (T Appendix 10–22; SBN 633–6). He also added some paragraphs where he tried to clarify some of the points that he made. Interestingly, none of these changes relate directly to the criticisms of the reviewers.

In *My Own Life*, Hume wrote that since he was by nature cheerful and optimistic, he soon recovered from his disappointment at the poor reception of the *Treatise* and went on to other projects. But there is good reason to think that Hume was affected by his disappointment throughout his life. The young age at which he wrote the book and his claims to make so many new discoveries became a constant refrain whenever Hume discussed the *Treatise* in later life. When he wrote to his friend Gilbert Elliot in 1751 advising him not to read the *Treatise*, he told him that he was influenced by "the Heat of Youth & Invention to publish too precipitately" and that he has "repented [his] Haste a hundred, & a hundred times" (Greig 1932, I: 158). Writing to a critic, Professor John Stewart, three years later, says he is particularly troubled by "the positive Air, which prevails in the [the *Treatise*] . . . which may be imputed to the Ardor of Youth" (p. 187). And in the famous Advertisement, written at the end of his life, when he publicly disowned the *Treatise*, he objected to critics who have directed their criticisms "against that juvenile work, which the Author never acknowledged . . ." So finally, for Hume himself, as for the first reviewer, the *Treatise* ended up being a *juvenile* piece of writing.

The Principal's Attack in 1745 and Hume's Defence in his *Letter from a Gentleman*

In 1745, Hume competed, unsuccessfully as it turned out, for the chair of moral philosophy at the University of Edinburgh. In order to defeat Hume's candidacy, the Principal of the University, William Wishart, circulated a collection of quotations from the *Treatise* along with six *charges* which, according to him, showed how Hume had undermined the principles of religion and morality (Hume 1967: 17–18). He accused Hume of holding each of the following: (1) "Universal Scepticism"; (2) Principles which lead to "downright Atheism," because he denied "the Doctrine of Causes and Effects"; (3) Errors concerning the "very Being and Existence of a God," because he denied that we have a distinct idea of existence apart from the idea of any other being; (4) Errors concerning the claim that God is "the first Cause, and prime Mover of the Universe"; (5) "Denying the

Immateriality of the Soul, and the Consequences flowing from this Denial"; (6) Undermining "the Foundations of Morality" because he made the difference between justice and injustice to be merely "artificial, and to arise from human Conventions and Compacts."

Hume was in England, in Hertfordshire, when he received a copy of Wishart's objections, and he hurriedly answered each of them in a letter sent to his friend Henry Home back in Edinburgh. His answers, along with Wishart's quotations and charges, were printed in an anonymous pamphlet entitled *A Letter from a Gentlemen to his Friend in Edinburgh.* In his answers, Hume supplied the context for the quotations, and showed how his meaning had been perverted by Wishart through "broken and partial Citations" (Hume 1967: 33). He did admit, however, that he ought to have "delayed the publishing of that Book [i.e. the *Treatise*]; 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 revisions. Hume's responses to the Principal's charges are worth considering in detail.

First, Hume denies Wishart's charge of universal skepticism, and he denies that, in any case, skepticism leads to irreligion. He states that a "universal Doubt" is impossible from a practical point of view, and the purpose of skeptical arguments such as he had given in the *Treatise* to show the weakness of "common Reason, and even of the Senses" was to check "the pride of *mere human Reasoners*" (p. 19). It is true that he had shown that those beliefs to which we "are necessitated from the strongest Instincts of Nature" are not fully consistent and not absolutely certain. But the aim of skeptical argument is to create "Modesty . . . and Humility."

"How," he asks, "is such a Frame of Mind prejudicial to Piety?" (p. 20). Hume goes on to point out that the early fathers of the Church, as well as the first Protestant reformers, sought to show "the Weakness and Uncertainty of *mere* human Reason" and that most heresies have resulted "from too great a Confidence in mere human Reason" and the unwillingness to submit to the superior Light of Revelation" (p. 21).

Hume is certainly on solid ground in showing the close connection between skepticism and religious fideism. Nevertheless, his own acceptance of fideism cannot be taken at face value, given other things we know about his religious views. This is a subject he was going to explore with some care in the first Dialogue of his *Dialogues Concerning Natural Religion*, written five or six years later.

Wishart's second objection was that he had denied the causal maxim, that every beginning of existence has a cause, and that this leads to atheism. Hume's first response to this charge was to point out, in effect, that his discussion of the causal maxim was merely epistemological – that is, that he was only concerned with the type of evidence which supports the maxim. He had denied that it was based on either intuition or demonstration; that is to say, it is known neither through immediate comparison of ideas, nor is it deduced from propositions which are

known intuitively. (Roughly speaking, it is neither self-evident nor deduced from what is self-evident.) But, Hume goes on to claim, there are other forms of evidence no less strong than intuition or demonstration, such as the evidence of the senses or "*moral Evidence*" (p. 22). All he had claimed in the *Treatise* was that the causal maxim was not supported, as most people thought, by intuition or demonstration. He had not claimed that moral evidence was inferior to these other two kinds.

Further, Hume asks how, if he had really doubted the causal maxim, this would lead to atheism. There is only one argument for the existence of God that would be affected, namely Samuel Clarke's form of the cosmological argument, "the *metaphysical* Argument *a priori*," an argument that has been rejected by many pious theologians (p. 23). Other forms of the argument, such as that of Descartes, remain intact, says Hume. Further, he argues that the argument from Design, "from the Order and Course of Nature," retains its full force "because it is so sensible, so convincing, and so obvious." Indeed, he claims that the whole thrust of his analysis of causality in the *Treatise* supports the Design argument. It shows that "we can judge, only of the Operations of Causes by Experience, and that, reasoning *a priori*, any thing might be able to produce any thing" (p. 25). "Whenever I see Order, I infer from Experience that *there*, there hath been Design and Contrivance." From this principle we are obliged "to infer an infinitely perfect Architect."

Is this discussion of the Design argument, in his response to Principal Wishart, really candid? This is a difficult question that does not admit of an easy answer. On the one hand, in his later writings, in section 11 of his *Enquiry Concerning Human Understanding* and in his posthumously published *Dialogues Concerning Natural Religion*, Hume showed how a full understanding of the implications of the principles of causality laid down in his *Treatise* do weaken the Design argument. On the other hand, nowhere in his published writings does he attack this argument, the central support of moderate religion in the Church of Scotland, in his own voice; he only attacks it through the voices of characters in these dialogues. Indeed, in the Appendix to the *Treatise* he had added a footnote in which he explicitly endorsed the Design argument: "The order of the universe proves an omnipotent mind . . . [and] nothing more is requisite to give a foundation to all the articles of religion" (T 1.1.14.12n.; SBN 633n). In the light of this ambiguity, it is interesting to note that in his response to the Principal, Hume accuses him of lack of candor when "the Inferences of others are ascribed to the Author as his real Opinion" (p. 24).

Principal Wishart's third charge was that Hume made a mistake about the very existence of God in the *Treatise* because he denied that existence is a distinct idea apart from the idea of the thing which is said to exist. Hume states that he does not know what to make of this charge. He has followed "the present pious and Learned Bishop of Cloyne," George Berkeley, in denying that we have "*abstract* or *general Ideas*, properly . . . speaking" (p. 26). It is true that he has denied that

we "form a general abstract Idea of Existence, which we unite with the Idea of God" but "this is the Case with regard to every Proposition concerning existence," even the idea of oneself, and Wishart surely will admit that he is convinced of his own existence.

Hume's answer to Wishart's fourth accusation, which involved a complete mis-construction of Hume's criticism of the occasionalist theory of causation, came to be incorporated into his *Enquiry Concerning Human Understanding* which he probably began writing at this time. This answer is very important, I believe, in interpreting his theory of causation. Wishart implied that Hume had denied that God is the first cause of the universe, created matter, gave it "its original Impulse, and likewise supports its existence" (p. 18). Hume explains that the quotation from the *Treatise* on which the Principal based this claim was related to his criticism of occasionalism. He argues that very few philosophers in England ever held this theory, which denies that there are real causes in nature, and maintains that God was the only cause in the universe. He says specifically that "Sir *Isaac Newton* (tho' some of his Followers have taken a different Turn of thinking) plainly rejects [the occasionalist theory], by substituting the Hypothesis of an Aetherial Fluid, not the immediate Volition of the Deity, as the Cause of Attraction" (pp. 28–9). Here and in the *Enquiry Concerning Human Understanding*, Hume implicitly endorses the view he ascribes to Newton, that there are real, though unknown forces in nature which we can speculate about (EHU 7.25 n.16; SBN 73–4n).

Wishart's fifth charge was that Hume denied the immateriality of the soul and accepted "the Consequences flowing from this Denial" (p. 18). Hume states that he nowhere denied that the soul was immaterial "in the common Sense of the Word" (p. 30). The view he did hold, namely that the whole question did not have a "distinct Meaning" because we have "no distinct Idea of Substance," had already been stated by Locke and Berkeley.

In answer to the sixth charge that he had undermined the "Foundations of Morality," Hume points out that he shares the view of "all the antient Moralists, as well as . . . Mr. *Hutchison* Professor of Moral Philosophy in the University of *Glasgow*" that "the Propositions of Morality" are objects of "our internal *Tastes* and *Sentiments*" and not "*merely* of Reason" (p. 30). Hume was being very politic in pointing out the commonality between his own view and that of Hutcheson: Wishart and Hutcheson were politically and philosophically allied, and Hume had good reason to believe that Hutcheson himself was opposed to his candidacy (Greig 1932, I: 58).

Wishart had specifically attacked Hume's view that justice is an artificial virtue, arising from "human Conventions and Compacts" (Hume 1967: 18), a part of his moral theory that certainly did oppose Hutcheson. In two long paragraphs, Hume defended his view. He had maintained that, "along with a *natural Instinct*," artificial virtues such as justice require "a certain Reflection on the general Interests of Human Society, and a Combination with others" (p. 31). He points out that Hutcheson had been criticized for basing virtue too much on instinct, and that

he (Hume) was correcting that by showing that the virtue of justice requires the use of "*Reason* and *Reflection*." He stressed that society is required before human beings form contracts, that contracts are unintelligible without society, and that even if they were intelligible we would be under no obligation to fulfill them without established conventions. Nevertheless, he maintained that "the Laws of Justice are universal, and perfectly inflexible."

Hume's *Letter to a Gentleman* marks a transition in his philosophical writing. After his failure to obtain the chair of moral philosophy at the University of Edinburgh, he became less defensive regarding his views on religion and went on the attack. In his *Philosophical Essays Concerning Human Understanding* (later renamed the *Enquiry Concerning Human Understanding*), published in 1748, Hume "cast the first part of [his *Treatise*] anew" (Greig 1932, I: 3). But in so doing he added back in discussions of religion that he had castrated from the earlier work, as well as leaving out and abridging some of the more difficult sections. Two sections of the *Enquiry* explicitly apply the principles of Hume's philosophy to matters of religion: section 10 "Of miracles" to revealed religion, and section 11, originally called "Of the practical consequences of natural religion," to rational religion. His recasting of Book 3 of the *Treatise*, published in 1751 as *An Enquiry Concerning the Principles of Morals* concluded with a section in which he contrasted the virtues approved of by his system of philosophy with the "monkish virtues" approved of by traditional Christianity.

Criticisms of the *Treatise* after Publication of the *Enquiries*

Much to Hume's chagrin, the publication of these new philosophical works did not stop his critics from going back to the *Treatise of Human Nature* for his philosophical views. In 1754, in a paper read to the Edinburgh Philosophical Society, John Stewart, Professor of Natural Philosophy (i.e. physics) at the University of Edinburgh, attacked the *Treatise* in terms reminiscent of the early reviewers. Stewart's paper was a reply to a paper by Hume's friend Henry Home; Home had argued for the activity of matter, and against the common philosophical view that new motion can only result from the action of a thinking being with a will. Stewart's reference to David Hume's philosophy entered into his article when he asserted that Kames's thesis that motion can arise from matter itself is "much the same thing as to allow, that motion may begin without any cause at all" (Stewart 1754: 116). This led to the accusation that the principle that "something may begin to begin to exist, or start into being without a cause, hath indeed been advanced in a very ingenious and profound system of the sceptical philosophy." He adds sarcastically that this view has "not yet been adopted by any of the societies for the improvement of natural knowledge." In a footnote referring to Hume's *Treatise of Human Nature*, he describes it as a book "suited only to the comprehension of Adepts"

and states that the *Enquiry Concerning Human Understanding* is a "compend or summary" of it "for the benefit of vulgar capacities."

Stewart's charge that Hume maintained that something can exist without a cause was based on section 3 of Part 3 of Book 1 of the *Treatise* where Hume had discussed the causal maxim "that whatever begins to exist, must have a cause of existence." This was a discussion which was not reproduced in his later philosophical writings. As co-secretary of the Philosophical Society of Edinburgh, Hume was responsible for editing the volume of papers in which those of Kames and Stewart were printed and he wrote a letter to Stewart in which he agreed not to discuss the dispute in the Preface to the volume. Nevertheless, he told him that he had "never asserted so absurd a Proposition, as *that any thing might arise without a Cause*" (Greig 1932, I: 187). He went on to give the same explanation that he had given in response to the same charge by Principal Wishart in the *Letter from a Gentleman*: his discussion had been epistemological and he had only "maintain'd, that our Certainty" that nothing can begin to exist without a cause "proceeded neither from Intuition nor Demonstration; but from another Source."

At the same time, Hume told Stewart that he was angry "where a man of Sense mistakes [his] Meaning." But he added that his anger was only directed at himself for having expressed himself so badly. Again, he expresses regret at having published so young and having claimed to "innovate in all the sublimest Parts of Philosophy." He hopes that "the same Doctrines, better illustrated & exprest" in his later philosophical writings will be more successful.

Thomas Reid's Criticisms of Hume's Philosophy and Hume's Response

Undoubtedly, the most important and influential response to Hume's philosophy published in his lifetime was that of his countryman, Thomas Reid. Published in 1764, long after Hume's two *Enquiries*, Reid's book *An Inquiry into the Human Mind on the Principles of Common Sense* went back and focused on the *Treatise*. Reid directed his criticisms to Book 1, arguing that Hume was led into the most extreme form of Pyrrhonian skepticism in that book. Hume had carefully guarded himself against this interpretation of his views in Part 12 of his *Enquiry Concerning Human Understanding* where he identified himself as a proponent of a "*mitigated* scepticism or ACADEMICAL philosophy" in which philosophical doubts of "PYR-RHONIANISM or excessive scepticism" are "corrected by common sense and reflection" (EHU 12.3.24; SBN 161). It was no coincidence that Reid went back to the first book of the *Treatise*, the Conclusion of which can be read as expressing a kind of despairing skepticism.[1] Moreover, Hume had eliminated from his first *Enquiry* the most paradoxical claims of the *Treatise*, such as the claim that perceptions are substances (T 1.4.5.5; SBN 233), and that the self or mind is "nothing

but a bundle of . . . different perceptions, which . . . are in a perpetual flux and movement" (T 1.4.6.4; SBN 252).

In the Dedication which served as a Preface to his book, Reid wrote that he had

> never thought of calling in question the principles commonly received with regard to the human understanding, until the *Treatise of human nature* was published, in the year 1739. The ingenious author of that treatise, upon the principles of Locke, who was no sceptic, hath built a system of scepticism, which leaves no ground to believe any one thing rather than its contrary. His reasoning appeared to me just: there was therefore a necessity to call in question the principles upon which it was founded, or to admit the conclusion. (Reid 1997: 3–4)

Reid claimed that Hume's skepticism resulted from the hypothesis that "nothing is perceived but what is in the mind which perceives it: That we do not really perceive things that are external, but only certain images and pictures of them imprinted on the mind" (1997: 4). While he began by ascribing this hypothesis to Locke, on the next page of the Dedication he claims that it "is ancient . . . and hath been very generally received by philosophers." In the body of the book he ascribes its origin to Descartes (p. 23), first calling it the "*the ideal system*" and then "the theory of ideas" (p. 75), the name which has stuck through the subsequent history of philosophy. Hume is supposed to have drawn from the theory the skeptical conclusion that "the whole universe about me, bodies and spirits, sun, moon, stars, and earth, friends and relations, all things without exception, which I imagined to have a permanent existence, whether I thought of them or not, vanish at once" (p. 4). Reid stated that he "thought it unreasonable . . . to admit a hypothesis, which . . . overturns all philosophy, all religion and virtue, and all common sense" (p. 5).

Reid's view that Hume had in effect produced a *reductio ad absurdum* of a "theory of ideas" assumed by previous philosophers has proved very resilient in the history of philosophy (e.g. Green 1968; Stroud 1977). As we have seen, Reid is ambiguous about how far back in history this assumption is supposed to have been made: sometimes the theory is ascribed to ancient philosophers, sometimes to modern philosophers since the time of Descartes, and sometimes it is an assumption which derives from the philosophy of Locke.[2] It is only the last of these that has any plausibility; but even here one must be very cautious in accepting Reid's historical claim (Yolton 1984).

What precisely is this the "theory of ideas" that is supposed to have led Hume to the conclusion that there is no world of external objects? Reid wrote:

> It is a fundamental principle of the ideal system, That every object of thought must be an impression, or an idea, that is, a faint copy of some preceding impression. This is a principle so commonly received, that the author [of the *Treatise of Human Nature*], although his whole system is built upon it, never offers the least proof of it. (1997: 33)

What Reid appears to have in mind here when he writes of the "ideal system" is the theory introduced in Book 1, Part 1, section 1 of the *Treatise*, though it is expressed in a rather confusing way. Hume's general term for all contents of the human mind is "perceptions" and these are divided into impressions and ideas. The former are identified as feelings and the latter as *thoughts*. The thesis that concerns Reid has to do with Hume's account of the *resemblance* of feelings and thoughts, impressions and ideas, that is, the thesis that "every simple idea has a simple impression which resembles it" (T 1.1.1.5; SBN 3). Reid claims that it is this principle which Hume merely accepted from his predecessors and did not argue for.

In order to understand the metaphysics which Reid identified as underlying Hume's philosophical project it is useful to consider exactly what principle Reid sought to put in its place. What exactly would it be, according to Reid, to perceive "things that are external" and not merely "images and pictures in the mind"?

Consider Reid's discussion of hardness in his chapter on touch in the *Inquiry into the Human Mind*. Hardness, for Reid, is a primary quality like motion, extension, or space, which belongs to objects external to us. He defines hardness as "the cohesion of the parts of body with more or less force" (1997: 61). Reid recognizes (and here he is in agreement with Hume) that our conception of and belief in hardness arises from a sensation that we have when we touch a hard object. Both Reid and Hume hold that sensations (i.e. Hume's sensory impressions) are causes of our conceptions of external objects. However, Reid thinks that we tend to confuse our conceptions of these external objects – in this case our conception of the property of hardness – with the sensations which give rise to them. The problem is that our minds become aware of the thought of hardness and we don't notice the sensation which occasions it, except in rare instances like those where we knock against a hard object with a great deal of force, causing pain. But, Reid claims, when we do attend to the sensation which causes our conception of hardness, we recognize that it is nothing like the quality of hardness itself. There is no resemblance between them, and the sensation is merely the "natural sign" that makes us think of the quality in the object (see Reid 1997: 58–61).

Reid is sometimes thought to be opposed to a representative theory of perception. But, in fact, his own theory holds that we are immediately aware of sensations which then make us think of external objects and their qualities. What he insists upon is that our conceptions of the latter are not themselves sensations, nor do they resemble the sensations which occasion them. In this way, Reid thinks that we are able to conceive of something which is completely different from a sensation, namely external bodies and their primary qualities.

What Reid takes issue with is the hypothesis in Hume's philosophy that the ideas or conceptions we have of external objects – like that of hardness – resemble the feelings or sensations from which they arise. Reid accepts only half of the theory that Hume put forward in the first section of Book 1 of the *Treatise of Human Nature*: he accepts the principle that ideas arise from sensations or impressions which cause them. However, he objects to Hume's other principle that ideas

resemble their corresponding impressions, and have the same basic properties as the impressions.

All Humean impressions, including sensations (the other kinds of impressions are passions and emotions), have the property of only existing while they are perceived (T 1.4.2.45–6; SBN 210–11). They are, as Reid put it at the beginning of his *Inquiry into the Human Mind*, only "fleeting and transitory beings, that . . . have no existence at all, any longer than I am conscious of them" (1997: 4). On Hume's theory that all ideas must resemble impressions, ideas must have these same properties, and so cannot legitimately represent anything like an external object or mind that is supposed to last over an extended time. However, according to Reid, the ideas which are occasioned by the sensations or impressions are entirely unlike them and so can represent what lasts – an enduring material object or mind.

This is only one example of the skeptical consequences that Hume draws from the theory of ideas, rightly understood. Throughout the *Treatise* Hume draws conclusions about limits to our knowledge and understanding in accordance with the principle that ideas are derived from impressions which they resemble.

Reid's interpretation of Hume's *Treatise* goes to the heart of Hume's metaphysics in a way that had not been done by the interpretations of earlier commentators on the book. Nevertheless, there are three ways in which Reid's analysis of Hume has proved to be very misleading. The first has to do with Reid's claim that Hume drew the skeptical implications of a theory of ideas which had been generally held by philosophers, and even more particularly by Locke and Berkeley. The second point, which is related to the first, has to do with Reid's claim that Hume merely assumed this theory and never argued for it. Thirdly, Reid misleadingly treated this theory as though it constituted the whole of Hume's metaphysics. Let us consider each of these three points.

First, we have already seen that Reid gives a quite misleading statement of the generality of the theory of ideas on which Hume's skepticism is founded. It is not a theory of ideas generally received by philosophers. When one goes beyond his polemic it becomes clear that Reid is concerned with a very specific theory of perception which maintains that ideas of external objects resemble the sensations which cause them (DeRose 1989). Moreover, Reid himself is committed to his own version of the theory of ideas, one which puts forward the hypothesis that the contents of our ideas are entirely different from that of the sensations which give naturally give rise to them. After reading a draft of the *Inquiry into the Human Mind* in manuscript two years before it was published, Hume wrote to their common friend Hugh Blair that as far as he could understand Reid's theory, it led back to "Innate Ideas" (Wood 1986: 416). This does not seem far off when one reflects on Reid's view that our conceptions of external objects arise from an "original principle" of our nature, and are entirely unlike sensations (Reid 1997: 60; cf. Wright 1987).

While Reid's more focused claim that Hume had drawn skeptical consequences from a theory of ideas adopted from Locke and Berkeley is more accurate from

both an historical and philosophical point of view, it still needs to be considered with a great deal of caution. Locke did not treat his thesis that all ideas are derived from sensation or reflection in the restrictive way that Hume was later to do, denying that we can have any conceptions of lasting external objects or objective power; this suggests that he had a different theory of ideas, rather than, as Reid implied, that Locke failed to see the consequences of his own theory that were later discovered by Hume. The latter claim makes for a tidy Whig account of progress in the history of philosophy, but does little justice to the intelligence of either philosopher. Further, while Hume's debt to Berkeley, as he himself acknowledged, is far greater than to Locke, it is still important to consider carefully whether there are not important differences in their basic theory of ideas which allow them to draw radically different philosophical conclusions about the limits of human understanding.

This brings me to the second point. It has become a commonplace among writers on Hume, following Reid, to claim that he simply assumes the theory of ideas of his predecessors and does not argue for it (Stroud 1977: 17). Hume denies this latter claim in his letter to Hugh Blair and, I think, with good reason:

> I think the Author [i.e. Reid] affirms that I had been hasty, & not supported by any Colour of Argument when I affirm, that all our Ideas are copy'd from Impressions. I have endeavourd to build that Principle on two Arguments. The first is . . . desiring any one to make a particular Detail of all his Ideas, where he woud always find that every Idea had a correspondent & preceding Impression. If no Exception can ever be found, the Principle must remain incontestable. The second is, that if you exclude any . . . particular Impression . . . as Colours to the blind, Sound to the Deaf, you also exclude the Ideas. (Wood 1986: 416)

While we can dismiss the second of Hume's arguments, since Reid can acknowledge it while still insisting that the ideas of external objects which arise from the sensations of sight and sound are dissimilar to them, the first is more to the point. In the body of the *Treatise* Hume gives detailed accounts of the origin of the ideas of space, time, and causal power, which not only give evidence for his general thesis, but more importantly clarify it. Indeed, I would argue that it is only when we turn to his account of the senses in section 2 of Part 4 of Book 1 ("Of scepticism with regard to the senses") that we can develop a clear understanding of the properties of impressions of sensation from which Hume thinks that our beliefs in external objects must be derived. Here he reveals that his arguments about the nature of our ideas rest on premises that are not obviously shared with his predecessors, including Berkeley (e.g. T 1.4.2.45; SBN 210–11).

Finally, and perhaps most importantly, Hume's account of the origin of our ideas provides only the first step in his own theory of mind. Throughout the *Treatise*, Hume's aim is to show how our commonsense judgments and beliefs arise from the combining and creative power of the imagination. He gives accounts of how, through the association of ideas and certain systematic "mistakes" that

imagination engenders we are led to form beliefs in absolute space and time (T 1.2.5.19–22; SBN 60–1), objective causal power (T 1.3.14.25; SBN 167), external bodies that go on existing independently of our perceptions (T 1.4.2.20–35; SBN 195–210), and a continuous self to which we ascribe lasting moral characteristics (T 1.4.6.16–20; SBN 260–2). He explains how through the association of ideas, feelings are transferred from impressions to ideas in order to form natural beliefs and moral judgments.[3] In his *Abstract* Hume claims:

> If any thing can entitle the author to so glorious a name as that of an *inventor*, 'tis the use he makes of the principle of the association of ideas, which enters into most of his philosophy . . . For as it is by means of thought only that any thing operates upon our passions, and as these are the only ties of our thoughts, they are really *to us* the cement of the universe, and all the operations of the mind must, in great measure, depend on them. (T A 35; SBN 661–2)

The great value of the *Treatise* over his two *Enquiries* is that it shows in detail how Hume attempted to build up a general theory of mind by postulating principles of the imagination and passions to account for the phenomena – the beliefs that we all naturally hold in our everyday lives. What Reid's description of Hume's philosophy generally disregards is his project of showing how, in spite of the limits of our ideas, we naturally form the beliefs of everyday life which overcome doubts engendered by reason and philosophy.

Hume's Repudiation of the *Treatise*

In 1775, the year before his death, Hume asked his publisher William Strahan to place an Advertisement, in which he disowned the *Treatise*, in front of the volume of his collected works containing his later philosophical writings.[4] In it he called his readers' attention to the fact that he had never acknowledged writing the *Treatise* (it had been published anonymously) and asked that only these later writings – including his two *Enquiries* – be regarded as representing "his philosophical sentiments and principles." The *Treatise*, he explained, had been "projected before he left College", and written and published "not long after."[5] It was unsuccessful and he had become aware that he had made a mistake "in going to the press too early." His later writings, in which "he cast the [*Treatise*] . . . anew," corrected "some negligences in his former reasoning and more in the expression."

In this Advertisement Hume complained of "several writers" who had criticized his philosophy by directing "all their batteries against that juvenile work." These critics were identified in a letter Hume sent to Strahan as Reid and James Beattie. Hume wrote to Strahan that the Advertisement was "a compleat answer to Dr. Reid and that bigotted silly Fellow Beattie" (Greig 1932, II: 301). We have already considered some of the reasons why Hume would consider the disowning of the *Treatise* in favor of the *Enquiry Concerning Human Understanding* to be an answer to Reid.

Both Reid and Beattie argued that Hume's philosophical principles led to Pyrrhonian skepticism, and they claimed to counter it through "common sense." But in the first *Enquiry* Hume had already explicitly adopted a mitigated skepticism which was based on "common sense" though, unlike his Scottish critics, he required that it be "methodized and corrected" through critical reflection (EHU 12.24–5).

Hume's public disavowal of the *Treatise* was motivated, at least in part, by the personal attack on him in Beattie's very popular book, *An Essay on the Nature and Immutability of Truth; in Opposition to Sophistry and Scepticism*, published in 1770. Its popularity may be measured by the fact that in 1774 Sir Joshua Reynolds painted a picture entitled *The Triumph of Truth* which celebrated Beattie's "victory" over the skeptics as represented by Hume and Voltaire (Mossner 1980: 597). Beattie's criticisms were largely polemical and, if anything, he was more "abusive," to use Hume's own word, than the first reviewers of the *Treatise*. For example, in the Introduction to the *Essay on . . . Truth*, he stated that since Hume "was not much acquainted with human nature" he was "not well qualified to write a treatise upon it" (Beattie 1983: 10). Beattie claimed that that, as a moral philosopher, Hume "exerted his utmost ingenuity in searching after paradoxes" and only landed on those which were "on the side of licentiousness and scepticism" (1983: 11). More substantially, he sees Hume as undermining common sense in his discussion of the source of the causal maxim in his *Treatise* (pp. 102ff.). Beattie claimed in opposition to Hume that the causal maxim was known through *intuition*, rather than experience, and so is as certain as any fundamental claim in mathematics.[6] It is intuitively known by young children before experience, as shown by the curiosity of some of them "at a time when their experience is very scanty" (pp. 108–9). Moreover, Beattie argued that by basing the causal maxim on observation and experience Hume had undermined "the most important argument that ever employed human reason . . . [namely] that which from the works that are created, evinces the eternal power and godhead of the creator" (p. 111). By questioning the certainty of the causal maxim, Hume had attacked the cosmological argument for the existence of God (pp. 112–13).

Conclusion

The largely polemical nature of Beattie's attack on Hume may suggest that he over-reacted to it by disowning the *Treatise* in his Advertisement. However, this would not account for the fact that he became dissatisfied with Books 1 and 2 of the *Treatise* even before the first reviews appeared and that, after he had written his *Enquiries*, he discouraged his friends and acquaintances from reading the *Treatise*. As early as 1751 Hume wrote to his friend Gilbert Elliot, advising him not to read the *Treatise* because the first *Enquiry* contains "everything of Consequence relating to the Understanding" and that "by shortening & simplifying the Questions" in the latter book he was really making them "much more complete" (Greig 1932, I: 158).

Moreover, as we have seen in his reply to the criticisms of John Stewart, Hume expressed the hope that he would avoid misunderstanding by having his philosophical principles interpreted through first *Enquiry* rather than the *Treatise*.

The charge is sometimes made that Hume gave up serious philosophy after writing the *Treatise* and became more concerned with literary success. This charge is based, at least in part, on a remark in *My Own Life* that "love for literary fame" was the "ruling Passion" of his life (Greig 1932, I: 7). It is true that in the philosophical works written after the *Treatise* Hume was very concerned to present his philosophical ideas in a form that would be accessible to his readers and would not be misinterpreted. The whole first section of his *Enquiry Concerning Human Understanding* examines the distinction between popular philosophical writing and what Hume calls the "accurate and abstruse" philosophy, which deals with subtle philosophical distinctions (EHU 1.3; SBN 6). While he acknowledges the difficulty of any philosophy such as his own that gives "an accurate scrutiny into the powers and faculties of human nature" (EHU 1.13; SBN 13) he argues that such a philosophy is necessary to counter the many abstruse philosophical books that provide shelter to "popular superstitions" (EHU 1.11; SBN 11). Still, he thinks that he can overcome some of the difficulties that dogged the *Treatise* by presenting his philosophy by using more "care and art" and avoiding the "unnecessary detail" that makes his first book hard to read (EHU 1.17; SBN 16). In any case, it is clear that the *Enquiry* is intended as a serious and accurate account of Hume's philosophical principles (Buckle 2001; Stewart 2002).

The question then arises whether we should disregard the *Treatise* as its author requested and accept his word that everything worth saving has been "cast . . . anew" in his later philosophical writings. This does not seem satisfactory. Some of the most intriguing discussions of the *Treatise* were never incorporated into the later writings, and yet have been important in subsequent philosophy. These include the sections of the *Treatise* on personal identity, on the immateriality of the soul and on the causal maxim "Whatever begins to exist, must have a cause of existence." Other key sections, including those on the causes of our belief in an external world, on abstract ideas, on our ideas of space and time, and on the principles of probabilistic reasoning are drastically cut in his later *Enquiry Concerning Human Understanding*. Even though Hume's views in these sections are highly controversial, they continue to form the basis for contemporary philosophical discussion. Further, in the *Enquiries*, Hume never discusses the overall project that he announces at the beginning of the *Treatise*, that of providing a new foundation for the sciences, based on a new science of human nature. And he cuts back on his discussion of the central metaphysical principle on which that science was supposed to be based, namely the principle of association of ideas and the related principle of association of impressions. At the same time those principles are presupposed in his later writing as well as the earlier, especially in his theory of belief, the indirect passions, sympathy, and his theory of moral judgement. Isn't it to the *Treatise* that we must look to find the fundamental nature of

the psychological assumptions on which his philosophical conclusions are founded? Finally, in his later *Enquiries*, he never discusses a major problem that he confronts head on in the Treatise: how one is to distinguish between the natural principles of the human mind which people *ought* to accept (i.e. the normative principles of knowledge and morals), and those that should be rejected as "unphilosophical."

So we cannot follow Hume's advice and disregard the *Treatise*. But what do we do with his claim that "negligences" in his "reasoning," as well as his "expression," are corrected in the later philosophical writings? Unfortunately, except for a few points mentioned in the Appendix that he added to Book 3, Hume never tells us what these negligences were. So, do we use his later philosophical writings such as his *Enquiry Concerning Human Understanding* to illuminate the *Treatise*? This seems like a sensible policy where the topics of the books overlap. But what of the doctrines only discussed in the earlier book, which he did not repeat in the later writings and which seem basic to the foundations of his philosophy?

Hume has bequeathed a difficult legacy to subsequent interpreters of his *Treatise*, one which, as every Hume scholar knows, has resulted in competing and conflicting interpretations of his most basic philosophical principles.

Notes

1 For a different interpretation of Hume's Conclusion to Book 1, see 9: HUME'S CONCLUSIONS IN "CONCLUSION OF THIS BOOK."
2 It is supposed to be a fundamental feature of what was later called "empiricism."
3 See, for example, T 1.3.7.5–7 and T 1.3.8.2; SBN 96–9 and T 2.1.11.3–8 and T 3.3.7–10; SBN 316–20 and SBN 575–8.
4 See the page preceding the text of his *Enquiry Concerning Human Understanding* in both modern editions.
5 In fact, the evidence of letters shows that Hume's writing of the *Treatise* began in 1729 when he was 18; it was published *13 years* after Hume left college at age 15. Hume is clearly deprecating his critics by pointing out that they were directing their criticisms at this piece of juvenilia!
6 The term "intuition" was extended by Beattie and other philosophers of the Common Sense school, including Reid, to include self-evident propositions which are not ascertained through the immediate comparison of ideas. In so doing they were breaking with the usage of the term as established by Descartes, and followed by Locke and Hume among many other early modern philosophers (Descartes 1985: 14–15; Locke 1975: 530–1; T 1.3.1.2, 1.3.3.1–3; SBN 70, 78–9; cf. Owen 1999: esp. 83–112).

References

Beattie, J. (1983) *An Essay on the Nature and Immutability of Truth; in Opposition to Sophistry and Scepticism*. New York: Garland Press. First published 1770, Edinburgh: Kincaid & Bell.

Buckle, S. (2001) *Hume's Enlightenment Tract: the Unity and Purpose of* An Enquiry concerning Human Understanding. Oxford: Clarendon Press.

DeRose, K. (1989) "Reid's anti-sensationalism and his realism," *Philosophical Review* 98, pp. 313–48.

Descartes, R. (1985) "Rules for the direction of the mind, Rule 3," in J. Cottingham, A. Kenny, D. Murdoch, and R. Stoothoff (trans. and eds.) *The Philosophical Writings of Descartes*, vol. 1. Cambridge: Cambridge University Press.

Fieser, J. (ed.) (1999) *Early Responses to Hume*, vol. 1: *Moral, Literary and Political Writings – Hume's Moral Philosophy*. Bristol: Thoemmes Press.

——(ed.) (2000) *Early Responses to Hume*, vol. 3: *Metaphysical and Epistemological Writings – Eighteenth-Century Responses*. Bristol: Thoemmes Press.

Green, T. H. (1968). "Hume and Locke." Reproduced from the general introduction to *David Hume: the Philosophical Works* [1874], 4 vols., vol. 1. New York: Thomas Y. Crowell.

Greig, J. Y. T. (1932) *The Letters of David Hume*, 2 vols. Oxford: Clarendon Press.

Hume, D. (1967) *A Letter from a Gentleman to his Friend in Edinburgh*, ed. E. C. Mossner and J. V. Price. Edinburgh: Edinburgh University Press.

Locke, J. (1975) *An Essay Concerning Human Understanding*, ed. P. H. Nidditch. Oxford: Clarendon Press.

Mossner, E. (1980) [1958] *The Life of David Hume*. Oxford: Clarendon Press.

Owen, D. (1999) *Hume's Reason*. Oxford: Oxford University Press.

Reid, T. (1997) *An Enquiry into the Human Mind on the Principles of Common Sense*, ed. D. R. Brookes. Edinburgh: Edinburgh University Press.

Ross, I. S. (1966) "Hutcheson on Hume's *Treatise*: an unnoticed letter," *Journal of the History of Philosophy* 4, pp. 69–72.

Stewart, J. (1754) "Some remarks on the laws of motion, and the inertia of matter," in D. Hume and A. Munro (eds.) *Essays and Observations, Physical and Literary: Read before a Society in Edinburgh and Published by Them*. Edinburgh: G. Hamilton & J. Balfour.

Stewart, M. A. (2002) "Two species of philosophy: the historical significance of the first *Enquiry*," in P. Millican (ed.) *Reading Hume on Human Understanding: Essays on the First Enquiry*. Oxford: Clarendon Press.

Stroud, B. (1977) *Hume*. London: Routledge.

Wood, P. B. (1986) "David Hume on Thomas Reid's *An Inquiry into the Human Mind on the Principles of Common Sense*: a new letter to Hugh Blair from July 1762," *Mind* 95, pp. 411–16.

Wright, J. P. (1987) "Hume vs. Reid on ideas: the new Hume letter," *Mind* 96, pp. 392–8.

Yolton, J. (1984) *Perceptual Acquaintance from Descartes to Reid*. Minneapolis: University of Minnesota Press.

Further reading

Emerson, R. (1994) "The 'Affair' at Edinburgh and the 'Project' at Glasgow," in M. A. Stewart and and J. P. Wright (eds.) *Hume and Hume's Connexions* (pp. 1–22). Edinburgh: Edinburgh University Press.

Harris, J. (2003) "Answering Bayle's question: religious belief in the moral philosophy of the Scottish Enlightenment," in D. Garber and S. Nadler (eds.) *Oxford Studies in Early Modern Philosophy* 1 (pp. 229–54). Oxford: Clarendon Press.

Jones, P. (ed.) (2005) *The Reception of David Hume in Europe*. Bristol: Thoemmes Press.

Penelhum, T. (1983) *God and Skepticism*. Dordrecht: Reidel.

Reid, T. (2002) *Essays on the Intellectual Powers of Man*, ed. D. R. Brookes. Edinburgh: Edinburgh University Press.

Russell, P. (2003) "Butler's 'Future State' and Hume's 'Guide to Life'," *Journal of the History of Philosophy* 42, pp. 425–48.

Sher, R. B. (1990) "Professors of virtue: the Edinburgh chair," in M. A. Stewart (ed.) *Studies in the History of the Scottish Enlightenment* (pp. 87–126). Oxford: Clarendon Press.

Stewart, M. A. (1985) "Hume and the 'metaphysical argument *a priori*'," in A. J. Holland (ed.) *Philosophy, its History and Historiography* (pp. 243–70). Dordrecht: Reidel.

——(1994) *The Kirk and the Infidel*. Lancaster: University of Lancaster Press. Corrected edn. 2001.

Wright, J. P. (2005) "The scientific reception of Hume's theory of causation: establishing the positivist interpretation in early nineteenth-century Scotland," in P. Jones (ed.) *The Reception of David Hume in Europe* (pp. 327–47). Bristol: Thoemmes Press.

Wright, J. P., R. Stecker, and G. Fuller (eds.) (2003) "Hume and his critics," in D. Hume, *A Treatise of Human Nature*, abridged and edited (pp. 349–61). London: J. M. Dent.

2

Hume's Other Writings

Wade L. Robison

Hume's ambitions are extensive. In introducing experimental reasoning into moral subjects, he says he intends to provide the basis for "a compleat system of the sciences," covering philosophy, ethics, politics, history, economics, religion – all the human mind can encompass (T Introduction 6; SBN xx). What we may call the traditional view is that he gave up this grand ambition after the *Treatise* fell "*dead-born from the Press*," as he described it in *My Own Life* (1987: xxxiv). Perhaps, the view continues, he saw no way to resurrect it to continue with its enterprise. Perhaps he could not answer the skepticism he raised in the *Treatise* which seems to undermine his system's foundations. Whatever his reasons, he turned to advance his admitted ambition for "literary fame," according to the traditional view, writing essays in politics and economics, for instance, and the six volumes of the *The History of England*. He became famous in his time, the view concedes, but not for his philosophy.

This paper's title seems to presuppose this traditional view. There is the *Treatise* and perhaps parts of the two *Enquiries* and the *Dialogues concerning Natural Religion*, all rich with philosophical problems, and then there are his other writings, well-written, insightful, but beyond the philosophical pale. Or so the story goes.

But the story is mistaken. There is a unity to Hume's writings, from his philosophical writings to his *Histories* and *Essays* and other works. They are unified in two different ways. First, they each use and illustrate the power of causal reasoning to provide us with understanding. Second, as a set they show that the human capacity to understand is (relatively) unlimited: every subject we examine we can come to understand. Hume's writings all further his original enterprise.

That enterprise has had such an enormous impact we are not well-positioned to understand it. What he did new is now so natural, so ingrained into the very way we think about economics and history, for instance, that it is hard for us to imagine how things could ever have been different. Because of Hume's success we will fail to appreciate it. We need thus to begin with some sense of what existed before.

Hume's enterprise is best contrasted with Descartes's. A natural way of reading Descartes gives power and focus to what Hume was doing. An excursion through the Cartesian conception of knowledge, and especially its limitations, will allow us to see how through all Hume's writings he means to provide a new understanding of our place in the world, a world which is of a piece, all of human experience being amenable to the same sort of reasoning he thinks was used so successfully by Newton and others to understand the physical world.

Grounds of Cartesian Doubt

We act, and through our actions, betray beliefs about the world and our place in it that are so deeply embedded in the way we live it would never occur to us they are beliefs at all. We take a breath before we swim, evidence of our belief that we cannot breathe in water. We brace ourselves to lift a pail of water, evidence of our belief that it is heavy. We lift our hand to wave hello, evidence of our belief that moving through air is effortless. But Descartes thinks such beliefs come from our mind having been encased in a body since birth.

In a striking image suggested by Pierre Daniel Huet (1725), our minds are trapped in our bodies, peering out through eyes covered time and again with eyelids as we blink. Encapsulated in skulls stuck on top of bodies, our eyes its apertures, our mind can catch only fleeting glimpses of what is outside. Our other senses give us no leverage. Our sense of touch extends only as far as our fingertips, small pads of contact with the world. Our ears capture only a narrow range of sounds, and our sense of smell is even more limited. Both our ideas and our reason are molded by this limited access to the world. Our "reason," Descartes says, "extends no further than [our] fingertips" (Cottingham et al. 1991, I: 87), and our beliefs? We come to beliefs "about which doubt is quite impossible," but which may be false (ibid., II: 12–13).

The phrase is not a gesture to an abstract possibility. Galileo's *The Starry Messenger*, published in 1609, gave visual proof that the earth and moon can move in tandem around the sun since Jupiter with its moons revolves around the sun (Robison 1974) and so puts into doubt our firmest beliefs, "quite impossible" to doubt, it would seem. For if we are on earth are moving around the sun as is Jupiter, we are moving through space at over 19 miles per second. Yet we sense no such movement. Since our first breaths, all of our relevant senses – the movement we do not feel, the sun and the moon and the stars rising, the rush of air we do not hear, our capacity to toss objects in the air and have them come straight down – have conspired to the belief that we are not moving through space. The belief is implicit in how we carry ourselves and the worries we do not have.

Once we begin to hunt for beliefs "about which doubt is quite impossible," but which we have because embodied, we find other striking examples. If we raise our hands, we feel no resistance and have no sensation we are pushing against

pounds and pounds of air. Yet the earth has an atmosphere, the atmosphere has weight, and that weight is upon all of us. We come to believe that air weighs nothing. Our bodies tell us so (see Cottingham et al. 1991, I: 218–19).

We are all plagued, Descartes argues, by beliefs reached "before we had the complete use of our reason," beliefs that make strange and unbelievable such truths as that we are on an earth spinning like a top while moving through space at a very rapid rate or that "when a vessel is full of gold or lead, it contains no more matter than when we think it is empty" (ibid.: 87). Descartes thinks such beliefs false despite being "quite impossible" to doubt. So, he argued, if even our deepest beliefs are false, we cannot trust without examination anything we believe about the world – about its physical properties, about our mental powers, about history, geography, morality, political systems, everything and anything. We cannot trust our experience and must begin anew about everything, on a foundation different from that which caused the beliefs we have come to have because we are encased in a body.

The Cartesian Vision

If our bodies prevent us from coming to knowledge, we must use our minds, somehow purified of their long association with our bodies. Descartes must believe that we can remove whatever contamination may have occurred and come to think as we would had we never been embodied. He must believe the mind is capable of knowledge in such a state. It is as though our minds are like God's in the capacity for knowledge. We need only achieve a Godlike epistemic state, what Descartes calls a clear and distinct perception. We need no more, and God needs no less, to know that mind and body are distinct substances, that bodies are extended, that there is no vacuum, and other truths about the world.

The example of light will help us understand the Cartesian vision – and its limitations. Descartes says that "light is simply the pressure of the material of [what he calls] the second element" so that our sensation of light is caused by "the motion by which . . . [many small balls] approach our eyes in a straight line" (Cottingham et al. 1991, I: 323). There is contact – as when we move a billiard ball by hitting it with a cue – and it is that contact of our eye balls with certain "small balls" that causes us to see light. An object we perceive as lit causes sensations of light just as moving the handle of a cane causes the tip to move: nothing is transmitted from one end to the other, and changes at the handle occur instantaneously at the tip (ibid.: 153).

This example lays out the essential elements of the Cartesian vision: by considering the relationships between objects, we can come to perceive, or intuit, truths about them. We may well wonder how Descartes knows that light is simply the pressure on our eyes of "small balls." We might think experimental evidence needed, but it is not clear Descartes thought so. He thinks no empirical evidence

is necessary for the fundamental principles of physics. He considers seven ideal cases in the *Principles* – two bodies of equal size colliding head on, two bodies of different size colliding head on, and so on. Of these he says, "These matters do not need proof since they are self-evident."

Or, as the French version has it,

> the demonstrations are so certain that even if our experience seemed to show the opposite, we should still be obliged to have more faith in our reason than in our senses. (Cottingham et al. 1991, I: 247)

Not only do we not need empirical evidence for them, but empirical evidence showing them false would be irrelevant: they are self-evident.

The appeal to the mind's Godlike capacity to discern self-evident propositions pervades Descartes's views. When pressed on whether light moves instantaneously, or takes time, Descartes says that light

> instantaneously arrives at the eye from the body which emits it, and . . . so certain [am] I of this that if it could be shown that it was false, I would be prepared to admit that I knew absolutely nothing in philosophy. (Cottingham et al. 1991, III: 46)

He adds, "If . . . a time-lag were detected, my philosophy . . . would be completely overturned" (ibid.: 46). He then goes on to provide a "demonstration" that he is correct.

The Limits of Descartes's Vision

Descartes was mistaken about light: it takes time for sunlight to reach our eyes. He was mistaken in much of his physics. But even had he been successful in basing physics upon a new indubitable foundation, he would have failed to account for much of human experience. In *The Passions of the Soul*, for instance, when he considers such psychological states as caring and love, anger and angst, Descartes cannot satisfy his criteria for success.

Consider joy. Descartes says it "is a pleasant emotion which the soul has when it enjoys a good which impressions in the brain represent to it as its own" (Cottingham et al. 1991, I: 360). Whatever the relations between the variables here – the emotion, the good, the impressions, the brain, the representation – we cannot understand how we can be joyful in the way we understand how "the slightest motion of [the tip of a quill] cannot but be transmitted simultaneously to the whole pen" (ibid.: 41). Light is pressure against our eye balls. But joy? Joy comes from the soul's "enjoying" a good it represents to itself. Besides using what he is explaining in his explanation, Descartes has forsaken mechanistic explanation and the kind of certainty he demands for knowledge.

The full story is long and complicated, but suffice it to say that we have a nest of problems even regarding those matters in physics that Descartes's system of knowledge seems specially suited to handle. When we turn to other matters the human mind seems capable of understanding we find Descartes forcing the field of inquiry into a form that ill suits the subject matter however well it may fit within the confines his method requires. What is missing from such propositions as that the House of Commons will become the dominant legislative body because it controls the purse strings – a proposition Hume considers (1987: 44) – are any but analogical references to bodies and motion. History cannot rise to the level of Cartesian certainty because it requires a reference to human intentions and aims and is concerned with individual events. The conditions Descartes imposes for knowledge, that is, limit its scope to a small portion of what we seem capable of understanding.

Adam and Hume's Attack

So the Cartesian enterprise fails, and in attacking it, Hume strikes at the heart of Descartes's assumption that we must rid ourselves of experiences gained through being embodied since birth. Consider Adam. Adam represents the ideal Cartesian – bereft of any beliefs gained by that long association with the body which causes us, as Descartes claims, to reason with our "fingertips, and . . . suppose that there is nothing in the world except what [we] touch" (Cottingham et al. 1991, I: 87). Adam is created with a mature mind and without any experience that could corrupt that mind or its beliefs. Adam had no formative years and so has no need of meditations to subject his views to deep skepticism and clear his mind of "the prejudices of . . . childhood." His mind is already clear. He fits exactly Descartes's criteria for the perfect knower, someone who can discern the truth without having to penetrate the haze of beliefs produced by long habituation in a body. But, as Hume says in the *Abstract* about one billiard ball about to strike another, "*Adam*, created in the full vigor of understanding . . . would never to able to infer motion in the second ball from the motion and impulse of the first" (A 11; SBN 650).

In providing Adam as the perfect Cartesian, Hume means to show that knowledge of the world is not possible on Cartesian grounds. Thus, he says in the *Enquiry concerning Human Understanding*,

> Adam, though his rational faculties be supposed, from the first, entirely perfect [as Descartes would suppose], could not have inferred from the fluidity and transparency of water that it would suffocate him, or from the light and warmth of fire that it would consume him. No object ever discovers, by the qualities which appear to the senses, either the causes which produced it, or the effects which will arise from it; nor can our reason, unassisted by experience, ever draw any inference concerning real existence and matter of fact. (EHU 1.4.6; SBN 27)

We need experience to make causal inferences. It is only because we have the experience that Descartes thinks taints our reasoning that we can draw any causal inferences at all.

Causal relations cannot be the object of clear and distinct perceptions. Among their other problems, there is no simultaneity of cause and effect as in Descartes's quill example. God keeps the world going from instant to instant by preserving it again and again, Descartes claims, but nothing can guarantee that God will not just refuse to preserve the world some time. It is not necessary that the world continue to exist, that every slice of time be followed by another. Thus neither Adam nor we can intuit, upon perceiving one billiard ball about to hit a second, that the second will even continue to exist, let alone move when hit. Relying on his "strong natural reason" alone, Adam would be as puzzled about what will happen as he would be on Mother's Day.

Hume has found in Adam an exemplar of an individual who, if any human being does, ought to have clear and distinct ideas of causal connections. Adam began his life exemplifying the clarity of mind Descartes's *Meditations* are meant to achieve, but he can make no causal inferences. Hume has turned Descartes's views on their head. It is only through our long association with our body that we can make causal inferences at all. Custom and habit, Hume says, not reason, animate our causal inferences.

The Science of Human Nature

What Hume provides in place of the Cartesian conception of reason is a conception that makes it possible, about any inference beyond our immediate experience concerning real existence and matter of fact, as he would put it, for "any thing [to] produce any thing" (T 1.3.15.1; SBN 173). It is thus possible for there to be causal maxims for all of human experience, everything, that is, regarding real existence and matter of fact – from what is true of the passions of the soul, which are not extended and thus not subject to motion, to what is true regarding anything the human mind can encompass – from history to economics, from politics to passions.

The principles of association Hume articulates in the *Treatise* and *Enquiry* are causal laws that regulate the order of our perceptions. As he begins his long inquiry into the nature of the causal relationship and of causal inference in Book I of the *Treatise*, he is explicit that the passions are subject to causal laws and to the same kind of causal laws that operate in the physical world. He says that

> tho' the ideas of cause and effect be deriv'd from the impressions of reflection as well as from those of sensation, yet for brevity's sake, I commonly mention only the latter as the origin of these ideas; tho' I desire that whatever I say of them may also extend to the former. Passions are connected with their objects and with

one another; no less than external bodies are connected together. The same relation, then, of cause and effect, which belongs to one, must be common to all of them. (T 1.3.2.16; SBN 78)

What a different world from that of Descartes! A single kind of causal relation permeates Hume's world, and his paradigm is not a causal relation between two simultaneous events, connected mechanically, as in Descartes's example of the quill, but a causal law between what can be very disparate kinds of things. It is a law that could equally well be made about relations among the passions, or among them and their objects, or among nations, or among citizens – among anything at all in the world in which there is order and regularity. For the crucial feature that ought to trigger a hunt for causal laws is what triggers Hume's hunt for causal laws regarding the passions. He says that "however changeable our thoughts may be, they are not entirely without rule and method in their changes" (T 2.1.4.2; SBN 283). The mark by which we may find causal laws is "rule and method" in how something changes.

Everywhere in human experience there is "rule and method," and everywhere human experience is informed by causal maxims. Hume has displaced the Cartesian conception of reason with a conception that permits the mind to discover causal truth wherever we find causal relations, that is, everywhere about anything in the world. We are not limited to discovering truths such as 2 + 2 = 4 or "I think, therefore I am" – typical Cartesian examples.

One objection to this understanding of Hume's enterprise comes from Hume's *Treatise* itself. There he seems to raise skeptical doubts about our capacity ever to know anything about the world. We make causal inferences all the time, and we can and do correct those that lead us astray. We are capable, that is, of rising above what our long association with our body has make it "natural" for us to infer, and so are capable of critiquing the inferences we make that motivated Descartes to jettison experience as a source for knowledge. But we are still inferring that something will come into existence when God, as Descartes would put it, could just decide not to sustain the world in existence. So every causal inference is at risk of catastrophic failure, and it is thus difficult to understand how we can claim to know that one event will ever occur upon the occurrence of another.

However Hume scholars settle this matter, and it remains unsettled, one criterion for their success is that Hume went on to write about, and help create, new fields of intellectual inquiry. He is known as one of the founders of modern psychology, as the first modern historian, as one of the founders of modern economics, as a founder of political science, having proposed that there are laws in politics as certain as those in physics. Whatever his intentions may have been, and however consistent he may have been with what he called his "juvenile" work, the *Treatise*, he went on to sketch, at least, "a compleat system of the sciences," unified by a single understanding of how we humans can come to have knowledge. He opened up the entire world to inquiry.

Hume's "other writings" are all illustrative of this enterprise – not necessarily with the kind of thoroughness we might expect from someone investigating economic transactions in detail, for instance, like Adam Smith, but with the kind of broad brush we would expect from someone trying to create a new order in the intellectual world, showing others how to proceed in every area the human mind may venture. Hume's other writings are all of a piece, that is, each illustrating how causal reasoning can provide us with knowledge – what Hume calls "a proof" – within any field of intellectual endeavor.

If we turn to those other writings, we will find Hume engaged in causal analysis. In some cases he is trying to identify the causes of some phenomena as when he argues that a basic cause of the English Civil War in the 1600s was the instability of the structure of government. Sometimes he is trying to remove a causal misconception as when he argues that restricting free trade will not increase a country's wealth. The analyses are always causal, always presuming that whatever occurs has causes, that we can only understand an effect by understanding what causes it, and that changing the cause will change the effect. Those analyses thus presume what may strike some Hume scholars as particularly contentious:

1   We can distinguish between causal beliefs in terms of which are firmly grounded and which are not.
2   We can take a critical stance towards our causal beliefs, however natural and inevitable they may seem to us, and change them.

The belief that a country best maintains its resources by restricting trade is an example for both (1) and (2).

The reason why some Hume scholars may find (1) particularly contentious is that it seems to run counter to the skepticism that some think permeates Hume's early philosophy. But if (1) is true, then it would seem that (2) is true as well: how could we distinguish between causal beliefs if we did not take a critical stance towards them?

I will not argue for these presumptions here. I am laying out a way of looking at all of Hume's writings that allows us to see them as of a piece, all furthering the single end of showing that our world is amenable to reason. We can at best illustrate this view of Hume I am hawking by examining several of Hume's essays. The full proof rests in part on a detailed examination of each of Hume's later works.

Hume's Other Writings: Political Science

In his essay "That politics may be reduced to a science," Hume considers the question "whether every form [of government] may not become good or bad, according as it is well or ill administered." He rejects "this sentiment," arguing that

So great is the force of laws, and of particular forms of government, and so little dependence have they on the humours and tempers of men, that consequences almost as general and certain may sometimes be deduced from them, as any which the mathematical sciences afford us. (1987: 16)

As an example of how forms of association make a difference, he cites the Venetian Republic and Poland. Because each prince in Poland, "by means of his fiefs, has a distinct hereditary authority over his vassals . . . the whole body has no authority but what it receives from the concurrence of its parts." In Venice, on the contrary, no nobleman can have an interest independent of the interest of the whole and so "no nobleman has any authority which he receives not from the whole." One consequence for Venice is that the nobles "will preserve peace and order, both among themselves, and their subjects" (1987: 17). Each nobleman's power depends upon the power of the whole. The tide lifts and carries or beaches all boats together, and so each nobleman has an interest in ensuring "the interests of the whole body," including his own.

In Poland, by contrast, each nobleman had an interest in preserving his fiefdom, and so he may, in any situation, judge that interest to be independent of his interest in preserving the state. Each nobleman must concur with all the others for common action to ensue, and, perhaps more important, each nobleman must consider, whenever another is attacked, whether it is in his interest to come to the other's aid. That the answer was not obviously always in the affirmative is provided by the continual dismemberment of the Polish state throughout its history.

So as, Hume says, "to prove more fully, that politics admit of general truths, which are invariable by the humour or education either of subject or sovereign," he goes on to "observe some other principles of this science" – such as "that, though free governments have been commonly the most happy for those who partake of their freedom; yet are they the most ruinous and oppressive to their provinces" (1987: 18–19). We need not determine whether such "general truths" as these, as Hume calls them, are in fact true in order to appreciate what Hume is about.

He means to show that there are causal laws regarding politics, that political causes, so to speak, such as a form of government, have effects as invariable as those in the "mathematical sciences." He thus says,

Legislators . . . ought not to trust the future government of a state entirely to chance, but ought to provide a system of laws to regulate the administration of public affairs to the latest posterity. (1987: 24)

"Effects," he says, "will always correspond to causes" (1987: 24). And he then asks, rhetorically,

Can we ascribe the stability and wisdom of the VENETIAN government, through so many ages, to any thing but the form of government? And is it not easy to point out

those defects in the original constitution, which produced the tumultuous govern-ments of ATHENS and ROME, and ended at last in the ruin of these two famous republics? (1987: 24)

When we examine our political lives, the forms of association we have with one another as citizens, we find, Hume says, causal laws about the effects of the various ways we can associate with one another. We cannot change any causal truths: the Polish form of association will have its effects as much as gravity does. We also cannot change our having some form of political association. That we do is not an historical accident, but a natural consequence of humans living together, traced "plainly in the nature of man, and in the equality, or something approaching equality, which we find in all the individuals of that species" (1987: 468). But though it is in some sense part of what it is to be human to have a body subject to gravity, and part of what it is to be human to have some form of political asso-ciation, it is no part of what it is to be human to have any particular form of political association.

That we find ourselves, upon reaching the age of reason, living within some particular form of political association means that though it is as normative for us as is our native tongue, we can assess it. Indeed, what marks the difference between this realm of experience we are investigating and that of the mathematical sciences and psychology is that we find ourselves within some convention, a natural social artifact. To this we have the internal point of view and we can, upon reaching the age of reason, take towards that convention the external point of view, subjecting the convention to criticism; and where its ends – instability, for instance – are not what we want, we can work to change the causes that produce it. A political system's being normative for us does not mean that we cannot rise above it, as it were, and take towards it a critical, i.e., a philosophical, point of view. It is just this point about the role of philosophers that Hume is making when he remarks, when speaking of "philosophers who have embraced a party," that that may be "a contradiction in terms." Philosophers who embrace a party have subsumed their critical faculties into a normative system and are unable, then, to assess it or its implications critically. They need to keep aloof from ideologies so as to be able to take a critical stance.

Hume's Other Writings: Economics

We can find this switch in points of view at play in many of Hume's essays and other writings. Hume begins "Of the balance of trade," for instance, by remarking that "nations ignorant of the nature of commerce . . . prohibit the exportation of commodities, and . . . preserve among themselves whatever they think valuable and useful" (1987: 308). They intend to increase the amount of the commodity among themselves, but in prohibiting its export, they act directly contrary to that

interest, not realizing that "the more is exported of any commodity, the more will be raised at home, of which they themselves will always have the first offer" (ibid.).

Nations have the same worry about balances of trade, concerned that their money will all go abroad when they purchase from other countries which do not return the favor. But money between nations is like water: "wherever it communicates, [it] remains always at a level" (1987: 312). So if the money in Great Britain were suddenly diminished by four-fifths, the resulting poverty would depress wages and prices, and merchants from other nations would rush in, like water through a broken dam, until the influx of foreign money buoyed up the British economy to a state of equality with those of its foreign trading partners. Britain would then lose its financial advantage, and trade would equalize as well.

Hume is here again appealing to cause and effect. He thus says that "the same causes, which would correct these exorbitant inequalities [brought about by Britain's loss of four-fifths of its money], were they to happen miraculously, must prevent their happening in the common course of nature" (ibid.).

Similarly, we need not worry that trading with our neighbors and rivals will be "at our expense" so that as they gain, we lose (1987: 328). It is rather that "the encrease of riches and commerce in any one nation, instead of hurting, commonly promotes the riches and commerce of all its neighbours" (ibid.). "In this respect," Hume argues,

> states are in the same condition as individuals. A single man can scarcely be industrious, where all his fellow citizens are idle. The riches of the several members of a community contribute to encrease my riches, whatever profession I may follow. They consume the produce of my industry, and afford me the produce of theirs in return. (1987: 329)

In short, our economic relations, both with one another and nation to nation, are as much governed by causal laws as are our political associations. We will have some sort of relation of trade with those nations with which we engage in commerce, but we can, on reflection, change the ways in which we engage in trade so as to take advantage of the good effects of the causal maxims at work.

I suspect that the normative character of the particular form of trade one is in is less striking than it is, say, in regard to one's form of political association. Conventions differ in their power, but we should not ignore what normative character, whether slight or all-enveloping, a particular convention has. Otherwise we shall be unprepared, when we move for change, to cope with those who argue that the present order of things is "natural," i.e., the way things ought to be.

In any event, that we can take both the internal and the external points of view regarding our experience in such matters is a mark of the fundamental difference Hume uncovered in coming to understand how we can come to a scientific understanding of all of human experience, i.e., an understanding of the causal maxims

that mark all of what we experience. We can change a particular convention explicitly, as Hume suggests that nations ought to do regarding those goods they hold dear, but they also can change, evolve "naturally," whether we explicitly try to change them or not.

For example, in "Whether the British Government inclines more to absolute monarchy, or to a republic," Hume says that there as been "a sudden and sensible change in the opinions of men within these last fifty years." He says,

> The mere name of king commands little respect; and to talk of a king as GOD's viceregent on earth, or to give him any of those magnificent titles which formerly dazzled mankind, would but excite laughter in every one. (1987: 51)

In the seventeenth century, the forms of political association arguably demanded that the king be conceived as God's viceregent on earth. This claim is of the same form as the claim that the forms of political association in Venice and Poland had different effects. But the form of association had changed by the time Hume wrote his essay: the evolutionary result of an explicit agreement to change, namely, the Revolution of 1690. So the causal maxims that marked the king's relation to his subjects would have changed as well. So just as it would have been discourteous in the seventeeth century not to add to an introduction of the king "those magnificent titles," it later became discourteous to add them, since reciting them would only evoke laughter. We have here both an appeal to some underlying causal law about what is courteous and what is not, a law about social conventions, that is, and a recognition that conventions change and, with those changes, the forms of life we share change as well.

Critical Reflections

Examples will hardly suffice to prove that Hume's works are all of a piece. What is required is a sustained and thorough analysis of each of Hume's other writings, showing how he makes use of causal analysis to introduce into the world a way of understanding our experience that allowed for its full scope to be an object of critical reflection. But examples have their place. A reader cannot help but be struck by how often Hume begins his essays, for instance, by turning to the relevant causal relations. A striking example is from "Of the rise and progress of the arts and sciences," where Hume begins by saying, "Nothing requires greater nicety, in our enquiries concerning human affairs, than to distinguish exactly what is owing to *chance*, and what proceeds from *causes*" (1987: 110).

But we find the same appeal to causal factors when we turn to "Of public credit," for instance, where Hume explains how "the practice . . . of contracting debt will almost infallibly be abused, in every government" since, as he puts it,

> It is very tempting to a minister to employ such an expedient, as enables him to make a great figure during his administration, without overburdening the people with taxes, or exciting any immediate clamours against himself. (1987: 352)

Hume's long essay "Of the populousness of ancient nations" is to be read not only as an essay on the population of ancient nations, but as a treatise on how we are to make critical causal inferences about the past, when the evidence is patchy and we have only records by which to verify the inferences we make from the records we have. This is not just an historical exercise, by Hume the historian, but a philosophical one about the critical use of causal reasoning in regard to a subject of particular concern to historians.

We find this appeal to causal maxims, and an appeal as well to how conventions can change, in Hume's *History of England*. They are informed by causal laws. Indeed, it is difficult to find the treatment of any subject there which does not appeal to causal maxims. In recounting the early history of England, for example, Hume notes that though it is unclear "what species of civil government the Romans on their departure had left among the Britons," it is probable that "the great men in the different districts . . . lived in a great measure independant of each other" (1983, I: 14). The subsequent inability to form a common front against common enemies – along the lines of the Polish model of political association – was in no small measure, Hume argues, a cause in Briton's inability to resist the barbarian invaders who plagued them after the Romans left. It is Hume's use of such causal maxims to frame his understanding of English history that creates the objectivity one finds in that history, making Hume the first of the so-called "philosophical historians."

Critical reflection upon the operative causal maxims is particularly striking in Hume's explanation of the civil wars of the seventeenth century. Here it is clear that he thought the particular form of political association with which England began that century was unstable and that the revolution of Cromwell, if we may call it that, accomplished, in the end, what could have been understood to be needed had philosophers subjected that form of political association to critical scrutiny rather than adopt the tenets of one party or the other.

We can pile example upon example, but the story I am telling sets a course of work, the results of which will be at odds with much of the conventional wisdom about Hume. The whole of Hume's work after the *Treatise* is a series of extended causal analyses. The *Dialogues concerning National Religion*, for instance, can be read as an extended treatment of the criteria for causal inferences: what is permitted and what is not. Indeed, such analyses form the bulk of the *Treatise*. Consider the "experiments'" to "place this system [of the passions] beyond doubt both with regard to love and hatred, pride and humility'" (T 2.2.2.1; SBN 332). What unifies Hume's works is a demonstration, as it were, of how human understanding is possible, through causal analysis, over whatever is in the world: of "a compleat system of the sciences, built on a foundation almost entirely new, and

the only one upon which they can stand with any security" (T Introduction.4; SBN xvi).

The claim is not casual, and it is not a rhetorical flourish. It signals Hume's enterprise of displacing the Cartesian conception of the world and of how we come to understand it and ourselves in it. Hume's aim is ambitious, nothing less than to replace Descartes's complete system with his own and provide a new understanding of our place in the world, one founded on experience. Hume's other writings both secured his literary fame and showed how we can understand the world in a way radically different from that inherited from Descartes.

References

Cottingham, J., A. Kenny, D. Murdoch, and R. Stoothoff (eds.) (1991) [1985] *The Philosophical Writings of Descartes*, 4 vols. Cambridge: Cambridge University Press.

Huet, Peter (1725) *Philosophical Treatise Concerning the Weakness of Human Understanding*. London: Printed for Gysert Dommer.

Hume, David (1970) *Dialogues Concerning Natural Religion*, ed. Nelson Pike. Indianapolis: Bobbs-Merrill.

——(1983) *The History of England, in Six Volumes.* Indianapolis: Liberty Fund.

——(1987) *Essays: Moral, Political and Literary*, ed. Eugene F. Miller, rev. edn. Indianapolis: Liberty Fund.

Robison, Wade (1974) "Galileo on the moons of Jupiter," *Annals of Science* 31, pp. 165–9.

Further reading

Clarke, Desmond M. (1982) *Descartes' Philosophy of Science*. University Park, PA: Pennsylvania State University Press.

Descartes, René (1972) *Treatise of Man*, ed. Thomas Steele Hall. Cambridge, MA: Harvard University Press.

——(1979) *Le Monde*, trans. Michael Sean Mahoney. New York: Abaris Books.

——(1983) *Principles of Philosophy*, trans. Valentine Rodger Miller and Reese P. Miller. Dordrecht: Reidel.

Garber, Daniel (1992) *Descartes' Metaphysical Physics*. Chicago: University of Chicago Press.

Gaukroger, Stephen (1989) *Cartesian Logic: an Essay on Descartes's Conception of Inference*. Oxford: Clarendon Press.

——(1995) *Descartes: an Intellectual Biography*. Oxford: Clarendon Press.

Hall, Marie Boas (1965) *Robert Boyle on Natural Philosophy*. Bloomington: Indiana University Press.

Part II
The Understanding

3

Impressions and Ideas

Janet Broughton

Hume aimed to produce a "science of MAN" (T Introduction 4; SBN xv), a set of principles of human nature that would explain a wide variety of the phenomena of human life. The most basic principles that concerned him were causal principles that would explain features of our mental life, though he also thought that these would in turn help to explain some general aspects of our behavior that depend upon how our minds work. The most basic causal principles of human nature would be generalizations about types of mental states; they would describe the mental processes by which we come to have our thoughts and feelings.

Hume's term for our mental states is "perceptions": they are the states we are in as something "strike[s] upon the mind" and enters "our thought or consciousness" (T 1.1.1.1; SBN 1). So "perceptions" are conscious states. They are like and unlike one another in various ways, and some of these respects in which they resemble one another play crucial roles in Hume's general explanatory enterprise. He says at the outset that some of our perceptions are alike in entering the mind with great "force and violence" and that others are by contrast "faint"; he calls the lively and forceful perceptions "impressions" and the fainter ones "ideas" (ibid.). Impressions are our "sensations, passions and emotions, as they make their first appearance in the soul"; ideas are the "faint images of these in thinking and reasoning" (ibid.). Even though an impression and an idea will have quite different degrees of liveliness, they may resemble one another closely in a different respect: they may have the same content. For example, my visual sensation of a cat may have exactly the same content as my memory or daydream of a cat.

Hume does not try to explain everything about our perceptions. He completely disavows any ambition to explain our sense-impressions: they arise in the mind "from the constitution of the body, from the animal spirits, or from the application of objects to the external organs" (T 2.1.1.1; SBN 276). If anyone can explain something about them, it will be the "anatomists and natural philosophers" (T 1.1.2.1; SBN 8), though when Hume remarks that our sense-impressions arise

from "unknown causes" (T 1.1.2.1; SBN 7), he is perhaps suggesting that while the anatomists and natural philosophers might learn a great deal about our sense organs, nerves, and brains, this would not amount to a full explanation of why or how these physical things give rise to mental states. Despite this limitation, Hume's explanatory ambitions are large: by starting with the fact that we have sense-impressions, and that they typically occur in distinctive patterns within our experience, he thinks he can explain central features of our mental life.

Many aspects of Hume's theory of impressions and ideas come up for discussion in other essays in this volume. In this essay I will sketch a series of issues about impressions and ideas that cut across several areas of concern to Hume and to his readers. First I will raise some questions about the general distinction between impressions and ideas. Then I will lay out the general taxonomy of impressions, including impressions of reflection, and examine the division of our ideas into ideas of memory and imagination, raising some questions about Hume's treatment of memory and about his principle of the association of ideas. I will go on to discuss several issues connected with Hume's "copy principle" and explore his distinction between perceptions that are simple and those that are complex. I will conclude by examining his account of how general words are possible.

Impressions and Ideas

Hume distinguishes between our sensations and feelings, on the one hand, and our thoughts, on the other; he treats this distinction as a commonplace. He also distinguishes between perceptions that are very lively and those that are less lively, and he makes the innovative claim that this distinction lines up with the commonplace one. Our impressions are sensations and feelings, and they are very lively perceptions. Our ideas are thoughts, and they are less lively perceptions.

What exactly Hume means by liveliness is hard to pin down. He introduces the notion of liveliness by speaking of it as a way in which perceptions make their entry into the mind: the greatest liveliness attends our impressions because they elbow their way into our minds with the "most force and violence" (T 1.1.1.1; SBN 1). This suggests that when he speaks of impressions as having great liveliness (force, violence, vivacity, and the like), Hume is, at least in part, making the familiar philosophical point that our minds are receptive in affording us sensations and feelings. Because Hume is concerned to claim that in respect of liveliness, our thoughts differ from our sensations and feelings only by degree, he may also be implicitly suggesting from the start that we are in some way, or to some extent, passive in relation to our ideas as well as our impressions. This is a line of thought he explicitly develops later, in offering principles of the mind that explain both why various ideas come into our minds when they do and why we hold the beliefs that we have. For example, his account of causal inference represents our basic causal beliefs as the product of custom or habit. Indeed, Hume describes belief

itself as "a more vivid and intense conception of an idea, proceeding from its relation to a present impression" (T 1.3.8.11; SBN 103), and he adds that "[w]hen I am convinc'd of any principle, 'tis only an idea, which strikes more strongly upon me" (T 1.3.8.12; SBN 103).

He often seems to suggest that the varying degrees of liveliness in our perceptions are a felt aspect of our experience that we will all immediately recognize once he points it out to us. And when he remarks, for example, that some of our impressions are "so faint and low that we cannot distinguish them from our ideas" (T 1.1.1.1; SBN 2), liveliness does indeed seem to be something we readily recognize: here, the attentiveness with which we are aware of things. If I am deeply engaged by the sight of egrets in my back yard, for example, I may hear my cat meow and pay the sound so little attention that when the birds fly away, I wonder whether I really heard the cat or only imagined hearing him. But if attentiveness is part of what Hume means by liveliness, it had better not be the whole of what he means, because many ideas engage my attention much *more* than impressions do. The ideas excited in my mind by my reading a novel, for example, may engage my attention much more closely than the cat's meow does.

In still other places, Hume seems to identify liveliness with the capacity of a perception to influence our passions and actions. He claims that if an idea is a belief, it has the same influence on us that our impressions have, and "[t]his effect it can only have by making an idea approach an impression in force and vivacity" (T 1.3.10.3; SBN 119). This sounds as though Hume intends for the varying degrees of liveliness of perceptions to explain the varying effects of our impressions and ideas. Intuitively, a greater force can produce more changes than a lesser force, and in places Hume seems to think that there is a natural explanatory link between degrees of liveliness and degrees of influence. In a late addition to the *Treatise*, however, he appears to retreat from such a view and instead to say that "liveliness" is just his name for *whatever it is* that "renders realities more present to us than fictions, causes them to weigh more in the thought, and gives them a superior influence on the passions and imagination" (T 1.3.7.7; SBN 629).

It is by no means clear that all these ways of thinking about liveliness can be combined into a single, satisfying view. Although I barely pay attention to the cat's meow, my mind is receptive in relation to the sound. I may be giving full attention to an unfolding fantasy about a terrorist cell in my local library without being in the slightest moved to call the FBI. Hume clearly wants to be able to interlock various parts of his theory of the mind through the notion of liveliness so that he can, for example, forge a connection between perceptual judgment and causal belief; he also wants the intrinsic character of this feature of our perceptions to do explanatory work. But he has difficulty in making plausible his claim that a single feature of our perceptions, one that we all easily recognize, is at work in this wide range of mental phenomena, or that its character helps to explain why we act and feel as we do.

Hume's classification of dreams as ideas causes trouble of an additional kind. It appears he should classify dreams as impressions: I seem to be receptive in having

many of my dreams; many of them fill the horizon of my attention; and a night-mare, say, can provoke behavior (for example, a scream) and feelings (for example, terror). What seems to prevent Hume from classifying dreams as impressions is the fact that our dreams are not *sensations*, and this raises a broad question about how Hume sees the connection between the commonplace terms in which we may talk about our mental states and the austere terms in which he develops his theory of the mind.

According to the commonplace use of "sensation" – one that Hume himself often seems happy to adopt – sensations are experiences that acquaint us with objects in the world around us through the use of our senses. Does Hume think that sensations, understood in that way, *necessarily* have more force and liveliness than, say, memories or imaginings? This seems implausible, especially given his own concession that sometimes "our impressions are so faint and low, that we cannot distinguish them from our ideas" (T 1.1.1.1; SBN 2). This concession seems to indicate that Hume does not think there is a constant correlation between being a sensation and being a very lively perception, much less a necessary connection. A claim of necessary connection would also be hard for Hume to defend, because on his view of necessary connection, if such a claim were true, we would be unable to so much as conceive of a very faint sensation (see, for example, T 1.3.3.3; SBN 79).

On the other hand, if he thinks that being a sensation and having great liveliness are necessarily connected, that would explain why he insists that dreams are ideas and allows only that they "may *approach* to our impressions" (T 1.1.1.1; SBN 2; emphasis added) in their liveliness. It would also help to explain the conclusion he draws from the discovery that all simple faint mental states are derived from resembling lively mental states. He concludes from this that all our ideas are derived from sensation and feeling, and treats the discovery as thus settling the debate about "*innate ideas*, or whether all ideas be derived from sensation and reflexion" (T 1.1.1.12; SBN 7). The discovery about faint and lively mental states would not show that all our ideas come from our sensations and feelings – the point at stake in the debate – unless Hume were prepared to say that being a lively mental state is tightly connected with being a sensation or feeling.

Original and Secondary Impressions

Throughout much of the *Treatise*, Hume's aim is to explain our ideas, but he also believes that he can explain some of our impressions. He distinguishes between two kinds of impressions: original impressions, or impressions of sense, and secondary impressions, or impressions of reflection, which include our passions, desires, and emotions (T 2.1.1.1; SBN 276). Although he does not aim to explain our original impressions, he does think he can find principles of the mind that will explain our secondary impressions.

Original impressions include the impressions conveyed by our five senses, of course, and they also clearly include bodily pains and pleasures. Secondary impressions are unlike original impressions because they are caused in us by original impressions or by ideas copied from original impressions. For example, suppose a dog gives me a painful bite (original impression). When I next see that dog I may remember my prior painful experience (memory-idea copied from an original impression), and that will give rise to fear (secondary impression). My fear is an impression because it is a lively perception, and also perhaps because it is a perception that introduces a distinctive and simple content into my mental life. But it is a secondary impression, because its causes are other states of my mind: it is secondary to, or causally dependent upon, other perceptions.

"Direct passions" are secondary impressions that proceed immediately from the "pain and pleasure" we experience when we are presented with "some good or evil"; these include grief and joy, fear and hope, aversion and desire, and volition (T 2.3.9.1, 2; SBN 438). What the direct passions are "about" or "of" is the good or evil object that causes them; for example, my fear is fear of the dog's painful bite, and the dog's painful bite was the cause of my fear.

"Indirect" passions also proceed from the pain or pleasure we experience when we are presented with a good or evil thing, but they depend in addition upon "the conjunction of other qualities" (T 2.1.1.4; SBN 276). The indirect passions include pride and humility, and love and hatred. These arise from our experience of pleasure (or pain) along with our recognition of the pleasurable, good object (or the painful, evil object) as conjoined either with ourselves or with someone else. Where we recognize ourselves as conjoined with the object, pride (or humility) results; where we recognize someone else as conjoined with the object, love (or hatred) results. For example, an act of generosity, or a beautiful scarf, may cause me pleasure, but it will cause me pride only if I recognize it as mine. The cause of the passion of pride will be the act or the scarf; its object will be myself.

Hume's broad taxonomy of impressions categorizes some aspects of our experience unclearly. For example, in places Hume claims that we experience different pleasures with highly distinctive characters when we see someone beautiful (T 2.1.5.1; SBN 285) or hear a "good composition of music" (T 3.1.2.4; SBN 472), and he treats these as original impressions. But it is hard to see how they could be original impressions in the official sense, because they do not seem to be experiences that arise directly "from the constitution of the body, from the animal spirits, or from the application of objects to the external organs" (T 2.1.1.1; SBN 276). Complications also arise in Hume's treatment of the direct passions of desire, aversion, and volition: to explain these, he finds himself appealing not just to "pain and pleasure" but also to other experiences: "natural impulses" that "produce good and evil" (T 2.3.9.8; SBN 439). These include bodily appetites, like hunger and lust, and the desire to punish our enemies and make our friends happy (ibid.). In some passages in Book 2, Hume seems to classify these as direct passions (ibid.); but elsewhere in Book 2, he identifies our desire to make our

enemies unhappy and our friends happy as derivative from the indirect passions of hatred and love (T 2.2.6; SBN 366–8), and in Book 1, he identifies our bodily appetites as primary impressions (T 1.1.2.1; SBN 8).

Whether or not there are ways to resolve these difficulties, we can see Hume as concerned in this taxonomy with the explicability of our impressions. He does not think he could explain either why red is in our repertoire of sensory experiences or why fear is what we feel when we see the vicious dog. But there is a sense in which the fear is explicable and the sensation of red is not. This sense depends upon the availability or unavailability of relevant principles about our mental states. There are general principles concerning our mental states that can explain the occurrence of fear in terms of the occurrence of other mental states (feeling and then remembering the painful bite), but there are no analogous general principles of the mind to explain the occurrence in my mind of a sensation of red as a cardinal flies by me. My sensation is not, Hume believes, at all explicable in terms of other mental states I have had.

Ideas of Memory and Imagination

Hume's most basic classification of ideas is not elaborate: he distinguishes between ideas of memory and ideas of the imagination. An idea of memory repeats the "order and position" of the remembered experiences (T 1.1.3.3; SBN 9), but Hume argues that this by itself does not make an idea a memory (T 1.3.5.4; SBN 627–8). A person might be told about an event and "receive" a highly detailed idea of it "from the discourse" of his interlocutor, yet later, when his interlocutor mentions some particular circumstance of the event, suddenly recall that he himself experienced that very event (T 1.3.4.3; SBN 628). The content of his idea would not have changed, Hume says, but at first his idea would have been an idea of imagination, and later, after the flash of recollection, it would have been an idea of memory: "as soon as the circumstance is mention'd, that touches the memory, the very same ideas now appear in a new light, and have, in a manner, a different feeling from what they had before" (ibid.). Hume draws from this the conclusion that imagination and memory "are only distinguish'd by the different *feeling* of the ideas they present" (T 1.3.4.4; SBN 628), and he usually describes this feeling as liveliness (e.g., T 1.1.3.1; SBN 8–9), the same feature that distinguishes impressions from ideas. Because liveliness also distinguishes mere ideas from beliefs, by describing memories as more lively than ideas of the imagination, Hume is able to tie together his taxonomy of ideas with his theory of belief: we have a belief both when we "feel an immediate impression of the senses" and when there is a "repetition of that impression in the memory" (T 1.3.5.6; SBN 86).

In the passage about the person who suddenly remembers the event his interlocutor was describing, Hume also seems to be contrasting imagination and memory by distinguishing between the origins of their ideas. Before remembering,

the person received his ideas from his interlocutor's discourse, but an idea is a memory only if its origin is the experience of what is remembered: "what is the memory but a faculty, by which we raise up the images of past perceptions?" (T 1.4.6.18; SBN 260). But this leads to a difficulty for Hume's explanatory project, which requires him to explain the occurrence of some perceptions in a person's mind simply as the effects in that person's mind of earlier perceptions. When a person remembers a long-ago experience, it is hard to understand that past experience as causing the later memory if the *only* cause we can point to is the occurrence of the past experience. It is unclear how something that has not existed for many years can suddenly cause the occurrence of a state of mind here and now; indeed, Hume himself says that a causal relation between two things requires their temporal and spatial contiguity (T 1.3.2.6; SBN 75). Even if we waive the requirement of spatial contiguity for causal relations between perceptions, we are left with an unmet requirement of contiguity in time.

One way to solve the problem of contiguity would be to allow our explanations of perceptions to draw upon something non-mental: physical traces in the brain, for example. But Hume's official conception of a science of man does not allow him such an option; all he can appeal to are our mental states. At times he seems to be responding to the pressure of this problem by treating impressions as somehow persisting and then resurfacing in the mind at a later time. The original impression would then in some way have been there all along: the sense-impression that came into existence in the past would have persisted ever since and would be numerically identical with the memory-impression now. Indeed, Hume sometimes writes of memories as impressions rather than ideas (e.g., T 1.3.5.1; SBN 84; T 1.3.6.4; SBN 88). But if Hume really wants to claim that our sense-impressions can persist in this way, then he will have to make room in his science of man for unconscious states of mind, since what would be in the mind between the occurrence of the original experience and the occurrence of the memory would not be a conscious state. This in turn would require Hume to overhaul his account of the basic states of mind with which the science of man is concerned, since on his account they are the states we are in as something "strike[s] upon the mind" and enters "our thought or consciousness" (T 1.1.1.1; SBN 1). This is a serious problem for the general explanatory project Hume pursues in the *Treatise*: it is a problem not just for his account of memory but for all of the explanatory accounts he offers in which a perception is said to be the effect of earlier perceptions that occurred at a temporal remove from the effect.

All of our ideas that are not ideas of memory are ideas of the imagination. Although the imagination is not tied down to the order and position of past experiences in producing its ideas, it does tend to form them in accordance with a simple mental principle: the principle of the association of ideas. Hume claims that if a person has a perception of X, then his next idea is likely to be an idea of something that resembles X, is X's cause or effect, or is contiguous to X. He does not treat the principle of association as one that would allow us to predict from

a person's current perception exactly what object his next idea will have. If I am looking at a red rose, the principle of association would explain my next idea in a looser sense: it would explain the next idea whether it turned out to be an idea of a pink rose, a sweet scent, a box of Miracle-Gro, or the corner of my yard where the rosebush is.

Hume says that association through the relation of contiguity can itself be explained as the outcome of another powerful principle of the mind: the principle of custom. When a pattern of experiences has occurred in the mind a number of times, and then the mind has experiences that repeat part of this pattern, it will tend to form ideas that will complete the pattern. Contiguity produces association because "the senses, in changing their objects, are necessitated to change them regularly, and take them as they lie *contiguous* to each other"; thus "the imagination must by long custom acquire the same method of thinking, and run along the parts of space and time in conceiving its objects" (T 1.1.4.2; SBN 11).

Association by causal relation turns out to play a central role in Hume's explanatory projects, because he explains our causal inferences as a special type of association by causal relation. He says,

> [F]rom [their] constant conjunction . . . objects acquire an union in the imagination. When the impression of one becomes present to us, we immediately form an idea of its usual attendant; and consequently we may establish this as one part of the definition of an opinion or belief, that 'tis *an idea related to or associated with a present impression*. (T 1.3.6.15; SBN 93)

Again, however, he explains the association itself by appeal to the operation of custom. The association is the effect upon the mind of repeated patterns of perceptions (here, "constant conjunction"), and Hume says that only "custom operating upon the imagination" can allow us to "draw any inference" from our past experience (T 1.3.8.12; SBN 103).

Thus both contiguity and causal relation produce association in the imagination by the operation of custom, or the mind's tendency to complete repeated patterns of experience. Association by resemblance is different. Suppose I look at a red rose and by association think of a pink rose. Hume nowhere suggests that if association is at work, then I must have experienced red rose/pink rose patterns in the past: association by resemblance is itself explanatorily basic. It is also puzzling. A red rose resembles a red fire-engine (they are both red), a calico cat (they are both living), the Empire State Building (they are both bigger than an ant), and my neighbor's trash can (they are both 30 feet away from my car). In Hume's terminology, these are all "philosophical" relations of resemblance (T 1.1.5.1; SBN 13–14), and reflection suggests that some philosophical relation of resemblance will hold between the red rose and *anything* that I have experienced. It is thus important for Hume to identify which qualities are the ones that will tend to produce association of ideas by the relation of resemblance. If he cannot do

this, then we may worry that the principle of association by resemblance cannot really explain anything, since no matter what idea happened to come into my mind, it would be an idea of an object that in *some* way resembles the object of my current perception.

Hume seems to acknowledge something like this difficulty. He says that "no objects will admit of comparison, but what have some degree of resemblance," and he notes that only some of these resemblances will produce an association of ideas (T 1.1.5.3; SBN 14). He tries to explain this by saying

> [w]hen a quality becomes very general, and is common to a great many individuals, it leads not the mind directly to any one of them; but by presenting at once too great a choice, does thereby prevent the imagination from fixing on any single object. (ibid.)

But some qualities (for example, the colors of things) often produce an association of ideas even though they are very general and are common to a great many individuals, and some qualities (like being 30 feet from my car) do not produce an association of ideas even though they are common to relatively few individuals.

Faced with this sort of objection, Hume might be prepared to concede that we cannot explain why some resemblances do, and others do not, naturally associate ideas in the mind. At any rate, he is clearly prepared to make a more general concession: he describes the association of ideas in the imagination as a "kind of ATTRACTION, which in the mental world will be found to have as extraordinary effects as in the natural" and then adds that its "effects are every where conspicuous; but as to its causes, they are mostly unknown, and must be resolv'd into *original* qualities of human nature, which I pretend not to explain" (T 1.1.4.6; SBN 12–13). Here he surely means the reader to think of Newton's theory of gravitational force. Newton thought that "experimental philosophy" should not include untestable speculation, or "hypotheses," about the ultimate character of gravitational force, though he also thought that we gained genuine understanding of gravitational force by producing testable theoretical laws of attraction (Newton 1947: 547). In a similar spirit, Hume is saying both that we must not speculate about why some relations and not others tend to associate perceptions in the imagination, and that we do genuinely understand something about the mind by discovering the principles according to which it operates. It is an open question, however, whether Hume's principles of association really afford us understanding of how the mind works, given how very early in the explanatory game he is obliged to say, "We just can't explain that."

The Copy Principle

In the very first section of the *Treatise*, Hume claims to find "phaenomena" that "prove" the principle that "*all our simple ideas in their first appearance are deriv'd*

from simple impressions, which are correspondent to them, and which they exactly represent" (T 1.1.1.6, 7; SBN 4). This is often called Hume's copy principle. It is his way of making the broadly empiricist claim that the character of our experience is what explains the basic types of ideas we can have. For him, this broad claim takes a causal form: he is saying that our sensations and feelings are the causes of our ideas. He thinks that if he can establish this claim, he will have shown the incorrectness of innatism, which he understands to be a claim about the causal origins of our ideas.

Hume apparently does not think of the copy principle as a principle that is necessarily true of our ideas; he seems to leave open as a theoretical possibility that our minds should have basic types of ideas that cannot be explained by the character of our experiences. But he thinks there is overwhelming empirical evidence to support the claim that in fact all our basic types of ideas are derived from our experiences. He points especially to two kinds of facts. The first is that although we can easily give a person the idea of, say, sweetness if we produce an impression of sweetness in him (by feeding him honey, for example), we cannot give him the impression by producing in him the idea. The second is that if someone has never had a certain sort of impression, then he also lacks the corresponding idea: "[w]e cannot form to ourselves a just idea of the taste of a pine-apple, without having actually tasted it" (T 1.1.1.9; SBN 5).

Although Hume seems to think that the available evidence supports an exceptionless copy principle, at least once he seems to concede that we can have ideas whose content is not copied from our impressions. He imagines a person who has seen all the shades of blue except for one, and who is viewing all of the shades he has seen arranged in order from lightest to deepest. Could this viewer form an idea of the missing shade of blue, even though he has never had an impression of it? Hume says that most people would think the viewer could indeed form the idea; then he adds, puzzlingly, that "the instance is so particular and singular, that 'tis scarce worth our observing, and does not merit that for it alone we should alter our general maxim" (T 1.1.1.10; SBN 6). Of course, a genuine counterexample to a general claim does require altering the claim; what is not clear is whether Hume thinks the counterexample is not after all genuine, or whether he thinks the required alterations would be minor and would not alter the fundamental point that our experience is what can explain what we are able to think about. (See Fogelin 1984; Bennett 2001: 213–18.)

Hume uses the copy principle to great effect in criticizing the philosophical tradition. By appealing to it, he is able to insist that any clear idea is one whose simplest components can be traced back to the contents of our impressions. So if some philosopher invokes an unclear idea like that of substance or necessary connection, Hume will offer to clarify it by tracing it back to the impressions from which it arises. Of course, the clarified idea inevitably turns out not to have the content that Hume's target philosopher thought it had. Hume's readiness to treat the copy principle as an empirically supported general claim rather than a necessary

truth may seem to leave him vulnerable to counterattack from the philosophers with whom he is disagreeing. A philosopher who thinks, say, that we have an idea of empty space that is not derived from experience might insist that his idea of space is a counterexample to Hume's copy principle. Hume often strengthens his criticisms considerably by offering a diagnostic explanation of his opponent's claim, that is, an explanation of how someone might quite naturally come to believe something that is (as Hume is claiming) mistaken. A good example is his diagnostic explanation of the belief that we have an idea of empty space (T 1.2.5.14–21; SBN 58–62).

Simple and Complex Perceptions

In advancing the copy principle, Hume certainly does not mean to say that I cannot think of a "*New Jerusalem*, whose pavement is gold and walls are rubies" (T 1.1.1.4) if I have not seen such a city. Rather, he is thinking of such an idea as a complex that is built up out of simple components, and he is claiming that each of the simple components is available to my mind only because I have previously had an impression that included an exactly similar simple component. It is surprising to see that Hume believes not only that every simple idea has its origin in a corresponding simple impression, but also that "every simple impression is attended with a corresponding idea" (T 1.1.1.8; SBN 4; see also T 1.1.1.3; SBN 2–3). One reason he might have for saying this is that he thinks of the copy principle as a causal principle that identifies simple impressions as the cause of simple ideas. It would be awkward for him to concede that the cause (a simple impression) sometimes occurs without the effect (a simple idea). But this is not a very satisfying reason: without further argumentation, we might not see any reason to agree that every simple impression in a person's life is followed by a corresponding simple idea. Of course, we could not identify specific instances in our own lives of stand-alone simple impressions, since to identify them would be to have a corresponding idea in our own minds. Conceivably Hume had some such point in mind in making his claim, though if he did, he certainly did not make it explicit.

Although the distinction between simple and complex perceptions is crucial to the copy principle, it is not itself an easy distinction to understand. Simple perceptions are those that "admit of no distinction nor separation" (T 1.1.1.2; SBN 2). For example, the impression or idea of an apple is not simple, because "[t]ho' 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; SBN 2). Such an experience or thought of an apple can be decomposed: to be aware of the apple is to be aware of a distinctive taste, smell, and color, but it is different from being aware just of a distinctive taste, or just of a distinctive smell or color. The awareness of the taste, of the smell, and of the color are in that sense components of the awareness of the apple.

Some of our impressions and ideas cannot be decomposed. "*Blue* and *green* are different simple ideas," for example (T 1.1.7.5, Appendix; SBN 637): to think of being blue does not involve having any other thought than the thought of being blue. Hume adds that different simple ideas may resemble one another: *blue* and *green* "are more resembling than *blue* and *scarlet*" (ibid.). But this does not mean that the idea of blue and the idea of green have a component that the idea of scarlet lacks; none of these three ideas has components at all.

Hume claims that complex perceptions are those that allow a distinction and separation of component perceptions, and he says that "there are not any two impressions which are perfectly inseparable" and that wherever "the imagination perceives a difference among ideas, it can easily produce a separation" (T 1.1.3.4; SBN 10). This strongly suggests that the mind can entertain any simple perception while at the same time entertaining no other perception. This may seem plausible in some cases: Hume invites us, for example, to imagine a mind that is "reduc'd even below the life of an oyster" and to "[s]uppose it to have only one perception, as of thirst or hunger" (T Appendix 16; SBN 634). But can we have, say, a perception of a red color while at the same time entertaining no perception of any shape? Perceptions of colors seem to be Hume's paradigm example of simple perceptions, and so we might expect his answer to be a clear and resounding "yes." But if his answer is in the end "yes," it is by no means resounding or clear.

In presenting something he calls a distinction of reason, Hume seems to say that in order to consider just the color of a shaped object, we cannot simply form an idea of the color and leave out any idea of a shape. We must do something more complicated. Suppose we want to consider just the color of a white globe. What we have to do is to consider the resemblance between a white globe and (say) a white cube and separate that resemblance from the resemblance between the white globe and (say) a black globe. We draw a distinction of reason between the color and the shape by "view[ing] them in different aspects, according to the resemblances, of which they are susceptible" (T 1.1.7.18; SBN 25).

This seems to pose several difficulties for Hume's account of simple perceptions. Although the distinction of reason – the viewing of different resemblances – gives Hume a way to say color and shape are distinct even though they are inseparable, his account of simple perceptions seems to require that they be distinct *and* separable. Hume might want to say that a distinction of reason produces not just distinctness of a sort but also separation of a sort, thus preserving consistency with the account of simple perceptions. But then he would seem to be generating a conflict with what he has said about blue, green, and scarlet, because there he has said that noting different resemblances between blue and green and blue and scarlet does *not* show that the idea of blue has components that are in any sense separable.

Hume's account of simple perceptions may be even more complicated if we take into account his distinctive theory about our ideas of extension, that is, our

ideas of spatial extent. Strictly speaking, he says, our impressions and ideas of spatial extents are perceptions of colored or tangible points arrayed in a distinctive way. Each impression of a point is a minimum impression – a visual sensation of a color or a tactile sensation of a pressure – that does not amount to seeing or feeling anything as divisible, and that thus does not amount to seeing or feeling anything as extended. Some readers have thought that at least where visual and tactile perceptions are concerned, Hume's account of simple perceptions is incomplete without reference to his theory of minimum impressions and ideas. But other readers have thought that for Hume, the relation between perceptions of spatial extents and the minimum perceptions they comprise must be different from the relation between complex perceptions and the simple perceptions they comprise. (See Cummins 1996; Garrett 1997: ch. 3.)

Either way, Hume's account of our ideas of spatial extent raises a quite general question about the copy principle: just what aspects of our thought count as components of ideas? When I have an idea of a spatial extent, I have an idea of a multitude of colored or tangible points that are arrayed in a distinctive way, so that they are to the right and left of one another, or above and below one another. Is the distinctive "manner" or "disposition" (T 1.2.3.5; SBN 34) of the points itself a component of the idea, derived from a component of my visual or tactile impressions? Hume denies strenuously that we have any idea of extension that is devoid of visual or tactile content, but he also seems to be asserting that our ideas of spatial extent include something more than just the visual or tactile content provided by minimum perceptions. Turning to a different corner of Hume's philosophy, when I have an idea of my desk as something that has persisted unchanged between the time I saw it yesterday and the present moment, is the idea of persistence a component of the idea? Is it derived from a component of my impressions? Again, Hume would deny that I can have an idea of persistence that excludes all successive content, but he also seems to be asserting that our idea of the desk as persisting includes something more than just a successive content. And turning to a question Hume did not directly address, when I have the idea that the desk is brown, is the "S is P" structure in my idea a component of the idea? Is it derived from a structure that is presented to me in my impressions? More broadly still, exactly what about our ideas does Hume hope to explain? And exactly what does his empiricism commit him to?

General Terms

Drawing explicitly on Berkeley's work, Hume argues that when we have general ideas, their generality cannot consist in their having a special sort of abstract content. He then takes up the challenge of explaining how, if that is true, it is possible for us to have general ideas and use general terms. (See Berkeley 1988, Introduction.)

Berkeley and Hume understood themselves to be attacking a theory of general ideas proposed by Locke, though whether Locke held precisely the theory they attack is open to debate. General ideas are ideas that somehow manage to signify a plurality of particular things. If I have the general idea *dog*, then my idea signifies all of the actual dogs that there are, and indeed it signifies all of the merely possible dogs as well. The Lockean theory that Berkeley and Hume criticize says that the general idea *dog* is able to signify many particulars (not just one) and the right particulars (dogs, not buildings or animals or just pit bulls) because it has a distinctive sort of content. In principle, a person could have the general idea *dog* as his only idea. It would be general because it would have an abstract content, a content that included none of the details of color, size, shape, and so on, that characterize our particular ideas of Lassie, Toto, and Asta. Its only content would be what different particular dogs have in common by being dogs.

Berkeley and Hume argue that we cannot have ideas whose content is in this sense abstract. Hume gives several arguments for this claim. One says that we cannot have an impression, say, of a dog as having no particular size, and since ideas are derived from impressions, we cannot have an idea of a dog as having no particular size, either. Another argument says that, because it is impossible for (e.g.) a dog to exist that has no particular size, and because we cannot have ideas of things that cannot exist, we cannot have an abstract idea of a dog.

All of our ideas are particular, in the sense that none of them is abstract; how, then, is it possible for some of them to be general, in the sense that they signify many things, and the right things? In answering this question, Hume rejects the assumption that the generality of an idea is something that it has simply in virtue of its content. Instead, his theory ties the generality of an idea in a person's mind to the relations of that idea to other ideas in that person's mind.

Hume draws upon a broad picture of how people actually learn what general words mean. We hear a word (say, "dog") each time we encounter a particular of a certain kind (say, various dogs). (Here Hume must be thinking of the way adults repeat words to children in the presence of the things the words name.) This repeated word-thing pattern in our experience engenders a custom in the mind. When the custom has been set up, "the hearing of [the] name" does two things: it "revives the idea of one of [the] objects, and makes the imagination conceive it with all its particular circumstances," and it "revives that custom, which we have acquir'd by surveying [the objects]" (T 1.1.7.7; SBN 20). So once I have learned what "dog" means, when I hear the word, I have an idea of a particular dog (say, Lassie), and I am disposed to think of the other particular dogs I have "surveyed." This disposition is crucial to Hume's account of general ideas: it connects the idea of Lassie with ideas of other dogs and gives the idea of Lassie its generality. It is what, on Hume's view, explains how an idea can signify many particulars and the right particulars. Of course, we often do not actually think of any of the other dogs we are disposed to think about, but if we should carelessly start to think, for example, that all dogs have long noses, the custom revived by the word "dog"

ensures that ideas of, say, a pug and a bulldog will "immediately crowd in upon us, and make us perceive the falsehood of this proposition, tho' it be true with relation to that idea, which we had form'd" (T 1.1.7.8; SBN 21).

This has struck many readers as a theory of general ideas that explains our capacity for conceptual thought only by attributing to us the very capacity in question. It seems that if my mind is to form the custom of thinking of one or more dogs when I hear the word "dog," then as I am learning what "dog" means, I must be aware of individuals (Lassie, Toto, Asta) *as* dogs. Indeed, Hume begins his account of general words by saying that the first step is that "we have found a resemblance among several objects, that often occur to us" (T 1.1.7.7; SBN 20); and many readers have thought that in order to find a resemblance among several particular things (for example, to think that Lassie, Toto, and Asta are alike in respect of being dogs), we must already have the capacity for conceptual thought that Hume is trying to explain. (See Kemp Smith 1941: 257; Mounce 1999: 27–8.)

Perhaps Hume would deny that we must be able to *think* of these individuals as alike in respect of being dogs. On his view, perhaps all we need to be able to say is that the individuals are in fact alike in respect of being dogs, and that this respect of likeness is one that triggers association and custom-formation in the mind. Here, as in many other places in the *Treatise*, part of the challenge of understanding and assessing Hume's claims and arguments lies in gauging what it is that he is taking for granted, and what it is he hopes on that basis to explain. And that in turn is something that, as I have tried to stress, requires us to reflect upon the specific ways in which his thought is shaped by his ambitious effort to understand a great deal about our mental life simply by understanding how our minds work upon the materials afforded by our sense experience.

References

Bennett, Jonathan (2001) *Learning from Six Philosophers*, vol. 2 (chs. 22, 32, 33). New York: Oxford University Press.

Berkeley, George (1988) [1734] *A Treatise Concerning the Principles of Human Knowledge*, ed. Jonathan Dancy. New York: Oxford University Press.

Cummins, Phillip (1996) "Hume on qualities," *Hume Studies* 22(1), pp. 49–88.

Fogelin, Robert (1984) "Hume and the missing shade of blue," *Philosophy and Phenomenological Research* 45, pp. 263–72.

Garrett, Don (1997) *Cognition and Commitment in Hume's Philosophy* (chs. 1–3). New York: Oxford University Press.

Kemp Smith, Norman (1941) *The Philosophy of David Hume*. Facsimile published 1983, New York: Garland Publishing.

Mounce, H. O. (1999) *Hume's Naturalism*. London and New York: Routledge.

Newton, Isaac (1947) [1713] *Mathematical Principles of Natural Philosophy; The System of the World*, trans. A. Motte, ed. F. Cajori. Berkeley, CA: University of California Press. Motte/Cajori trans. first published 1934. First edn. (in Latin) of *Principia*, 1687.

Further reading

Broughton, Janet (2000) "Explaining general terms," *Hume Studies* 26(2), pp. 279–89.

Everson, Stephen (1988) "The difference between feeling and thinking," *Mind* 97, pp. 401–13.

Fodor, Jerry (2003) *Hume Variations*. New York: Oxford University Press.

Noxon, James (1976) "Remembering and imagining the past," in Donald Livingston and James King (eds.) *Hume: a Re-evaluation* (pp. 270–95). New York: Fordham University Press.

Passmore, John (1968) *Hume's Intentions* (chs. 5, 6). London: Duckworth.

Pears, David (1990) *Hume's System: an Examination of the First Book of his* Treatise (Part I). New York: Oxford University Press.

Sellars, Wilfrid (1997) [1956] *Empiricism and the Philosophy of Mind*, with introduction by Richard Rorty and study guide by Robert Brandom (ch. 6). Cambridge, MA: Harvard University Press.

Stroud, Barry (1977) *Hume* (chs. 2, 10). London: Routledge.

Tienson, John (1984) "Hume on universals and general terms," *Noûs* 18, pp. 311–30.

Williams, Michael (1985) "Hume's criterion of significance," *Canadian Journal of Philosophy* 15, pp. 273–304.

4

Space and Time

Lorne Falkenstein

Hume's account of space and time has received comparatively little attention from scholars. The classic treatments of this topic are those offered by Kemp Smith, Hendel, Broad, Flew, and Fogelin (see References). These accounts were generally critical and focused on what Hume had to say about the finite divisibility of space and time. More recent work by Baxter, Franklin, Frasca-Spada, Holden, Jacquette, and myself has been both more sympathetic and more wide ranging. Most notably, Franklin (1994) has argued that Hume's position on finite divisibility is more mathematically respectable than the older literature recognized, Holden (2002) has shown that his position was well motivated by deep metaphysical concerns shared by other early modern philosophers, and Baxter (2001) has expanded on the older literature with original and insightful accounts of Hume's views on time. I (1995, 2001) have argued that Hume's accounts of space and time have important bearing on his views on such other topics as representation, causality, and the immateriality of the soul. Viewed from this wider perspective, there is far more to Hume's accounts of space and time than just a non-standard position on infinite divisibility.

There are two parts to Hume's theory of space and time. One concerns the *properties* of extension and duration as exhibited in our perceptions, the other the influence of the *relations* of spatial and temporal contiguity on our beliefs and our passions. The first part of Hume's theory is presented over two widely separated sections of the *Treatise*: 1.2.1–5 and 1.4.5.7–16. The second is presented over portions of *Treatise* 1.1.4, 1.3.2, 1.3.9, 1.3.13, and 2.3.7–8.

Extension and Duration

Hume had little interest in mathematics or natural science, and consequently little interest in space or time as topics in mathematics or physics. As Flew (1976: 258)

has pointed out, his study was motivated by the desire to refute a then popular argument for fideism (the view that religious belief is legitimate even though not supported by reason). As I have pointed out (1995), it was also applied to undermine claims about the unity of consciousness and the immateriality of the soul.

Hume's Reply to the Paradox of Composition

According to arguments first formulated by Zeno of Elea in the sixth century BCE, and revived by early modern fideists such as Pierre Bayle (1697: 372) our concepts of space, of extended objects, of time, and of motion are incoherent. This incoherence proves that there is something radically defective about human reason, which fails to grasp what are apparently the simplest and most important things. This in turn means that reason is in no position to impugn the claims of faith.

The main argument for the incoherence of our concepts of space and time runs as follows: What is extended or enduring consists of parts. These parts cannot be of zero magnitude or they would not add up to anything. However, if they have some magnitude, they must be divisible, since any quantity not equal to zero is divisible. It would seem to follow that they must be infinitely divisible. But an infinity of parts, each of which has some magnitude, could not be contained in any finite volume. Consequently, and paradoxically, any finite space, time, or extended or enduring object, however small or brief it might be, must be infinitely large or long-lasting, because divisible into infinitely many yet smaller parts, each of which, still having some magnitude, must sum to an infinitely large magnitude.

Whether rightly or wrongly, Hume did not question this argument's second half, concerning the summation of an infinite number of parts. (The older Hume literature focuses on this apparent failing, whereas Holden [2002] has argued that there were compelling metaphysical reasons for early modern philosophers to accept the implication, even though it might be evaded within the field of pure mathematics.) Hume instead argued that, contrary to the argument's first half, a magnitude could be composed of a finite number of unextended parts (T 1.2.1–2 and 4; SBN 26–33 and 39–53). As Franklin (1994) has shown, this is an entirely respectable position, not merely metaphysically, but also mathematically. The account that follows builds on Franklin's result.

To make this case, Hume first set out to make a point, not about space or time or external objects, but about the perceptions contained in the bundle that constitutes a human mind. He argued that these perceptions are disposed in such a way as to constitute a phenomenal space and time that are immune to the composition paradox (T 1.2.1, 1.2.2.1–2, 1.2.4.6–7; SBN 26–30, 41–2).

Hume's Arguments for the Finite Divisibility of Perceptions (T 1.2.1)

Hume argued that since the capacity of the mind is finite, its ideas cannot be infinitely complex. There is a limit to the number of ideas there can be in a mind, and this limit applies both to the number of ideas that can be present at any one time and to the number of ideas that can occur in succession over any finite interval of time. Consequently, there is also a limit to the number of parts or simple ideas that can compose any one complex idea (T 1.2.1.2–3; SBN 26–7).

There are also no infinitely complex impressions. For, just as our minds have only finite capacities, so our sense organs have only finite powers of discrimination. Hume appealed to two main experiments to prove this point, one for space (T 1.2.1.4; SBN 27–8), concerning an ink blot seen at a distance, the other for time (T 1.2.3.7; SBN 35), involving a whirling coal. Both experiments are problematic, and the reader can refer to Jacquette 2001, and Larivière and Lennon 2002, respectively, for detailed discussions. However, the main point of each experiment can be stated quite briefly: The fact that a distant ink blot appears as a point that cannot be divided without disappearing proves that we experience visible points that have color and location, and so exist (somewhere on the visual field), even though they have no extension (since they cannot be further divided). The best explanation of the fact that a whirling coal appears as an unbroken circle is that there is a limit to the speed with which impressions can succeed each other, where speed of succession is defined as the number of successive impressions in a standard interval, such as that marked by two heartbeats. Beyond this limit impressions become simultaneous and so, if adjacent, form a line rather than a moving point. Accepting this explanation entails that between any two successive impressions there is not always a further impression.

Admittedly, to prove that visual impressions consist of indivisible parts is not to prove that there are only finitely many of these parts in any given compound impression. Hume appealed to a further experiment to yield this result. We can see two differently colored points to be so positioned that they are immediately adjacent to one another, without there being either any gap or any overlap between the two. As Hume put it, "A blue and a red point may surely lie contiguous without any penetration or annihilation" (T 1.2.4.6; SBN 41). Because the visual field is not infinitely divisible, one colored point can be set immediately alongside another to produce an unbroken, compound impression that is double the size of a single, pointal impression. Adding a third point produces a compound that is triple the size. And only a finite number of points need to be added to make the compound exceed any given limit.

These claims may seem paradoxical for the reason Zeno originally gave. If a single impression has, effectively, zero magnitude, how can placing a second one immediately adjacent to it produce a compound that is double the size of the

original? Hume was able to evade the paradox because on his account the size of a compound is not a product of the size of its parts. As explained more fully in the following section, it is rather a product of the manner of disposition of its parts. A pointal impression set beside another pointal impression does not add its volume to the volume of the first pointal impression; it rather marks a location immediately outside of the location of the first pointal impression, thus marking an interval consisting of two adjacent locations. While the magnitude of each pointal impression is zero, the number of locations picked out by two non-over-lapping impressions is double that of the number picked out by a single pointal impression.

The empirical nature of Hume's argument against infinite divisibility is worth underscoring. The fideists Hume was opposing had argued a priori that a finite extension cannot be composed of infinitely many extended parts, because however little extension these points may have, if there are infinitely many of them they must sum to an infinite magnitude. Hume, in contrast, argued a posteriori that a finite *visible* extension cannot be composed of an infinite number of *colored points*, since, as we see, differently colored points can be set immediately alongside one another without any observable overlap or separation, so that there is not always a further visible point between any two given visible points. It is this empirical premise that entails that the compound resulting from the successive addition of points immediately alongside points will necessarily surpass any finite limit as points continue to be added.

The mathematicians of Hume's day drew a distinction between two kinds of parts that extension can be divided into: proportional parts, which become pro-portionally smaller as the division continues (e.g., as the whole is divided into halves, quarters, eighths, and so on), and aliquot parts, or parts of the same size. Hume's empirical argument led him to declare this to be a "frivolous" distinction (T 1.2.2 n. 6; SBN 30 n. 1). Because it is factually false that between any two colored points there is a further colored point, the visual field can only be divided a finite number of times before indivisible points are reached. Beyond this limit, no further proportional parts can be generated. Mathematicians can invoke the notion of proportional parts if they want, but because that notion is not validly applied to points on the visual field it cannot be invoked to prove that the visual field is infinitely divisible.

The Coherence of Hume's Account

In making the observation that a blue and red point can lie contiguous without annihilation or penetration, Hume, who was no mathematician, inadvertently articulated a concept that only came to be rigorously described much later: that of a discrete space. A space is discrete only if (but not if[1]), for any line in that space, every point on that line except the first (if there is one) has an immediate succes-

sor, and every point on that line except the last (if there is one) has an immediate predecessor, where "immediate" means that there are no other intervening points. The set of integers ordered by the "greater than or equal to" relation,

$$\{. . . -3, -2, -1, 0, 1, 2, 3 . . .\}$$

is a model of a discretely ordered set, and the set of points in a two or three dimensional lattice that have integers as their Cartesian coordinates is a model of a discrete manifold. Hume appealed to an equally compelling model of a discrete manifold, the model of colored points on the visual field. In different words, Hume claimed that the visual field is a discretely ordered manifold of colored points.

This model undermines Zeno's case against an extension composed of extensionless points. Zeno had argued that an unextended point has zero magnitude, and that zero added to zero however many times is still zero. But Hume's model of extensionless colored points on the visual field illustrates that Zeno failed to appreciate that extension is not the product of the *mathematical summation* of its parts; it is the product of their *manner of disposition*. In a universe or a model where points are governed by an immediate successor relation, as is the case with the visual field, it is possible to place one unextended point immediately alongside another, without leaving any gap or separation, but also without having the two overlap, so that, even though the points are themselves unextended, their manner of disposition produces an unbroken, extended compound. I stress that the compound is unbroken because, for there to be a break between two points, there must be a location between them that is not occupied by either of them. But in a discrete space there simply is no further location between two immediately adjacent locations. This is why it is question-begging to attempt to attack Hume's position by observing that the visual field does not have granular look to it, as Broad (1961:166) tried to do (cf. Flew 1976: 265). A collection of points can only look granular if there are locations between points and that is just what we deny when we maintain that a space is discrete.

A similar reply can be made to an argument that, as Fogelin (1985) has noted, the fideists had employed to support their opposed position. They had argued that if a point, *b*, were taken to be immediately adjacent to a second point, *a*, to its left, and to a third point, *c*, to its right, then the point could not be unextended. For, they claimed, if *b* touches *a* on its left side, and *c* on its right side, these two sides must be distinct (otherwise *a* would touch *c* and *b* would in effect not exist between them). Consequently *b* must have distinct right and left sides, must be divisible into at least two parts, and so must be extended.

This argument only works if we presume what we are supposed to prove – that a space or ordered manifold is infinitely divisible. In that case there is always a further point between any two given points, so that the only way two bodies can touch, that is, not have a further point separating them, is by overlapping. But in a discrete manifold the notion of touching is equivocal. Two bodies can touch either by overlapping at a point or by being immediately adjacent without

overlapping. Either way there is no intervening point or gap. Consequently, in a discrete manifold, *b* can touch *a* by having *a* as its immediate predecessor, and can touch *c* by having *c* as its immediate successor without overlapping with either and without having to be ascribed distinct parts, one of which overlaps with *a* and the other of which overlaps with *c*.

The Idea of Equality (T 1.2.4)

Hume was aware of geometrical demonstrations of the incommensurability of line segments (such as the side and the hypotenuse of certain triangles) and of a one-to-one correspondence between points on a short line and an arbitrarily long one (such as the points on the circumferences of concentric circles defined by the radii of those circles, as alluded to by Bayle [1697: 366]). Both sorts of demonstration would appear to establish that space must be infinitely divisible. However, Hume charged that they beg the question by presuming the existence of things (perfectly straight lines between any two given points and miniscule differences in length) that we do not actually see and that can only be proven to exist by presupposing the very thing that we are trying to prove: that space is infinitely divisible (T 1.2.4.8, 17, 23–4, 29–31; SBN 42, 44–5, 47–9, 50–2, 638).

In somewhat more detail, Hume claimed that our judgments of such things as the flatness, straightness, curvature, and equality of figures are usually not precise (T 1.2.4.8–24, 25–7, 28; SBN 45–50, 637). Though our impressions are made up of a finite number of parts, and so have a size and shape that is a function of the number and manner of disposition of these parts, it is impossible to count them or ascertain their precise arrangement in most cases (T 1.2.4.19; SBN 45). Where there is a significant difference between figures, or where we only need a rough approximation, we can tell with certainty whether one figure is greater or lesser, flatter, straighter, more curved, or more or less equal to another (T 1.2.4.22; SBN 637, 47). Where the differences between figures are not as great or the standards more strict we can move a common measure over both to ascertain differences or equalities (T 1.2.4.23; SBN 47). But this is only a partial improvement. We can still conceive how the addition or removal of a single point would produce or destroy an equality, or a deflection by a single point destroy straightness of a line or flatness of a surface, below the level that we can normally notice (T 1.2.4.24; SBN 47–8).

Because such standards of equality, straightness, and flatness as we are able to employ are normally imprecise, they do not mandate the claim that between any two points a straight line may be drawn or the claim that there are incommensurable lines. All we can determine is that between any two points a more or less straight line can be drawn and that for any two lines a measure can be found that will fit an integral number of times over each without any notable overlap or shortfall (T 1.2.4.30–1; SBN 51–2, 638).

These features of Hume's account have been invoked by Baxter (forthcoming) to argue that Hume sought to show that all the claims of Euclidean geometry are correct within certain limits while discourse that exceeds these limits is illegitimate. However, following on Franklin (1994), I think it important also to note that Hume's commitment to finite divisibility does not logically entail an inconsistent account of the pointal world beyond these limits. Non-Euclidean consequences can be derived from that commitment, but what Baxter correctly identifies as an attempt to preserve the truth of Euclidean geometry within certain limits is not a smokescreen used to conceal logically entailed absurdities. In a discrete space there are no smooth curves (only step-shaped lines that approximate the shape of curves) and lines can only intersect at certain angles. Not every two points can be joined by a straight line, and one or more of the shortest paths from one point, a, to another, b, can overlap for some considerable distance with one or more of the shortest paths from a to some point c not on any of the shortest paths from a to b. So far as we can see, none of these features of the geometry of discrete manifolds is false of colored points as they exist on the visual field. The colored lines that we see are not so straight or so smoothly curved that we can say for sure that they are not composed of step-shaped arrangements of colored points. And the acute and obtuse angles we see are not so sharp that they could not possibly be composed of step-shaped configurations of colored points, or colored lines that overlap for some small distance. Therefore, what we actually see when we look at a geometrical construction is consistent (because of its imprecision) with both the geometry of discrete and the geometry of dense or continuous manifolds.

The geometrical proofs of infinite divisibility ignore this imprecision in our visual experience and simply assume that between any two given points there must be the further points needed to make a perfectly straight line or a perfectly smooth curve. By doing this they presume what they purport to prove: that between any two points there must be a further point.

Hume claimed that geometers have fallen into this error because of a natural tendency of the imagination to continue along a course of thought (T 1.2.4.24; SBN 47–9). Having discovered that initial judgments of equality, straightness, or flatness can be corrected by the employment of a measure, and those based on the employment of a measure corrected by those based on the employment of a more exact measure, we imagine that this process of correction could go on forever, and so formulate ideas of inequalities and incongruences that exceed the tolerances of geometrical construction in a discrete manifold.

This marks the first sustained discussion in the *Treatise* of the notion of a fiction. In its broadest sense a fiction is a complex idea that does not correspond to any complex impression. But in a special, technical sense a fiction is an incoherent idea that we manage to think by confusing two importantly distinct though closely related ideas, and that we are tempted to think and even believe by some natural impulse of the imagination. In the case at hand the ideas are the idea of a perfect standard of equality and congruence based on identical numbers and dispositions

of points, and the idea of making a correction to a previous estimate of equality or congruence by employing a more exact standard of measurement. The former idea is valid of the abstract geometry of points in a discrete manifold, though useless for any practical purposes given the impossibility of discriminating points on the visual field in almost all circumstances. The latter is an idea of something that happens when measuring figures on the visual field. Importing the latter concept into the former realm produces the fiction of a further point that can be found between any two given points by employing a more exact standard of measurement.

Fictions play an important role at many points in the *Treatise* (1.2.3.11 anticipating 1.2.5.29, 1.4.2.26–30, 1.4.2.36–40, 1.4.2.52, 1.4.3.2–5, 1.4.5.13, 1.4.6.6; SBN 37, 65, 200–1, 205–8, 214–16, 210–21, 237–8, 253–5).

The Infinite Divisibility of Objects (T 1.2.2)

As Hume himself observed, when we view an object, such as an ink blot, different collections of impressions arise in the mind depending on the distance of the object from the eye (T 1.2.1.4; SBN 27–8). The further the object is away, the fewer the number of impressions it produces. As we obtain closer and closer views, parts of the object that were previously represented by a single point come to be represented by distinct points. This experience naturally leads us to think that our impressions are seldom, if ever, adequate representations of objects, because multiple parts of objects are continually discovered to have been mapped onto single impressions. It can even serve as the basis for a kind of causal inference. Observing that a given view of an object such as a grain of sand produces a certain, necessarily finite number of impressions, we might infer from past experience that a yet closer view would produce even more (T 1.2.2.5; SBN 28). Consequently, even though our impressions are not infinitely divisible, we can readily think that objects might be. (Note that this conclusion is assisted by the same inertial tendency of the imagination that produces the idea of perfect equality.)

Hume rejected this inference. While he allowed that our impressions can be inadequate representations of highly complex objects, and that even our ideas can exhibit this failing (because we reach a point at which we can no longer keep track of all the parts we have discovered), he maintained that our ideas are adequate representations of the smallest parts of extension (T 1.2.2.1; SBN 29). Nothing, after all, can be smaller than a thing that has no parts, and our minimal ideas have no parts. They do not, therefore, fail to adequately copy the part/whole structure of smallest possible things. But, Hume continued, if our ideas are adequate copies of smallest possible things, and an infinite number of these ideas would necessarily constitute an infinite extension, then no finite extension can be infinitely divisible (T 1.2.2.2; SBN 29–30).

As Fogelin (1988: 54) has pointed out, this argument is unconvincing. While it might be true that our smallest ideas are adequate representations of points, it

is not obviously true that the discrete manner of disposition of our ideas is an adequate representation of the manner of disposition of the parts of objects. Our ideas are disposed in such a way that there is not always a further idea between any two given ideas, and that is why a finite number of them must necessarily compose an infinite extension. But we have some experience that indicates that this is not the case with the parts of objects (since we find that a closer view of the same object gives us a larger number of impressions). In light of this evidence, we need to have some reason to suppose that the parts of objects are only discretely disposed.

It might be thought that one such reason would be that objects just are collections of impressions that have been commonly observed to occur together. This is a view that Hume took up later in the *Treatise* (1.4.2.36–40; SBN 205–8), though he found problems with it (1.4.2.44–6 and 56–7; SBN 210–12 and 217–18). However, even if objects just are collections of impressions, it is a question what the bounds of those collections are. Objects appear different from different angles and distances, and they evolve over time. Consequently, the collection of impressions associated with the name of an object is seldom confined to those provided by a single view. The information provided by many different views informs the collection. But how many such views? Hume could have no a priori reason for ruling out the possibility of obtaining increasingly microscopic views of the same object, which would multiply the collection of ideas of its smaller parts and of the parts of those parts without end.

Hume thought it important to deny the infinite divisibility of objects because he accepted the common view that infinite divisibility is paradoxical. Consequently, he feared that allowing infinite divisibility would concede too much to fideists wishing to recommend belief in religious mysteries. But whether or not an infinitely divisible object really is paradoxical (a fuller and somewhat more sympathetic investigation of Hume's struggles with this topic is found in Frasca-Spada 1998: 11–56), Hume at least showed that we are not compelled to accept this notion. There is no geometrical proof of the infinite divisibility of space that does not beg the question, and there is an alternative way of conceiving space and time, modeled on the discrete manner of disposition of colored points on the visual field, that does not involve paradoxes.

Manners of Disposition (T 1.2.3)

Hume's position has one particularly striking aspect: it involves a commitment to the tenet that some impressions and ideas have spatial qualities (as opposed to simply representing spatial qualities). Our impressions and ideas do not just occur successively over time. Some of them, those of vision and touch, consist of pointal parts that are disposed alongside one another in space (T 1.2.3.4; SBN 34). Insofar as this is the case they compose compound perceptions that have shape and size.

The impression or idea of a colored square is not so much a thought "of" a square as a thought that is square or that consists of a square configuration of unextended, colored points (T 1.4.5.9–10; SBN 235–6). It represents a square by quite literally being a square (T 1.4.5.15; SBN 239–40). Correspondingly, the impression or idea of a time represents time by quite literally taking time to occur. As Hume put it, five notes played upon a flute give us an impression of time, but the impression of time is not a sixth impression added to the other five. It is rather the compound impression that consists of these five notes occurring successively in that particular order (T 1.2.3.10; SBN 36–7).

As I have argued at some length elsewhere (1997) in taking this position Hume was not introducing a third thing, "manners of disposition," in violation of the tenets that all the perceptions of the mind resolve themselves into impressions and ideas, that all simple ideas are copied from simple impressions, and that all ideas that are distinguishable are separable. Instead, he was introducing a special kind of impression, compound impressions, that exhibit something no simple impression does: a collection of component impressions disposed in a certain manner. Our ideas of spaces and times are copied from these compounds. They are not copied from any antecedent *simple* impression. But they are copied from antecedent *compound* impressions. They are, initially, ideas of *particular* spaces and times made up of particular colored or tangible points disposed in particular fashions and ideas of particular perceptions occurring in particular sequences. While we may subsequently draw a distinction of reason between the manner in which these perceptions are disposed and the perceptions (T 1.1.7.18; SBN 25), such distinctions involve no separation, but merely join particular compound ideas, consisting of particular perceptions disposed in particular manners to a habit to conceive related compound ideas.

The fact that manners of disposition are inseparable from the perceptions that are disposed in those manners poses a challenge when joined to Hume's "separability principle," his claim that wherever there is a difference between ideas, those ideas can be separately conceived (*Treatise* 1.1.7.3; SBN 18). Since there is a difference between a blue point and a red point, we should be able to conceive the one separately from another. But if their manner of disposition is not a third impression that can be separated from either of them, then it becomes a question how they can be separated from one another. The answer to this challenge rests with taking the blue point, the red point, and the compound impression consisting of the two points disposed in a certain fashion to be three mutually distinct and separable perceptions rather than taking the latter to be a derivative product of the former two.

Consider the compound idea, "bLr," consisting of a blue point set immediately to the left of a red point. I can conceive of a blue point without having to conceive of it as being adjacent to a red point. So "b" is distinguishable and separable from "r." But it is also distinguishable and separable from "bLr." The idea of a blue point by itself is a distinct idea from the idea of a blue point set to the left of a

red point. The reverse holds as well. The idea of a blue point set to the left of a red point is distinct and separable from the idea of a blue point by itself. For, while I may not be able to conceive of a blue point immediately to the left of a red point without conceiving of a blue point, I am certainly able to conceive of a blue point immediately to the left of a red point without having to conceive of a blue point on its own, or of a blue point with something else immediately to its left. The conception of a blue point that is involved in "bLr" is therefore importantly distinct from the conception of a blue point on its own. The two conceptions are as separable from one another as either is from the conception of a red point.

What I cannot do, however, is conceive of "Lr" on its own. I cannot conceive of "r" as being disposed on the right without conceiving of it as being on the right of something. So while "b" is separable from "bLr," "Lr" and "L" are not. Insofar as "L" figures in an idea it brings two simples with it.

As Weinberg (1965) classically pointed out, early modern philosophers had difficulties accepting that there could be relations that cannot be reduced to some non-relational property of things. But Hume should not be assimilated with this group. The views he articulated at the outset of *Treatise* 1.3 on the distinction between the relations of contiguity, cause, and identity on the one hand, and those of resemblance, difference, and degrees of quantity and quality on the other, will not permit such an assimilation. Were spatiotemporal relations reducible to non-relational properties of the relata, it would be possible to determine that "b" is to the left of "r" simply by inspecting "b" and "r" and comparing our ideas of those notions, just as it is possible to determine that "b" is brighter or more saturated than "r" simply by inspecting our ideas of "b" and "r." There would then be no difference between spatiotemporal relations and those relations having to do with degrees of quality. Both would be discerned by inspecting and comparing the relata. But Hume denied this. He maintained that spatiotemporal relations can only be discovered by taking note of the manner in which the impressions our ideas copy were originally disposed in compound impressions. His notion of a compound impression consisting of simple impressions disposed in a certain fashion opens the door to the notion of irreducible relations – relations that can only be discerned in the compound impression and not discovered by comparing ideas of the relata. To that extent, it challenges the view of commentators such as Green (1886: 170–228), Laird (1932: 29–30), and Pears (1990: 101), who have maintained that Hume was a "cognitive atomist" who believed that information can only be conveyed by simple impressions.

The Simplicity of the Soul (T 1.4.5)

Hume appealed to the spatiality of visual and tangible impressions and ideas to reply to the "Achilles" argument for the simplicity and consequent indivisibility and incorruptibility of the soul. (Mijuskovic 1974 is the standard study of the

history of this argument.) As McIntyre (1994) and Russell (1995) have noted, this argument had been at the core of a prominent controversy between Samuel Clarke and Anthony Collins in the earlier part of the eighteenth century (see Clarke 1738), though Hume would also have found it in Bayle (1697: 130–4). The argument rests on two premises. The first is that what is extended is an aggregate of distinct, independent substances and so not a true unit. The second is that a thought cannot be divided into parts each of which is assigned to a different substance. For, in that case, each substance would know only the part of the thought assigned to it and no one of them would know the whole. Putting these two premises together yields the conclusion that whatever it is in us that has a thought, the mind or soul, must be something that is simple and indivisible, and hence not subject to corruption or decay.

Recognizing that some thoughts are themselves extended allowed Hume to hoist this argument with its own petard.

> The free-thinker may now triumph in his turn; and having found there are impressions and ideas really extended, may ask his antagonists, how they can incorporate a simple and indivisible subject with an extended perception? All the arguments of the theologians may here be retorted upon them. Is the indivisible subject, or immaterial substance, if you will, on the left or on the right hand of the perception? Is it in this particular part, or in that other? Is it in every part without being extended? Or is it entire in any one part without deserting the rest? 'Tis impossible to give any answer to these questions, but what will both be absurd in itself, and will account for the union of our indivisible perceptions with an extended substance. (T 1.4.5.16; SBN 240)

How, Hume asked, could extended thoughts occur within an unextended soul? If just one part of the thought occurs within the soul then it does not know the whole thought. If every part occurs in the soul then, since the parts occur at different, adjacent locations in space, the soul must be spread out over these locations, and it must therefore itself be extended. To maintain that the soul manages to exist completely and entirely in every part of the extended thought at once without determent to its simplicity is absurd. There is simply no coherent way of accounting for how an extended thought might be "incorporated" with an unextended soul.

But, Hume proceeded to claim, not all thoughts are extended (our passions and our impressions of taste and smell are not extended, for example [T 1.4.5.9– 10; SBN 235–6). The proponents of the Achilles argument are right that these unextended thoughts cannot be incorporated with an extended soul. Since we have some thoughts that are extended and some that are not, and neither a material nor an immaterial soul can coherently be supposed to be the subject of all of them, we are compelled to reject the supposition that our thoughts inhere in any sort of substance, material or immaterial (T 1.4.5.33; SBN 250).

In this way, Hume's account of space and the spatiality of ideas leads, via a critique of the Achilles argument, to the rejection of the notion of a thinking substance and so sets the stage for Hume's alternative theory of the self in *Treatise* 1.4.6.

The Idea of Vacuum (T 1.2.5)

Hume maintained that the idea of vacuum is just a "fiction," in the technical sense described earlier. He took this to be a further consequence of his position on space.

According to that position, the space we perceive consists of points disposed in a certain fashion. But the points are unextended. If they lack extension, then they must possess some other property on pain of being nothing at all. Nothing cannot be disposed in any particular fashion or constitute a space by being so disposed.

Hume dealt with this worry by claiming that though the points are not extended, they do possess other properties. Some are colored. Others have tactile qualities. They are therefore something real and not nothing (T 1.2.3.13–16; SBN 38–9).

Hume took this to entail that there could be no true idea of vacuum or empty space. For, if points cannot exist, even in imagination, unless they are colored or tangible, and space can only be composed of points that exist and not of points that do not exist, then any space we sense and any idea of space we form in the imagination must be a plenum (T 1.2.4.2, 1.2.5.1; SBN 39–40, 53).

Admittedly, people do talk about vacuum, and so must have some sort of idea of it. But this idea is a fiction produced by an unnoticed conflation of two distinct ideas. One of these ideas is the idea of the kinaesthetic sensation we experience when we contract the muscles of our eyes or limbs. The other is the idea of two colored or tangible points separated by a number of intervening colored or tangible points.

The impressions these two ideas copy usually go together. We usually move our eyes or our limbs from one point to another while moving over a series of intervening points. And we discover a proportion between these two ideas. The longer the kinaesthetic sensations endure, the more points pass in review. But on some occasions we experience the kinaesthetic sensations without experiencing any succession of intervening points, as when moving our eyes through perfect darkness or our arms through a medium that offers no resistance. In these circumstances, the associative tendencies of the imagination, established by past experience, lead us to think of a quantity of intervening points of the sort that usually accompanies a kinaesthetic sensation of that duration (T 1.2.5.15–19; SBN 58–60). But no points are seen or felt (T 1.2.5.5–6; SBN 55–6). This sets up a conflict between our cognitive faculties. The imagination produces a vivacious idea of intervening points. The senses tell us that no such points exist. To simultaneously reconcile the conflicting ideas, we suppose that the points do exist, but that they are colorless or intangible. This, according to Hume, is patent nonsense, since a colorless

or intangible point is nothing at all and therefore cannot exist. But the nonsense is concealed by our tendency to confuse closely resembling ideas and use words without thinking of what they really mean (T 1.2.5.21; SBN 60–1).

Clever though this account is, it was motivated by a false difficulty. Hume supposed, perhaps rightly enough, that a point cannot exist unless it exhibits some quality that distinguishes it from nothing (T 1.2.4.2; SBN 39). But his own theory entails that points are not just colored or tangible. They are also disposed in a certain fashion within a compound impression. They therefore have a further real feature in addition to color or solidity: location.

When, for example, I look at the night sky and see two stars separated by a region of darkness, I do not have three impressions: one of one star, one of the other, and one of the contraction of the eye muscles as they move from the one to the other. I have one compound visual impression that consists of a number of parts, two of which are colored and disposed in such a way that they are separated from one another. But to recognize the colored points as disposed at separated locations within a compound impression is *ipso facto* to recognize the existence of a number of unoccupied locations between and around them.

There is an important difference, not recognized by Hume, between the experience that a blind person has of perfect darkness and the experience that a sighted person has of a number of colored points separated by perfect darkness. Hume considered the experience of the blind person to be an experience of nothing or not really an experience at all, and argued that therefore the experience a sighted person has of perfect darkness must be the same (T 1.2.5.5–5 and 11–13; SBN 55–6 and 57–8). But the experience of the sighted person is not just an experience of those points that have color, but of a compound impression consisting of all the locations around and between the colored points. These locations are unoccupied by anything colored or tangible. But that does not mean that they do not exist. An unoccupied location is still a location that exists relative to occupied locations, and therefore is not nothing.

Hume's Account of Contiguity
(T 1.1.5, 1.3.8, 2.3.7)

According to Hume's account, spatiotemporal structure is given. Impressions consist of parts that occur after and alongside one another (T 1.2.3.2–11; SBN 33–7). These spatial and temporal relations are not invented or even first discovered by the imagination (T 1.2.3.4 and 10; SBN 34 and 36–7). On the contrary, they are one of the factors that influence the imagination and determine the way it operates on subsequent occasions (T 1.3.8.5, 1.3.9.5, 1.3.13.1–2, 2.3.7–8; SBN 100, 109, 143–4, 427–38).

It is a law of the workings of the mind, discovered by observation, that once an impression consisting of various spatially adjacent parts has occurred, the imagi-

nation develops a tendency to imagine these parts in that conjunction. The same holds for impressions consisting of temporally successive parts. Accordingly, when a sensation occurs on a subsequent occasion, ideas that copy the previously contiguous sensations will tend to occur. Moreover, these ideas will so readily take the place of the actual sensation that a quantity of the vivacity of the original sensation will be confused with them, and they will therefore be more vivacious than ideas arising from pure fantasy.

Hume took this phenomenon to constitute independent empirical confirmation of his account of causal inference (T 1.3.8.5; SBN 100). The fact that other relations between sensations observed in the past can also be seen to communicate vivacity to associated ideas gives us a further reason to accept that constant conjunction (the relation at the root of causal inference) does so as well.

He also appealed to the influence that spatiotemporal contiguity relations have on the imagination to account for why certain patently irrational inferences are nonetheless psychologically very compelling (T 1.3.9.9, 1.3.13.1–2; SBN 110–11, 143–4) and for why passions and sympathy can vary in strength even though the circumstances producing them are otherwise identical (T 2.1.9.4–8, 2.1.11.6 and 14–17, 2.3.7–8; SBN 305–7, 318, 322–3, 427–38). The vivacity communicated to associated ideas by spatiotemporal contiguity relations can boost or diminish that communicated by constant conjunction, thus leading us to do such things as believe the conclusions of recent experiments more firmly than those established by an equal or greater number of older ones. Because sympathy and indirect passions such as humility involve an association of ideas, they are likewise intensified by contiguity.

But the influence of contiguity relations also poses a problem. Hume recognized that these relations are at best able to modify the strength of beliefs formed on other grounds, not induce belief on their own. He also recognized this to be the case with resemblance relations (T 1.3.2.2, 1.3.9.6 and 8–15; SBN 73–4, 109–10 and 110–15). But then why maintain that the one remaining natural relation, constant conjunction, produces belief?

Bell (2002) has claimed that this problem was so serious that it led Hume to abandon the view that belief is the product of a transmission of vivacity from impressions to associated ideas and maintain instead (in the *Enquiry*) that it is brute-factually instinctive, but Welton and I (2001) have contested this claim. Hume was able to resolve the problem adequately in the *Treatise* by invoking a number of salient differences between constant conjunction, on the one hand, and resemblance and contiguity, on the other (T 1.3.9.7; SBN 110). Where there is a constant conjunction, two types of impression are repeatedly witnessed to occur one after and one immediately alongside another. The double contiguity relation, together with the repetition of like cases, makes multiple links that can be expected to have a stronger effect on the operations of the imagination than the observation of a single, isolated case of spatial or temporal contiguity alone. The repetition also trains the imagination to develop a habit of

thought. Moreover, and in contrast, the contiguity and resemblance relations are diffuse, in the sense that many different impressions can all be contiguous to or can all resemble any given impression. As a consequence, the vivacity communicated from an impression to an associated idea by these relations is, as it were, dispersed in a number of different directions. In the case of constant conjunction, in contrast, one type of impression is related to just one other type of impression (except, of course, in the case of "inconstant causes" where a similar dispersal of vivacity occurs [T 1.3.11–12 esp. 1.3.11.9–13 and 1.3.12.19; SBN 124–42 esp. 127–30 and 137–8).

Notes

1　There must also be a boundary point at any division. A boundary point is an element that is either the last point of the part of the line before the division, or the first point of the part after, or both. This further condition is necessary to prove that setting points immediately adjacent to points will eventually surpass any finite bound, and for drawing a distinction between dense and continuous manifolds (dense manifolds do not satisfy it). See Huntington 1917, which provides a good introduction to all the mathematical notions discussed in this chapter.

References

Baxter, D. L. M. (2001) "Hume on steadfast objects and time," *Hume Studies* 27, pp. 127–46.
——(forthcoming) "Hume's theory of space and time in its skeptical context," in David Norton (ed.) *The Cambridge Companion to Hume*, 2nd edn. Cambridge: Cambridge University Press.
Bayle, P. (1697) *Dictionnaire historique et critique*. Rotterdam: R. Leers. Cited from P. Bayle (1991) *Historical and Critical Dictionary: Selections*, trans. R. H. Popkin. Indianapolis: Hackett.
Bell, M. (2002) "Belief and instinct in Hume's first *Enquiry*," in P. Millican (ed.) *Reading Hume on Human Understanding*. Oxford: Clarendon Press.
Broad, C. D. (1961) "Hume's doctrine of space," *Proceedings of the British Academy* 47, pp. 161–76.
Clarke, S. (1738) "A letter to Mr. Dodwell concerning the immortality of the soul," in *The Works of Samuel Clarke, D.D.*, vol. 3. London: John and Paul Knapton. Reprinted 1978 in facsimile, New York: Garland Press.
Falkenstein, L. (1995) "Hume and Reid on the simplicity of the soul," *Hume Studies* 21, pp. 25–45.
——(1997) "Hume on manners of disposition and the ideas of space and time," *Archiv für Geschichte der Philosophie* 79, pp. 179–201.

Falkenstein, L., and D. Welton (2001) "Humean contiguity," *History of Philosophy Quarterly* 18, pp. 279–96.

Flew, A. (1976) "Infinite divisibility in Hume's *Treatise*," in D. W. Livingston and J. T. King (eds.) *Hume: a Re-evaluation* (pp. 257–69). New York: Fordham.

Fogelin, R. (1985) *Hume's Skepticism in the* Treatise of Human Nature (ch. 3). London: Routledge.

—— (1988) "Hume and Berkeley on the proofs of infinite divisibility," *Philosophical Review* 97, pp. 47–69.

Franklin, J. (1994) "Achievements and fallacies in Hume's account of infinite divisibility," *Hume Studies* 20, pp. 85–101.

Frasca-Spada, M. (1998) *Space and the Self in Hume's* Treatise. Cambridge: Cambridge University Press.

Green, T. H. (1886) General introduction, in *David Hume: the Philosophical Works,* vol. 1. London: Longmans, Green. Reprinted 1964 in facsimile, Aalen: Scientia Verlag.

Hendel, C. W. (1963) *Studies in the Philosophy of David Hume* (pp. 409–41 and 498–504). Indianapolis: Bobbs-Merrill.

Holden, T. (2002) "Infinite divisibility and actual parts in Hume's *Treatise*," *Hume Studies* 28, pp. 3–25.

Huntington, E. V. (1917) *The Continuum and Other Types of Serial Order*, 2nd edn. Cambridge, MA: Harvard University Press.

Jacquette, D. (2001) *David Hume's Critique of Infinity*. Leiden: Brill.

Kemp Smith, N. (1941) *The Philosophy of David Hume* (pp. 272–346). London: Macmillan. Facsimile published 1983, New York: Garland Publishing.

Laird, J. (1932) *Hume's Philosophy of Human Nature*. London: Methuen.

Larivière, D. A., and T. M. Lennon (2002) "The history and significance of Hume's burning coal example: time, identity, and individuation," *Journal of Philosophical Research* 27, pp. 511–26.

McIntyre, J. L. (1994) "Hume: second Newton of the moral sciences," *Hume Studies* 20, pp. 3–18.

Mijuskovic, B. L. (1974) *The Achilles of Rationalist Arguments: the Simplicity, Unity, and Identity of Thought and Soul from the Cambridge Platonists to Kant*. The Hague: Martinus Nijhoff.

Pears, D. (1990) *Hume's System: an Examination of the First Book of his* Treatise. Oxford: Oxford University Press.

Russell, P. (1995) "Hume's *Treatise* and the Clarke–Collins controversy," *Hume Studies* 21, pp. 95–115.

Weinberg, J. (1965) "The concept of relation: some observations on its history," in Julius Weinberg, *Abstraction, Relation, Induction*. Madison: University of Wisconsin Press.

Further reading

Baxter, D. L. M. (1988) "Hume on infinite divisibility," *History of Philosophy Quarterly* 5, pp. 133–40.

Cummins, P. (1990) "Bayle, Leibniz, Hume and Reid on extension, composites and simples," *History of Philosophy Quarterly* 7, pp. 299–314.

Frasca-Spada, M. (1990) "Some features of Hume's conception of space," *Studies in History and Philosophy of Science* 21, pp. 371–411.

Grünbaum, A. (1968) *Modern Science and Zeno's Paradoxes.* London: Allen & Unwin.

Slownik, E. (2004) "Hume on length, space, and geometry," *Canadian Journal of Philosophy* 34, pp. 355–74.

Tweyman, S. (1974) "Hume on separating the inseparable," in W. B. Todd (ed.) *Hume and the Enlightenment* (pp. 30–42). Edinburgh: University of Edinburgh Press.

Van Steenburgh, E. W. (1977) "Durationless moments in Hume's *Treatise*," in G. P. Morice (ed.) *David Hume: Bicentennary Papers* (pp. 181–5). Austin: University of Texas Press.

Waxman, W. (1996) "The psychologistic foundations of Hume's critique of mathematical philosophy," *Hume Studies* 22, pp. 123–67.

5

Belief, Probability, Normativity

William Edward Morris

Although Hume gave the title, "Of knowledge and probability," to Part 3 of Book 1, he discusses knowledge only in its first section, and only cursorily. The real topic is *probability*. In Part 3, Hume is primarily interested in beliefs that go beyond our immediate perceptions and memories: beliefs that are the product of inferences from what we have observed to what we haven't observed.

Hume accepts the traditional absolute categorial distinction between knowledge (*scientia*) and belief (*opinio*).[1] Unlike our current conception of empirical knowledge as justified true belief, on this scheme, the two categories have different objects. In Hume's terminology, knowledge is concerned with *relations of ideas*; belief is concerned with *matters of fact*.

While Hume follows Locke in calling the category of belief, "probability," he argues that Locke's scheme isn't fine-grained enough. It requires us "to comprehend all our arguments from causes or effects under the general term of probability." But it isn't *merely probable* that the sun will rise tomorrow, or that all human beings will die, so "in common discourse we readily affirm, that many arguments from causation exceed probability, and may be receiv'd as a superior kind of evidence."

To capture this feature of ordinary language, Hume creates a special subdivision in the category of probability for "those arguments, which are deriv'd from the relation of cause and effect, and which are entirely free from doubt and uncertainty."[2] Hume calls these probabilities, "proofs." They are based on constant conjunctions, while (mere) probabilities – "that evidence, which is still attended with uncertainty" (T 1.3.11.2; SBN 124) – are based on variable or "irregular" conjunctions: rhubarb doesn't always purge; opium isn't always a soporific. But he notes that

> the gradation . . . from probabilities to proofs is in many cases insensible; and the difference betwixt these kinds of evidence is more easily perceiv'd in the remote degrees, than in the near and contiguous. (T 1.3.12.2; SBN 131)

Hume emphasizes that the distinction between proofs and probabilities isn't absolute: appropriate experiences may elevate what was formerly a probability to a proof, while others may demote a proof to a probability.[3]

Hume maintains that we are "determined by custom to transfer the past to the future in all our inferences." If we're "wise," we transfer events in the same proportions as we have experienced them in the past, assigning to each "a particular weight and authority" (EHU 6.1.4; SBN 58). But where we've experienced that different effects follow from what appear to be similar causes, we factor in all of them, in proportion as we have found them to be more or less frequent. For Hume, "A wise man . . . proportions his belief to the evidence" (EHU 10.1.4; SBN 110).

Hume's discussion of probability begins in section 2 and culminates in his account of the "Rules by which to judge of causes and effects" in section 15. The pivotal argument of Part 3, however, occurs in section 6, "Of the inference from the impression to the idea," where Hume argues that our inferences from the observed to the unobserved are not "determin'd by reason," but are solely products of custom and habit.

This argument, whose details are discussed in 6: CAUSATION, is perhaps the most familiar piece of philosophy ever written in English. Hume considered it the centerpiece of Book I, and perhaps of the entire *Treatise*, for on the title page of the *Abstract*, clearly referring to the *Treatise* as a whole, he called it "the CHIEF ARGUMENT of that BOOK" (A title page; SBN 641).

Long taken to be an entirely negative argument, it is largely responsible for Hume's reputation as a skeptic. On this reading, inaugurated by his contemporary, Thomas Reid, Hume exposed the latent skepticism in "the way of ideas" he inherited from Locke and Berkeley, in order to push empiricism to its logical, absurd limits. Thomas Hill Green ensured that this picture became orthodoxy by making it the focal point of his and Thomas Grose's edition of Hume's *Philosophical Works* (1874–5).

No Hume scholar today entirely accepts this traditional reading, but there is by no means a consensus as to how his famous argument should be read. Its structure, content, and conclusion are still widely debated. In addition to debates about its details, there is a further question about the argument's aim and intent, even for those who acknowledge that Hume's purpose isn't purely negative: If Hume is a skeptic, what is the nature of his skepticism?

Answering this question is made more difficult by the fact that Hume's contemporary readers use "skeptic" and "skepticism" in a variety of incompatible ways. And even though the terms "skeptic" or "skepticism" don't appear in Part 3 until much later (in T 1.3.13), Hume is nonetheless partially responsible for encouraging those who see his argument as in some sense skeptical. When he recast the argument for the first *Enquiry*, he called the section in which it appears, "Sceptical doubts concerning the operations of the understanding." In the *Abstract*, just after summarizing it, he states:

> By all that has been said the reader will easily perceive, that the philosophy contain'd in this book is very sceptical, and tends to give us a notion of the imperfections and narrow limits of human understanding. (A 27; SBN 657)

Whether Hume means by our "imperfections and narrow limits" nothing more than the cautious fallibilism, or "mitigated skepticism," he endorses in the *Enquiry*, or whether he has something stronger in mind, remains a source of considerable controversy.

Hume's argument is certainly *negative* in that it purports to show that our causal inferences aren't "determin'd by reason." But in arguing for this conclusion, he also maintains that our causal inferences have "no just foundation" (T 1.3.6.10; SBN 91), that there are no just inferences from either demonstration or probability that could yield the Principle of the Uniformity of Nature as a "just conclusion" (T 1.3.6.4–7; SBN 88–90), and that no "conclusions from causes and effects are based on solid reasoning" (T 1.3.6.8; SBN 90).

Hume's use of "just" conforms to two prominent uses of that term. In the *Oxford English Dictionary*, one prominent entry for "just" is "having reasonable or adequate grounds; well founded." Another closely related entry is "in accordance with reason, truth, or fact. Right; true; correct," as in this couplet from Alexander Pope's 1725 translation of Homer's *Odyssey*:

> Much he knows, and just conclusions draws,
> From various precedents, and various laws.

In using "just" this way, Hume is telling us that there are no inferences, foundations, or conclusions concerning reasoning from causes and effects that have adequate grounds or are otherwise well founded. In so doing, he moves beyond *description* to *prescription*. Hume doesn't endorse these inferences and he believes we shouldn't *shouldn't* as well.[4] In this sense, Hume's argument in T 1.3.6 has a negative normative conclusion. Understanding this feature of Hume's argument also helps us see how it might also be regarded as skeptical.

Locke's normative theory of probable reasoning in the *Essay concerning Human Understanding* sets out the principles by which he believes we "ought to regulate our Assent, and moderate our Perswasions" (Locke 1975: 1.1.3). It is a theory that Hume clearly has in mind throughout Part 3.[5] If Hume's argument is correct, Locke has failed to provide such a theory. Locke's theory isn't based on "just foundations" and "just inferences"; it yields no "just conclusions" based on "solid reasoning." A Lockean confronted with Hume's argument might well conclude that skepticism about probable reasoning is an unavoidable consequence of the argument.

But we *do* form beliefs as the result of causal reasoning. Hume's conclusion in the argument of T 1.3.6 also contains an account of *how* we form them. This signals a positive turn, with which he is mostly concerned in the remainder of Part 3.

Even Hume's positive theory of belief, however, is a subject of contemporary debate. Are Hume's results simply descriptive contributions to cognitive psychology, or do they sketch a normative epistemology? Hume's avowed objective in the *Treatise* – "An ATTEMPT to introduce the experimental Method of Reasoning into MORAL SUBJECTS" (T title page; SBN xi) in order to develop a science of human nature – seems purely descriptive. His discussion of belief, however, also seems to go beyond mere description, endorsing some patterns of belief formation while condemning others. He even specifies some "general rules by which we *ought* to regulate our judgment concerning causes and effects" (T 1.3.13.11; SBN 149; my emphasis). Thus his theory of belief seems to include prescriptive or normative elements. How do these elements square with his descriptive theory of belief formation? Is it even possible for a naturalistic theory like Hume's to be genuinely normative? And if it *is* possible, what is the *source* of normativity in his theory?

Hume's Theory of Belief

The nature of belief

The positive descriptive conclusion of Hume's argument in T 1.3.6 is "one part of the definition of an opinion or belief, that 'tis *an idea related to or associated with a present impression*" (T 1.3.6.15; SBN 93).

Hume's definition is partial because it is incomplete. As it stands, it fails to distinguish between mere conception and genuine belief. While an idea is an essential component in any belief, it can't be all there is to belief, since an idea is also involved whenever we conceive of something we don't believe. We need to be able to distinguish between simply conceiving of Petra and believing it to be an ancient Nabatean trading center located in what is now southwestern Jordan.

Hume maintains that in trying to make this distinction, the following dilemma is inevitable. Either the belief is some new idea, such as that of *reality* or *existence*, which we join to the simple conception of an object, or it is merely a peculiar *feeling* or *sentiment* (T Appendix 2; SBN 623).

Hume argues that we don't have an abstract idea of existence that is distinguishable and separable from our ideas of particular objects, so there is in fact no idea available to add to my idea of Petra to convert my conception of it into a belief. Further, since we can permute and combine ideas in imagination as we please, if belief consisted simply in adding a new idea to our original conception, we would be free to believe anything we choose to believe. Finally, if belief consisted in altering the parts or the composition of an idea, the resulting belief would be a different idea altogether from what we originally conceived (T 1.3.7.1–2; Appendix 2; A 19–22).

The situation is similar with distinguishing disbelieving and believing, or incredulity and belief. If belief added to or otherwise altered some of the *qualities* of the

idea involved, it would make disagreements about existence or other matters of fact impossible. If I believe that Petra was an ancient Nabatean trading center and you think it merely the product of legend, our disagreement depends on our having similar ideas of Petra (T 1.3.7.3; SBN 95).

Hume concludes that the differences between mere conception, incredulity, and belief must lie in the different *manners* in which we conceive the idea involved. He initially explains this difference in terms of force and vivacity. Beliefs are more forceful and vivid ways of conceiving ideas. To capture this difference, he maintains that "belief may be most accurately defin'd: A LIVELY IDEA RELATED TO OR ASSO-CIATED WITH A PRESENT IMPRESSION" (T 1.3.7.5; SBN 96).

But Hume realizes his definition isn't sufficient for an adequate account of belief. While having coffee with my colleague, Petra Visscher, I happen to think of the ancient Nabatean trading center whose name she bears. Here I have a lively idea of the city associated with present impressions of my friend, but in a way that fails to capture what Hume wants his definition to capture – the *connection* between the impression that generates the belief and the belief itself. My idea of Petra the city was already lively. The definition fails to explain how it *became* enlivened.

Hume responds by introducing a causal element into his account: beliefs are ideas enlivened by force and vivacity transmitted from an associated impression, making the idea almost as vivid as an impression:

> I wou'd willingly establish it as a general maxim in the science of human nature, *that when any impression becomes present to us, it not only transports the mind to such ideas as are related to it, but likewise communicates to them a share of its force and vivacity.* (T 1.3.8.1; SBN 98)

The impressions that enlivened my idea of Petra were those I received when I read Guzzo and Schneider's (2002) *Petra*. Before I read this book, I thought Petra was a legendary city. But taking *Petra* to reliably present factual information, my perceptions formed an associative bond with my idea of Petra, enlivening it. This makes the causal element explicit. Hume tacitly modifies his definition (first at T 1.3.7.6; SBN 97, then at T 1.3.8.14; SBN 105) to include it.

Because Hume, following Locke, uses "probability" as a name for the category of belief, any belief that isn't produced by the senses and memory is "determin'd by custom" in this manner, and

> as we call every thing CUSTOM, which proceeds from a past repetition, without any new reasoning or conclusion, we may establish it as a certain truth, that all the belief, which follows upon any present impression, is deriv'd wholly from that operation. (T 1.3.8.9; SBN 102)

Only through "custom operating on the imagination" can we "draw any infer-ence from the appearance of one [object] to the existence of another." The result

of such inferences is belief, and belief is a change in the manner in which the idea in question is conceived, which Hume in turn identifies with a change in feeling or sentiment. Therefore it follows, he adds provocatively, that "all probable reasoning is nothing but a species of sensation" (T 1.3.8.11; SBN 103).

Extending the theory

Although Hume thinks his theory of belief is the only "satisfactory and consistent explication" of our idea of belief, many of his readers have disagreed. They've criticized his reliance on force and vivacity as both empirically inadequate and incapable of doing the work the theory requires (see 3: IMPRESSIONS AND IDEAS). In his afterthoughts in the Appendix, even Hume himself seems ambivalent about the adequacy of his characterization of belief (T 1.3.7.7; SBN 628–9).

But careful readers of the *Treatise* should consider that the remaining sections of Part 3 considerably augment Hume's initial theory, providing it with additional resources that don't force him to rely exclusively on terms of phenomenal intensity, such as force and vivacity. Hume's presentation of his extended theory also shows that he doesn't take it to be dealing exclusively with belief formation. He is equally interested in the conditions under which beliefs are legitimately or illegitimately formed, which adds a prescriptive or normative dimension to the theory.

While my discussion cannot exhaustively examine all its elements, in what follows I sketch the central components of Hume's extended theory of belief, highlighting the prescriptive dimension he adds to it.

Testimony

Hume maintains that his definition of belief is "entirely conformable to everyone's feeling and experience." He illustrates the point with a contrast between reading a book *as* fiction and *as* history. Although I "receive the same ideas, and in the same order" in both cases, my "manner of conceiving" them is significantly different. I *relate* differently to them. In reading *Petra as* history, I *take* Guzzo and Schneider's words in a different way than I would have had I thought I was reading a work of fiction. What they say is not just phenomenally more intense for me; it has more "solidity," more "steadiness," more "influence."

Hume's emphasis on how we *take* the work signals that we take a more active role in belief formation than his initial account suggests.[6] Not only must I make a decision as to how *Petra* is intended, I must also consider its authors' "characters and motives." I must assess their competence, reliability, and accuracy. While we often make these decisions tacitly, doing so requires that we accept further testimonial evidence. My belief that *Petra* is trustworthy is largely based on my beliefs about its publisher, the University of Chicago Press, beliefs that depend on the testimony of others. Even with a reliable work, critically accepting its particular claims may require considerable reflection in assessing what others have said on

these matters as well as the authors' evidence for them. *Taking* is in this sense a far more complex process than is generally recognized.[7]

Assessment is evaluation, and evaluation contains a prescriptive or normative component. It is something I can do carefully, thoroughly, and critically, or carelessly, cursorily, and uncritically.

Hume emphasizes the normative element in evaluating testimony in his discussion of *credulity*, our "too easy faith in the testimony of others" (T 1.3.9.12; SBN 112). We have "a remarkable propensity to believe whatever is reported . . . however contrary to daily experience and observation," a "weakness" Hume thinks the influence of resemblance explains. Since words are intimately connected with ideas in the mind of the person giving testimony, and these ideas are connected with the facts or objects they represent, this latter connection "commands our assent beyond what experience will justify." But Hume argues that we are capable of correcting this "rash" tendency by regulating our acceptance of testimony by our "experience of the governing principles of human nature." That is, we can correct our initial rash judgment by reminding ourselves of our own propensity to want to believe and where this propensity, if unchecked, has led us in the past, as well as by checking the current testimony with our own experience and the experience of others, since "experience be the true standard of this, as well as of all other judgments" (T 1.3.9.12; SBN 113).

Not only do we take an active part in "receiving" and assessing testimony, it is a part we can play in a better or worse manner. Evaluation of testimony is itself subject to evaluation and thus criticism and correction. Others may criticize my assessments, and I may criticize theirs. Since assessment is criteria- or standards-based, it has a public component: my assessments aren't just what I *prefer*; they are what I *commend* to others as well. Since testimony is an unavoidable and pervasive cause of belief, assessing or evaluating testimony is equally unavoidable and pervasive, which necessarily brings a prescriptive or normative element into virtually all areas of our lives.

Two systems of "*realities*"

In addition to the reflective and normative elements involved in receiving and accepting testimony, Hume considerably augments his theory of belief by emphasizing the systemic elements involved in distinguishing belief in matters of fact from "other ideas, which are merely the offspring of the imagination."

In taking *Petra* to be "a true history," I not only assess the credentials of its authors and the accuracy of its claims, but also I take it to depict accurately what is real. I decide to give the ideas I get from it a place in the system of interconnected perceptions that form my picture of the world.

Whatever is present to the memory, Hume maintains, "must easily distinguish itself above the mere fictions of the imagination" because of its superior force and vivacity, "which resembles an immediate impression." We "form a kind of system"

of our memories. When it is enlarged by our present perceptions, this "system" constitutes what "we are pleas'd to call a *reality*."

Through custom and causation, our system of the memory and senses becomes intimately connected with a second system that "peoples the world, and brings us acquainted with such existences, as by their removal in time and place, lie beyond the reaches of the senses and memory." Since we feel that we are

> in a manner . . . necessarily determin'd to view these particular ideas, and that the custom or relation, by which [we are] determin'd, admits not of the least change, [we] form them into a new system . . . (T 1.3.9.3; SBN 108)

Because these ideas are involuntary, "precise," "solid and real," "certain and invariable," and "fixed and unalterable" (T 1.3.9.7; SBN 110), we also "dignify" this system "with the title of *realities*."

Petra is beyond the direct reaches of my senses and memory, but the perceptions I received in reading *Petra* have become connected through custom and causation with my idea of Petra. While these connections make my idea livelier through transfer of force and vivacity, the involuntary character of those connections reminds me that my idea is no mere "fiction of the imagination." These connections help *fix* the idea in my mind, giving it a firmness and solidity it didn't have before, as well as providing details I can't alter at will, making my idea more determinate. Now that I know that Petra was located in what is now southwestern Jordan, and that it flourished at the time of Alexander the Great, Petra *fits* into my general picture of the ancient world in a way it didn't before.

Nonetheless, Hume insists, "all this, and every thing else, which I believe, are nothing but ideas." Even so

> by their force and settled order, arising from custom and the relation of cause and effect, they distinguish themselves from the other ideas, which are merely the offspring of the imagination. (T 1.3.9.4; SBN 108)

Here "settled order" becomes at least as important as "force." With his emphasis on the systemic character of "*realities*," Hume introduces an element of *coherence* into his theory of belief. While ideas of objects are representative, and therefore capable of truth and falsity for Hume, their truth or falsity isn't decided by comparison with "objects," conceived as things that are independent of my perceptions. Deciding whether I regard a candidate for belief as true or false is to decide whether it fits – or *should* fit – into the settled order of my belief system. Making this decision sometimes requires that I reflectively assess the evidence for including it. Hume emphasizes the need for decision here by calling the second system of *realities*, "the object . . . of the judgment." For although causation is responsible for enlivening the ideas that form this second system, the causes are my perceptions of the words and images in *Petra* that constitute my testimonial

evidence about Petra's existence. Accepting this evidence, as we've seen, is not automatic. It involves assessment and evaluation, adding another element of prescriptivity or normativity to Hume's account.[8]

Legitimate versus illegitimate belief

At this point, Hume is satisfied that he has shown how his theory successfully distinguishes between genuine belief and "the mere fictions of the imagination." He is, however, aware that others may object, since on his theory,

> our assent to all probable reasonings is founded on the vivacity of ideas, it resembles many of those whimsies and prejudices, which are rejected under the opprobrious character of being the offspring of the imagination. (T 1.3.9.19n.; SBN 117)

Prejudices may be just as vivid as any other belief. "Philosophers" like Locke reject them as products of the imagination. But how can *Hume* reject them, since on his theory probable reasoning is also the product of the imagination?

In the footnote at T 1.3.9.19; SBN 117–18, Hume considers one way of distinguishing probable reasoning from "the offspring of the imagination" *within* the imagination: use "reason" to cover demonstrative and probable reasoning, and assign all the other deliverances of the fancy to "imagination." Hume admits that he has "often been oblig'd to fall into" using this distinction, but he makes clear that doing so is a serious mistake, for "nothing is more contrary to true philosophy than this inaccuracy."

There are at least two reasons why Hume thinks this distinction is "contrary to true philosophy." The first is that making the distinction in terms of "reason" gives the misleading and mistaken impression that he is committed to a *faculty* of reason, which he isn't (see Garrett 1997).

But a more serious problem is that Hume thinks that many of the beliefs "philosophers" dismiss as "offspring of the imagination" are products of probable reasoning themselves. This is nowhere more evident than in the way philosophers treat beliefs that arise from *education*. They reject education "as a fallacious ground of assent to any opinion" (T 1.3.10.1; SBN 118), because it is "an artificial and not a natural cause," one whose "maxims are frequently contrary to reason, and even to themselves."

Hume's response is that beliefs arising from education are "built almost on the same foundation of custom and repetition as our reasonings from causes and effects" (T 1.3.9.18; SBN 117), so there is no principled way of rejecting them on the basis of the mechanisms that formed them. If we are to reject false, unreasonable, and even inconsistent beliefs that arise from education, we need to find some other way of picking them out.

Hume's discussion of "unphilosophical probabilities" further reinforces his response. After considering "probabilities of chance" in 1.3.11 and "probabilities of causes" in 1.3.12, he begins the next section by remarking:

> ALL these kinds of probability are receiv'd by philosophers, and allow'd to be reasonable foundations of belief and opinion. But there are others, that are deriv'd from the same principles, tho' they have not had the good fortune to obtain the same sanction. (T 1.3.13.1; SBN 143)

The "others" include the products of prejudice, bias, salience, and of course "education." They are *unphilosophical* because philosophers don't endorse them. They are *probabilities* because they involve the transfer of force and vivacity from a present impression to the ideas involved. They always have "a considerable influence on the understanding" and not infrequently exert "a superior influence on the judgment" (T 1.3.13.1–2; SBN 143).

Locke, of course, would deny that unphilosophical probabilities are derived from the same principles as the probabilities of chances and causes. His account of probable reasoning in the *Essay* treats chances and causes as products of the understanding and therefore "reasonable foundations of belief and assent," while the associations Hume calls unphilosophical probabilities are merely deliverances of the imagination. "Association," for Locke, "is as frequent a cause of mistake and error in us as anything else that can be named, and is a disease of the mind as hard to be cured as any" (Locke 1993: 101). He regards association as an unnatural process akin to "madness," which "gives Sence to Jargon, Demonstration to Absurdities, and Consistency to Nonsense" (Locke 1975: 2.23.18; see also Locke 1993: 106).

But Hume has already argued that probable inferences can't be established by reason, but only by the very associative mechanisms Locke dismisses. Hume explains the probabilities of chances and causes in terms of the operations of those same mechanisms. "All reasonings," he stresses repeatedly, "are nothing but the effects of custom" (T 1.3.13.11; SBN 149). Philosophical and unphilosophical probabilities are thus both "deriv'd from the same principles." This doesn't mean that Hume endorses the products of education, bias, and prejudice, but it does mean that he must find another way to reject them. Hume's account of general rules provides the vehicle that shows how he can distinguish beliefs we should endorse from those we should reject. Prejudices, Hume maintains, result from our tendency to "rashly" form general rules, as when someone concludes, from a few cases, that Irishmen are witless. But

> shou'd it be demanded why men form general rules, and allow them to influence their judgment, even contrary to present observation and experience, I shou'd reply that, in my opinion it proceeds from those very principles, on which all judgments concerning causes and effects depend. (T 1.3.13.8; SBN 147)

Custom operates, not just when there is constant conjunction, but when objects *resemble* those we have experienced. The weaker the resemblance, the weaker the belief, but custom still has some force as long as traces of resemblance remain.

Paradoxically, even though custom is "the foundation of all our judgments," still it sometimes affects the imagination "in opposition to the judgment," producing "a contrariety in our sentiments concerning the same object."

A Lockean might argue that this points to a fatal flaw in Hume's account. On Locke's theory, there is no problem in explaining a conflict between the two faculties of judgment and imagination: probable reasoning caused the judgment, while association caused the enlivening of the imagination. But since there is nothing but custom on Hume's account, why doesn't the most vivid idea simply win out? How can there really be a conflict?

Hume responds:

> In almost all kinds of causes there is a complication of circumstances, of which some are essential, and others superfluous; some are absolutely requisite to the production of the effect, and others are conjoin'd only by accident. Now we may observe, that when these superfluous circumstances are numerous, and remarkable, and frequently conjoin'd with the essential, they have such an influence on the imagination, that even in the absence of the latter they carry us on to the conception of the usual effect, and give to that conception a force and vivacity, which make it superior to the mere fictions of the fancy. (T 1.3.13.9; SBN 148)

Consider someone securely strapped into an ascending ski lift chair.[9] He can't control his vertigo because his fear of falling, reinforced by the enlivening effect of his fear upon the idea of falling, powerfully affects his imagination. "The circumstances of depth and descent strike so strongly upon him" that he is convinced he will fall, despite "the contrary circumstances of support and solidity, which ought to give him a perfect security" (T 1.3.13.10; SBN 148). Hume calls this application of custom and experience to his present situation "the first influence of general rules." He thinks it is possible for the person to resolve the "contrariety" that results from the circumstances of depth and descent on the one hand, and his awareness of the safety and security of his situation on the other, by "reflection on the nature of those circumstances." Reflection leads him to correct his initial conviction. When we compare that judgment "with the more general and authentic operations of the understanding, we find it to be of an irregular nature, which is the cause of our rejecting it" (T 1.3.13.12; SBN 150).

Hume calls this "a second influence of general rules, alluding to the general rules he introduces in T 1.3.15 as "rules by which to judge of causes and effects" (hereafter, "c&e rules"). He describes them as "rules that are form'd on the nature of our understanding, and on our experience of its operations in the judgments we form." The eight rules Hume sketches in T 1.3.15 range from summaries of the conditions for causation (rules 1–3) and a statement of what Mill calls "the method of concomitant variation" (rule 7), to variations on the theme that "like causes like" (rules 4–6 and 8). Hume admits that the rules have very little specific content, so little that it is appropriate to regard them as "formal" (Dauer 2000). Applying them

"requires the utmost stretch of human judgment," because the rules "are very easy in their invention but extremely difficult in their application" (T 1.3.15.11; SBN 175). They are difficult to apply because they demand that those who use them pay close attention to the particular circumstances of the case at hand, and make careful judgments about the similarities to and differences from other cases.

Nevertheless, if the skier is "wise," when he applies the c&e rules Hume thinks he will realize that his initial reaction was the product of an "irregular" projection of his experience, "irregular" because it ignored relevant circumstances in the situation – the "support and solidity" the lift provides. Had he taken these circumstances into consideration at the beginning, he would have made a different judgment. Applying the c&e rules thus results in a "more authentic," regulated judgment.

But both his initial reaction and his revised reflective judgment arise from "the influence of general rules"; both are products of the operation of the associative mechanisms, and are derived solely from custom and experience. "Sometimes the one, sometimes the other prevails, according to the disposition and character of the person. The vulgar are commonly guided by the first, and wise men by the second" (T 1.3.13.12; SBN 150). However, when "our general rules are set in opposition to each other," as in the skier's case, "philosophers" attribute the initial inference to the imagination because it is "more capricious and uncertain," and the second to the judgment, "as being more extensive and constant." But in Hume's view, nothing is "more contrary to true philosophy," since both general rules are the products of the same mechanisms.

Because of this, Hume is aware that his account will appear paradoxical, even skeptical, to traditional philosophers:

> Mean while the skeptics may here have the pleasure of observing a new and signal contradiction in our reason, and of seeing all philosophy ready to be subverted by a principle of human nature, and again sav'd by a new direction of the very same principle. The following of general rules is a very unphilosophical species of probability; and yet 'tis only by following them than we can correct this, and all other unphilosophical probabilities. (T 1.3.13.12; SBN 150)

Hume, of course, doesn't think he has really unearthed "a new and signal contradiction in our reason," although to a philosopher like Locke it might appear that he has. He has explained, in his own terms, how the alleged conflict between judgment and imagination arises, and how it is sometimes resolved. If he is correct, then the attempts of philosophers to provide an absolute means of partitioning off the legitimate principles of belief-formation are doomed to failure. Whether we regard the partition as being between two faculties, or between two compartments of the imagination, the partitioning itself isn't sufficient to determine which beliefs are legitimate and which are not.

Hume's explanation, however, invites another question. He rejects certain general rules, such as the belief that all Irishmen are witless, while endorsing the c&e rules. What is the basis for his prescriptive preference for them?

For Hume, no a priori principles of abstract reasoning can successfully deter-mine how we should regulate our causal judgments, which is why

> our scholastic headpieces and logicians shew no superiority above the merely vulgar in their reason and ability, as to give us any inclination to imitate them in delivering a long system of rules and precepts to direct our judgment, in philosophy. (T 1.3.15.11; SBN 175)

Since "any thing may produce any thing" (T 1.3.15.1; SBN 173), the only way we can determine whether something is "really" a cause or an effect is through experience, which takes us to the heart of Hume's often misunderstood naturalism.

Hume intends us to take the c&e rules as informal codifications of the causal inferences that we regard as having been successful in our collective experience. He believes they are reliable guides to forming more fine-grained beliefs. They help us "learn to distinguish the accidental circumstances from the efficacious causes" (T 1.3.13.11; SBN 149), by which we can accurately distinguish the cir-cumstances that "are absolutely requisite to the production of the effect," from others "only conjoin'd by accident" (T 1.3.13.9; SBN 148). This is why Hume believes that the c&e rules are

> all the LOGIC I think proper to employ in my reasoning; and perhaps even this was not necessary, but might have been supply'd by the natural principles of our under-standing. (T 1.3.15.11; SBN 175)

Making causal judgments – whether tacitly or explicitly – in accordance with the c&e rules is the only way, Hume believes, that we can accurately distinguish beliefs that are formed legitimately from those that are not.

Normativity

The c&e rules are clearly norms for causal reasoning. Hume doesn't merely *describe* how we regulate our conduct: he *endorses* certain ways of forming beliefs and *rejects* others. Then he goes one step further and recommends that *we* endorse them as well. He characterizes the c&e rules as "general rules, by which *we ought* to regulate our judgments concerning causes and effects" (T 1.3.13.11; SBN 149; my emphasis).

But there is still a question about whether his theory is *genuinely normative*. Is Hume correct in thinking that we too should place our confidence in the c&e rules and regulate our judgments in accordance with them?

To decide the issue, we must appeal to some normative perspective from which we assess the norms for causal reasoning and either endorse or reject them. But since we have no access to the world except through the process of causal

reasoning, there is no transcendent point of view from which we can assess these norms of causal reasoning. The only point of view from which we can assess the norms of causal reasoning is that of causal reasoning itself.

While he agrees there is "undeniable evidence of normativity" in Part 3, David Owen (1999) has argued that Hume's account is normative only in the limited sense that it provides standards for correct and incorrect ways of making causal judgments. The c&e rules are criteria for the right way to participate in the practice of probable reasoning, but they don't tell us why we should prefer that practice to others, such as superstition.

Owen draws an analogy with the practice of augury, where Roman priests observed and interpreted sacrificed sheep's entrails in order to make predictions that directed public policy. Although there are criteria for when the entrails have been read correctly and when they haven't, those standards don't tell us why we should prefer augury to other methods of prognostication.

For Hume, however, the questions Owen tries to distinguish – the question of the standards governing causal reasoning and the question of the reasons for preferring the practice of causal reasoning to other practices, such as augury – aren't really separate. Even though augury is practiced according to a set of standards, reading the entrails and making predictions from them is itself a form of causal reasoning. The practice aims at accurate prediction, so it should be judged by the standards that govern causal reasoning. Even the best–conducted auguries will be "irregular," for they systematically ignore relevant circumstances experience has shown to be significant for successful prediction.

Similar irregularities are found in liars, in the superstitious, and in beliefs that are inculcated through "education." Pathological liars may eventually believe their lies, but their beliefs don't fit into a coherent and stable system of "realities." The superstitious uncritically accept the testimony of unreliable sources and questionable authorities. Educators may say what they will when indoctrinating their hapless captive pupils, but what they say can ultimately be rejected if it is inconsistent, or runs counter to well-founded causal reasoning.

By acting in accordance with the c&e rules, explicitly or tacitly, we refine our causal expectations in the light of experience. Successful predictions will be brought into our system of "realities." When our expectations are "methodized and corrected" in the light of further experience and by their fit with a set of coherent and well-confirmed beliefs, they become at first practical wisdom and eventually the basis for Humean science. Little more is needed for "the wise" to reject belief in miracles, the products of education, and unreliable testimony. Prejudices, whimsies, and superstitions may die hard, but their failure to satisfy the norms of causal reasoning will weaken their hold. Once we make a habit of regulating our beliefs by the c&e rules, there is little chance our prejudices, whimsies, and superstitions will be "reinvigorated" (Rawls 2000).

Fortunately, this process is not something we must do alone. Our range of observations and experiences is widened considerably by testimony, as well as by

cooperative, collaborative, and collective endeavors in everyday practical matters, and for those whose talents run in that direction, in a scientific community. The results of these endeavors are, in a very real sense, a social construction.

But natural as all this may appear, it is not a simple process, especially as our causal beliefs become more refined and complex. Hume acknowledges and accommodates this complexity:

> All the rules of this nature are very easy in their invention, but extremely difficult in their operation; and even experimental philosophy, which seems to be the most natural and simple of any, requires the utmost stretch of human judgment. There is no phenomenon in nature, but what is compounded and modify'd by so many different circumstances, that in order to arrive at the decisive point, we must carefully separate whatever is superfluous, and enquire by new experiments, if every particular circumstance of the first experiment was essential to it. These new experiments are liable to a discussion of the same kind; so the utmost constancy is requir'd to make us persevere in our enquiry, and the utmost sagacity to choose the right way among so many that present themselves. (T 1.3.15.11; SBN 175)

Satisfying these standards, fallible and dynamic as they are, will nonetheless ultimately suffice to undermine a wise person's misplaced confidence in prejudices, whimsies, and superstitions. Since scientific and superstitious expectations are products of the same sorts of processes, they stand or fall by the same standards. Hume provides us with a set of standards by which science succeeds and superstition fails. If he is correct, there is no *separate* problem of providing support for our practice of causal reasoning over and above the problem of providing standards for causal reasoning itself.

The only way to test causal reasoning is by causal reflection. When we reflect on the origins and processes of causal reasoning, beliefs that we formed in accordance with the c&e rules will increase in their force and settled order, while those that were not so formed lose force and stability. Beliefs formed in accordance with the c&e rules will also cohere with my "system of *realities*," while those that are not will not have that "fixed and unalterable" character. Applying reflective causal reasoning to our reflection-produced standards for causal reasoning doesn't undermine our confidence in those standards, it can only increase our confidence in them. In this sense, Hume's account of the norms of causal reasoning is genuinely normative.[10]

See also 6: CAUSATION; 9: HUME'S CONCLUSIONS IN "CONCLUSION OF THIS BOOK".

Notes

1 See Hacking (1975, esp. chs. 1–4, 19), for a helpful discussion of the pervasiveness of this distinction in early modern philosophy, and for a provocative thesis about how

the transformation in the notion of *opinio* made possible the emergence of the skeptical problem of induction.

2   Hume takes this distinction over into the first *Enquiry*, where he defines proofs as "such arguments from experience as leave no room for doubt or opposition," explicitly mentioning "Mr. Locke" (EHU 6.1 n. 1; SBN 56 n. 1).

3   Hume emphasizes this feature of proofs and probabilities in section 10, "Of miracles," of the first *Enquiry* more than he does in the *Treatise*. However, it is generally agreed that section 10 is very similar, if not identical, with a portion of the text Hume deleted from the manuscript of the *Treatise* when "castrating my work, that is, cutting off its nobler parts; that is, endeavouring it shall give as little offence as possible" (Hume 1969, I: 6, letter to Henry Home) to (Bishop) Joseph Butler, in hopes of receiving a favorable reading from one of the leading moral philosophers of his day. In that letter, Hume mentions that he is enclosing "some *Reasonings concerning Miracles*," which he says he has removed from the manuscript. It is generally agreed that the section would have appeared in Part 3 of Book 1, although opinions differ about where it would have been placed. David Wootton (1990) places it at the end of 1.3.13, "Of unphilosophical probabilities," but I agree with Traiger's (1994) conjecture that a more plausible location is in or following 1.3.9, "Of the influence of belief," where there seems to be a clear allusion at T 1.3.9.4; SBN 120 to these arguments. Hume says, in a discussion of credulity, belief, and the passions, that we are fascinated with the "magnificent pretensions" and "miraculous relations" of "quacks and projectors," to the extent that our "astonishment" "so vivifies and enlivens the idea, that it resembles the inferences we draw from experience." He then adds: "This is a mystery, with . . . which we shall have farther occasion to be led into in the progress of this treatise."

4   *Pace* Don Garrett, who argues that Hume's conclusion in the argument of 1.3.6 "is a claim in cognitive psychology, not in evaluative epistemology," and that Hume doesn't engage in evaluative epistemology until the very end of 1.4.7 (Garrett 1997: 214; see also 9: HUME'S CONCLUSIONS IN "CONCLUSION OF THIS BOOK"). While I agree that *part* of Hume's conclusion is a claim about how our causal expectations are formed, which may thus be regarded as a claim in cognitive psychology, I find it difficult to see how Hume's repeated assertions that there is no "just foundation" in reason and no "just inference" to a "just conclusion" for our causal inferences, bear a purely descriptive reading. David Owen defends a view similar to Garrett's by maintaining that these phrases simply refer to the production of belief by reason or inference (Owen 1999: 139–40 n. 38). Loeb (2002: 19 n. 23; 43–4 n. 13) and I, for similar reasons, find this reading of "just" implausible.

5   While it is clear that Locke is perhaps Hume's most prominent target in Part 3, his arguments are generic enough that their effectiveness isn't limited to Locke alone. As Morris (1988) and Baier (1991) have emphasized, Hume's aim is broad enough to cover any "intellectualist" attempt to show that our causal inferences are "determin'd by reason," and thus include not only Locke but also his rationalist predecessors.

6   Traiger (1994: 249) also stresses the importance of "how one takes words or utterances" in "fixing the relation between impressions of language and the ideas which we form," even down to "the most basic 'taking' – taking a sound or set of marks as a word."

7 Traiger (1994) is of course a notable exception.

8 The importance of Hume's often-neglected discussion of the "two systems of *realities*" is given its due in Loeb (2002: chs. II.1, II.3, and III.1), who emphasizes the role of stability in the two "systems."

9 Hume's example is of the "familiar instance" of someone who is "hung out from a high tower in a cage of iron" and who "cannot forebear trembling, when he surveys the precipice below him, tho' he knows himself to be perfectly secure from falling" (T 1.3.13.10; SBN 148). The updated, ski lift version of the example is due to Traiger 2005.

10 Annette Baier has argued that the c&e rules are the result of a "successful turn of self-consciously sensitive and custom-dependent causal reasoning on itself, a reflexive turn that leads to endorsement, as rules, of those habits that had survived the test of reflexive employment" (1991: 97). The inferences "we endorse are the ones that can become successfully reflexive. *Successful reflexivity is normativity*" (pp. 99–100). Baier, however, does not really explain how the reflexive endorsement test works.

Don Garrett, on the other hand, maintains that "while it is clear that" the c&e rules "are rules for engaging in induction, and that they are produced and justified by their inductive success, any overall endorsement of induction is at best implicit" (1997: 249 n. 2). Garrett argues that Hume makes no normative claims until very late in Book 1 (at T 1.4.7.11; SBN 270), where Hume announces what Garrett dubs "the Title Principle": "Where reason is lively, and mixes itself with some propensity, it ought to be assented to. Where it does not, it never can have any title to operate upon us" (Garrett 1997: ch. 10, esp. pp. 232–7. See also 9: HUME'S CONCLUSIONS IN "CONCLUSION OF THIS BOOK"). If I understand Garrett correctly, the reason induction is not normatively successful is not just that Hume offers no overall endorsement of induction in Part 3, it is because one central component of the "dangerous dilemma" Hume constructs in T 1.4.7 challenges the account of Part 3. Garrett believes that the Title Principle successfully defuses the threat posed by the dangerous dilemma, giving Hume a basis for "a skeptical recommitment to reason." I am not convinced that the Title Principle has enough content to play this role. For an alternative reading of Part 4, which doesn't read Hume as challenging his results in Part 3, see Morris 1989, 2000a, and 2000b.

References

Baier, Annette C. (1991) *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge, MA: Harvard University Press.

Dauer, Francis W. (2002) "Humean naturalism and the problem of induction," *Ratio* 13(2), pp. 123–37.

Garrett, Don (1997) *Cognition and Commitment in Hume's Philosophy*. New York: Oxford University Press.

Guzzo, Maria Giulia Amadasi, and Eugenia Equini Schneider (2002) *Petra*, trans. Lydia G. Cochrane. Chicago: University of Chicago Press.

Hacking, Ian (1975) *The Emergence of Probability*. Cambridge: Cambridge University Press.

Hume, David (1969) [1932] *The Letters of David Hume*, 2 vols., ed. J. Y. T. Grieg. Oxford: Clarendon Press.

Locke, John (1975) *An Essay concerning Human Understanding*, ed. Peter H. Nidditch. Oxford: Clarendon Press.

——(1993) [1706] *Of the Conduct of the Understanding*. Bristol: Thoemmes Press.

Loeb, Louis E. (2002) *Stability and Justification in Hume's* Treatise. New York: Oxford University Press.

Morris, William Edward (1988) "Hume's refutation of inductive probabilism," in James H. Fetzer (ed.) *Probability and Causality* (pp. 43–77). Dordrecht: Reidel.

——(1989) "Hume's scepticism about reason," *Hume Studies* 15, pp. 39–60.

——(2000a) "Humean reason and the problem of warrant," *Hume Studies* 26(2), pp. 305–21.

——(2000b) "Hume's Conclusion," *Philosophical Studies* 99, pp. 89–110.

Owen, David (1999) *Hume's Reason*. Oxford: Oxford University Press.

Rawls, John (2000) *Lectures on the History of Moral Philosophy*, ed. Barbara Herman. Cambridge, MA: Harvard University Press.

Traiger, Saul (1994) "Beyond our senses: recasting Book I, Part 3 of Hume's *Treatise*," *Hume Studies* 20(2), pp. 241–59.

——(2005) "Reason unhinged," in Joyce Jenkins, Jennifer Whiting, and Christopher Williams (eds.) *Persons and Passions: Essays in Honor of Annette Baier* (pp. 100–16). South Bend, IN: University of Notre Dame.

Wootton, David (1990) "Hume's 'Of miracles': probability and irreligion," in M. A. Stewart (ed.) *Studies in the Philosophy of the Scottish Enlightenment* (pp. 191–229). Oxford: Clarendon Press.

Further reading

Bennett, Jonathan (2001) *Learning from Six Philosophers*. Oxford: Clarendon Press.

Falkenstein, Lorne (1997) "Naturalism, normativity, and skepticism in Hume's account of belief," *Hume Studies* 23(1), pp. 29–72.

Korsgaard, Christine M. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.

Passmore, John (1968) [1952] *Hume's Intentions*. London: Duckworth.

Stroud, Barry (1977) *Hume*. London: Routledge.

6

Causation

Abraham Sesshu Roth

The part of Hume's *Treatise* that we look to when we're interested in his views about causation turns out to be not so much about causation *per se*, but about causal reasoning of one form or another. Hume seems to think that the concept or *idea* of causation figures in this sort of reasoning. What is the idea, and what role, if any, does it play in this reasoning? These questions concern Hume's psychology and epistemology.

We might also ask, in a more metaphysical vein, about Hume's views concerning causation itself. Is it that objects or events are causally related only insofar as they are instances of a regularity in the course of nature? Must there be, rather, some further connection between the objects for one to count as a cause of the other – a connection that is not presented in experience, but which underlies the regularity of objects or events presented in experience?

Hume's definitions of "cause" would seem to speak against attributing to him any such realist view attributing "secret powers" or "ultimate principles" in a cause that would account for its effect. The first definition is in terms of regularity:

> We may define a CAUSE to be "An object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedency and contiguity to those objects, that resemble the latter." (T 1.3.14.31, 35; SBN 170, 172)

The second definition adverts to dispositions of the mind:

> A CAUSE is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other. (T 1.3.14.31, 35; SBN 170, 172)

These definitions of "cause" are not equivalent; indeed, they do not even yield the same extensions (Robinson 1962). However, there is some hope of making the

extensions coincide by reading the second "definition" as a subjunctive conditional that makes reference to an *ideal* mind assumed to be in ideal circumstances.[1]

In any case, neither definition gives support to anything but a regularity reading of causation. The first obviously does not. And it becomes clear that the second does not either, once we recognize that the disposition in question is shaped by exposure to just the sort of regularity mentioned in the first definition.

These definitions notwithstanding, there has been much debate over Hume's metaphysics of causation.[2] But whatever evidence there may be for attributing to Hume a robust conception of causation involving "secret" connections or powers lies more with the *Enquiry* than with the *Treatise*. Since our concern here is with the latter, my focus will be on the aforementioned epistemological and psychological issues arising from Hume's views about causal inference.

Causal Inference and Justification

Hume observes that in order for two objects to be related as cause and effect, they must be successive and contiguous. He goes on to ask,

> Shall we then rest contented with these two relations of contiguity and succession, as affording a compleat idea of causation? By no means. An object may be contiguous and prior to another, without being consider'd as its cause. There is a NECESSARY CONNEXION to be taken into consideration; and that relation is of much greater importance, than any of the other two above-mention'd. (T 1.3.2.11; SBN 77; see also T 1.3.6.3; SBN 87)

But what is this necessary connection? Hume's usual empiricist procedure for getting clear on some idea is to examine the impression from which it derives. Our sense impressions of causally related objects, however, reveal nothing beyond contiguity and succession (T 1.3.6.3; SBN 88; T 1.3.14.16; SBN 163). But in contrast to his discussion of the philosophical notion of substance (T 1.4.3.6–8; SBN 221–2), Hume does *not* conclude that there is no impression and no corresponding idea of necessary connection. Instead, Hume starts "beating about the neighbouring fields" with the aim of finding the missing impression (T 1.3.2.13; SBN 78). What we get as a result is a very lengthy discussion of *probable reasoning*.

Probable reasoning is so called by Hume to contrast with the certainty of demonstrative reasoning found for example in mathematical deduction. Two features of probable reasoning are important for our purposes (T 1.3.2.2; SBN 73–4; T 1.3.6.7; SBN 89). First, it is the basis for all of our beliefs in unobserved matters of fact, i.e. those ideas that go beyond what is given in current experience or yielded by memory. For example, I see footprints, and infer that some creature has walked by recently. Or I notice an egg rolling toward the edge of the table

and then predict that it will fall and break on the floor. The second notable feature of probable reasoning is that it involves relations of cause and effect: the footprints are the *effect* of the animal's having walked by, the egg breaks *because* it strikes the floor. No such inferences can be made unless the objects in question are taken to be causally related. Without a causal connection between them, a footprint would yield no evidence, justification, or basis for concluding that an animal had passed through. For Hume, probable reasoning *is* causal reasoning.

It *appears* that Hume emphasizes necessary connection as a part of the idea of causation because of the *justificatory role* he sees for it in causal or probable inference. He says,

> the only [relation], that can be trac'd beyond our senses, and informs us of existences and objects, which we do not see or feel, is *causation*. This relation, therefore, we shall endeavour to explain fully before we leave the subject of the understanding. (T 1.3.2.3; SBN 74)

Notice that Hume doesn't think that reasoning merely *brings about* beliefs regarding matters of fact beyond what has been perceived. Reasoning, rather, *informs* us of those matters of fact. This suggests that the reasoning potentially amounts to justification for the beliefs that it yields. This interpretation receives some corroboration from another passage where Hume says that the causal relation is the only relation that gives "*assurance* from the existence or action of one object, that 'twas follow'd or preceded by any other existence or action" (T 1.3.2.2; SBN 73–4, my emphasis). If causation figures centrally in reasoning that *informs* and *assures* us of some conclusions, then it's hard not to think that Hume's point is that causation (or the idea thereof) has a role in *warranting* or *justifying* these conclusions for us.

Shortly thereafter (T 1.3.2.11, SBN 77 quoted above), Hume asserts that necessary connection is the most important element of the idea of causation – presumably because of its role in probable reasoning. This becomes more explicit later on, when Hume says that necessary connection is the foundation of our inference from cause to effect (T 1.3.14.21; SBN 165). So Hume seems to be advocating a semantic claim regarding our concept or idea of causation, namely, that necessary connection is a part of it. This claim seems to rest on the following:

(J) The causal relation plays a justificatory role in probable reasoning, and can only do so if it is understood as involving necessary connection.

The Traditional View: Skepticism and Unreflective Inference

We are in for a surprise, then, when we turn to perhaps the most important part of the discussion of probable or causal inference, namely, section 1.3.6 ("Of the

inference from the impression to the idea"). There, Hume defends two theses about this inference, one positive and one negative, each of which appears to be at odds with the justificatory role accorded to necessary connection.

The negative thesis seems to reveal a deep skepticism about probable reasoning. Hume argues that causal inference and all the resulting beliefs are not based on any reason, that it is "impossible for us to satisfy ourselves by our reason, why we shou'd extend that experience beyond those particular instances, which have fallen under our observation" (T 1.3.6.11; SBN 91). Such a skepticism undermines the rationale for having necessary connection as part of our concept of causation. Necessary connection was to play a justificatory role in probable inference. But now we have no need to locate such an element in this reasoning – because we now know that *nothing* will legitimize (in any epistemic sense) the causal inferences that constitute this reasoning. Thus, even if our prior understanding of causation is one that includes necessary connection, it now seems possible – indeed, advisable – to replace it with a conception of causation that would dispense with necessary connection. (Hume also expresses skepticism about probable reasoning later on, in T 1.4. But the tension between the alleged skepticism of T 1.3.6 and (J) is especially problematic, since 1.3.6 falls within the ambit of the discussion the express purpose of which is to articulate the notion of necessary connection presented as essential for causation.)

The positive thesis of T 1.3.6 is that causal inference is immediate: nothing intervenes between an impression given in experience and the idea inferred. The transition is described in a way that makes it sound mechanical. Moreover, the image of this cognitive behavior as immediate and reflexive is reinforced by Hume's explanation of it as the upshot of a natural disposition combined with exposure to a constant conjunction of objects in experience (T 1.3.6.2–3; SBN 87; T 1.3.6.15; SBN 93; see also T 1.3.8.10–13; SBN 102–4). Thus, if our experience is that of an A always followed by a B, then when we perceive or think of A, the idea of B will immediately come to mind. Like the negative thesis, this positive thesis also appears to be in tension with (J). For if the inferential transition is as immediate and as unreflective as this thesis appears to suggest, there seems not to be *any* role for the idea of necessary connection to play, let alone the justificatory role (J) would accord to it.

The traditional strategy for resolving these tensions is simply not to take at face value the remarks about the importance of necessary connection for causation. On this view, (J) is a provisional claim, introduced for rhetorical purposes only. It is presented as a part of tradition or a common prejudice, and can be discarded once Hume's skeptical and psychological insights are appreciated.

The traditional understanding can be challenged on two grounds, corresponding to how it interprets the negative and positive parts of Hume's discussion here at T 1.3.6 and elsewhere. The negative thesis that probable inference is not based on reason is traditionally understood to be a defense of skepticism about probable reasoning. But T 1.3.6 has more recently been the subject of an extensive and rich

literature that seeks to show that Hume is not the skeptic that tradition has made him out to be.

Second, the traditional view interprets Hume's positive thesis to be that causal inference is, without exception, the result of automatic habit. This is taken to imply that the inference from one object to the other is *always* immediate. While this is not meant to rule out a *chain* of such inferences, there is no room in this picture for a role to be played by the idea of causation or necessary connection. The second challenge notes, on the contrary, that Hume *does* acknowledge causal inferences that are much less immediate than the traditional view would allow. And it contends that these more deliberate probable inferences suggest a role for necessary connection in causal inference.

Both of these challenges to the traditional understanding need to be investigated. I turn first to the challenge that focuses on Hume's positive picture of probable inference via custom/habit.

Causal Inference is not always Immediate and Unreflective

It is undeniable that Hume often describes causal inference as the automatic and unreflective exercise of habit built up through experience of constant conjunction.[3] This prompts some interpreters to conclude that necessary connection (or the idea thereof) is epiphenomenal, playing no role in probable reasoning.[4] According to this view, the later remarks of 1.3.6, presented subsequently to the introduction of the thesis about the role of custom and habit, express what Hume really thinks should be a part of our idea of causation.

The traditional interpretation is surely right up to a point; Hume does think that there are causal inferences that are immediate and unreflective, and that these inferences offer little opportunity for necessary connection to play a substantive role. But it is mistaken to think that Hume described *all* causal inferences in this way. Some inferences are not immediate, and there are several reasons for thinking that Hume saw a role for the idea of necessary connection amongst these.[5]

First, even late in his discussion, Hume asserts that necessity is the basis of the inference (T 1.3.14.21; SBN 165). But more worrying for the traditional view is the following puzzle. If Hume thinks that our causal inferences are just the product of custom mechanically conceived, then it's not clear why he's concerned (as he seems to be in T 1.3.6 and elsewhere) to find a *connection* between objects we take to be related as cause and effect. Why would we need to discern or be shown a connection between the objects if we're just automatically making a transition from one to the other – that is, if it's just a basic or primitive psychological fact that such a transition occurs?

Hume's picture, rather, seems to be that without some sort of connection, causal inferences, *or at least some significant class of them*, would not be undertaken.

Although the ideas are not related in such a way as to be connected by the understanding, we can draw the inferences because of their connection in the imagination. Thus,

> Had ideas no more union in the fancy than objects seem to have to the understanding, we cou'd never draw any inference from causes to effects, nor repose belief in any matter of fact. (T 1.3.6.12; SBN 92)

It might be objected that when Hume speaks in this context of a connection or union of ideas, he means nothing more than the fact that there is an association or transition between the ideas in the imagination. But the problem with identifying connection with association or transition is that Hume seems to want to *explain* at least *some* instances of the transition between ideas *in terms of the connection between them*. Hume says,

> When the mind, therefore, passes from the idea or impression of one object to the idea or belief of another, it is not determin'd by reason, but by certain principles, which associate together the ideas of these objects, and unite them in the imagination . . . [T]he principles of union among ideas, I have reduc'd to three general ones, and have asserted, that the idea or impression of any object naturally introduces the idea of any other object, that is resembling, contiguous to, or *connected* with it. (T 1.3.6.12–13; SBN 92, my emphasis)

For the sort of association of ideas we have in causal inference, the relevant relation is connection, which Hume elsewhere refers to as cause and effect. And this connection will hardly explain the transition if it simply *is* the transition. So we cannot defend the traditional interpretation of Hume on causal inference by identifying connection with transition or association. We still need an explanation of why Hume would be interested in a connection, and the traditional interpretation that the inference is automatic and immediate has nothing to say about this.

I should reiterate that Hume holds that some causal inferences *can* be described, more or less correctly, as automatic and immediate. For these inferences, the association in the imagination is all there is to the inference; it would be accurate enough to say that connection plays no explanatory role in these cases. But the interest in a connection suggests that Hume thought that there are *other* causal inferences (associations or unions of ideas in the imagination) that cannot be described as automatic and immediate, and for which the idea of a connection is significant.

A final consideration against the traditional reading of Hume on causal inference is the ample evidence that Hume indeed thinks that there are instances of probable reasoning/causal inference that fail to fit the model of unreflective, automatic habit. This is particularly evident in Hume's discussion of reasoning about probabilities of causes. For this sort of reasoning, Hume explicitly describes

a model of immediate and unreflective inference based on custom, something he calls "imperfect habit." While *some* inferences proceed this way, Hume says

> this method of proceeding we have but few instances of in our probable reasonings; and even fewer than in those, which are deriv'd from the uninterrupted conjunction of objects. (T 1.3.12.7; SBN 133)

Rather, we usually proceed in a way that is more reflective:

> we commonly take knowingly into consideration the contrariety of past events; we compare the different sides of the contrariety, and carefully weigh the experiments, which we have on each side: Whence we may conclude, that our reasonings of this kind arise not *directly* from the habit, but in an *oblique* manner; which we must now endeavour to explain. (T 1.3.12.7; SBN 133)[6]

Reflective Causal Inference

Causal inference thus is not always the automatic and unreflective upshot of habit. Hume recognizes much causal reasoning to be more reflective and sophisticated. This suggests an alternative to the traditional understanding, one that does accord to the idea of necessary connection a role in causal inference. To see better what this role might be, we need to examine further what Hume takes to be involved in reflective causal inference.

Reflective causal inferences are set apart from those that are immediate and automatic by the reasoning subject's concern with the *evidence* or *reasons* for the conclusions drawn. Care, resourcefulness, and persistence are generally required to discern the evidence, reasons, or justification needed for arriving at what the subject takes to be a legitimate conclusion. One cannot engage in this sort of endeavor without having some notion of *argument* or *inference* that connects the evidence or reasons recognized as such with the conclusion. Conscious and reflective causal inference requires the subject to take there to be a connection between objects related as cause and effect so as to be able to draw an inference from one to the other. He or she must, in effect, be able to construct an argument (or something like it) with a premise concerning one, and a conclusion concerning the other.

Taking there to be some such intelligible connection does not, of course, entail that the inference is legitimate or justified. I could be mistaken in thinking that one idea is (sufficient) evidence or reason for the other, or that there is anything like a valid or otherwise legitimate argument from one to the other. Indeed, it may even be that inductive skepticism holds sway and no causal inference is legitimate. The connection underlying the inference, then, is not so much a relation between the ideas themselves. To the extent the ideas are connected, it is only because of my taking there to be some sort of evidential, justificatory relation

between them. *Without my taking there to be a rational or evidential connection between cause and effect, I will not be able to draw the inference, at least not reflectively.* And if I happen to draw the inference in a more immediate fashion but upon reflection do not take there to be a connection, I would then withdraw the inference, or suspend belief in the conclusion.

In short, the connection Hume has in mind is that of (taking there to be) an intelligible connection, i.e. an argument, or something being a reason for something else. The resulting inference is understood by the subject as normative, and not as a mere mechanical transition between ideas or beliefs. If a reasoner is not equipped with relevant ideas or concepts of *argument*, *justification*, *evidence*, or *something being a reason for something else*, etc., then he or she will be unable to undertake the sort of reflective and deliberate causal inference that Hume thinks constitutes much of the probable reasoning that engages us.

Let me hasten to add that the sort of connection that Hume has in mind is not the demonstrative variety found, for example, in mathematics, which he would later place under the heading of relations of ideas. It is not clear *why* Hume thinks that we commonly take it to be that a non-demonstrative form of intelligibility is involved in causal inference. After all, Hume rightly thinks that many of his philosophical predecessors take the connection to be demonstrative. But *that* it is Hume's view that we commonly take the connection as non-demonstrative is clear from where he looks – or, rather, where he doesn't look – to find the impression corresponding to the idea of connection. He doesn't look to mathematical or other demonstrative forms of reasoning as yielding the idea of necessary connection.

So, causal inferences of the reflective sort involve the subject taking there to be an intelligible albeit non-demonstrative connection underwriting our causal inferences. An inductive skepticism puts into question this kind of intelligibility. But irrespective of the matter of skepticism, it seems that a satisfactory psychological account of *reflective* causal inference would require the reasoning subject to possess some such concept of connection.

It is at this point that one might wonder whether the conceptual resources necessary for reflective causal inference might reside in our idea of necessary connection. If this is indeed the case, then it should become clearer why one would want to maintain, as against the traditional understanding, that Hume's early remarks in 1.3 concerning the importance of necessary connection for causal inference are not merely provisional. That is, (J) would represent a view to which Hume is committed.

According to Hume, of course, to think of evidence or non-demonstrative argument in terms of a necessary connection between objects involves a distortion stemming from its misapplication (T 1.3.14.14, 25; SBN 162, 167). Exactly how we are to understand this "projective error" is a difficult matter, but *part* of the mistake, I think, is that the seemingly epistemological notions of evidence, argument, or something being a reason for something else, are confused for the notion of a metaphysical relation holding between objects.

If necessary connection does in fact play a role in causal inference, there will be constraints on how we are to interpret Hume's discussion in 1.3.14. There, Hume identifies the source of our idea of necessary connection *not* as an impression of sensation (which results from the sense organs and is taken to indicate something outside of us), but as an impression of reflection (an inner impression comprising emotions and passions). He further characterizes the impression as a *determination of mind*. (Indeed, when Hume discusses the projective error (T 1.3.14.25; SBN 167) he says that it is this determination of mind which is misapplied to objects; he does not explicitly mention the idea of non-demonstrative evidential connection.) It is not at all obvious what Hume could have meant by a "determination of mind," nor is it clear how exactly to fit it into his taxonomy of the mental. So it's unclear just *which* impression he thought was the source of the idea of necessary connection. This issue cannot be addressed here. But any such interpretive undertaking will require one to keep in mind that, if the current suggestion is correct, the resulting idea of necessary connection – that is, the idea of non-demonstrative reason or argument – must play an important role in reflective probable reasoning. Meeting this condition would pose a difficulty for the familiar proposal that would identify the determination of mind with a simple, unanalyzable feeling, on analogy with the "intrinsic," non-representational quality some find in the experience had e.g. when one looks at a red object in normal conditions (see Stroud 1977: 86).

What was Hume arguing in *Treatise* 1.3.6?

I turn to the second challenge to the traditional view. The traditional view, recall, seeks to reconcile the tensions in T 1.3 by arguing that the remarks attributing a justificatory role to necessary connection are merely provisional and ultimately rejected in favor of the skepticism defended in the negative part of 1.3.6. The second challenge contends that Hume is *not* arguing there that causal inferences are unwarranted; there is no skepticism in this section that would put in question the justificatory role of necessary connection. But is the reading of Hume that underwrites this challenge correct? And if Hume is not arguing for skepticism, what is he trying to do here?

In the relevant text, Hume argues that we cannot draw a causal inference from an object given in experience (the impression) to one unobserved (the idea) unless we've had experience of a constant conjunction between those sorts of objects (T 1.3.6.2–3; SBN 87). Hume next asks whether experience produces this transition by means of reason or understanding on the one hand, or by an association of ideas in the imagination on the other (T 1.3.6.4; SBN 88). He continues,

> If reason determin'd us, it wou'd proceed upon that principle, *that instances, of which we have had no experience, must resemble those, of which we have had experience, and*

that the course of nature continues always uniformly the same. In order therefore to clear up this matter, let us consider all the arguments, upon which such a proposition may be suppos'd to be founded; and as these must be deriv'd either from *knowledge* or *probability*, let us cast our eye on each of these degrees of evidence, and see whether they afford any just conclusion of this nature. (T 1.3.6.4; SBN 89)

Hume proceeds to rule out both sorts of evidence or argument for the *principle of uniformity*. Knowledge (or *certainty* as Hume sometimes calls it; T 1.3.1.2; SBN 70) consists of what is known either *intuitively* (immediately apprehended because of features intrinsic to the ideas themselves), or *demonstrably* (discoverable through a chain of propositions/ideas, each link of which is intuitively comprehended). But the conceivability, hence possibility, of nature not being uniform shows, for Hume, that the principle of uniformity is neither intuitively nor demonstrably known.

The principle of uniformity is not founded on probability, either. Probability here is understood broadly to cover causal reasoning in general. (This includes both cases for which our experience makes us certain enough to count them as proofs, and those probabilities [in the narrow sense] for which some doubt remains [T 1.3.11.2; SBN 124].) As causal or probable inference is drawn from experience, it presupposes the very principle of uniformity that we're seeking to establish as a "just conclusion" (T 1.3.6.4; SBN 89). Hume concludes that this principle cannot "arise" from probability, saying that the "same principle cannot be both cause and effect of another" (T 1.3.6.11; SBN 90).

The upshot, then, is that the principle of uniformity is derived neither from knowledge nor from probability. It is therefore unavailable for use by reason or the understanding in basing conclusions on experience. Thus, "even after experience has inform'd us of their *constant conjunction*, 'tis impossible for us to satisfy ourselves by our reason, why we should extend that experience beyond those particular instances, which have fallen under our observation" (T 1.3.6.11; SBN 91).

Let us consider some of the readings that have been offered of this famous discussion.

The Skeptical Reading

Hume asks whether reason, given the experience of constant conjunction, determines one to draw a causal inference. According to the skeptical reading, Hume's concern here is with epistemological warrant, justification, and rationality. This much is suggested by his effort to seek "evidence," or a "foundation," for the principle of uniformity, to see whether it is a "just conclusion" (T 1.3.6.5–6; SBN 89). And the outcome of his investigation is skeptical: reason does not determine the transition; we cannot reasonably ("by reason") account for drawing inferences based on experience (Flew 1959; Stove 1973; Stroud 1977; for criticism of Stove, see Morris 1988).

But there are reasons not to regard Hume as a straightforward skeptic about causal inference. First, the remarks about the "science of man" (T Introduction 6–7; SBN xvi–xvii), as well as the discussion in T 1.1.1, suggest that Hume sees himself as engaged in a scientific enterprise consisting fundamentally of just the sort of ampliative reasoning the skeptical interpretation would have Hume regard as unwarranted: that which draws from experience particular and general conclusions that go beyond that experience. Second, there are a number of places where Hume says things that at least imply that causal inference *can* be warranted. For example, he discriminates between good and bad inferences in contrasting philosophical and unphilosophical probability in T 1.3.8.1; SBN 143. And in T 1.3.15, he offers rules by which to judge of cause and effect, and by which we may know when objects really are related by cause and effect (T 1.3.15.2; SBN 173). These rules are offered so that we are better able to draw inferences, and to "choose the right way among so many that present themselves" (T 1.3.15.11; SBN 175). Even within T 1.3.6, Hume says of the relation of cause and effect that "'tis the only one, on which we can found a just inference from one object to another" (T 1.3.6.7; SBN 89).

Though unsympathetic to the skeptical view, Garrett (1997: 79) argues that Hume's use of causal reasoning is not conclusive evidence against the skeptical interpretation. After all, it's also a part of Hume's view that even if unwarranted, we naturally and inevitably will draw these inferences. Hume's reasoning in the *Treatise*, it is suggested, is just a manifestation of this psychologically irresistible tendency. But against this, there is nothing in Hume's psychological theory that would require him to write a book filled with causal reasonings. However much his causal reasonings might be inevitable in ordinary life, the *Treatise* need not have exhibited them.

Why, then, would Hume present causal reasoning in the *Treatise* if he were skeptical and not compelled to do so? Millican (2002, citing Noxon 1973) suggests an answer, one that might also account for Hume's discrimination between good and bad inferences. It is a matter of methodological consistency. Causal inference is natural and inevitable: to the extent that one is presented with the "evidence," then one will draw inferences as a result of this natural tendency. It is not inevitable for one to seek out all the evidence or to think through all the implications. Some, such as philosophers and scientists, are inclined to follow through on this sort of reasoning. Others are too busy or unable to carry on with it beyond the extent to which they are so determined. But the inability or unwillingness to follow through on this sort of reasoning and, for example, seek out as much evidence as one can on all varieties of matters would count as *rational failing* (Millican 2002: 165, citing Hume's *Dialogues* [1949: 134]). Hume's distinction between good and bad (or less good) reasoning is based on this sort of consideration. And this is also supposed to explain why Hume would also exhibit probable reasoning throughout the *Treatise*.

However, it is unclear how Hume, skeptically interpreted, could be entitled to the aforementioned judgment of rational failing. It may be natural and inevitable to think in a certain way given a particular context (i.e., those circumstances where

one has faced some body of experience). But it doesn't follow that it would be a rational failing not to seek out more such contexts in which to display what would be natural and inevitable in such contexts. Moreover, this sort of answer seems not to account for the guileless and unironic language both in Hume's discussion of the Rules in T 1.3.15 and the remarks concerning a new science of man in the *Treatise* introduction. (However, for further defense of the more traditional skeptical view, see Winkler 1999.)

The difficulties facing the skeptical interpretation can be seen as motivating non-skeptical readings of the negative argument in T 1.3.6. Let us consider some of them.

The Anti-rationalist Reading

According to this reading, Hume is simply seeking to show that the kind of warrant involved in probable or causal inference is not the sort to be had in reason construed in some more narrow, demonstrative, sense (Beauchamp and Rosenberg 1981; Arnold 1983; Baier 1991; and in part Broughton 1983). On one version, we might try to understand Hume's notion of demonstration as that of a deductively valid argument with a priori premises. On this view, when Hume says that *reason would "proceed upon" the principle of uniformity*, he is saying that if reason determined us, the principle of uniformity would be a premise in a demonstrative argument to establish that experience serves as evidence upon which to form beliefs about unobserved matters of fact. Such an argument would be a foundation for our causal inferences or probable reasoning. According to this view, Hume's point is that there is no such demonstrative argument, and therefore *reason* does not determine us. The anti-rationalist view concludes that causal or probable inference must therefore confer a *different* sort of warrant than that had in demonstrative argument (i.e., reason, narrowly construed).[7]

The objection to this approach is that it fails to account for a central element in Hume's argument. If Hume's point is simply to show that probable or causal inference is not founded on a deductive argument from a priori premises, why would he go to the trouble of considering a *probable* justification for the principle of uniformity and rule it out on grounds of circularity? Such an argument is manifestly unsuited to serve as a rationalist foundation of probable argument. That Hume was willing to consider this sort of non-demonstrative argument demonstrates that his aim was not or not merely to refute a rationalist foundation for probable inference.

The Causal/Explanatory Reading (1)

Non-skeptical interpretations contend that Hume in T 1.3.6 is not really concerned with epistemological issues of warrant and skepticism. His aim, rather, is

psychological or explanatory: it is to explain what is responsible (in the causal sense) for the probable reasoning that we undertake. (See Garrett 1997; Owen 1999; Broughton 1983 (for a hint of this); Connon 1979; Noonan 1999) More directly relevant for our purposes, the *negative* part of the discussion seeks to rule something out as a cause or explanation of why we make these inferences. Thus, when Hume says that reason does not determine us to make the inference, he's not saying that the inference is unreasonable, unjustified, or unwarranted. He's just saying (on one version) that a particular faculty – that of reason or the understanding – does not issue in or cause the inference. This is meant to leave open whether the inference is legitimate. As Garrett (1997: 94–5) has put it, what is being argued for is a thesis in cognitive psychology. If Hume's concerns are simply psychological, then we avoid the sorts of interpretive problems faced by the skeptical reading.

Before I elaborate upon this view, we should observe that it would be mistaken to think that Hume's interest in T 1.3.6 is purely psychological or causal, and not at all epistemological. While there is reason (as we have seen above) not to read Hume as straightforwardly skeptical, this doesn't entail that we should read Hume as not even *concerned* with warrant in T 1.3.6. Hume might, after all, be posing in the negative part of that discussion a skeptical challenge *that he thinks he can answer*. So, Hume's endorsement of a "science of man," his discrimination between good and bad reasoning, and the other obstacles to reading Hume as a skeptic leave open the possibility that the negative discussion of T 1.3.6 *is* concerned with warrant, and *does* mean to raise a skeptical challenge.

Owen (1999: 136) tries to show that the negative part of the discussion must be, fundamentally, causal/explanatory and *not* concerned with warrant (or the lack thereof). He argues that Hume's positive, psychological story of the inference simply *would miss the point* if the negative part of the story were concerned with warrant/justification. The positive story is that the association of ideas from custom or habit is causally responsible for probable reasoning. If telling this story isn't simply changing the subject, then the negative story that precedes it must have been saying what is *not* the source or cause of probable inference and belief.[8]

However, there might be a way to read the negative story as concerned with warrant and skepticism, while reconciling it with the positive story being psychological. We might think that the skepticism concerning Hume in this part of the *Treatise* is a skepticism stemming from an *absence* of a working psychological model for the sort of cognition in question. Suppose that the only available model of cognition is the broadly perceptual model that applies to demonstrative reasoning, and which Hume inherits from Locke (Owen 1999; Millican 2002). According to the causal/explanatory interpretation, in showing that the understanding is not responsible for causal inference, Hume is in effect showing that this perceptual model fails to apply to probable reasoning. This then raises worries about the epistemic credentials of the sort of cognition – in this case, probable/causal

inference – for which one had hoped to have a psychological model. If probable reasoning falls outside of the only available model of cognition, then a fear is that what we took to be probable belief and reasoning is not the systematic output of any single type of cognition but, rather, a random collection of beliefs and inferences stemming from disparate and unrelated prejudices, fantasies, biases, delusions, etc. By analogy, our skepticism about clairvoyance has much to do with the absence of any plausible psychological model for how one's allegedly clairvoyant beliefs systematically correspond to matters of fact for which one has no standard causal and sensory access. The absence of such a model can lead us to conclude that there is no such correspondence: that the beliefs are false or unreliable, or that they are beliefs of mundane provenance wrongly taken to be clairvoyant.

Now, if *this* is the source of skeptical concern about causal inference underlying the negative discussion of T 1.3.6, then it would be entirely appropriate to respond with a psychological model for this form of cognition. And since Hume does exactly that, it would be charitable to read in this way whatever concern with warrant and skepticism there may be in T 1.3.6. It should be noted, however, that there are different sources and kinds of skeptical worry, and that not necessarily all of them will be addressed by a psychological model of probable reasoning. I think that whatever concern with skepticism that might be found in T 1.3.6 and T 1.3 more generally is reconcilable with – indeed calls for – the sort of psychology that Hume presents.[9] It may be that there are other parts of the *Treatise*, such as 1.4, where Hume expresses skeptical worries that might not be addressed by the proposed psychology.[10]

The Causal/Explanatory Reading (2)

While epistemological issues such as warrant and skepticism do figure in the negative argument in T 1.3.6, views such as those defended by Garrett and Owen make a strong case for thinking that Hume *is* fundamentally concerned there with defending a (negative) thesis that could be described as causal, explanatory, or psychological. There is, however, a serious objection to this sort of view. According to this criticism, *if* Hume is trying to establish a causal thesis, then his argument would be clearly inadequate, and in a way that should and would have been apparent to Hume.

To see this, we need to fill out the causal/explanatory interpretation in some more detail. Hume is trying to show that the faculty of reason or understanding does not cause us to draw the relevant inferences. For if it did, it would have to "proceed upon" the uniformity principle. The first thing to notice is that implicit here is the thought that if reason or understanding is responsible for the inference, there must be some sort of *argument* to support drawing the inference. This *supporting argument* will be understood differently depending on which version of the causal explanatory view one defends. According to Garrett (1997), it is a

meta-argument for the legitimacy of the inductive practice in general (though Garrett [1998, 2001] refines and generalizes the proposal so that supporting arguments need not be limited to those *about* inductive inference). Alternatively, Owen (1999) thinks of the supporting argument as part of the first-order argument connecting some relevant body of experience with the drawing of the particular inference on this occasion.

However we understand the structure and function of the supporting argument, Hume's strategy is to show that there cannot be any such argument, and hence reason/understanding is not responsible for the inference. Thus, Garrett (citing Fogelin 1985) has referred to the negative discussion of T 1.3.6 as the "no-argument argument."

Hume contends that the principle of uniformity would have to figure in any supporting argument underlying causal inference (perhaps as a premise, or intermediate idea or step). Hume argues, however, that there is no satisfactory way to establish this principle so that it may be put to service in the supporting argument. Hume considers and rejects the only two forms of evidence or argument that can be given for the principle. Any *probable* argument for the principle would be circular, presupposing the very principle we're seeking to establish. Thus, such an argument for the principle of uniformity would not explain why someone engages in probable reasoning, for it presupposes that they already do. So the principle cannot "arise from probability." And the fact that we can conceive the denial of the principle of uniformity shows, for Hume, that there is no demonstrative argument "to prove" it. The principle of uniformity cannot be established, and so the supporting argument cannot get off the ground. Hence, reason or the understanding is not responsible for the causal or probable reasoning we undertake.

But there is a serious gap in Hume's argument, at least as understood by proponents of the causal/explanatory account (Millican 2002). Conceivability considerations are used by Hume to rule out any good demonstrative arguments for the principle of uniformity, and consequently, to rule out good demonstrative arguments that might serve as the supporting argument underlying the inference. But who is to say that the understanding or reason is moved only by *good* arguments? Might there be some *bad* argument that nevertheless causes reason to draw the probable inference? Hume has not ruled out this possibility. Thus, if Hume's interest is simply to identify what does or does not *cause* us to undertake probable inferences yielding probable belief, then his argument is clearly inadequate. It may be that Hume just missed this point. However, it is more likely (and more charitable) to think that the problem lies with the causal/explanatory interpretation of Hume's argument.

One response to this criticism suggests that an *infallibilist* conception of reason or understanding figures in Hume's argument. Hume therefore need not consider *bad* arguments when he is trying to argue that reason is not responsible for probable inferences. The infallibility of the understanding already rules out the possibility that it would be swayed by bad arguments.

But Hume does not think that the faculty of reason or understanding is infallible. He says,

> In all demonstrative sciences the rules are certain and infallible; but when we apply them, our fallible and uncertain faculties are very apt to depart from them, and fall into error . . . Our reason must be consider'd as a kind of cause, of which truth is the natural effect; but such-a-one as by the irruption of other causes, and by the inconstancy of our mental powers, may frequently be prevented. (T 1.4.1.1; SBN 180)[11]

Thus, even if reason is understood as a faculty whose *proper* function can attain something approaching infallibility, Hume recognizes that its performance can fall far short of the ideal, and that there is a very real possibility for error. Thus, if Hume wants to rule out reason or understanding as causally responsible for our probable inferences, he must consider the possibility of error and malfunction. That is, he must rule out the possibility that probable inferences are undertaken because the faculty accepts a *bad* argument. And he doesn't seem to do this in the negative argument of T 1.3.6. So, if we hold the causal/explanatory interpretation, it seems that we must admit that there is a gap in Hume's argument – a gap, moreover, that Hume could not have missed.

One promising avenue of response to this important criticism would be to reply that Hume *does* in effect consider the possibility that bad arguments might move reason/understanding to undertake causal inferences. Hume does not explicitly address the possibility. But his use of conceivability considerations to see whether there is a legitimate argument for the principle of uniformity presumably employs the very same faculty of reason or understanding that would construct or accept such an argument. Hume thinks that when we use our faculty of reason, we arrive at the result that there is no good argument for the principle. Notice, however, that a bad argument that's convincing for the faculty would also have to appear good from the perspective of that faculty. So, if, in employing the faculty, no argument appears good enough, then no argument – good or bad – is compelling for the faculty. Thus, in employing the (fallible) faculty to show that there are no good arguments, Hume will have in effect canvassed and eliminated the possibility that the understanding might *accept* a bad argument for the principle of uniformity.

As it stands, this reply on behalf of the causal/explanatory interpretation raises questions that cannot be addressed here. For instance, why think that Hume held that one could have a test for *compelling* arguments? I take this to follow from the fallibility of the understanding, and something that Hume is committed to: the conceivability test for possibility. It does seem that when the faculty judges that a *particular* argument is good or bad, this will in effect be a test for whether that particular argument is compelling or not for the faculty. But Hume needs a stronger and more *general* conclusion, namely, that there is no compelling argument.

It's not yet clear that the faculty's general conclusion that there is no good argument will establish that there is no compelling argument.

So, there are obstacles for taking this approach in reply to the Millican criticism. But the causal/explanatory interpretation has significant merits. Moreover, this reply on its behalf is intriguing and has not, to my knowledge, been considered in the literature. It warrants further investigation.

Traditionally, Hume's epistemology of causal inference has been interpreted as skeptical, and his psychology of causal inference as countenancing only inferential transitions that are uniformly immediate and mechanical. Neither claim sits well with the thought, suggested in T 1.3.2 and elsewhere, that the idea of necessary connection has an important role to play in causal inference. But doubts about the traditional view can be raised on both counts. There is a great deal of textual evidence to show that Hume allowed for reflective and deliberate probable reasoning. And while the negative discussion of T 1.3.6 surely reflects some concern with epistemological issues of warrant and skepticism, there is reason to think that Hume's theory of causal inference, within T 1.3.6 but also in T 1.3 more generally, is fundamentally causal and explanatory.

See also 5: BELIEF, PROBABILITY, NORMATIVITY; 9: HUME'S CONCLUSIONS IN "CONCLUSION OF THIS BOOK."

Notes

1 See Garrett 1997 for a sophisticated proposal along these lines. The relation between the two definitions is discussed in many places, including: Beauchamp and Rosenberg 1981; Russell 1984; Costa 1986; Craig 1987; Baier 1993; and Blackburn 1993.
2 See Wright 1983; Strawson 1989. For criticism, see Winkler 1991. A number of articles by figures central to this debate are collected in Read and Richman 2000.
3 T 1.3.6.12; SBN 92. Immediate and automatic: T 1.3.6.2; SBN 87; T 1.3.6.14–15; SBN 93; esp. T 1.3.7n.; SBN 97n. Custom: T 1.3.7.6; SBN 97; T 1.3.8.10–13; SBN 102–4; T 1.3.14.31; SBN 170; T 1.3.16.8; SBN 178. Habit: T 1.3.16.19; SBN 179.
4 Noonan (1999: 153) gives a particularly strong statement of this view. Stroud holds that Hume is not really in a position to attribute a role to necessary connection (Stroud 1977: 88, 226–32, esp. 227, 230). Owen (1999: 152–3) also seems to rule it out, though at p. 128 he comes close to giving it a role. Garrett (forthcoming) bucks this trend and identifies a role for the *impression* of necessary connection to play, though along lines quite different from what will be suggested below.
5 Owen (1999: 149) notes the distinction between reflexive and reflective inferences, though not to find a role for necessary connection to play in inference. However, he does see a role for the principle of uniformity in more reflective inferences (p. 142).

6   Further evidence for inference that is not immediate lies in Hume's discussion of causal inference based on a single case (T 1.3.8.14; SBN 104–5), and his treatment of inferences involving judgments of causality (T 1.3.15.11; SBN 175).

7   Whereas Stove (1973) thinks that Hume only recognizes deductively valid arguments as legitimate, and so draws the skeptical conclusion that causal inference is unwarranted. See also Fogelin 1985.

8   Owen's criticism would seem to target Stroud (1977), who takes the negative story as a skeptical clearing of the ground that would then highlight the need for a positive psychological account of the inferences.

9   Nevertheless, there are passages within 1.3 that would need to be addressed; e.g. at T 1.3.12.20; SBN 139.

10 In this regard, see Garrett 1997: 80. One might try to reconcile the psychology with the concern with skepticism and warrant by going beyond the relatively minimal suggestion here by proposing the ambitious thesis that attributes to Hume a developed theory of epistemic justification understood as stability of belief (Loeb 2002).

11 See also how Hume contrasts arithmetic and algebra on the one hand with geometry on the other (T 1.3.1.4–5; SBN 71), as well as the remarks about the merchants at T 1.4.1.3; SBN 181.

References and further reading

Arnold, N. Scott (1983) "Hume's skepticism about inductive inferences," *Journal of the History of Philosophy* 21, pp. 31–55.

Baier, Annette (1991) *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge, MA: Harvard University Press.

——(1993) "David Hume, Spinozist," *Hume Studies* 19, pp. 237–52.

Beauchamp, Tom L., and Alexander Rosenberg (1981) *Hume and the Problem of Causation.* Oxford: Oxford University Press.

Blackburn, Simon (1993) "Hume on the mezzanine level," *Hume Studies* 19, 273–88.

Broughton, Janet (1983) "Hume's skepticism about causal inferences," *Pacific Philosophical Quarterly* 64, pp. 3–18.

Connon, R. (1979. "The naturalism of Hume revisited," in D. F. Norton, N. Capaldi, and W. L. Robison (eds.) *McGill Hume Studies* (pp. 121–45). San Diego, CA: Austin Hill Press.

Costa, Michael J. (1986) "Hume and causal inference," *Hume Studies* 12, pp. 141–59.

Craig, E. J. (1987) *The Mind of God and the Works of Man.* Oxford: Clarendon Press.

Flew, Anthony (1959) *Hume's Philosophy of Belief.* London: Routledge & Kegan Paul.

Fogelin, Robert L. (1985) *Hume's Skepticism in the* Treatise of Human Nature. London: Routledge.

Garrett, Don (1997) *Cognition and Commitment in Hume's Philosophy.* Oxford: Oxford University Press.

——(1998) "Ideas, reason, and scepticism" (symposium on Garrett 1997), *Hume Studies* 24, pp. 171–94.

——(2001) "Reply to my critics" (symposium on Garrett 1997), *Philosophy and Phenomenological Research* 61, pp. 205–15.

——(forthcoming) "Hume's theory of causation: inference, judgment, and the causal sense" in a volume of essays on Hume's *Treatise*, ed. Donald Ainslie.

Hume, David (1949) [1779] *Dialogues concerning Natural Religion*, 2nd edn., ed. Norman Kemp Smith. Edinburgh: Nelson.

Loeb, Louis E. (2002) *Stability and Justification in Hume's* Treatise. Oxford: Oxford University Press.

Millican, Peter (2002) "Hume's sceptical doubts concerning induction," in Peter Millican (ed.) *Reading Hume on Human Understanding: Essays on the First* Enquiry (pp. 107–73). Oxford: Oxford University Press.

Morris, William Edward (1988) "Hume's refutation of inductive probabilism," in James Fetzer (ed.) *Probability and Causality* (pp. 43–77). Dordrecht: Reidel.

Noonan, Harold W. (1999) *Hume on Knowledge*. London: Routledge.

Noxon, J. (1973) *Hume's Philosophical Development*. Oxford: Clarendon Press.

Owen, David (1999) *Hume's Reason*. Oxford: Oxford University Press.

Read, Rupert, and Kenneth A. Richman (2000) *The New Hume Debate*. London: Routledge.

Robinson, J. A. (1962) "Hume's two definitions of 'cause'," *Philosophical Quarterly* 12, pp. 162–71.

Robison, Wade (1977) "Hume's causal skepticism," in G. P. Morice (ed.) *David Hume: Bicentennary Papers* (pp. 156–66). Edinburgh: University of Edinburgh Press.

Russell, Paul (1984) "Hume's 'Two Definitions' of cause and the ontology of 'Double Existence'," *Hume Studies* 10, pp. 1–25.

Stove, D. C. (1973) *Probability and Hume's Inductive Scepticism*. Oxford: Clarendon Press.

Strawson, Galen (1989) *The Secret Connection: Causation, Realism, and David Hume*. Oxford: Clarendon Press.

Stroud, Barry (1977) *Hume*. London: Routledge.

Winkler, Kenneth P. (1991) "The new Hume," *Philosophical Review* 100, pp. 541–80.

——(1999) "Hume's inductive skepticism," in M. Atherton (ed.) *The Empiricists: Critical Essays on Locke, Berkeley, and Hume* (pp. 183–212). Lanham, MD: Rowman & Littlefield.

Wright, John P. (1983) *The Sceptical Realism of David Hume*. Minneapolis: Minnesota University Press.

7

Identity, Continued Existence, and the External World

Donald L. M. Baxter

To the question whether Hume believed in mind-independent physical objects (or as he would put it, bodies), the answer is Yes and No. It is Yes when Hume writes "We may well ask, *What causes induce us to believe in the existence of body?* but 'tis in vain to ask, *Whether there be body or not?* That is a point, which we must take for granted in all our reasonings" (T 1.4.2.1; SBN 187). However, the answer is No when, after inquiring into the causes of that belief, he writes, "What then can we look for from this confusion of groundless and extraordinary opinions but error and falsehood? And how can we justify to ourselves any belief we repose in them?"(T 1.4.2.56; SBN 218).

When faced with this apparent contradiction commentators have explored the various ways to resolve it. Some say that Hume is simply and irretrievably inconsistent. For instance Passmore, though admiring of the *Treatise*, finds it shot through with substantive contradictions (Passmore 1952: 1–2, 87, 152–3). Stroud sees Hume's position as ineluctably "paradoxical" (Stroud 1977: 245–50).

Other commentators have relied on the "No" answer and concluded that Hume restricts his attention to characterizing our impressions and ideas. Some of these interpreters think of Hume as a negative dogmatic "skeptic" who believes only in perceptions and tries to destroy the ordinary belief in an external world. This view is put more moderately by Thomas Reid and more vituperatively by James Beattie (see Kemp Smith 1941: 4–5). Others see Hume not as a destructive skeptic but as a phenomenalist who takes the external world simply to be composed of our sense impressions, whether perceived or unperceived. The classic example is Price who means in his book to be presenting "Hume's own theory" despite some corrections and developments (Price 1940: 227).

Other commentators have focused on the "Yes" answer and concluded that Hume is a realist about the external world, and not in the end a skeptic. These are the naturalists. Kemp Smith is the classic source. "Certain beliefs or judgments . . . can be shown to be 'natural', 'inevitable', 'indispensible', and are thus removed beyond the reach of our sceptical doubts" (1941: 87). The naturalist

position has been highly refined by the proponents of the "New Hume" interpretation in recent years. They take the natural belief to involve a lot of extra content about the direct perception and the nature of external objects, take Hume to have rejected all of this extra content for philosophical reasons; and they take Hume to have retained and believed the bare supposition that there exists some cause for our perceptions utterly unlike them. The philosophical correction of our natural belief reduces, but does not eliminate, the natural confidence we have in its truth. This residual naturalness provides some epistemic justification for the belief (Wright 1983: 225; Read and Richman 2000: 1–15). A different development of the naturalist interpretation is found in Garrett. For Hume, skeptical doubt is the result of various conclusions of reasoning that conflict with various beliefs which are in accord with our natural inclinations. Garrett points out that Hume's doubts about the competence of our faculty of reason raise doubts about these conclusions of reasoning. Thus, Garrett claims, his skepticism about our faculty of reason restrains his skepticism about these beliefs. That they are in accord with our natural inclinations, and serve certain desires, gives them title to be believed (Garrett 1997: 232–7).

My own view is that Hume's "Yes" and "No" answers can best be accounted for by seeing him as a Pyrrhonian skeptic in the tradition of Sextus Empiricus (Mates 1996). This interpretation, while finding considerable truth in all the various positions canvassed above, is closest to those of Norton, Fogelin, and especially Popkin.

Skepticism

The ancient Pyrrhonists found themselves unable to rest content with any philosophical arguments. For any argument for any conclusion, they always found equally compelling arguments for the opposite conclusion. Thus they considered the question still to be open and they continued their inquiry after the truth of the matter in a way intentionally reminiscent of Socrates. In the meantime they suspended judgment, not endorsing one position or its opposite.

The Pyrrhonists were reproached in ancient times and ever after for following a program that would bring life to an early end. To suspend judgment across the board would be to live without belief, the critics charged. It would be to emulate the namesake of these inquirers, Pyrrho, who according to legend would have died if not for the watchfulness of his friends. Having no belief one way or the other about what was dangerous, he took no care to avoid being run over by chariots, falling off cliffs, etc.

The legend seems unlikely on the face of it, given human nature, and the Pyrrhonists certainly were not so heedless. Their answer to the reproach required distinguishing two kinds of assent: actively endorsing a view as true on the basis of reasons, on the one hand, and passively acquiescing in a view impressed on them,

on the other (see Popkin 1966; Frede 1987). Someone who had carefully gathered arguments and evidence in favor of a view about what's true, and for whom assent was solely the result of reasoning, would assent in the former way. Someone who simply went along with the ways things appeared to him to be would assent in the latter way. It would be somewhat like passively supposing. With this distinction in hand, the Pyrrhonists could grant that they lived without belief in the more demanding sense of assenting to a view. Belief, in this more demanding sense, was an ideal of their dogmatic opponents – an ideal the Pyrrhonists didn't see how anyone could live up to. However, in assenting to views in the less demanding sense, they had enough belief to live life as mindfully as anyone else.

By following in these Pyrrhonian footsteps, Hume can both assent to the existence of the external world, and not assent to it, without inconsistency. He does passively acquiesce in the natural view thrust upon him by appearance, that there is an external world. However, he does not actively endorse as true the view that there is an external world. Sometimes even his passive acquiescence wavers. When he inquires into the causes of the view he finds temporarily that it does not even appear to be true. It appears to be the result of "trivial qualities of the fancy, conducted by . . . false suppositions" (T 1.4.2.56; SBN 217). Carelessness and inattention rectify this temporary anomaly, however, and return him to the view it is natural for him to acquiesce in. Keeping the kind of assent carefully in mind, we can call such a view a natural belief.

Surprisingly Hume himself would not defend the Pyrrhonists by distinguishing kinds of assent. He has a crude view of Pyrrhonism and echoes the above reproach himself. However, in pursuing philosophy he proceeds as a Pyrrhonist even if he wouldn't characterize himself as doing so. He naturally holds and acts on the belief in the external world because it almost irresistibly appears true, even though careful inquiry into the causes of the belief make it appear unjustified and even inconsistent. These opposing views of the belief keep him from actively endorsing it or its negation as true. In fact he thinks if we were able actively to endorse beliefs as true, that is, if our process of belief formation enabled us to believe only what we had good reason to believe, we wouldn't endorse anything (T 1.4.1.8; SBN 183–4). Nonetheless he for the most part acquiesces in the belief because it costs him "too much pains to do otherwise" (T 1.4.7.11; SBN 270).

In this skeptical way Hume takes for granted the existence of body. His real interest is in the workings of the human mind. Just as he has the natural belief in the existence of body, so do we all. The fact that we assent in any way to such an ill-grounded view calls for an explanation. He concludes that it is part of our nature passively to assent to it. We believe it instinctively. Were we ideally rational, we would fail to have the belief. So other causes besides justifying reasons must explain the fact that we have it. The bulk of T 1.4.2 is an inquiry into those causes.

It may seem odd that a skeptic goes on to theorize about causes. However, assent to views that seem forced on one by appearances need not be restricted to our fundamental natural beliefs, such as in the existence of body. In fact such a

restriction would be unnatural and unpleasant for someone of an inquisitive turn of mind. Hume on the whole finds himself inclined to "refin'd reasoning" (phrase from T 1.4.7.7; SBN 267–8), inclined to weigh opposing arguments and to see which results strike him more forcefully. In this he follows the ancient academic skeptic Carneades. Nonetheless, in my opinion Hume remains more Pyrrhonist than academic by finding no merit in this procedure as a cause of true belief. A relentless application of reason would undermine any of its results. Fortunately such severity seems beyond him most of the time, and beyond most people all of the time (T 1.4.1.7; SBN 183). Hume summarizes his position as follows:

> Thus the skeptic continues to reason and believe, even tho' he asserts, that he cannot defend his reason by reason; and by the same rule he must assent to the principle concerning the existence of body, tho' he cannot pretend by any arguments of philosophy to maintain its veracity. Nature has not left this to his choice, and has doubtless esteem'd it an affair of too great importance to be trusted to our uncertain reasonings and speculations. (T 1.4.2.1; SBN 187)

The Imagination

Why, then, do we naturally believe in external physical objects? To begin an answer Hume makes a distinction he will find useful between "continu'd" existence and "distinct" existence. For an object to have continued existence is for it to continue to exist unsensed by anyone. For an object to have distinct existence is for it to exist and function in a different location than that of the perceiver and independently of being perceived. Hume makes the distinction because although distinct existence is what most people think of when they think of external objects as such, the belief in continued existence is prior in the order of acquisition of beliefs.

With this distinction in mind, Hume poses his main question, "whether it be the *senses, reason,* or the *imagination,* that produces the opinion of a *continu'd* or of a *distinct* existence" (T 1.4.2.2; SBN 188).

Before beginning his answer, he reminds the reader that his concern is with our belief in objects that are not "specifically different" from perceptions. His point is that he is not presently concerned with the ancient philosopher's belief in pure substance, some mysterious principle of unity and identity underlying difference and change in ordinary objects – a principle which in itself has no sensible qualities. In such a case philosophers try to imagine something "unknown and invisible," an "unintelligible something" (T 1.4.3.4; SBN 220). In Locke's famous phrase a pure substance is simply "an uncertain supposition of we know not what" (Locke 1975: 1.4.18). But no idea could be used to think of such a thing. Since all ideas are copied from perceptions, no idea can represent something "specifically different" from perceptions (T 1.2.6.8; SBN 67–8). The best one could do is to try to form a "relative idea" of the causes of our perceptions "without trying to

comprehend the related objects" (T 1.2.6.9; SBN 68. See Locke 1975: 2.23.2–3). But this would be to conceive "an external object merely as a relation without a relative" (T 1.4.5.19; SBN 241). Such an idea could not be used to think of an object's external existence. The idea of an object's external existence, like the idea of its existence in general, is not different from the idea of the thing itself. If there is no idea of the thing, there is no idea of its existence, external or otherwise. Thus Hume says, "For as to for the notion of external existence, when taken for something specifically different from our perceptions, we have already shown its absurdity" (T 1.4.2.2; SBN 188).

Hume turns to explaining the belief in objects *not* specifically different from perceptions. His concern is with our belief in external objects that in many ways resemble our perceptions, for instance in shape, solidity, motion, color, taste, smell, sound, etc. Note that for Hume perceptions are not perceiv*ings*, but perceiv*eds*. They are not acts of mind, but mental objects on which acts of mind are directed. They differ from external objects mainly by being "perishing and internal." That is, perceptions cease to exist when not perceived, and have no location in external space. So when we distinguish external objects from perceptions we only "attribute to them different relations, connexions, and durations" (T 1.2.6.9; SBN 68).

Back to the main question then. Hume says it is obvious that the senses by themselves "are incapable of giving rise to the notion of the *continu'd* existence of their objects, after they no longer appear to the senses." They would have to continue to operate after they had ceased to operate. Nor can they produce belief in *distinct* existence. They neither display perceptions and objects separately nor fool us into thinking our perceptions are distinct from us (T 1.4.2.3–13; SBN 188–93).

Reason is not the cause, either. People innocent of the use of reason, such as children, nonetheless believe in continued and distinct existence. Further, even the seasoned reasoner, wise in the ways of induction, cannot observe a constant conjunction of perceptions and objects from which to conclude that the former are caused by the latter. He cannot observe any conjunction at all, for again, objects and perceptions are not presented as distinct things (T 1.2.4.14; SBN 43–4).

That leaves the imagination to be responsible for the belief in continued and distinct existence. Hume often uses "the imagination" to refer broadly to "the faculty, by which we form our fainter ideas," in contrast to the faculty of memory by which we form more vivid ideas. In this larger sense, the imagination is the mental faculty by which we associate ideas according to the principles discussed in T 1.1.4. In the present discussion, however, Hume uses "the imagination" in the "more limited sense" of excluding not only memory but also "our demonstrative and probable reasonings" (T 1.3.9.19n.; SBN 117–18n.)

As Hume notes, "all impressions are internal and perishing existences, and appear as such," yet only some provoke the imagination to yield belief in body (T 1.4.2.15; SBN 194). There must be certain qualities of these impressions that the imagination is susceptible to. Neither "involuntariness" nor "superior force and violence" will do (cf. Berkeley 1982: Pt. 1, sects. 29, 30, 33). Pains exhibit these.

More promising, according to Hume are "constancy," which is always a factor at least to some extent, and "coherence," which is often a factor. He talks as if these are qualities of external objects, but they are primarily qualities of impressions. More accurately, they are qualities of impressions in sequences. Impressions exhibiting constancy are closely resembling ones interrupted by others. Impressions exhibiting coherence are ones in a sequence with many of the elements of a familiar causal sequence. For example, repeated views of the scenery from one's window give constant impressions. Repeated glances at the logs burning in the hearth give coherent impressions.

Hume begins by discussing how coherence operates on the imagination. For some kinds of impressions – "our pains and pleasures, our passions, and affections" (T 1.4.2.16; SBN 194) – it has no notable effect. Their regular changes (see, for example, T 2.1.4.3; SBN 283) don't seem to require that they continue to exist unperceived. On the other hand, some sequences exhibiting coherence would be contrary to our experience of causal sequences unless the missing elements were assumed to have existed unperceived. For example, the sound of feet on the stair followed by the creak of the door is most naturally explained by appeal to the fuller sequence of impressions one gets when both watching and listening to a person mount the stairs and push open the door. Thus we assume more regularity than we observe. This is not mere causal reasoning, which is constrained to observed regularity. It is rather an inertia of the mind in continuing a way of thinking once begun. The mind notices a ragged regularity contrary to the neat regularity it is used to, but continues its habit of expecting the neat regularity by assuming it took place partly unperceived (T 1.4.2.20–2; SBN 195–8). Hume earlier discussed a precedent for this mental inertia when explaining how we come to the fiction of perfect equality in geometry (T 1.2.4.24; SBN 199).

Although Hume regards the effects of coherence to be important, he says, "I am afraid 'tis too weak to support alone so vast an edifice, as is that of the continu'd existence of all external bodies" (T 1.4.2.23; SBN 198–9). Constancy is needed as well. Why? Hume gives no further answer. My guess is that coherence could at best give us the belief of the continued and distinct existence of types of objects. It is types, not durable individuals, that figure into causation. Only constancy gives us belief in objects that retain their individual identity through time.

Hume gives a brief summary of the process by which constancy seduces the imagination into producing the belief in external body. We have a visual impression of for instance "the sun or ocean." After looking away we look back and get another such impression. These impressions are interrupted – others occurred in between – and so are numerically distinct, but they closely resemble; they have "like parts in a like order." The imagination is susceptible to regarding distinct, closely resembling things as the numerically same thing. This is the source of equivocation, for example (T 1.2.5.21; SBN 61–2). So we take the second impression to be numerically identical with the first. Yet the interruption makes it obvious that the impressions are numerically distinct. We are uncomfortable with the

contradiction, a discomfort from which we are eager to free ourselves. To do so we suppose that the first impression continued to exist unperceived and reappeared as the second impression. In other words one and the same impression was present to mind, then absent, then present again. In supposing so, we don't think of it as an impression, but rather as a continued existence. This supposition acquires the force and vivacity of a belief from its relation to the remembered impressions and the propensity to consider them the same. So we come to believe of distinct, closely resembling impressions that they are one and the same continued existence (T 1.4.2.24; SBN 199). It is a short step to believing of them that they are therefore a distinct existence, an external body.

His summary prepares the reader for the "considerable compass of very profound reasoning" needed to explain the role of constancy (T 1.4.2.23; SBN 199). He distinguishes four requisites for understanding and justifying his system, which need elaboration.

Identity

The first is to "explain the *principium individuationis*, or principle of identity." Whatever Hume took this phrase to mean, his purpose is to discover some of "the essential qualities of identity" (T 1.4.2.31; SBN 202), that is, qualities that are entailed by an object's being identical and without which there cannot be the same object. Note that the sense of "identity" relevant here is that of being "individually the same" (T 1.4.2.24; SBN 199) as opposed to being resembling or to being the same sort of thing. The qualities Hume will discover will be "the invariableness and uninterruptedness of any object, thro' a suppos'd variation of time" (T 1.4.2.30; SBN 201). He later confirms that both of these are essential by saying that variation is "evidently contrary" to identity (T 1.4.3.2; SBN 219), and " 'tis a false opinion that any of our objects, or perceptions, are identically the same after an interruption" [in existence as opposed to an interruption just in appearance] (T 1.4.2.43; SBN 209). Whether Hume is right about their essentiality is a controversial matter, but why he would take it for granted is easy enough to understand for both cases. Variation amounts to having, then lacking, some non-relational quality. If what has the quality is numerically identical with what lacks it, then a contradiction is true of that thing. So they are not identical. Interruptedness amounts to a temporal separation. So if something were numerically identical after the interruption, it would be separated from itself. So they are not identical. So identity requires invariableness and uninterruptedness. The "suppos'd variation of time" is also required, as will be seen.

Hume doesn't say why discovering these essential qualities is germane to his explanation of our belief in body. My guess is that uninterruptedness is the link between identity and continued existence unperceived. Constancy causes us to attribute identity, which causes us to attribute uninterruptedness, which causes

us to attribute continued existence, which then causes us to attribute distinct existence.

Hume arrives at invariableness and uninterruptedness by means of a remarkably compressed and opaque discussion. In my opinion it is also remarkably enlightening for anyone thinking seriously about the concept of identity. A more standard view, however, is that Hume was just naively and confusedly trying to come to grips with Frege's puzzle about the truth and informativeness of statements of identity (Bennett 2001: 298).

Hume is concerned with a different problem, however, not amenable to Frege's solution. Despite his mention of a proposition and words, Hume is not directly concerned with language and its meaning. He is directly concerned with ideas and what they represent there as being. Propositions, for Hume, are composed of ideas. Ideas for Hume are not like words or sentences. Ideas are like actors in the mind that, generally, portray objects and events outside the mind (cf. T 1.4.6.4; SBN 252–3). By having certain characteristics, ideas represent there as being objects with those characteristics, just as by being stout an actor represents there as being a man – Falstaff, say – who is stout (cf. Wittgenstein 1974: 2.171). (I use the odd phrase "represent there as being" to avoid the implication that the represented object really exists.) Hume's problem concerns which facts about ideas would represent there as being two things that are identical. In particular he is concerned with how many ideas it takes to do this.

It might seem obvious that in order to represent there as being two things that are identical, one must use a single idea to represent there as being a single thing. After all, identical things are just a single thing. But there is more to the idea of identity than just being an idea of unity (singleness, oneness). Questions whether things are identical can arise, even though they are in fact identical. I may wonder whether Cicero is Tully, even though he is. Such questions do not arise for something which we think of simply as a single thing. The question whether it is identical with itself seems silly (T 1.4.2.26; SBN 200).

To differentiate the idea of identity from the idea of unity, we might be tempted to represent there as being an identity by means of two ideas. But it is immediately obvious that this suggestion won't work. Two ideas represent there as being two distinct things, not one and the same thing (T 1.4.2.27; SBN 200).

If not one, yet if not more than one, then how many? Somehow we need more than a single idea but less than multiple ideas – a requirement that is impossible to fulfill (T 1.4.2.28; SBN 200). Hume says, "To remove this difficulty, let us have recourse to the idea of time or duration." This may seem a bit of bravado. However, he does not propose to do the impossible by finding an idea that literally is "a medium betwixt unity and number." Rather he will find an idea which we treat as if it were such a medium. That will be the idea of identity. By finding out which experience is necessary for generating that idea, he will find the essential qualities of identity. That experience will require applying the previously acquired idea of time or duration (T 1.4.2.29; SBN 200–1).

It may seem odd to think that the idea of time is required for the idea of identity. Paradigm cases of identity don't seem to involve time. For instance, in saying the number 4 is identical with the number 4 there seems to be no reference to time or duration. To see why he focuses on time, however, consider the specific problem he is addressing. It is as if he were a director trying to stage a play of the following form. Scene 1: A man and a man enter the stage in a way that leaves the audience wondering whether or not they are the same man. Scene 2: It turns out that they are the same man. The problem is how many actors to use to in Scene 1. If the director uses one actor, what is there for the audience to wonder about? They are obviously the same man. Furthermore Scene 2 becomes impossible to stage. Likewise, if he uses two actors they are obviously distinct men, and again Scene 2 becomes impossible to stage. But what other choice is there? The natural answer to this question is that the director use one actor and draws attention to him at different times.

In addition to the naturalness of appealing to time, the appeal is perfect for finding an idea involving both unity and number. On Hume's theory there are two ways to take up time: (1) having duration, which entails being many things in succession, and (2) being steadfast, which entails being a single thing (as opposed to being a succession) coexisting with a succession. The idea of identity will amount to the fiction of a steadfast object with duration – something that is both one and yet many.

Hume holds that "time, in a strict sense, implies succession" (T 1.4.2.29; SBN 200). In other words, the idea of time or duration is a general idea of the manner in which temporal successions are arrayed. It is just the idea of temporal successiveness. When Hume uses the word "time" he is not speaking of a dimension in which objects are located. That is why he uses "time" and "duration" interchangeably. He just means to refer to what temporal successions have in common.

A succession is a number of distinct things, one after the other. In contrast are what Hume calls "stedfast" and "unchangeable" objects. These are single objects that coexist with successions but which are not successions themselves. Thus they lack duration despite coexisting with successions that, as such, have duration. A mantelpiece would be an example of a steadfast object coexisting with the flickering succession of flames below. It would coexist likewise with the less predictable succession of thoughts in the mind of the homeowner relaxing alongside (T 1.2.3.6–11; SBN 34–7).

Even though steadfast objects really lack duration, we tend falsely to attribute it to them and to ignore their steadfastness. When we do so " 'tis only by a fiction of the imagination, by which the unchangeable object is suppos'd to participate of the changes of the co-existent objects, and in particular of that of our perceptions" (T 1.4.2.29; SBN 200–1). A non-succession lacks successiveness, but we attribute it anyway. We can't help it. Because we rarely give sustained attention to anything, so distracted are we by our tumbling thoughts and impressions, we treat everything as having the successiveness we continually experience. There are other

reasons as well that make treating everything as a succession an almost irresistible habit of mind (T 1.2.5.29; SBN 65).

Think of the fiction this way: we confuse (1) coexisting with all members of a succession with (2) successively coexisting with each member and not the others. The former is how a steadfast object coexists with a succession. The latter is how several things in succession coexist with another succession: the first member of one succession coexists with just the first member of the other, then the second coexists with just the second, then the third coexists with just the third, etc. If you are having trouble keeping these two ways of coexisting straight, then you can understand why ordinarily we don't bother to.

One day after the habit of treating everything as a succession is firmly established we attend closely to a steadfast object, such as a cup, accompanied by some succession, such as the ticks of a clock. We discover no variation or interruption in the object. It is patently a single thing as we contemplate it. However, in retrospect we can't resist the habit of considering it to have duration and therefore of being many things in succession. We find, therefore, that we represent the object as being a single thing or as being more than one thing, depending on the view we take. When we consider the moments of the ticks one after the other, we think of the steadfast object as unitary. Its invariableness and uninterruptedness are what are salient. However when we consider different moments of different ticks simultaneously, we think of the object as a multiplicity. We distinguish in our mind the object insofar as it coexists with the earlier tick (and not the later one), from the object insofar as it coexists with the later tick (and not the earlier one). The object's successively existing at different moments is what is salient. That it does so is just our habitual fiction; really it coexists with all the ticks, and not successively with each to the exclusion of the others. But we are well past such nicety by this point.

The result is the idea of identity. "Here then is an idea, which is a medium betwixt unity and number; or more properly speaking, is either of them, according to the view, in which we take it" (T 1.4.2.29; SBN 201). The idea of the steadfast object with duration is an idea of something that is one thing when looked at one way and many things when looked at another.

Now, again, Hume has not done the impossible. He has not given us an idea of something that is more than one but less than many. In fact the idea he has given us involves the inconsistency of lacking duration yet having it. However, the switching of views masks the inconsistency and the mongrel idea comes to play an important role in our mental economy, especially the sector concerned with the external world.

Hume now has an idea he can use to address the problem he began with. When we say that an object is identical with itself, what we mean is that the object existent at one time is identical with itself existent at another. That way we make there be a difference (a difference in time) between the idea meant by the word "object" and the idea meant by the word "itself" (T 1.4.2.29; SBN 201). We don't simply represent there as being more than one thing, however, because we can always

switch views and represent them as being one. Nor do we simply represent there as being one thing, because we can always switch views again. Nonetheless, being one thing is the stronger view since it is directly copied from experience.

It is essential to getting the idea of the oneness of the object that it be invariable and uninterrupted through a change in time that we suppose applies to it. Thus Hume concludes that the essential qualities of identity are "*invariableness* and *uninterruptedness*" (T 1.4.2.30; SBN 201). He now has in hand the link between constancy and continued existence unperceived.

Continued Existence

The next step, "the *second* part of my system," is to explain how their constancy, i.e., invariableness, causes us to ascribe numerical identity to interrupted perceptions. They have only one of the essential qualities of identity, but we somehow end up ascribing both. So the question concerns "the source of the error and deception with regard to identity, when we attribute it to our resembling perceptions, notwithstanding their interruption" (T 1.4.2.32; SBN 201–2). Note the falseness of the attribution. Another layer of fiction is being added.

Hume's explanation for the fiction appeals to a two-part "general maxim" in his science of human nature (T 1.4.2.33; SBN 203). The first sub-maxim is, "wherever there is a close relation betwixt two ideas, the mind is very apt to mistake them, and in all its discourses and reasonings to use the one for the other" (T 1.2.5.19; SBN 60). The second is, "wherever the actions of the mind in forming any two ideas are the same or resembling, we are very apt to confound these ideas, and take the one for the other" (T 1.2.5.21; SBN 61). In proposing the second sub-maxim, Hume takes for granted that we know what he means by "action of the mind." He later calls it "that certain *je-ne-sçai-quoi*, of which 'tis impossible to give any definition or description, but which every one sufficiently understands" (T 1.3.8.16; SBN 106)! My guess is that he is at least asking the reader to recognize that it can feel different to the mind to think of different things. For example it takes more effort to think of something complicated or painful than it does to think of something easy or pleasant. In any event, a consequence of these two maxims is that we have an almost irresistible tendency to confuse closely resembling ideas. Not only are the ideas closely related, but the actions of the mind in conceiving each of them are almost the same (T 1.2.5.19, 21; SBN 60, 61; T 1.4.2.32; SBN 202–3). As a result we fail to distinguish the resembling perceptions in a series exhibiting constancy.

Attributing identity to them is the result of an additional application of the second sub-maxim. The disposition of the mind when thinking of a constant though interrupted sequence of perceptions is very, very similar to the mind's disposition when thinking of a steadfast object, or as Hume says, an object "which preserves a perfect identity" (T 1.4.2.33; SBN 203). (He uses this phrase presum-

ably because a steadfast object displays both invariableness and uninterruptedness through a supposed change in time, the essential qualities of identity.[1]) When thinking of a steadfast object the mind merely continues the action of having an idea of the object instead of forming new ideas. No special effort distinguishes one moment from the next. Thinking of a succession of related objects is almost as easy. The relation facilitates the transition from one idea to the next. When the relation is close resemblance, the sequence of actions of the mind is "so smooth and easy, that it produces little alteration on the mind, and seems like the continuation of the same action" (T 1.4.2.34; SBN 204). As a result of this similarity in the actions of the mind, the mind regards the succession of related objects as "only one object." In the case at hand the mind is thinking of impressions exhibiting constancy. So in addition to failing to distinguish them from one another it explicitly regards them as identical.

Hume says that later in the *Treatise* he will mention many other instances of the tendency to regard distinct, closely related things as the same thing (T 1.4.2.35; SBN 204). These instances come under the discussions of the ancient philosophy and of personal identity (T 1.4.3 and T 1.4.6). Right now, though, the concern is with the impressions, the images, in a succession exhibiting constancy. He says, "'tis to these interrupted images we ascribe a perfect identity" (T 1.4.2.36; SBN 205).

Hume is on his way to explaining how we come mistakenly to think that perceptions exhibiting constancy are continued and distinct objects. It is the perceptions themselves that we come to regard this way. In coming to believe in body, we do not distinguish external bodies from internal perceptions. Here "we" refers to "the unthinking and unphilosophical part of mankind (that is, all of us, at one time or other)" (T 1.4.2.36; SBN 205), which Hume also calls "the vulgar." The vulgar ordinarily have no reason to distinguish perceptions from the objects they represent. "Those very sensations, which enter by the eye or ear, are with them the true objects, nor can they readily conceive that this pen or paper, which is immediately perceiv'd, represents another, which is different from, but resembling it" (T 1.4.2.31; SBN 202). It is only after someone arrives at a firm belief in external body, that he can then go on and make the philosophical distinction between internal perception and external objects. Prior to that, "the vulgar *suppose* their perceptions to be their only objects" (T 1.4.2.43; SBN 209). It is not that they explicitly suppose this. They simply take it for granted without a thought.

So far in the narrative we have attributed identity to distinct, interrupted perceptions. At this point a tension arises. The interruption of the perceptions naturally causes us to regard them as numerically distinct, in opposition to the strong tendency to regard them as numerically identical. This contradiction makes the mind uneasy, and it seeks relief by giving up one of the opposing views. The resolution of the tension is the "third part of that hypothesis I propos'd to explain" (T 1.4.2.36; SBN 205).

The tendency to attribute identity is too strong, so the opposite view must be given up. Thus we "suppose that our perceptions are no longer interrupted, but

preserve a continu'd as well as an invariable existence, and are by that means entirely the same." However, this resolution of the tension would seem immediately to raise another one. After all, the interruptions are too "long and frequent" to over-look. It couldn't be plainer that one of the essential qualities of identity is not present. So in order to suppose that there is really one invariable, uninterrupted perception, we have to suppose that it is often not present to mind. But how can a perception exist without being present to mind (T 1.4.2.37; SBN 205–6)?

This question is out of order, however. It would occur only to a philosopher who had long ago acquired the idea of body, and had gone on to distinguish internal perceptions from external body. The vulgar make no assumption that the very things they see and feel are mind-dependent. The only problem, then, is to explain to philosophers how we can ordinarily overlook the absurdity of supposing that mind-dependent perceptions be separated from the mind.

Hume's answer is to deny the philosopher's assumption that perceptions are dependent on the mind. Perceptions are distinct from the mind. Distinct things are separable. So perceptions are separable from the mind. There really is no con-tradiction in thinking perceptions take their leave of the mind, nor in thinking that they later come to be reunited with it. In his discussion of these points Hume anticipates his bundle theory of the self to be presented in T 1.4.6; SBN 251–63. In any event, there is no obstacle to his account of the mind's resolution of the first tension:

> When the exact resemblance of our perceptions makes us ascribe to them an identity, we may remove the seeming interruption by feigning a continu'd being, which may fill those intervals, and preserve a perfect and entire identity to our perceptions. (T 1.4.2.40; SBN 208).

In other words we take the perceptions in the constant but interrupted series to be a steadfast perception, invariable and uninterrupted, which is alternately present and absent to mind

> But as we here not only *feign* but *believe* this continu'd existence, the question is, *From whence arises such a belief*? and this question leads us to the *fourth* member of this system. (T 1.4.2.41; SBN 208)

Hume has earlier argued that belief consists in the vivacity, force, liveliness of an idea and that a belief in general is "A LIVELY IDEA RELATED TO OR ASSOCIATED WITH A PRESENT IMPRESSION" (T 1.3.7.5; SBN 96). The mind is excited by a lively impression, the impression attracts to mind a related idea. The relation between the perceptions gives the mind a propensity to form the idea and also smooths the way for the idea's appearance to mind. The excitement from the impressions lingers through the easy transition to the idea and is thus conveyed to the idea (T 1.4.2.41; SBN 208).

The case at hand is not exactly an instance of the general case, but the belief arises in much the same way. The relevant impressions are "impressions of the memory" (so-called because ideas of memory are "equivalent to impressions" [T 1.3.4.1; SBN 82]). It is the resemblance between actions of the mind that directly gives the mind a propensity to regard the interrupted perceptions as numerically the same. But the propensity and easy transition have the same effect of conveying the mind's excitement from the memory impressions to the idea of their identity. So the propensity "bestows a vivacity on that fiction; or in other words, makes us believe the continu'd existence of body" (T 1.4.2.42; SBN 209).

As noted at the beginning, "continu'd existence" for Hume is shorthand for "continued existence unperceived." So it is an immediate and obvious inference that a continued existence exists "independent of and distinct from the perception," i.e. is a body, an external object (T 1.4.2.2; SBN 188). This completes Hume's account of how constancy causes us to believe in objects that retain their identity through time.

Hume gives a very similar, though compressed, account of the causes of our belief in external objects that retain their identity not only through change of time, but other changes as well. The close relations between successive qualities in an altering object can cause us to believe the fiction that there is an identical, altering object (T 1.4.3.3; SBN 220). Ultimately we will attribute identity to sequences of perceptions displaying neither invariableness nor uninterruptedness. Sometimes the attribution is not even hindered by sudden great changes, if they are common enough (T 1.4.6.14; SBN 258). We end up very far afield from the steadfast object that precipitated the idea of identity.

The Philosophical System

Hume called our collection of beliefs in body a "vast . . . edifice." The weakness of its foundation of falsehoods and fictions is apparent on the least reflection. A few simple experiments suffice to undermine it. For example, gently push one eye to separate the focal points of the eyes to get double vision. Instantly, all the objects one directly sees are doubled. Since changing our perceptual apparatus is what caused the second member of each pair to appear, and since both members of any pair are exactly the same sort of thing, it follows that both depend for their existence on being perceived. They are clearly perceptions, not existences distinct from the mind (T 1.4.2.45; SBN 210–11).

If we reasoned justly, such results would lead us to abandon our belief in continued and distinct existence. Instead, philosophers respond by distinguishing internal perceptions, which are not continued and distinct, from external objects, which are. He calls this "*philosophical hypothesis*" the doctrine of "double existence." The doctrine, Hume argues, "*has no primary recommendation either to reason or to the imagination*" (T 1.4.2.46–7; SBN 211–12).

Not to reason. "The only existences of which we are certain, are perceptions" because they are "immediately present to us by consciousness." The only way to conclude by reason that anything else exists is by appeal to what must exist to cause our impressions of sense or memory (T 1.3.4). However, knowledge of causal connections requires past observation of constant conjunction between things of one sort and things of another. Such a constant conjunction can never be observed between external objects and internal perceptions, because all we directly perceive are the internal perceptions. So causal reasoning to the existence of external objects as their causes is not possible (T 1.4.2.47; SBN 212).

Not to the imagination either. Starting from the supposition that "our perceptions are broken, and interrupted, and however like, are still different from each other" there seems to be no way to explain how we would arrive at the conclusion that there are additional entities, much like perceptions, yet "continu'd, and uninterrupted, and identical" and not immediately perceived. Furthermore, any such argument would be unnaturally subtle and complex. What is more natural, as witnessed by the fact that it is so much more common, is the supposition that "our perceptions are our only objects, and continue to exist even when they are not perceiv'd." False, Hume says, but much more natural (T 1.4.2.48; SBN 213).

Thus, "*the philosophical system acquires all its influence on the imagination from the vulgar one.*" Clearly the doctrine of double existence has some authority, since it is common among those who consider the simple experiments against the vulgar system. However, it must derive its authority, having "no original authority of its own." It does so as a compromise between two unshakeable beliefs. We unphilosophically come to believe that our constant perceptions are a continued and distinct existence immediately perceived. This belief is so natural and stubborn, that nothing can cause us to give it up for long. A little philosophical reflection easily makes us see the dependence on our perception of what we immediately perceive. We find ourselves obliged to reject our unphilosophical opinion. However it "has taken such deep root in the imagination, that 'tis impossible ever to eradicate it." To relieve ourselves from the struggle between the stubborn unphilosophical opinion and its obvious falsehood, "we contrive a new hypothesis" which seems to satisfy both reason and the imagination: "the double existence of perceptions and objects." We take perceptions to have the obvious interruption and lack of continuity and identity between them. We take objects to be continued and uninterrupted and identical though time. "This philosophical system, therefore, is the monstrous offspring of two principles, which are contrary to each other, which are both embrac'd by the mind, and which are unable mutually to destroy each other." The doctrine of double existence is a pretext under which we in fact are alternately believing of our immediate perceptions, that they are continued and distinct and yet interrupted and dependent (T 1.4.2.49–53; SBN 213–16).

When faced with a contradiction, the imagination makes a distinction with nothing to recommend it but the fact that it eases our minds. Two other consid-

erations show that "feigning a double existence" (T 1.4.2.52; SBN 215) is merely imaginary makeshift. First, philosophers tend to suppose that external objects resemble internal perceptions. Having no way to compare them, we have no evidence for this. It is merely a result of the fact that the imagination "*borrows all its ideas from some precedent perceptions*." So everything it conceives in any detail must resemble them. Second, philosophers tend to suppose that a particular object resembles the perception it causes. The imagination is prone, when ideas are united by some relation "to compleat every union" by adding further relation. In this case, to causation we add further resemblance. The philosopher's doctrine of double existence is shot through with fictions of the imagination (T 1.4.2.54–5; SBN 216–17).

This ends Hume's explanation of the causes both of the common, "popular" belief in continued and distinct existence, as well as of the "philosophical" doctrine of double existence. As he considers the explanations he not only finds no reason actively to assent to these beliefs, but for the moment feels moved passively to acquiesce in the view that they are absurd. "I cannot conceive how such trivial qualities of the fancy, conducted by such false suppositions, can ever lead to any solid and rational system." The popular belief in continued and distinct existence is based on the coherence and constancy of our perceptions, which in themselves give no reason to believe in body. Constancy has the greatest influence but by means of false supposition. " 'Tis a gross illusion to suppose, that our resembling perceptions are numerically the same, and 'tis this illusion, which leads us into the opinion, that these perceptions are uninterrupted, and are still existent, even when they are not present to the senses." The philosophical system retains these difficulties and adds further "absurdity" by additionally denying that our perceptions are numerically the same, then inventing a distinction between perceptions and objects to escape the contradiction. "What then can we look for from this confusion of groundless and extraordinary opinions but error and falshood? And how can we justify to ourselves any belief we repose in them?" The confusion and absurdity of both the popular and philosophical views, when displayed so clearly, moves Hume at this point to withhold any sort of belief in them. (At T 1.4.5.1; SBN 232 he speaks of "such contradictions and difficulties in every system concerning external objects.") He finds himself "more inclin'd to repose no faith at all in my senses, or rather imagination" (T 1.4.2.56; SBN 217). Thus Hume finds himself with irrefutable "skeptical doubt," a malady naturally caused by "profound and intense reflection on those subjects."

Active endorsement of the view that body exists is forever precluded. The arguments against body "*admit of no answer*." (EHU 12.15 n. 1; SBN 155) The malady "can never be radically cur'd." However, passive acquiescence in the existence of body returns once we have been distracted. "Carelessness and in-attention alone can afford us any remedy." For those who work through Hume's arguments then, after taking an hour off, continue to read Hume's book, the remedy amounts to returning to the doctrine of double existence (T 1.4.2.57; SBN 218). Outside

the study when dining or playing backgammon, the remedy is more effective and one returns to the vulgar belief in body alone. One passively acquiesces in appearances. "Nature has not left this to his choice, and has doubtless esteem'd it an affair of too great importance to be trusted to our uncertain reasonings and speculations" (T 1.4.2.1; SBN 187). As Hume puts it in the *Abstract*, "Philosophy would render us entirely Pyrrhonian, were not nature too strong for it" (A 27; SBN 657). As I've argued, in acquiescing to nature's insistence without epistemic warrant, we are Pyrrhonian as well.

Value of Hume's Account

It remains to inquire about the value of Hume's just-so story about the genesis of the belief in the external world. There is little evidence that our belief is acquired in the particular way Hume describes. Very young infants expect objects gradually hidden by a screen still to be there (Baillargeon 1987). It may perhaps be that we have a built-in ability to respond to perceived objects as capable of existing unperceived (Spelke 1985). Even if we acquire this response in the course of experience, it seems to occur at a low level, certainly below consciousness (see Pylyshyn 2001). Conscious conflict in attributing both identity and distinctness to certain impressions, resolved by positing their continued and distinct existence, is not part of the story of contemporary psychology.

Nonetheless, the question Hume raised about how we move from various scattered inputs to representing there as being a unified object remains important and unresolved. His question survives in part in the "binding problem" in cognitive science: the problem of how features separately detected by the brain come to be experienced as united in a single object. Likewise survives his general solution of appealing to something added by the mind to the data (Hardcastle 1998). Attempts to address Hume's question are also found in studies of the "tunnel effect," a case of "amodal completion" in which, if the motion and timing are right, observers perceive there to be a single moving object even when part of the trajectory is hidden by a barrier (Cary and Xu 2001: 186). It has also been argued that connectionist models of the acquisition of the concept of object permanence bear important similarities to Hume's account (Collier 1999).

Furthermore, there are philosophical fruits of Hume's own proposal. Regardless of the actual causes of the belief in continued existence unperceived, Hume seems right that the interrupted evidence we receive from the world does not warrant it. There is nothing in the patterns of irradiation of the retinas of Baillargeon's infants that decides in favor of a continued existence that will reappear when the screen is removed. (cf. Quine 1960: 22). Such lack of warrant should make us wary about our judgments concerning external objects. Hume's way of being a skeptic, his manner of combining wariness about what's really true with assent to what appears plausible, gives us a model for coping with our perplexities. His

lessons about why and how to be a skeptic about the external world are among the most valuable and instructive we have.

Note

1   "Imperfect identity," in contrast, is not a species of identity but rather is identity naturally but falsely attributed to something that lacks one or both the essential qualities of identity (T 1.4.6.9; SBN 256).

References

Baillargeon, R. (1987) "Object permanence in 3.5- and 4.5-month-old infants," *Developmental Psychology* 23, pp. 655–64.

Bennett, J. (2001) *Learning from Six Philosophers*, vol. 2. Oxford: Clarendon Press.

Berkeley, G. (1982) [1710] *A Treatise concerning the Principles of Human Knowledge*, ed. K. Winkler. Indianapolis: Hackett.

Carey, S., and F. Xu (2001) "Infants' knowledge of objects: beyond object files and object tracking," *Cognition* 80, pp. 179–213.

Collier, Mark (1999) "Filling the gaps: Hume and connectionism on the continued existence of unperceived objects," *Hume Studies* 25, pp. 155–70.

Fogelin, R. J. (1985) *Hume's Skepticism in the* Treatise of Human Nature. London: Routledge.

Frede, M. (1987) "The skeptic's two kinds of assent and the question of the possibility of knowledge," in *Essays in Ancient Philosophy* (pp. 201–22). Minneapolis: University of Minnesota Press.

Garrett, D. (1997) *Cognition and Commitment in Hume's Philosophy*. Oxford: Oxford University Press.

Hardcastle, V. G. (1998) "The binding problem," in W. Bechtel and G. Graham (eds.) *A Companion to Cognitive Science*. Oxford: Blackwell.

Kemp Smith, N. (1941) *The Philosophy of David Hume*. London: Macmillan.

Locke, J. (1975) *An Essay concerning Human Understanding*, ed. P. H. Nidditch. Oxford: Clarendon Press.

Mates, B. (trans.) (1996) *The Skeptic Way: Sextus Empiricus's* Outlines of Pyrrhonism. Oxford and New York: Oxford University Press.

Norton, D. F. (1982) *David Hume: Common-sense Moralist, Skeptical Metaphysician*. Princeton, NJ: Princeton University Press.

Passmore, J. A. (1952) *Hume's Intentions*. Cambridge: Cambridge University Press.

Popkin, R. H. (1966) "David Hume: his Pyrrhonism and his critique of Pyrrhonism," in V. C. Chappell (ed.) *Hume: a Collection of Critical Essays* (pp. 53–98). Garden City, NY: Doubleday.

Price, H. H. (1940) *Hume's Theory of the External World*. Oxford: Clarendon Press.

Pylyshyn, Z. (2001) "Visual indexes, preconceptual objects, and situated vision," *Cognition* 80, pp. 127–58.

Quine, W. V. O. (1960) *Word and Object*. Cambridge, MA: MIT Press.

Read, R., and K. A. Richman (2000) *The New Hume Debate*. London and New York: Routledge.

Spelke, E. S. (1985) "Perception of unity, persistence, and identity: thoughts on infants' conceptions of objects," in J. Mehler and R. Fox (eds.) *Neonate Cognition: Beyond the Blooming, Buzzing Confusion*. Hillsdale, NJ: Erlbaum.

Stroud, B. (1977) *Hume*. London: Routledge.

Wittgenstein, L. (1974) [1921] *Tractatus Logico-philosophicus*, trans. D. F. Pears and B. F. McGuiness. London: Routledge & Kegan Paul.

Wright, J. P. (1983) *The Sceptical Realism of David Hume*. Minneapolis: University of Minnesota Press.

8

Personal Identity and the Skeptical System of Philosophy

Corliss Gayda Swain

The discussion of personal identity occupies so few pages of the *Treatise* and its Appendix that mastering Hume's views on this topic may appear easy. This appearance would, however, be deceiving. To understand Hume's account of personal identity one must, as Hume says, "take the matter pretty deep" (T 1.4.6.5; SBN 253). Because Hume, often implicitly, but sometimes explicitly, compares his views on personal identity with his views on other metaphysical issues, especially causation and external objects, one cannot achieve a full understanding of Hume's theory of personal identity without understanding those views, the discussion of which takes up much of Book I of the *Treatise*. Understanding Hume on these other matters is also necessary for understanding the puzzling comments Hume makes in the Appendix about his theory of personal identity. While this approach requires considerable effort, the effort is well worth it: for it leads to a deeper and more unified understanding of Hume's skepticism – its nature, grounds, and limits.

Interpretations of Hume's skepticism have evolved. Traditionally, Hume was thought to be skeptical of causation, external objects, and persons, and it was thought that Hume wanted to demonstrate the irrationality of our everyday beliefs about these things. The interpretation I offer below continues the reassessment of that tradition, a reassessment traceable to Norman Kemp Smith's *The Philosophy of David Hume* (1941). New interpretations take Hume's skepticism to be far more limited than the tradition suggests, aimed more at philosophical views of causation, external objects, and the self than at these realities themselves. (See Strawson 1989; Baier 1991; Garrett 1997.)

Influential representatives of the traditional interpretation of Hume's theory of personal identity include Thomas Reid, Terence Penelhum, Roderick Chisholm, and Derek Parfit. Robert Fogelin and Wade Robison hold intermediate views. Fogelin (1985) attempts to reconcile Hume's naturalism with his skepticism, but he nonetheless sees Hume as an extreme skeptic. Robison (1974) interprets Hume as explaining the content and origin of popular beliefs about the self while denying the truth of these beliefs. The traditional interpretation is often presented as uncontroversial in discussions of personal identity that include an historical survey. See for example, Noonan 2003: ch. 4 and Olson 1997: ch. 7. However, most Hume scholars today see Hume as offering a positive account of the enduring self (e.g. Passmore, Pike, Noxon, Biro, Stroud, MacIntyre, Beauchamp, Bricke, Traiger, Flage, Swain, Loeb, Waxman, and Baxter to name only a few). Their disagreements concern the nature and adequacy of Hume's account.

Hume's views on personal identity are not difficult to summarize. The first element of his view is negative: Hume rejects a common philosophical view of the self according to which the self has both perfect simplicity and perfect identity. (Hume uses the terms *self*, *person*, and *mind* interchangeably, and in this article I follow his practice.) A being has *perfect simplicity* just in case it is not composed of parts at any given time; it has *perfect identity* just in case it is composed of exactly the same parts from one time to the next. When articulating his view of personal identity, Hume distinguishes between objects that have perfect identity and other objects that continue in existence over time. To illustrate the notion of perfect identity he uses the example, familiar from Locke (1975: 2.27.3), of a mass of particles. Such a mass becomes a different entity when and only when particles are subtracted from the mass or added to it. Whether they are simple or complex, whether made up of bits of matter or composed of something immaterial, only those beings whose composition remains unchanged through time have what Hume calls "perfect identity" (T 1.4.6.8; SBN 255–6). Hume denies that the self has this kind identity. Instead, he argues that the identity we attribute to the self is an "imperfect identity," the same kind of identity that we attribute to other objects whose parts change over time, including plants, animals (T 1.4.6.5; SBN 253 and T 1.4.6.12; SBN 257), ships (T 1.4.6.11; SBN 257), and republics (T 1.4.6.19; SBN 261). Even though we think of these things as unified, singular entities, continuing in existence over time, none of them has "perfect simplicity" or "perfect identity." Hume's positive view, then, is that the self is a composite (T 1.4.6.22; SBN 263), and, regarding its identity, Hume holds that the parts that constitute the self do not remain the same throughout a person's existence, but "succeed each other with an inconceivable rapidity . . . in a perpetual flux and movement" (T 1.4.6.4; SBN 252).

On Hume's positive theory, the parts that make up the self, or mind, are perceptions. These perceptions include not only what Hume calls "IMPRESSIONS OF SENSATION" (sensory perceptions) and "IMPRESSIONS OF REFLECTION" (passions and emotions), but all other kinds of conscious mental states as well (memories,

ideas, thoughts, and occurrent beliefs). (See 10: THE POWERS AND MECHANISMS OF THE PASSIONS.) The mind or self is not, according to Hume, something distinct from its thoughts, emotions, and sensory experiences. It is, instead, a composite comprising all these things.

Commentators disagree about the proper interpretation of the passages in which Hume criticizes the view that the self is simple and unchanging (Traiger 1985 and Ainslie 2004 summarize alternative interpretations). We can, nonetheless, find in these passages at least three arguments for the claim that the self is composed of its perceptions. One argument is that we could make no sense of the fact that one's perceptions *belong* to one's self given the further fact, insisted on by Hume, that each perception is a distinct being that is capable of existing on its own, unless we granted that these distinct perceptions *constitute* the self as parts of the whole (T 1.4.6.3; SBN 252). A second argument rests on two premises: (1) that we can never perceive our own minds except by perceiving our thoughts, sensations, and emotions; and (2) that thoughts, sensations, and emotions are *all* we perceive when we perceive our own minds. Proof that we cannot perceive our minds except by perceiving thoughts, sensations, and emotions is supposedly found in the fact that "[w]hen my perceptions are remov'd for any time, as by sound sleep; so long am I insensible of *myself*" (T 1.4.6.3; SBN 252). Finally, the constitutive relation between one's perceptions and one's self is supposed to be evidenced by the fact that "were all my perceptions remov'd by death, and cou'd I neither think, nor feel, nor see, nor love, nor hate after the dissolution of my body, I shou'd be entirely annihilated" (T 1.4.6.3; SBN 252). These arguments provide some grounds, though perhaps not conclusive reasons, for believing that there is no such thing as a mind that is completely devoid of mental contents and that there is nothing to the mind over and above its contents.

Hume sometimes describes the self as a "bundle or collection of different perceptions" (T 1.4.6.4; SBN 252), and even as a "heap" (T 1.4.2.39; SBN 207). For this reason some commentators have called Hume's theory a "bundle theory of the self" (see Pike 1967; Patten 1976; and Baier 1979). However, to the extent that bundles and heaps are arbitrary aggregates of parts, this name is somewhat misleading. Hume also describes the self as a system. He claims that "the true idea of the human mind" is an idea of "a system of different perceptions or different existences, which are link'd together by the relation of cause and effect, and mutually produce, destroy, influence, and modify each other" (T 1.4.6.19; SBN 261). Instead of calling his positive theory a bundle theory, we might less misleadingly characterize it as a causal theory of the self. Hume holds that the self's continuance through time is a matter of its being organized as a system. This is not to deny that the self continues in existence through time. It simply affirms that the continued existence of the self, like the continued existence of almost any organized system, including that of organisms, does not depend on the continued existence of the parts that constitute that system at any moment. While Hume calls the identity of all such changing objects "fictitious" (T 1.4.6.15; SBN 259), he

opposes fictitious identity not to real identity but only to perfect identity. (Swoyer 1982 and Traiger 1987 discuss Hume's use of the term "fiction" in his accounts of identity.)

Although Hume does not mention him specifically in this discussion, his theory of personal identity owes much to Locke. Locke had already made the distinction between the kind of identity we attribute to unchanging objects like the mass of matter and the kind of identity we attribute to objects whose constituents change over time (1975: 2.27.4–6). Locke also developed a systemic or causal theory of identity to account for the identity of plants and animals (2.27.4–6). Hume accepts Locke's account of the identity of plants and animals, and he believes there is "a great analogy betwixt [that identity] and the identity of a self or person" (T 1.4.6.5; SBN 253). Hume does not, however, accept Locke's account of personal identity. Locke believed that memory links together the distinct experiences that make up the self, forming a single consciousness that stretches over time (1975: 2.27.9–10 and 2.27.17–23). Hume rejected this hypothesis about the connecting principle, arguing as follows:

> Who can tell me, for instance, what were his thoughts and actions on the first of *January* 1715, the 11th of *March* 1719, and the 3d of *August* 1733? Or will he affirm, because he has entirely forgot the incidents of these days, that the present self is not the same person with the self of that time; and by that means overturn all the most establish'd notions of personal identity? (T 1.4.7.20; SBN 262)

Relying on "established notions" of personal identity, according to which the self – not just the man, as Locke had claimed (1975: 2.27.20) – extends beyond its own memories, Hume proposes that the perceptions that make up the self are linked not by memory alone but by many different causal relations.

On Hume's theory, memory is one connecting principle, insofar as it causes new perceptions that resemble earlier perceptions (T 1.4.6.18; SBN 260–1), and resemblance as well as cause and effect can produce the mental associations that are sources of beliefs about the enduring self. However, memory is not *the* connecting principle linking the perceptions that make up the self. Once causal laws governing our psychology have been discovered, we can "extend the same chain of causes, and consequently the identity of our persons, beyond our memory, and can comprehend times, and circumstances, and actions, which we have entirely forgot, but suppose in general to have existed" (T 1.4.6.20; SBN 262). By taking into account causal connections between mental states in addition to those involved in memory, Hume's theory can accommodate the fact that some irremediably forgotten experiences are considered as much a part of one's self as the experiences one remembers.

Hume first restricts his attention to "personal identity, as it regards our thought or imagination" (T 1.4.6.5; SBN 253) when looking for the relation that grounds our judgments of identity. There he finds a number of causal relations between

perceptions: "Our impressions give rise to their correspondent ideas; and these ideas in their turn produce other impressions. One thought chaces another, and draws after it a third, by which it is expell'd in its turn" (T 1.4.6.19, SBN 261). When Hume expands his view beyond the thought and imagination, considering, in addition, "our passions and the concern we take for ourselves" (T 1.4.6.5; SBN 253), he finds further evidence in support of his causal theory of the self. Echoing Locke once again (1975: 2.27.23, 25–6), Hume says, "our identity with regard to the passions serves to corroborate that with regard to the imagination, by the making our distant perceptions influence each other, and by giving us a present concern for our past or future pains or pleasures" (T 1.4.6.19; SBN 261). These newly considered causal connections between mental states provide further reasons for rejecting a memory account of personal identity: For it is obvious that the causal connections between present decisions and future happiness, in the case of prudential thoughts, and between past actions and present misery, in the case of regrets, are not produced by memory alone (T 1.4.6.20; SBN 261–2). (For other views concerning the relation between Hume's view of personal identity with respect to the understanding and with respect to the passions see Capaldi 1989; Henderson 1990; Baier 1991; ch. 6; McIntyre 1989 and 1993; Chazan 1992; and Purviance 1997.

According to Hume, "the true idea of the human mind, is to consider it as a system of different perceptions . . . which are link'd together by the relation of cause and effect" (T 1.4.6.19; SBN 261). He thus seems to embrace a psychological connectedness theory of personal identity. However, Hume's theory of personal identity is complicated by the further fact that, according to him, "the union of cause and effect . . . resolves itself into a customary association of ideas" (T 1.4.6.16; SBN 260). This has important implications.

Hume asks two related questions about the identity that we attribute to persons: "whether [this relation of identity] be something that really binds our several perceptions together or only associates their ideas in the imagination" (T 1.4.6.16; SBN 259) and "whether in pronouncing concerning the identity of a person, we observe some real bond among his perceptions, or only feel one among the ideas we form of them" (T 1.4.6.16; SBN 259). His answer is informed by his theory of causation. Since he accepts the principle that "the understanding never observes any real connexion among objects" (T 1.4.6.16; SBN 259–60; and T Appendix. 20; SBN 635), including connections between causally related objects, Hume concludes that "identity is nothing really belonging to these different perceptions, and uniting them together; but is merely a quality, which we attribute to them, because of the union of their ideas in the imagination, when we reflect on them" (T 1.4.6.16; SBN 260; and T Appendix 20; SBN 635). For Hume, the link between the perceptions that constitute the self is to be found in the observer, who connects together her ideas of those perceptions, rather than where we would expect it – in the observed person, where it would connect the perceptions that actually constitute the self.

Although he concedes that with his causal theory of the self, there are no known – or knowable – connections among the perceptions that constitute a self, Hume does not take judgments of personal identity to be arbitrary and unjustified, since those judgments must be grounded in real, observed regularities. On Hume's theory of the self, if the mind's activities lacked the regularities that ground causal judgments when observed, we could make no correct attributions of personal identity. Because the mind's activities do exhibit the kind of regularity that grounds causal judgments, we can justifiably and non-arbitrarily connect ideas of the perceptions that make up the self when we reflect on them.

Correctly characterizing Hume's theory of personal identity is not easy. Although we earlier called his theory of personal identity a causal theory, we may, on reflection, decide that Hume does not actually offer an account of personal identity because he never explains what connects the connected perceptions that constitute a mind. If the self is a composite, and its parts "form a whole only by being connected together" (T Appendix 20; SBN 635), then it seems reasonable to insist that a satisfactory account of the self explain what connects these parts. Because Hume argues that no such explanation is possible, his account could be characterized as a "no-theory" or skeptical account of personal identity.

Hume's views on personal identity are like his views on causation in this respect. Just as we may be tempted to say that Hume has no metaphysical account of causal connections because he believes that we have no idea of the connecting principle that binds together every cause and effect, so also we may be tempted to say that Hume has no metaphysical account of the self because he believes that we have no idea of the connecting principle that binds together the perceptions that make up a self. In one sense that is true, but in another sense it is quite far from the truth.

Consider the case of causation: Some philosophers would say that Hume does have a coherent, novel, and even plausible metaphysical account of causation. However, for Hume, the causal link is not a link between the causally related objects; it is, instead, a link in perceivers' minds between their perceptions of the causally related objects. Cause and effect are not directly connected; instead, perceptions – ideas and impressions – of causes and effects are connected in the minds of perceivers. If we insist that a satisfactory metaphysical account of causal relations must identify a direct bond between the causally related objects, or that any proposed link must be a link between the causally related objects themselves, we must deny that Hume offers a satisfactory metaphysical account of causation. From this perspective, Hume appears to be a skeptic about metaphysical accounts of causation (though not necessarily about causation itself). The grounds of this skepticism are that *cause and effect are distinct existences*, and that *the mind never perceives connections among distinct existences*. We could, however, with equal justification insist that a satisfactory metaphysical account of causation must explain what the actual connection is between causally connected objects. If the connection is, in fact, an indirect connection that exists only in the minds of perceivers when they

observe (or think about) the causally related objects, then an account that says this would satisfy the actuality requirement despite its failure to say what directly connects the supposedly connected objects. Viewing matters from both perspectives enables us to see how Hume could be both a skeptic and a naturalist when it comes to causation. He is skeptical of metaphysical accounts of causal connections, but such skepticism is compatible with belief in causation and (indirect) causal connections.

Returning to Hume's account of personal identity, if we insist that a satisfactory metaphysical account of the self must explain what actually binds together the perceptions that constitute a self, then we ought to take Hume to be arguing for skepticism about metaphysical accounts of the self. His grounds are very like the grounds appealed to in connection with causation: "*that all our distinct perceptions are distinct existences* and *that the mind never perceives any connections among distinct existences*" (T Appendix 21; SBN 636). However, if we allow that a satisfactory metaphysical account of the self need only identify the actual connection between the perceptions that make up the self, and if that connection is one that exists only in the minds of perceivers when they reflect on the constituting perceptions, then an account that says this would be a satisfactory metaphysical account according to the actuality requirement.

The words we use to characterize Hume's account – whether we call it a *satisfactory metaphysical* account or not – are unimportant, but there is something at stake. Whether we accept a proposed account of causation or the self depends on what constraints we impose on such accounts. Hume found himself forced to reject some plausible constraints because he believed they could never be satisfied. In explaining why they cannot be satisfied, Hume provided grounds for a limited and principled skepticism. His grounds rule out the possibility of accounts that try to explain connected phenomena in terms of connecting principles. (We could call these "metaphysical glue" accounts.) After rejecting these constraints, Hume proposes alternative accounts of causation and the self which do not even attempt to supply the metaphysical connecting principle. But if the only defect of these accounts is that they do not satisfy the a priori constraints, and if Hume has proved that these constraints cannot be satisfied anyway, then one might reasonably conclude that the constraints themselves, rather than Hume's alternative accounts, should be rejected. (A version of this argument is presented in Swain 1991.) This is a powerful sceptical conclusion concerning the possibility of metaphysics and its scope.

Hume's account of personal identity involves three main ideas: (1) that the self is a composite with the kind of identity – an imperfect identity – shared by all objects that are not composed of exactly the same constitutive parts throughout their existence, (2) that the self forms a whole over time by being composed of successive parts linked together by causal relations, and (3) that the only connection we can observe (or even conceive of) with respect to the parts that successively make up the self is a mental connection between, not the parts themselves, but

only our ideas of these parts, a connection forged when reflecting on perceptions that exhibit causal regularities. Hume's account, or at least this set of three ideas, is consistent. This account nevertheless has an air of paradox, since it affirms that the self is a system of connected perceptions but goes on to say that the only knowable link between the supposedly connected parts is a link between *ideas* of these parts rather than a link between the parts themselves.

Hume's account of the self shares this feature with his account of causation. Hume characterizes his views concerning a necessary connection between cause and effect as paradoxical (T 1.3.14.24; SBN 166), because he accepts the paradoxical conclusion that the "connexion and necessity" of causes "belongs entirely to the soul, which considers the union of two or more objects in all past instances" (T 1.3.14.23; SBN 166). This "paradox" perfectly mirrors the "paradox" concerning the self outlined above. Hume agrees that causes and their effects are necessarily connected, but he observes that the only link he can discover between these connected objects is a link between our *ideas* of the objects rather than a link between the objects themselves. While Hume's account explains why we justifiably think of certain objects as necessarily connected, it does so by explaining our grounds for connecting the *ideas* of these causally related objects (namely, their regular conjunction) rather than by identifying the link that connects causes to their effects and explaining our epistemic access to this link. Hume anticipates his critics' complaining that he would "remove [the causal power] from all causes, and bestow it on a being that is no ways related to the cause or effect, but by perceiving them" (T 1.3.14.26; SBN 168), and sees them ridiculing his position as "a gross absurdity, and contrary to the most certain principles of human reason" (T 1.3.14.26; SBN 168). Hume also foresees his critics' accusing him of rejecting the truism that "[t]hought may well depend on causes for its operation, but not causes on thought" (T 1.3.14.26; SBN 167). Although he is well aware of these paradoxical features of his account of causation, Hume nonetheless offers this account as a true account. Given that it *is* paradoxical, the only hope Hume says he has of its acceptance is "by dint of solid proof and reasoning" (T 1.3.14.24; SBN 166).

When assessing Hume's accounts of the self and of causation, it is important to note that, for Hume, the materials out of which a unified system is constructed exist and are related to each other in definite ways prior to the mind's synthesizing activities. On Hume's view, the mind does not invent fire, wood, or ashes. Nor does it invent the constant conjunction expressed in the proposition, "Ashes are the product of burning wood." This much is perfectly objective. But insofar as the causal connection is supposed to involve more than constant conjunction, there is an extra bit – the necessary connection – which is, according to Hume, a product of the mind's synthesizing activity. Philosophers who take Hume's account of causation to be a simple regularity theory deny that there is more to Humean causation than constant conjunction. However, while that interpretation is consistent with Hume's first definition of a cause, it does not comport well with

Hume's second definition, which refers to the synthesizing activity of the mind: the way that ideas are united in the imagination (T 1.3.15.35; SBN 172). To do justice to both definitions, we should take Hume's theory of causation to involve both an objective element that consists of facts about regularities and a subjective element, consisting of facts about the mind's synthesizing activities. (See 6: CAUSATION for discussion of Hume's two definitions.)

Similarly, on Hume's view of the self, the thoughts, sensations, and emotions which constitute the self exist independently of any synthesizing activity of the mind. These thoughts chase, draw after, and expel each other (T 1.4.6.19; SBN 261) with the regularity characteristic of causally connected things prior to and independently of reflection. If they did not, we would have no grounds for connecting our ideas of these perceptions together. Persons are thus not *completely* self-made. But they are self-made to some extent: For although reflection discovers, rather than produces, the thoughts and their regular conjunctions, reflection on mental contents and their regularities does produce a sense of a unified and continuing self by producing the feeling that these conjunctions are necessary and that the conjoined items are connected.

Interpreters who take Hume's theory of causation to be a straightforward regularity theory tend to ignore what Hume says about the synthesizing activities of the mind and the subjective side of causal connections. With interpretations of Hume's view of the self, we find a tendency to focus exclusively on the synthesizing activities of the mind or the subjective side of self-construction, while the objective regularities that ground this activity tend to get ignored. The resulting view is that persons are almost completely self-made, and the only facts about what belongs to any given mind are those that depend on the synthesizing activity of the mind. To do justice to Hume's account of the self, however, we must recognize that the possibilities of self-construction are heavily constrained by the stubborn facts of human psychology.

Any assessment of Hume's theory of the self must take into account Hume's remarks in the Appendix to the *Treatise*. In the Appendix, Hume was concerned to clarify his positions and to correct any mistakes he had made. He identifies only three mistakes, one of which he characterizes as a "considerable mistake" (T Appendix 1; SBN 623), while the other two are called "errors of less importance" (T Appendix 22; SBN 636). The considerable mistake appears to implicate Hume's theory of personal identity. The section of the Appendix where Hume discusses this mistake begins as follows:

> I had entertain'd some hopes, that however deficient our theory of the intellectual world might be, it wou'd be free from those contradictions, and absurdities, which seem to attend every explication, that human reason can give of the material world. But upon a more strict review of the section concerning *personal identity,* I find myself involv'd in such a labyrinth, that, I must confess, I neither know how to correct my former opinions, nor how to render them consistent. (T Appendix 10; SBN 631)

This passage admits of two interpretations. The familiar interpretation holds that Hume discovers a mistake in his theory of personal identity. On a different interpretation, which I develop below, Hume's considerable mistake does not involve the theory of personal identity; instead, the mistake involves Hume's earlier claim that accounts of the intellectual world (or the self) are special in that they alone are, or can be, free of contradictions. (Fieser [1989:105–6] and Waxman [1992: 233] also suggest that the passage in the Appendix refers to Hume's claim about the intellectual world in T 1.4.5.1.)

In the Appendix passage, Hume appears to claim that his theory of personal identity involves unavoidable inconsistencies. Because finding unavoidable inconsistencies in one's theory is generally held to be a bad thing, most commentators who have grappled with Hume's theory of personal identity have interpreted this passage as evidence that Hume was dissatisfied with his earlier account and had, upon review of it, found reasons for rejecting it. Once they accept that Hume found something wrong with his account of personal identity and *only* in his account of personal identity, since he mentions only this section, their next task is to identify the precise source of Hume's dissatisfaction. (Both Stroud [1977: ch. 6] and Pears [1990: 120] argue persuasively that the problem must be unique to Hume's account of personal identity.) This has proved difficult.

Commentators agree that if Hume did find some new problem when he reviewed his section on personal identity, he wasn't forthcoming about its nature in the Appendix. Hume mentions inconsistencies, but the supposedly inconsistent principles that he refers to are logically compatible and are the grounds for his own skeptical accounts of the self, external objects, and causation. Hume also points out as a deficiency in his account of personal identity the fact that it does not provide the principle that connects the perceptions that make up the self, but this deficiency was already remarked in the *Treatise*. It is not something that would have struck Hume only after he reviewed his section on personal identity. While much ingenuity and philosophical acumen has been exhibited by commentators who believe Hume did find some new problem with his account of personal identity, none of the many attempts to locate the problem and to show that it infects only Hume's account of personal identity has won widespread acceptance. See Stroud 1977: ch. 6; Garrett 1981 and 1997: ch. 8; and Fogelin 1985: ch. 8 for overviews and criticisms of various interpretations. Stroud, Garrett, and Fogelin also offer their own interpretations, but none has gone unchallenged.

Given the difficulty of finding a problem that infects only Hume's account of personal identity and fits with what Hume says in the Appendix, we have reason to welcome an alternative interpretation of Hume's "considerable mistake." On this interpretation, the mistake is not found in Hume's section on personal identity but rather in the statement he made in the immediately preceding section, "Of the immateriality of the soul." That section begins by drawing a contrast between accounts of the external world – accounts that explicate our idea of matter – and accounts of the mind, which explicate our idea of the self. Hume says the former

contain "contradictions" as well as "difficulties" (T 1.4.5.1; SBN 232), while the latter, although they involve difficulties, need contain no contradictions. He draws the contrast as follows:

> Having found such contradictions and difficulties in every system concerning external objects, and in the idea of matter . . . we shall naturally expect still greater difficulties and contradictions in every hypothesis concerning our internal perceptions, and the nature of the mind . . . But in this we shou'd deceive ourselves. *The intellectual world, tho' involved in infinite obscurities, is not perplex'd with any such contradictions*, as those we have discover'd in the natural. *What is known concerning it, agrees with itself*; and what is unknown, we must be contented to leave so. (T 1.4.5.1; SBN 232; my italics)

This is the mistaken claim that Hume is concerned to correct when he says, "I had entertain'd some hopes, that however deficient our theory of the intellectual world might be, it wou'd be free from those contradictions, and absurdities, which seem to attend every explication that human reason can give of the material world" (T Appendix 10; SBN 633). Interpretations of the significance of Hume's mistake will differ depending on how Hume's claims about contradictions are understood. Fieser (1989) and Waxman (1992), for example, take the contradictions to signal something quite different from what I propose. We nonetheless agree that "strict review of the section concerning *personal identity*" (T Appendix 10; SBN 631) made Hume realize that his claim about the possibility of a consistent, though incomplete, metaphysical account of the "intellectual world" was mistaken. The interpretation I present below argues that the contradictions Hume refers to teach us only about the futility of trying to give certain kinds of metaphysical accounts; they do not indicate that there is some special problem with Hume's positive account of personal identity. (For an alternative interpretation, see Winkler 2000.)

On my interpretation, the vexed and seemingly unanswerable question, "What new problem concerning his theory of the self was Hume alluding to in the Appendix?" need not be answered, since on this interpretation there is no reason to think that Hume *did* allude to some new problem in his account. Rather than looking for new problems, we can focus, instead, on the more general and important task of understanding the kinds of contradictions and absurdities that Hume believes infect every attempt to account for our most central and important beliefs: What are the contradictions? Why do these contradictions lead to skepticism? and What are we supposed to be skeptical about?

The only "inconsistency" or "contradiction" that Hume mentions in the Appendix is between two principles: "*that all our perceptions are distinct existences,* and *that the mind never perceives any real connexion among distinct existences*" (T Appendix 21; SBN 636). These principles play an important role in Hume's accounts of causation and of external objects as well as in his account of the self.

It is obvious that the principles are not inconsistent with each other. But if they are both true, the implications are far-reaching and radical. If they are true, then whenever we find distinct existences, even if we have good reasons to believe that they are connected so as to form some sort of system, we will be frustrated in our attempts to discover the links that actually connect them. And we find distinct existences in very important places. Hume points out that causes and effects are distinct existences that we nonetheless take to be connected. The regularity of their conjunction is our grounds for believing that there is a connection, but we can find nothing in the causally related objects – or anywhere else – that would explain this regularity. Hume argues that anyone who has tried to give an account of causal powers or to identify the connecting link between cause and effect has run afoul of one or the other of the two principles: either by failing to acknowledge the distinctness of cause and effect, or by speaking of some sort of metaphysical bond without having any idea of the bonding agent (T 1.4.3.9–10; SBN 223–4). In Hume's words, "it appears, that when we say we desire to know the ultimate and operating principle, as something, which resides in the external object, we either contradict ourselves, or talk without a meaning" (T 1.4.7.5; SBN 267).

Hume's discussion of our idea of external bodies has a similar structure. He describes our belief that objects persist over time as a belief about connected objects or perceptions. To believe that the desk I see now is the same as the desk I saw earlier is simply to believe that there is a special kind of connection between the desk I see now and the desk I saw earlier. Hume does not take this belief to be arbitrary or unjustified. The observed coherence and consistency of our experiences of external objects constitute our grounds for believing that the objects are connected. However, we can find nothing – in the connected object themselves or anywhere else – that would explain this coherence and consistency or provide the principle of connection. Those who have tried to give an account of the connecting principle (linking together the parts that constitute the enduring objects of perception) have run afoul of one or the other of the two principles: either they fail to recognize the distinctness of the distinct objects of perception, which is the case with the vulgar (T 1.4.2.43; SBN 209), or they speak of some sort of metaphysical bonding agent (like a substance or substratum) without having any coherent idea of what they are talking about (T 1.4.3.7–8; SBN 222–3).

When he tries to explain personal identity, Hume encounters a similar problem. A certain group of observed causal regularities among perceptions lead us to believe that these perceptions are connected together to form one self. Our belief is justified insofar as these regularities really exist. However, we can find nothing, no connecting principle, that would *explain* why these regularities exist. Those who have tried to provide an account of the connecting principle (linking the perceptions that constitute the self) have run afoul of one or the other of the two principles: either they fail to recognize the distinctness of the distinct parts of the self, or they speak of some metaphysical bonding agent (either a substantial self

in which perceptions inhere or a real connection, like that provided by consciousness when consciousness is supposed to connect the distinct perceptions) without having any coherent idea of what connects the connected parts.

Hume himself speaks of connections among the perceptions that constitute the self, but he makes it clear that his connecting principle does not do what is required of a metaphysical account of the self. As Hume says in the Appendix, summing up his Book I discussion: "If perceptions are distinct existences, they form a whole only by being connected together. But no connexions among distinct existences are ever discoverable by human understanding. We only *feel* a connexion, or a determination of the thought, to pass from one object to another" (T Appendix 20; SBN 635). And this feeling is just one perception among the many that make up the self. It is not capable of binding together distinct perceptions, nor can it explain the observed, and very real regularities we find in the psychological realm. Instead, the feeling itself depends on the prior existence of those regularities.

In the *Appendix*, Hume asserts that the task of explaining "the principles, that unite our successive perceptions in our thought or consciousness" is hopeless. He also says that problems for his account of personal identity arise *only* "when [he] proceed[s] to explain the principle of connexion, which binds [all our particular perceptions] together." When it comes to that explanation, Hume acknowledges that his account is "very defective" (T Appendix 20; SBN 635). That should not come as news to careful readers, since this "defect" was acknowledged in the original Book I account. The news is that if Hume is right about the implications of his two principles, then this defect is one that *any* coherent and intelligible account of the self must have. Hume's mention of the only possible solutions: "Did our perceptions either inhere in something . . . individual, or did the mind perceive some real connexion among them" (T Appendix.21; SBN 636) makes it quite clear that the problem he finds with his own account of personal identity is that it does not satisfy the constraint mentioned earlier: that a satisfactory metaphysical account of a composite self should say what directly connects the parts of the self together.

The problem is general. This is the same problem Hume encountered when trying to account for cause and effect and when trying to account for external objects. Here are general grounds for skepticism about metaphysical accounts of any complex phenomena. All these metaphysical explanations require that the mind conceive of a connection between distinct existences. However, if the mind never conceives of connexions among distinct existences, then no satisfactory metaphysical account of any complex thing is possible.

Hume does offer grounds for skepticism about metaphysical accounts of causation, external objects, and the self – accounts that try to *explain* the observed and undeniable regularities that form the basis for our beliefs about these things. Those same grounds, however, do not support skepticism about the existence of causal relations, external objects, and selves, or about our ability to know that when two objects are causally related, which objects are mind-independent and which are

not, and about where my self leaves off and someone else begins. Metaphysical theories of connecting principles would be nice, but they are not required either for the existence of connections or for justified beliefs about them. (See Strawson 1989 for a similar interpretation of the scope of Hume's skepticism.) Hume argues that our experiences of real regularities are the only justification we have for those beliefs, and these experiences provide all the justification we really need, especially since any attempt to go beyond experience in order to provide a metaphysical explanation of experienced regularities leads only to confusion, contradiction, and error – in the intellectual world as well as in the natural.

The fact that the same "contradictions" can be found in philosophical theories of the intellectual world as are found in the natural world would make the case for skepticism about metaphysical theories of *anything* even stronger than Hume had originally thought. That may be why Hume sees his review of the section on personal identity as providing a good *general* reason for skepticism (T Appendix 10; SBN 633). He already had good reasons to doubt that we can have knowledge of the principles that connect together the diverse elements that constitute the system of nature (via cause and effect) and the distinct parts that constitute external objects. Now, when he realizes that the same doubts arise concerning the principles that connect together the perceptions that constitute a mind, the case for skepticism is complete. No connecting principles are knowable, but satisfactory metaphysical accounts of both the mind and the external world require knowledge of connecting principles.

On this reading, Hume's skepticism is general but its target is limited. This skepticism is grounded in the principle that the mind never perceives connections among distinct existences. This is the reason it is limited, extending only to explanations of observed regularities in terms of connecting principles. Once we understand the reasons for skepticism about metaphysical accounts that posit connecting principles, we can see that these reasons need not undermine our beliefs in connections. Such beliefs, which Hume holds are inevitable and indispensable in the conduct of life, can be justified in the absence of satisfactory metaphysical accounts. The justification may be limited, incomplete, and subject to philosophical objections and doubts, but a limited and incomplete justification is better than no justification at all. It is certainly better than a proposed justification that obscures the facts or invents fictions that cannot be maintained consistently. And some philosophical objections and doubts are of such a nature that once raised, they can never be satisfactorily answered. (For example, Descartes's antecedent skepticism, which raises doubts about the reliability our all epistemic faculties, cannot be answered, since such an answer would require reliance on the very faculties whose reliability has been called into question.)

In assessing Hume's views on personal identity, one question that remains is whether he can account for our beliefs about other persons. While Hume himself does not give such an account, it appears his account can be extended to cover the case of other selves. Although we cannot directly perceive other people's

thoughts and feelings, these thoughts and feelings are made manifest in many ways. Hume notes that they are made manifest in speech, gesture, and action (T 3.2.1.2; SBN 477). If we can perceive our own thoughts, sensations, and feelings and note their external manifestations, and if we can detect causal patterns among both so that we have a sense of self as well as an understanding of the external manifestations and causes of these thoughts and feelings, then as long as we accept the principle that "the same effect never arises but from the same cause" (T 1.3.15, 6; SBN 173), which is one of Hume's rules for judging cause and effect, we can discover the causal relations that obtain between other people's perceptions sufficiently to ground judgments concerning other minds.

While much remains to be said about Hume's theory of personal identity, I close with a brief examination of the practical implications of Hume's theory of the self. Derek Parfit (1971) has argued that accepting an account of the self that explains selfhood in terms of psychological connections between changing mental contents could effect salutary changes by blurring the boundaries of self-regard and concern for others. While a full discussion of a Humean response to this intriguing suggestion is impossible within the limits of this article, a few sobering remarks seem in order. One of these remarks concerns Hume's views on the limits of self-construction. For Hume, the self is not a mere illusion, since there really are real regularities among the perceptions that constitute each individual mind. Moreover, the distinctions we make between one self and another are also based on facts about regularities. Whether we are composites of causally connected mental states or singular centers of consciousness, these facts remain as they are and form the basis for the distinctions we make between one self and another. If the facts are undeniable and the synthesizing activities are unavoidable, reflection will do little to undermine our common beliefs about the boundaries of the self and the proper objects of prudential concern.

As regards the therapeutic effects of any philosophical account of the self, Hume held that, generally, and happily for us, very refined reflections have little effect on practice (T 1.4.7.7 and 1.4.7.8; SBN 268). He would likely take this line with respect to metaphysical reflections on the nature of the self. Our prudential concerns are unlikely to be countered by such abstract reflections. If philosophy recommends a widening of concerns beyond the self, it may succeed, but only if something in human nature supports this recommendation. Sympathy is, for Hume, the natural principle in human beings that opens us to the sufferings and happiness of others and gives us a lively concern for them (T 2.1.11.2; SBN 316). But sympathy is based, in part, on our ability to imagine what another person thinks and feels. We often do this by imagining what we would think and feel under similar circumstances (T 2.1.11.5; SBN 317). Sympathy also requires that these imaginings have an affective element that spills over into our thoughts about other people's feelings. For this to occur, we must first care about our own pains and pleasures, our own happiness and misery. So, here is another paradox for paradox lovers: the borders between self and other must be drawn clearly before

sympathy can soften them. And sympathy, rather than metaphysics, represents our best chance at this self-expansion.

See also 5: BELIEF, PROBABILITY, NORMATIVITY; 6: CAUSATION; 7: IDENTITY, CONTINUED EXISTENCE, AND THE EXTERNAL WORLD.

References

Ainslie, Donald (2004) "Hume's anti-cogito." Paper presented at the 31st Annual Hume Conference, Tokyo, Japan.

Baier, Annette (1979) "Hume on heaps and bundles," *American Philosophical Quarterly* 16, pp. 285–95.

——(1991) *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge, MA: Harvard University Press.

Capaldi, Nicholas (1989) *Hume's Place in Moral Philosophy*. New York: Peter Lang.

Chazan, Pauline (1992) "Pride, virtue, and self-hood: a reconstruction of Hume," *Canadian Journal of Philosophy* 22, pp. 45–64.

Fieser, James (1989) "Hume's pyrrhonianism: a developmental interpretation," *Hume Studies* 15, pp. 93–119.

Fogelin, Robert J. (1985) *Hume's Skepticism in the* Treatise of Human Nature. London: Routledge.

Garrett, Don (1981) "Hume's self-doubts about personal identity," *Philosophical Review* 90, pp. 337–58.

——(1997) *Cognition and Commitment in Hume's Philosophy*. New York: Oxford University Press.

Henderson, Robert S. (1990) "David Hume on personal identity and the indirect passions," *Hume Studies* 16, pp. 33–44.

Kemp Smith, Norman (1941) *The Philosophy of David Hume*. London: Macmillan.

Locke, John (1975) *An Essay concerning Human Understanding*, ed. P. H. Nidditch. Oxford: Clarendon Press.

McIntyre, Jane L. (1989) "Personal identity and the passions," *Journal of the History of Philosophy* 27, pp. 545–57.

——(1993) "Norms for a reflective naturalist: a review of Annette C. Baier's *A Progress of Sentiments*," *Hume Studies* 19, pp. 317–23.

Noonan, Harold W. (2003) *Personal Identity*, 2nd edn. London and New York: Routledge.

Olson, Eric T. (1997) *The Human Animal: Personal Identity without Psychology*. New York: Oxford University Press.

Parfit, Derek (1971) *Reasons and Persons*. Oxford: Clarendon Press.

Patten, S. C. (1976) "Hume's bundles, self-consciousness and Kant," *Hume Studies* 2, pp. 59–75.

Pears, David (1990) *Hume's System: an Examination of the First Book of his* Treatise. Oxford and New York: Oxford University Press.

Pike, Nelson (1967) "Hume's bundle theory of the self: a limited defense," *American Philosophical Quarterly* 4, pp. 159–65.

Purviance, Susan M. (1997) "The moral self and the indirect passions," *Hume Studies* 23, pp. 195–212.

Robison, Wade L. (1974) "Hume on personal identity," *Journal of the History of Philosophy* 12, pp. 181–93.

Strawson, Galen (1989) *The Secret Connexion: Causation, Realism, and David Hume*. Oxford: Clarendon Press.

Stroud, Barry (1977) *Hume*. London: Routledge.

Swain, Corliss (1991) "Being sure of one's self: Hume on personal identity," *Hume Studies* 17, pp. 107–24.

Swoyer, Chris (1982) "Hume and the three views of the self," *Hume Studies* 8, pp. 43–61.

Traiger, Saul (1985) "Hume on finding an impression of the self," *Hume Studies* 11, pp. 47–68.

——(1987) "Impressions, ideas, and fictions," *Hume Studies* 13, pp. 381–99.

Waxman, Wayne (1992) "Hume's quandary concerning personal identity," *Hume Studies* 18, pp. 233–54.

Winkler, Kenneth P. (2000) "'All is revolution in us': personal identity in Shaftesbury and Hume," *Hume Studies* 26, pp. 33–40.

Further reading

Ainslie, Donald (2001) "Hume's reflections on the simplicity and identity of mind," *Philosophy and Phenomenological Research* 62, pp. 557–78.

Ashley, L., and M. Stack (1974) "Hume's theory of the self and its identity," *Dialogue* 13, pp. 239–54.

Baxter, Donald L. M. (1998) "Hume's labyrinth concerning the idea of personal identity," *Hume Studies* 24, pp. 203–34.

Beauchamp, Tom L. (1979) "Self-inconsistency or mere perplexity?," *Hume Studies* 5, pp. 37–44.

Biro, John I. (1976) "Hume on self-identity and memory," *Review of Metaphysics* 30, pp. 19–38.

——(1979) "Hume's difficulties with the self," *Hume Studies* 5, pp. 45–54.

Bricke, John (1977) "Hume on self-identity, memory and causality," in G. P. Morice (ed.) *David Hume: Bicentenary Papers* (pp. 167–74). Austin: University of Texas Press.

——(1980) *Hume's Philosophy of Mind*. Princeton, NJ: Princeton University Press.

Chisholm, Roderick (1976) *Person and Object*. LaSalle, IL: Open Court.

Cummins, Phillip D. (1973) "Hume's disavowal of the *Treatise*," *Philosophical Review* 82, pp. 371–9.

Flage, Daniel E. (1990) *David Hume's Theory of Mind*. London and New York: Routledge.

Giles, James (1993) "The no-self theory: Hume, Buddhism, and personal identity," *Philosophy East and West* 43, pp. 175–200.

Immerwahr, John (1977) "The failure of Hume's *Treatise*," *Hume Studies* 3, pp. 57–71.

Loeb, Louis E. (1992) "Causation, extrinsic relations, and Hume's second thoughts about personal identity," *Hume Studies* 18, pp. 219–32.

MacNabb, D. G. C. (1951) *David Hume*. London: Hutchinson University Library.

Martin, Raymond, and John Barresi (2000) *Naturalization of the Soul: Self and Personal Identity in the Eighteenth Century*. London: Routledge.

McIntyre, Jane L. (1979a) "Further remarks on the consistency of Hume's account of the self," *Hume Studies* 5, pp. 55–61.

——(1979b) "Hume on the self," in D. F. Norton, N. Capaldi, and W. L. Robison (eds.) *McGill Hume Studies* (pp. 79–88). San Diego: Austin Hill Press.

——(1993) "Hume's underground self," *Studies in Early Modern Philosophy* 3, pp. 110–26.

——(2003) "'So great a question': a critical study of Raymond Martin and John Barresi, *Naturalization of the Soul: Self and Personal Identity in the Eighteenth Century*," *Hume Studies* 29, pp. 363–73.

Nathanson, Stephen (1976) "Hume's second thoughts on the self," *Hume Studies* 2, pp. 36–45.

Noxon, James (1969) "Senses of identity in Hume's *Treatise*," *Dialogue* 8, pp. 367–84.

——(1973) *Hume's Philosophical Development*. Oxford and New York: Clarendon Press.

Passmore, John (1952) *Hume's Intentions*. Cambridge: Cambridge University Press.

Pears, David F. (1966) "Hume on personal identity," in D. F. Pears (ed.) *David Hume: a Symposium* (pp. 43–54). New York: St Martin's Press.

——(1993) "Hume on personal identity," *Hume Studies* 19, pp.289–300.

Penelhum, Terence (1955) "Hume on personal identity," *Philosophical Review* 54, pp. 571–89.

——(1967) "Personal identity," in P. Edwards (ed.) *Encyclopedia of Philosophy*, vol. VI (pp. 95–107). New York: Macmillan.

——(1975a) *Hume* (esp. ch. 4). London: Macmillan.

——(1975b) "Hume's theory of the self re-visited," *Dialogue* 14, pp. 389–409.

Pitson, A. E. (2002) *Hume's Philosophy of the Self*. London: Routledge.

Reid, Thomas (1892) *The Philosophy of Reid*, *as contained in the* Inquiry into the Human Mind on the Principles of Common Sense, ed. E. H. Sneath. New York: Henry Holt.

——(1969) [1785] *Essays on the Intellectual Powers of Man*. Cambridge, MA: MIT Press.

Robison, Wade L. (1979) "In defense of Hume's *Appendix*," in D. F. Norton, N. Capaldi, and W. L. Robison (eds.) *McGill Hume Studies* (pp. 89–99). San Diego: Austin Hill Press.

Roth, Abraham Sesshu (2000) "What was Hume's problem with personal identity?," *Philosophy and Phenomenological Research* 61, pp. 91–114.

Yang, Xiaomei (1998) "Hume's view on personal identity: scepticism or nonscepticism?," *History of Philosophy Quarterly* 15, pp. 37–56.

9

Hume's Conclusions in "Conclusion of this Book"

Don Garrett

In the Introduction to *A Treatise of Human Nature*, Hume proposes to remedy what he calls "the present imperfection of the sciences" by undertaking to establish a "science of man" or human nature – what would now be called cognitive and conative psychology – that will be "the only solid foundation for the other sciences" (T Introduction 6; SBN xvi). The prospects for this approach seem to him bright. In Book 1 of the *Treatise* (entitled "Of the Understanding"), Hume examines in considerable detail "ideas, their origin, composition, abstraction, connexion, &c." (Part 1); "the ideas of space and time" (Part 2); "knowledge and probability" (Part 3); and "the sceptical and other systems of philosophy" (Part 4). However, as he begins Book 1's final section – T 1.4.7, entitled "Conclusion of this book"[1] – he is almost immediately engulfed by what he calls "desponding reflections." In the course of considering some of his previous findings, he comes to conclude that "we have, therefore, no choice left but betwixt a false reason and none at all" (T 1.4.7.7; SBN 268). He goes on to declare himself "ready to reject all belief and reasoning," explaining that he "can look upon no opinion even as more probable or likely than another" (T 1.4.7.8; SBN 268–9); and in the concluding sentence of the section, he clearly identifies himself as "a sceptic" (T.1.4.7.15; SBN 274). Yet he then goes on in Books 2 and 3 of the *Treatise* to pursue – with confidence seemingly unabated – detailed investigations "Of the Passions" and "Of Morals"; and in doing so, he regularly makes use of the main theses of Book 1 without so much as a mention of the extraordinary "desponding reflections" of its concluding section. The puzzle of how to understand the skepticism expressed in the concluding section of Book 1 (on some points anticipated in earlier sections of *Treatise* 1.4) in relation to the optimistic conduct of the science of man that dominates the rest of the work constitutes perhaps the most fundamental problem in the interpretation of Hume's *Treatise*.[2]

This fundamental problem often finds particular expression in the question of whether Hume in the *Treatise* regards at least some human beliefs – including of course his own, as expressed in that very book – as *rationally justified*. Many readers have understood the *Treatise*, despite its vigorous defense of numerous doctrines in the science of man, as straightforwardly denying that any human beliefs are rationally justified (Fogelin 1985; Waxman 1994; Singer 1995; Broughton 2003). It has also been suggested that he regards some human beliefs as justified, but only in a *practical* rather than a *theoretical* way (Ridge 2003); that he regards some beliefs of the "vulgar" as justified but not those of philosophers (Loeb 2002); and that he alternates unstably between regarding human beliefs as rationally justified and regarding them as rationally unjustified (Fogelin 1993, 1998). Others have proposed that the vivid skeptical remarks of "Conclusion of this book" do not represent Hume's own worries at all but instead dramatize a skepticism that he thinks must plague those who accept a rationalist orientation that he himself rejects (Baier 1991; Morris 2000).

The same fundamental problem also finds particular expression in another common question: that of the relation between Hume's *skepticism* and his *naturalism* in the *Treatise*. Thus, some have interpreted the *Treatise* as reaching skeptical conclusions that necessarily undermine all positive claims, including those of his own naturalism (Meeker 2000). Others have understood its philosophy as a successful and principled subordination of skepticism to naturalism (Kemp Smith 1941). Some have proposed that Hume's skepticism and his naturalism stand in an uneasy and problematic relation (Singer 2000), while others regard them as straightforwardly incompatible elements combined by Hume only at the price of internal contradiction (Passmore 1952). Still others have argued that Hume's skepticism and his naturalism are compatible and even mutually supporting (Fogelin 1985).

In order to understand the philosophical role of "Conclusion of this book" in Hume's philosophy, it is important to answer both of these common but contentious questions. Complicating the attempt to answer them, however, is the fact that their key terms – "rationally justified," "Hume's skepticism," and "Hume's naturalism" – are highly ambiguous. In order to see how these terms can be disambiguated and the questions containing them answered, it is necessary to understand, first, some features of Hume's approach in the *Treatise* to the science of man; and, second, the particular structure and content of "Conclusion of this book" itself.

Some Features of Hume's Approach to the Science of Man

The investigation of faculties

Book I of the *Treatise* is an attempt to address important philosophical questions through an investigation of the manner of operation of human psychological

faculties. It is sometimes supposed that Hume is hostile to the very notion of faculties (see, for example, Owen 1999), for he criticizes philosophers who "say, that any phenomenon which puzzles them arises from a faculty or an occult quality, and there [make] an end of all dispute and enquiry upon the matter" (T.1.4.3.10; SBN 224). In fact, however, Hume himself makes central and extensive use of the term "faculty" in the *Treatise*; it occurs in that work over sixty times, beginning in the fourth paragraph of the "Introduction." From the fact that the mind engages in a given activity, he readily and repeatedly infers that it has a faculty or power for engaging in that activity. His objection is not to characterizing things as having faculties; rather, it is to supposing that by doing so one has already isolated and provided ultimate explanations – that is, explanations grounded ontologically in supposed "real necessary connections" in nature – for the behavior of things. Hume's own aim as a philosopher, in contrast, is to provide adequate explanations of the manners of operation of mental faculties by discovering true causal generalizations that encompass those operations, thereby rendering "all our principles as universal as possible, by tracing up our experiments to the utmost, and explaining all effects from the simplest and fewest causes" (T Introduction 8; SBN xvii).

Hume divides the mind's perceptions into *impressions* and *ideas*, with the former distinguished from the latter by their greater liveliness, or "force and vivacity." He recognizes two primary faculties for having impressions: *sensation* is the faculty of having impressions not produced by other perceptions, while *reflection* is the faculty of having impressions that are produced by other perceptions. In addition to faculties for having impressions, he recognizes two primary faculties for having ideas: the *memory* and the *imagination*. Memory, as he understands it, is a faculty for having imagistic ideas that retain both the order and a distinctive share of the force and vivacity of the impressions from which they were copied, while imagination is the faculty for having all other ideas that are images of impressions. He argues that the cognitive faculties of conception, judgment, and reasoning are all ultimately particular ways of having ideas (T 1.3.7.5n; SBN 96–7), and hence he regards them all as aspects or sub-faculties of the imagination as just defined. (Even more specific faculties include "genius," which he explains as the ability – greater in some minds than in others – to summon precisely the ideas that are most useful for thought in a particular situation.) In recognizing memory and the imagination as the only two primary faculties for having ideas, Hume is in effect denying the existence of a separate faculty of intellect – that is, a representational faculty of having ideas that are not imagistic – of the kind assigned such central importance by such "rationalist" seventeenth-century philosophers as Descartes, Spinoza, Malebranche, and Leibniz.

In addition to this broad sense of "imagination," however, in which imagination is contrasted with memory, Hume also takes care to distinguish a narrower sense of the term, in which imagination is contrasted with *reason* and comprises all of the same mental operations that are included under the broader sense "excluding

only our demonstrative and probable reasonings" (T 1.3.9.19n; SBN 117–18). As this explanation of the narrower sense implies, he uses the term "reason" to designate the faculty of reasoning; indeed, he identifies it as a "faculty" and as the capacity to engage in reasoning many times in the course of the *Treatise*. As the same explanation also indicates, he recognizes two different kinds of exercises of this faculty: *demonstrative* reasoning and *probable* reasoning. Instances of the former depend only on the unchangeable relations among ideas, and the denials of their conclusions are literally inconceivable; instances of the latter depend ultimately on the relation of cause and effect, as discovered by experience of constant conjunction, and their denials are conceivable. Probable reasoning, in this broad Lockean sense, includes inferences conveying full assurance based on uniform and exceptionless experience, which he calls "proofs"; and it also includes "probability" in a narrower (and more familiar) sense, in which some uncertainty or lack of full assurance remains as a result either of ignorance of relevant causes that would determine an outcome from among seemingly similar "chances" (as in "the probability of chances"), or of mixed experience (as in "the probability of causes"), or of only partial resemblance with previously observed circumstances (as in "analogy") (T 1.3.11–12).

It is often said (see, for example, Winters 1979; Baier 1991; Ridge 2003) that Hume uses the term "reason" in multiple senses in the *Treatise*. In fact, however, aside from an obviously different use of the term in such phrases as "having a reason" or "give a reason" (where it means something like "an approvable consideration in favor of believing or doing something"), he always uses the term quite univocally to designate precisely the inferential faculty in his faculty-based psychology. This feature of his usage is often missed by commentators who, understandably eager to identify the normative content of his philosophy, interpret "reason" itself as primarily a normative term. It can also be obscured by an understandable tendency to interpret his uses of the term "reason" in accordance with a much broader (and more contemporary) faculty sense of "reason" according to which it designates something like an appropriate responsiveness to reasons generally. But although Hume certainly allows that there *are* reasons – that is, approvable considerations for believing or doing things – and he sometimes uses the distinct term "reasonable" to describe those who recognize and respond well to them,[3] he never describes such broad responsiveness as "reason." Moreover, while he recognizes reasons and alludes to them throughout the *Treatise*, reasons are not explanatorily basic for him: it is not that some considerations are approvable because they are (mysteriously and non-naturally) *reasons*; rather, they constitute *reasons* because they are (naturally and non-mysteriously) humanly approvable considerations. The natural operations by which they achieve and maintain this status of approvability as considerations – by Hume and by others – are among the primary targets of his investigation.

Those who suggest that Hume uses the term "reason" in multiple senses in the *Treatise* are often motivated at least in part by the apparent difficulty of interpret-

ing univocally his famous conclusion in *Treatise* 1.3.6: that the inferences he calls "probable," "causal," and "founded on experience" – and which we now call "inductive" – are not "determin'd by reason" (T 1.3.6.12; SBN 92). Since these inferences clearly *are* exercises of the inferential faculty – as he consistently allows whenever he specifies the scope of reason, and as he also implies by calling them "inferences" and "reasonings" throughout the very argument in question – it appears to many that Hume must be using the term "reason" in some other sense in his famous conclusion: for example, to deny that inductive inference bestows rational warrant or justification (Stove 1973); to deny that such inference satisfies a Cartesian model of *demonstrative* or *deductive* or *a priori* reason (Beauchamp and Rosenberg 1981; Baier 1991); to deny that it satisfies a Lockean model of reasoning that essentially involves "intermediate ideas" (Owen 1999); or to deny that it involves a specifically "rational perception" or "rational insight" (Millican 2002). Yet far from announcing any of these suggested changes in meaning, Hume never gives the slightest sign that he is introducing *any* new sense of "reason" at this or any other crucial point in the *Treatise*. This conduct stands in stark contrast to his usual practice of making an explicit acknowledgment when he uses a key term (e.g., "imagination" or "probability") in two different senses.. Hume does describe other common uses of the term "reason" (T 2.3.3.6–9: SBN 414–18), but he is careful to indicate that these are not approved uses.

The solution to the interpretive problem is not to distinguish two (or more) senses of "reason" but rather to distinguish between an inference's being an *exercise* of reason and its being *causally determined by* reason. For in general, when Hume raises the question of whether something is *caused*, *produced*, or *determined* by a particular faculty (for example, whether the faculty of reason can alone produce action), he is asking not whether it *is* an exercise of that faculty – a point that is generally quite obvious, one way or the other – but rather whether it is *caused by* an exercise of that faculty. (Indeed, quite generally, he regards faculties and abilities as causes *through* the effects of their exercises, not as causes *of* their exercises.) And since nothing can cause itself, the obvious fact that an inference *is* by its nature an exercise of reason does not entail that it is *caused by* an exercise of reason. Hume argues that all probable inference involves exposure to the constant conjunction of objects of one type with objects of a second type, followed by a mental transition from an impression or lively idea of one to a lively idea of the other; and he calls this transition the "presumption of a resemblance betwixt those objects of which we have had experience, and those of which we have had none," "transferring the past to the future," or "putting trust in past experience." The mechanism that produces this transition, he argues, is not, as one might think, any further argument or reasoning about the relation between what has been observed in the past and what is observed at present, but is rather an immediate mechanism of "custom or habit" that bypasses any reasoning or conclusion about that relation. Thus, in claiming that reason does not "determine" probable (i.e., inductive) inferences, he means that these inferences – while of course they are

themselves exercises of reason and examples of reasoning – require a crucial transition of thought that is not itself mediated by any component process of reasoning (as Lockean and Cartesian "demonstrations" *are*, in contrast, often mediated by component demonstrations; see Garrett 1998 and 2001).

The investigative use of faculties

For Hume, psychological faculties are among the primary objects of investigation; but they are also, of course, among the primary instruments with which any investigation must be carried out. In his later *An Enquiry concerning Human Understanding*, he describes Descartes's method as requiring at the outset:

> an universal doubt, not only of all our former opinions and principles, but also of our very faculties; of whose veracity, say they, we must assure ourselves, by a chain of reasoning, deduced from some original principle, which cannot possibly be fallacious or deceitful. (EHU 12.1.3; SBN 149)

Hume unequivocally rejects this requirement, on the grounds that it is necessarily self-defeating:

> But neither is there any such original principle, which has a prerogative above others, that are self-evident and convincing: Or if there were, could we advance a step beyond it, but by the use of those very faculties, of which we are supposed to be already diffident. The Cartesian doubt, therefore, were it ever possible to be attained by any human creature (as it plainly is not) would be entirely incurable; and no reasoning could ever bring us to a state of assurance and conviction upon any subject. (EHU 12.1.3; SBN 150)

In Humean inquiry, by contrast, at least some of our faculties must begin with at least some prima facie authority; and if the inquiry is to proceed by way of argument and inference, then at least some aspects of reason must surely be among those faculties (as must memory, we may assume). However, the fact that at least some faculties have provisional authority does not entail that any faculties are immune to criticism, revision, or even outright rejection. It entails only that any skepticism about them will not be "antecedent" skepticism of the kind he attributes to Descartes but rather "consequent": the result of, not preliminary to, particular investigations into the nature of those faculties (investigations conducted, of course, with those very faculties).

Thus it is the naturally accepted operations of human faculties themselves that constitute the starting point of inquiry for Hume. He notably does *not* begin the *Treatise* by proposing a priori epistemic normative principles. This absence is no oversight, for mere *relations of ideas* – the only possible source of genuine a priori

knowledge in his science of man – are incapable of providing normative epistemic principles much beyond a basic principle of non-contradiction. (The principle of non-contradiction – that is, the principle that the conjunction of a proposition and its denial cannot be true – might perhaps be considered both epistemically normative and a priori; for judging a proposition true or untrue arguably involves a particular kind of epistemic approval or disapproval of believing or assenting to it; and "relations of ideas" involving the idea of truth itself guarantee that a judgment and its denial cannot both be true. Yet this alone is not even enough to establish, on an a priori basis, the further normative principle that one *should never assent both to a proposition and to its denial*; for it might turn out that, as a matter of psychological necessity or contextual exigency, one could successfully affirm some particularly important truth – a valuable thing to do – *only* by affirming both it and its denial.) Indeed, even the normative *moral* principles that he discusses later in the *Treatise*, in Book 3, are not a priori on his view; they are, instead, the empirical deliverances of a moral sense that is, in many important respects, similar to the five ordinary senses.

Although some kinds of exercises of one's faculties may have, at the beginning of a project of Humean inquiry, greater or lesser authority than others, such differences of authority are thus not the consequences of an a priori theory of epistemic warrant, but rather the natural results of previous exercises of one's faculties in and on experience. For example, although Hume begins the *Treatise* with a commitment to the "experimental method," his Introduction makes it clear that this commitment is based on a confidence in the method's reliability that results naturally from the observation of its apparent past successes in natural philosophy (i.e., natural science), not on any a priori intuitive or demonstrative insight that it *must* be the proper method. Again, he notes in the course of his investigations of probable reasoning in the narrow sense (i.e., reasoning producing less than the full assurance of a uniform experiential "proof") that some operations of such reasoning are accepted as "philosophical," while others are rejected as "unphilosophical" (T 1.3.13); and in explaining the psychological origins of these contrasting evaluations – which he clearly already shares as the result of the very originating operations he describes – he finds no reason to quarrel with them. Finally, at the end of his investigation of probable reasoning (in the broad sense) in Part 3 of Book 1, he puts forward for adoption a set of "rules by which to judge of causes and effects" (T 1.3.15). Once again however, these are not ascertained a priori, but "are formed on the nature of our understanding, and on our experience of its operations in the judgments we form concerning objects" (T 1.3.13.12; SBN 149), that is, by reviewing the apparent past successes and failures of particular kinds of probable arguments. (For a different view of the role of normative principles in Humean inquiry, see Broughton 2003.)

In keeping with this lack of reliance on substantive a priori epistemic norms, Hume in the *Treatise* generally and quite reasonably seeks, as much as possible,

to defer addressing normative questions about the faculties he is investigating until most or all of the results about their operations have been assembled. Thus, for example, his famous discussion of the mechanism of inductive inference in *Treatise* 1.3.6 ("Of the inference from the impression to the idea") contains no hint of skepticism, despite the qualms that its conclusions about the nature of this mechanism might naturally arouse (and have aroused) in readers; on the contrary, no consideration or even mention of the question of what effect this discovery should have on our *confidence* in inductive inference occurs until the very final section of Book 1, "Conclusion of this book." Even his discovery in the first section of Book 1, Part 4 ("Of skepticism with regard to reason") of a way in which "all the rules of logic require a continual diminution, and at last a total extinction of belief and evidence" (T 1.4.1.6; SBN 183) there provides only confirmation for his prized theory of the nature of belief, as consisting in the relative liveliness of ideas (a degree of liveliness below that of impressions and memories), against a competing theory; again, it is only in "Conclusion of this book" that the discovery is brought to bear on the critical *evaluation* of reason as a source of belief.[4] This is not, of course, because he is interested only in the descriptive at the expense of the normative, nor is it because he initially failed to notice the bearing of his discoveries on normative questions; rather, he is simply aiming to defer the normative evaluation of our faculties until after they have delivered as *much* descriptive data about themselves as possible.

These two facts about Hume's method in the *Treatise* – that he begins not with normative principles governing epistemic warrant but with the naturally accepted exercises of our faculties, and that he generally seeks to defer the normative evaluation of the faculties being investigated as long as possible – suggest two corresponding cautions about the interpretation of the *Treatise*. First, it should not be assumed that Hume himself intends or endorses unstated normative conclusions when these conclusions would not follow from his descriptive claims without the assistance of equally unstated normative principles – even if those normative principles happen to strike the *interpreter* as evident and a priori. Second, it should not be assumed that Hume's psychological *description* of the way an argument or discovery makes him *feel* at a particular time necessarily constitutes normative endorsement of the response so described.

Structure and Content of "Conclusion of this book"

"Conclusion of this book" has three main parts. The first is a consideration of sources of perplexity and doubt, which might be called Hume's "skeptical recital"; the second is a narrative description of a succession of states of mind that follow upon this consideration; and the third is a discussion of the appropriate role of skepticism in philosophical inquiry.[5]

Causes of doubt

In "Conclusion of this book," Hume stops to assess his epistemic situation before entering upon the new topics of Books 2 and 3. In a famous metaphor, he finds his ship of inquiry to be a "leaky weather-beaten vessel" that has "narrowly escaped shipwreck," thus raising doubts about its suitability for further use. He mentions his memory of "past errors and perplexities" as one cause of his "diffidence for the future," but he finds particular cause for apprehension in the "wretched condition, weakness, and disorder of the faculties" that he must employ. After drawing a distinction between the "numberless infirmities" that are "peculiar to" himself and those that are "common to human nature," he goes on to detail five sources of errors and perplexities indicative of the cognitive infirmities that are common to human nature – all of them suggested by the specific results of his previous investigations.

The first of these sources of perplexity concerns the manner in which "the memory, the senses, and the understanding are all of them founded on the imagination, or the vivacity of our ideas" (T 1.4.7.3; SBN 265) – that is, the manner in which all of these faculties are dependent on the feature of the imagination by which ideas acquire increased liveliness, or "force and vivacity," under certain circumstances. The source of this worry in relation to "the understanding" lies in his famous investigation of probable (i.e., inductive) reasoning in *Treatise* 1.3, which concludes that, in all such reasoning, past experience of a conjunction between two types of events has produced a mental association whereby an impression or lively idea of an event of one type gives rise to a lively idea of an event of the other type – with this liveliness of the second idea constituting the assent to the conclusion. Because the mechanism that explains this associative transition is "custom or habit," rather than a separate mediating inference about the relation between past experience and the present instance, Hume now remarks, in "Conclusion of this book," that he is unable to "give a reason" why he should assent to such reasoning; that is, he is unable to cite a consideration in favor of the claim that assenting to the conclusions of reasoning based on such a mechanism will yield beliefs in conclusions that are true. It is presumably because of this inability that he describes the imagination's tendency to "enliven some ideas beyond others" as "seemingly so trivial." In bringing "the senses" within the scope of this source of perplexity, Hume is drawing on the conclusions of *Treatise* 1.4.2 ("Of scepticism with regard to the senses"), in which he argues that the belief in the existence of *bodies*, as objects that are distinct from the mind and continue to exist when unperceived, also depends on the enlivening of certain ideas – an enlivening that occurs only as the result of the mind's conflation of different kinds of perceptual series. Finally, in bringing "the memory" within the scope of this cause of perplexity as well, Hume is relying on his earlier claim that the assent characteristic of memory, too, is a matter of its liveliness, or force and vivacity (T 1.1.3.3; SBN 9).

The second source of perplexity lies in a conflict discovered in *Treatise* 1.4.4 ("Of the modern philosophy") between our "reasoning from cause and effect" and the belief in bodies as continued and distinct existences. There is, Hume argues, a "satisfactory" causal argument, discovered by "the modern philosophy," from the relativity of sense perception to the conclusion that nothing in bodies resembles our perceptions of such "secondary" qualities as colors, sounds, tastes, smells, heat, and cold (T 1.4.4.3–4; SBN 226–7). Yet, he goes on to argue, we can conceive of the other, "primary," qualities of bodies, such as extension, solidity, and motion, only by attributing to those bodies spatially distributed secondary qualities, such as color or tactile qualities, to occupy the locations of these bodies and distinguish them from their surroundings. Hence, if the "satisfactory" causal argument is correct, the only determinate way in which we can conceive of bodies will be an erroneous one. This outcome provides further cause to question the reliability of the idea-enlivening feature of the imagination that produces both causal reasoning and the belief in bodies.

The third source of perplexity lies in the existence of a persistent illusion, uncovered in Treatise 1.3.14 ("Of the idea of necessary connexion"), by which the mind supposes itself to perceive an "energy," "tie," or "efficacious quality" in objects that shows how and why causes are necessarily connected with their effects. The illusion arises, Hume argues, because the mind mislocates an internal feeling of the necessity of its own mental transition between perceptions of causes and effects, interpreting this feeling as an impression of a quality in (or relation between) the causally related objects themselves. The result is that although we in fact do not even really understand what we mean by the claim to have discovered such an "energy," "tie," or "efficacy" – because the necessity of causal relations is not the sort of thing that could be a feature of the objects themselves – this defect is generally hidden from us. The problematic tendency of the mind to "spread itself on external objects" (T 1.3.14.25; SBN 167) that produces the illusory mislocation is another feature of the imagination, and thus raises even more generally the question of the extent to which we should accede to features of the imagination.

This question leads ultimately to what Hume calls a "very dangerous dilemma" concerning the question of the attitude we should take toward the various operations of the imagination. The full force of the dilemma, however, depends on what may be regarded as itself a fourth source of perplexity: reason's natural tendency toward reflexive self-subversion. As Hume now observes, he had shown in *Treatise* 1.4.1 ("Of skepticism with regard to reason") that

> the understanding, when it acts alone, and according to its most general principles, entirely subverts itself, and leaves not the lowest degree of evidence in any proposition, either in philosophy or common life. (T 1.4.7.7; SBN 267–8)

This is not itself a claim about rational justification; throughout the *Treatise* Hume consistently uses the term "evidence" to designate psychological "evident-

ness" or "assent." (For particularly clear cases in which "evidence" can only mean "evident-ness or assent" and not "rational justification," see T 1.3.13.3–6; SBN 144–6.) Rather, it is a claim about the natural result of the iterated application of probable reasoning, specifically (it seems clear) "the probability of causes," an ordinary aspect of the operation of reason discussed at length in *Treatise* 1.3.12 ("Of the probability of causes"). In the probability of causes, one's confidence that a particular outcome will occur in given circumstances varies directly (through a natural psychological mechanism that Hume details) with the proportion of such outcomes previously observed in similar circumstances.

By means of the probability of causes, the certainty of any demonstrative reasoning is naturally replaced by the less-than-maximal assurance characteristic of "probability" in the narrow sense: for we recognize that reason is itself

> a kind of cause, of which truth is the usual effect; but such-a-one as by the irruption of other causes, and by means of the inconstancy of our mental powers, may frequently be prevented. (T 1.4.1.1; SBN 180)

When, in conformity with this reflection, we "enlarge our view, to comprehend a kind of history of all the instances, wherein our understanding has deceiv'd us, compar'd with those in which its testimony was just and true," we find some instances in which, although we thought that we had successfully performed a demonstration, we later discovered that our faculties had in fact failed us. Reflection on this fact, performed in accordance with the probability of causes, naturally lessens the degree of confidence in the conclusion of any supposed demonstration to that of mere probability.

That is not all, however. For similar reasons, the degree of assurance attached to the conclusion of any *probable* reasoning (in the broader sense) – including but by no means limited to a conclusion that we have successfully performed a demonstration – will *also*, in accordance with the probability of causes, naturally decrease upon higher-order review of the fallibility of the faculties by which we have performed that probable reasoning. Furthermore, reflexive higher-order reviews – assessing the reliability of the faculties by which we have reached conclusions about the reliability of the faculties by which we have reached conclusions, and so on – can be iterated indefinitely; and, Hume claims, each iteration would naturally further diminish the degree of assent of the original conclusion (as well as that of all of the judgments at intermediate levels) until no assent at all remained in any judgment.

In order to see more specifically how Hume can suppose that this reduction would occur, it will be useful to introduce some terminology. Let "standard degree of assent" designate the degree of assent (i.e., force and vivacity) that a particular judgment would have *if* it were produced by straightforward reasoning in normal circumstances. Thus, if the judgment is the result of a genuine demonstration, the standard degree of assent will be the certainty characteristic of full demonstration;

if the result of a failed demonstration, then no assent at all; if the result of a genuine "proof," then the full degree of assent characteristic of proof; if the result of what Hume calls "the probability of chances," then a degree of assent proportionate to the number of positive chances in the total number of chances; if the result of analogy, then a degree of assent proportionate to the closeness of the analogy; and if the judgment is the result of the probability of causes, then the degree of assent proportionate to the number of experienced positive instances in the total number of experienced instances. Let "actual degree of assent" designate the degree of assent with which a judgment is actually made.

Now consider a judgment J1, reached either by a purported demonstration or by probable reasoning, having an actual degree of assent d1. Thus, the individual in question judges:

J1 (held with actual degree of assent d1)

But the probability of causes, Hume has argued, requires a survey of past instances of judgments like J1 with respect to their reliability, i.e., with respect to the concordance or lack of concordance between the *actual degree* of assent with which they are reached and the *standard degree* of assent. This review results in a second judgment, which is (of course) held with its own actual degree of assent, d2:

J2: Degree of assent d1 is not higher than the standard degree of assent for J1 (held with actual degree of assent d2).

Degree of assent d2 may be very high. But if d2 is less than full assent – as it will be if J2 is made in the awareness of past instances in which we sometimes feel an actual degree of assent for a judgment that is higher than its standard degree – then, Hume argues, the probability of causes requires that the occurrence of J2 will lower the actual degree of assent to J1 to a level *below* d1. But for the same reasons that the probability of causes required the making of J2, it also requires a third judgment, with its own actual degree of assent:

J3: Degree of assent d2 is not higher than the standard degree of assent for J2 (held with actual degree of assent d3).

Presumably d2 = d3, since both J2 and J3 will be made by reviewing the same set of cases. But in any event, d3 will be, for the same reason, less than full assurance. The fact that J3 occurs with less than full assurance will, by the probability of causes, decrease the actual degree of assent to J2 below its original d2; and this decrease in assent to J2 from d2 will result in a *further* decrease in the degree of assent to J1 (which was initially d1 but has *already* been diminished once). Likewise, the probability of causes requires a fourth judgment, this one concerning J3:

J4: Degree of assent d3 is not higher than the standard degree of assent for J3 (held with actual degree of assent d4).

As before, since d4 will be less than full assent (presumably, in fact, equal to d2 and d3), the result will be a decrease in the actual degree of assent to J3, which will produce a second diminution in the actual degree of assent to J2, which in turn will produce a third diminution in the actual degree of assent to J1. Identical

considerations will again apply at each higher level. Since Hume holds that the original actual degree of assent to J1 is a finite quantity of liveliness and that there is a finite minimum quantity of liveliness lost in any diminution of it, he concludes that a finite series of diminutions would entirely exhaust the assent to any judgment.

As a claim about what repeated standard reasoning *would* in fact do, given experience of past errors in reasoning, this conclusion does not invoke any a priori probability theory, nor is the diminution of all assent that he mentions to be identified with reaching a probability of zero (equivalent to certainty of the denial) as understood in the now-established mathematical probability calculus. Indeed, Hume's principles of probability are not intended to be a priori at all; rather, they are projected from the basic and philosophically approved operations of probabilistic reasoning that he identifies empirically in *Treatise* 1.3. On Hume's view, the natural self-subverting outcome of probabilistic reasoning that he identifies is *prevented* from occurring only by a "trivial feature of the imagination" whereby "after the first and second decision . . . the action of mind becomes force'd . . . and the ideas faint and obscure," so that "the mind reaches not its objects with easiness and facility" and "the posture of the mind is uneasy." For this he proposes a physiological explanation: "the spirits being diverted from their natural course, are not govern'd in the movements by the same laws, at least not to the same degree, as when they flow in their usual channel" (T 1.4.1.10; SBN 185). Because of this feature of the imagination, the higher-level iterated reflexive judgments lose their force and, hence, in this case, their ability to sap the force of previous judgments, leaving the assent to the original judgment largely intact. Thus, whereas Hume's *first* source of perplexity about our faculties lay in the mind's dependence on a seemingly trivial feature of the imagination that makes assent possible, *this* source of perplexity lies in the mind's dependence on a quite different feature of the imagination, one that serves to prevent the subsequent loss of that very assent by iterated reflexive reasoning.

Reason's potential self-subversion is a source of perplexity for Hume in its own right; but it also contributes essentially to a crowning source of perplexity, which lies in the "dangerous dilemma" that confronts us when we try to specify just what features of the imagination we should accept and which we should reject. One alternative, Hume notes, would be to accept and endorse every trivial feature of the imagination. But to this proposal he raises two obvious objections. First, it would lead us into so many "errors, absurdities, and obscurities" (including those analyzed and ridiculed in T 1.4.3, "Of the antient philosophy") that we would become ashamed of our credulity. Second, the beliefs that result from these features of the imagination are often directly contrary to one another, so that they cannot all be accepted together. Another alternative would be to reject the trivial features of the imagination and resolve to accept only the features constituting "the understanding, that is . . . the general and more establish'd properties of the imagination" (in the broad sense). However, this would require us to reject the

feature of the imagination by which reason is saved from self-subversion. Realizing this, one might hope to amend the alternative by adopting a maxim not to engage in "refined and elaborate reasoning"; but this is unsatisfactory, Hume argues, for three reasons. First, it would cut off all science, which depends on refined and elaborate reasoning. Second, if one trivial feature of the imagination (namely, the assent-salvaging feature) is to have weight, parity of treatment suggests that we must allow all of the others as well. Finally, the reasoning by which we conclude that we must adopt this principle is itself elaborate and refined, so that we violate the rule in the very act of establishing it. The result, he reports, is that we have "no choice left, but betwixt a false reason and none at all"; for if we accept and endorse the seemingly trivial features of the imagination, we run into absurdities and contradictions; and if we endorse none of them, then we are left with the conclusion that we should ultimately maintain no assent to any conclusion of reasoning (T 1.4.7.7; SBN 267–8).

Three states of mind

Rather than continue to address the dangerous dilemma directly, Hume reports that

> for my part, I know not what ought to be done in the present case. I can only observe what is commonly done; which is, that this difficulty is seldom or never thought of; and even where it has once been present to the mind, is quickly forgot, and leaves but a small impression behind it. Very refined reflections have little or no influence upon us; and yet we do not, and cannot establish it for a rule, that they ought not to have any influence; which implies a manifest contradiction. (T 1.4.7.7; SBN 268)

Without yet knowing what "ought" to be done about the dangerous dilemma, he proceeds to narrate his progress through a sequence of three states of mind that it initiates. The first state of mind is one of "melancholy and delirium":

> But what have I here said, that reflections very refined and metaphysical have little or no influence upon us? This opinion I can scarce forbear retracting, and condemning from my present feeling and experience. The intense view of these manifold contradictions and imperfections in human reason has so wrought upon me, and heated my brain, that I am ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another. (T 1.4.7.8; SBN 268–9)

In this condition, he observes, he finds himself unable either to decide on or to carry out any action.

Hume finds the "cure" for this condition not in reason but in nature, which either "relax[es] this bent of mind" or displaces it through "some avocation, and lively impression of my senses." Thus, after dining, playing backgammon, or con-

versing with friends, he finds that his doubts appear "so cold, and strain'd and ridiculous, that I cannot find it in my heart to enter into them any farther" (T 1.4.7.9; SBN 269).

The natural successor to philosophical melancholy and delirium is a state of "indolence and spleen." For although Hume is now willing to allow himself an "indolent belief in the general maxims of the world" with which he can participate in and enjoy "the common affairs of life," he still remembers the painful melancholy and delirium produced by philosophical inquiry and describes himself as unwilling to "torture my brain with subtilities and sophistries, at the very time that I cannot satisfy myself concerning the reasonableness of so painful an application." In thus "submitting to my senses and understanding" in common life while foreswearing philosophical inquiry, he at first seems to himself to be showing "most perfectly my sceptical disposition and principles" (T 1.4.7.10; SBN 269–70); for he is unwilling to act in a way contrary to his natural inclination without "good reason," and he is now doubtful of finding any reason for doing so that will appear to him to be sufficiently good. Indeed, were we ever to engage in philosophy again, he remarks, it would have to be "on sceptical principles, and from an inclination, which we feel to the employing ourselves after that manner." The principle to which he feels currently inclined, he realizes, is what we may call the "Title Principle" (Garrett 1997):

> Where reason is lively, and mixes itself with some propensity, it ought to be assented to. Where it does not, it can have no title to operate on us. (T 1.4.7.11; SBN 270)

The Title Principle seems initially to exclude philosophical inquiry; but as Hume tires of "amusement and company," he finds that he does feel a renewed inclination to employ himself in philosophy. This inclination arises primarily as the result of two passions: "curiosity" concerning basic principles and distinctions, and "ambition" to contribute "to the instruction of mankind" and to acquire "a name by my inventions and discoveries." (It is notable that these passions of ambition, as "the love of fame," and curiosity, as "curiosity, or the love of truth," are given prominent roles in Book 2 of the *Treatise*.) In addition, he reflects that, even without curiosity and ambition, it is difficult for "the mind of man to rest, like those of beasts, in that narrow circle of objects, which are the subject of daily conversation and action," especially because superstitious religion is bold in its speculative systems and arises "naturally and easily from the popular opinions of mankind" (T 1.4.7.13; SBN 271). If the question is whether to prefer religion or philosophy as a guide to speculative topics, he declares his clear preference for philosophy, for it is less dangerous than religion and less able to disturb us in the conduct of our lives. Given the passions of curiosity and ambition, the Title Principle itself allows and approves an end to indolence and spleen and a return to philosophy.

The role of skepticism

"Conclusion of this book" ends with reflections about the proper way to proceed in philosophy "upon skeptical principles," that is, in awareness of the infirmities of human cognition that he has discovered. Hume allows that the establishment of a true system of philosophy "may" be too much to be hoped for. Yet he aims at least to establish a system that is "satisfactory to the human mind, and might stand the test of the most critical examination"; and in promoting the study of human nature, he intends to contribute to the "advancement of knowledge" by pointing out to philosophers "more distinctly, those subjects, where alone they can expect assurance and conviction." He notes that the passions stirred by this project serve to overcome the indolence and spleen that "sometimes prevail" upon him. For those who have a similar "easy disposition," philosophy can be in accordance with skeptical principles despite the disturbing features of human cognitive faculties previously uncovered, for "a true sceptic will be diffident of his philosophical doubts, as well as of his philosophical conviction" (T 1.4.7.14; SBN 273).

Finally, Hume urges, "sceptical principles" allow us not only to conduct "elaborate philosophical researches" but even to "yield to that propensity, which inclines us to be positive and certain in *particular points*, according to the light, in which we survey them in any *particular instant*" – so long as the momentary expressions of certainty that naturally result are not understood as implying any "dogmatical spirit" (T 1.4.7.15; SBN 273).

The Rational Justification of Belief

Does Hume conclude in "Conclusion of this book" that no beliefs are rationally justified? The question is ambiguous in two different dimensions. First, a belief's being "justified" may be understood either as its being the object of a successful *process of justification* or as its being *justly deserving* of acceptance. Of course, for those who hold it as an a priori principle that a belief is justly deserving of acceptance if and only if it is the object of a successful process of justification, the distinction may be unimportant; but Hume announces no such principle. Second, the "rationality" of the justification of a belief may be understood either in terms of some relation to a specified *process of reasoning* (for example, in terms of generation *by* a process of theoretical or practical reasoning) or in relation to a specified *dimension of evaluation capable of providing or sustaining reasons* (for example, evaluation for conduciveness to truth). Again, for those who hold it as an a priori principle that all and only those beliefs that are capable of sustaining reasons stand in a suitable relation to reasoning, such a distinction may be unimportant; but Hume announces no such principle.

In seeking to understand Hume's conclusions about the status of human beliefs, we may distinguish at least three properties that beliefs may have. First, they may

be *generated by reason*; that is, they may arise as the conclusion of reasoning. Second, they may have *epistemic merit*; that is, they may be *justly deserving of belief* from the standpoint of evaluation that is concerned with the value of *truth*. Third, they may have *rational support*; that is, they may have epistemic merit as a *consequence* of the way in which they are generated by reason. In asking whether Hume regards any beliefs as rationally justified, then, we may be asking any of a number of different questions, depending on how "justification" and "rationality" are understood; but among them, at least, are these:

1   Does Hume regard any beliefs as having rational support?
2   Does Hume regard any beliefs as having epistemic merit?[6]

Rational support

Generation by reason is a precondition for rational support, and in the course of his investigations Hume discovers that the role of reason in generating certain key beliefs is at least problematic. For example, in uncovering the causes of the belief that there are *bodies* (i.e., objects having a "continued and distinct existence" apart from their being perceived by any mind) Hume concludes that, although a certain kind of "irregular reasoning" plays a role in a secondary mechanism that can contribute to the belief, it is nevertheless

> impossible . . . that from the existence of any of the qualities of [perceptions], we can ever form any conclusion concerning the existence of [continued and distinct bodies beyond our perceptions], or ever satisfy our reason in this particular. (T 1.4.2.47; SBN 212)

Moreover, his investigation of induction concludes that the "presupposition that nature is uniform" is not generated by reason either, although it must be noted that this "presupposition" is not itself, in the first instance, an explicit belief (i.e., a lively idea) at all, but rather a disposition to draw conclusions about unobserved cases on the basis of observed ones. When an explicit belief in the uniformity of nature does arise – as it eventually does for philosophers, at least – it *is* the outcome of a process of (probable) reasoning, albeit one that continues to depend (as all probable reasoning does) on the original "presupposition" that nature is uniform.

Epistemic merit

Given the antecedent provisional authority of our faculties, however, an absence of rational support does not by itself settle the question of the epistemic merit of beliefs for Hume. On the contrary, "Conclusion of this book" is, in large measure, an attempt to reach a final assessment of the epistemic merit of our beliefs in light

of his perplexing empirical discoveries about the nature of our cognitive faculties. A positive outcome is far from pre-ordained, as is clear from his striking remark (already cited), immediately after first confronting the "dangerous dilemma," that he is "ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another."

When first confronting the "dangerous dilemma," Hume reports an extreme loss of assurance (which he speculatively ascribes to a physiological cause, "heating of the brain") and appears to be considering the view that no beliefs have any epistemic merit. However, this is clearly a report of a temporary state of mind, and not an authorial endorsement of an epistemic principle. In *A Letter from a Gentleman to his Friend in Edinburgh*, a later pamphlet in which he responds to charges of radical skepticism, he warns explicitly against the error of interpreting this passage as an expression of his final view (writing of himself, as author of the *Treatise*, in the third person):

> All those Principles, cited . . . as Proofs of his Scepticism, are positively renounced in a few Pages afterwards, and called the Effects of Philosophical Melancholy and Delusion. These are his very Words; and his Accuser's overlooking them may be thought very prudent, but is a Degree of Unfairness which appears to me altogether astonishing. (Hume 1967)

Hume's final position is not a rejection of all belief and reasoning, but rather a return to philosophy with a renewed aim of contributing to "the advancement of knowledge," although now with a chastening sense of the limitations of his faculties and the likelihood that his particular conclusions may well be superseded by the discoveries of subsequent generations. He acknowledges, and even emphasizes as a valuable discovery, that reasoning alone is unable to produce this return. For it is the force of human nature, and not any reasoning, that first displaces philosophical melancholy and delirium in favor of indolence and spleen; and it is then the passions of curiosity and ambition, plus the contrast with superstition, that make possible the replacement of indolence and spleen with a renewed appetite for philosophy. But none of this entails that Hume regarded the denial of all epistemic merit (notwithstanding his report of its transitory appeal) as itself ultimately having epistemic merit; indeed, that would be highly paradoxical. Nor does it entail that he ultimately found no epistemic merit in the view that many beliefs do have epistemic merit. He not only allows but insists that our most central beliefs may, in light of the limitations and infirmities of our faculties, have somewhat *less* epistemic merit than we might optimistically have supposed. For example, in his later *Abstract of a Treatise of Human Nature*, he refers to the *Treatise* as tending "to give us a notion of the imperfections and narrow limits of human understanding" (A 17), but he never denies that our most central beliefs have merit. On the contrary, he quite clearly proposes in the Title Principle – which he accepts and never renounces – that they do.

Of course, if we ultimately have "no choice left, but between a false reason and none at all," then it is hard to see how any beliefs – or any generated by reason, at least – *could* have epistemic merit. The Title Principle, however, offers a third choice, and hence a way out of the dangerous dilemma. Because the elaborate and refined reasoning by which the probability of causes would undermine all belief is not lively but strained, and mixes itself with no propensity or interest, it may, according to the Title Principle, be ignored. Other reasoning, however, ought still be assented to; and since even the most elaborate and refined philosophical reasoning turns out to mix itself with the propensities of curiosity and ambition, this includes philosophical reasoning. In effect, the Title Principle excises from the domain of assent-worthy reasoning precisely the reason-undermining iterations of the probability of causes that were the original source of the dangerous dilemma, while leaving other reasoning intact. This principle was not available in Hume's original reasoning about the dangerous dilemma because the principle is not generated by reason itself; rather it is a natural product of the indolence and spleen that follows upon philosophical melancholy and delirium. Yet once we find ourselves adopting it, we may conclude through reasoning that the principle is conducive to the satisfaction of our ends. And, since this reasoning is itself in accordance with the principle, the application of the principle effectively approves of itself. Moreover, once we are reasoning in accordance with the Title Principle, we can also come to approve of it as conducive to the truth; for given any conclusion p reached by such reasoning, it follows in a single step of further such reasoning that p *is true*.

As Hume recognizes, a serious question nevertheless remains: can acceptance of the Title Principle survive reconsideration of the original sources of perplexity that caused such anxiety? His answer is that it can. Although the original causes of perplexity remain as reminders of the infirmities of our faculties, they need not entirely destroy either our confidence in those faculties or their authority; for we may turn our newfound modesty about our faculties against their very tendency to produce perplexity. It is for this reason that he can assert:

> A true sceptic will be diffident of his philosophical doubts, as well as of his philosophical conviction; and will never refuse any innocent satisfaction, which offers itself, upon account of either of them. (T 1.4.7.14; 273)

It might be thought that Hume's invocation of passions and his reference to "innocent satisfaction" indicate that he has after all abandoned the view that his beliefs have any genuine epistemic merit, in favor of the view that they are to be accepted *only* in virtue of their practical merit (e.g., their merit from the standpoint of pleasure, rather than truth, as a value). Clearly the Title Principle and the return to philosophy in accordance with it are not originally generated by reasoning about their truth-conduciveness, and the argument that they are conducive to the pleasurable satisfaction of our passions appeals to practical, rather than epistemic, ends. Yet assent to the conclusion that philosophy *will* bring us satisfaction if we

continue to pursue it in accordance with the Title Principle depends on premises that must themselves be accepted as true or likely to be true, and not merely as pleasurable. Moreover, the very taking of pleasure in philosophy through the satisfaction of the passions of curiosity (see T 2.3.10, "Of curiosity, or love of truth") and ambition ("to contribute to the *instruction* of mankind . . . by my inventions and *discoveries*") requires taking oneself to be discovering philosophical *truths* in the process. But if one regards one's beliefs as true or as likely to be true, then one will necessarily regard that as a mark in their favor from the point of view of truth as a value; that is, one will regard them as having some epistemic merit. Thus if Hume were to reject the epistemic merit of all beliefs, he would also lose the belief that they had any justifying practical merit. While the pursuit of philosophy in accordance with the Title Principle requires passions, the "ought" contained in the principle itself is epistemic, and not merely practical.

Nor does Hume limit the epistemic merit of beliefs to the vulgar, who never reflect sufficiently to encounter the infirmities of human cognition and thus can retain stable structures of belief. He seems to allow that doubt, indolence, and spleen may occasionally recur for philosophers, particularly when they consider the infirmities of human cognition intensely, just as their complete certainty may reappear when those infirmities are entirely forgotten and an object of consideration appears in a convincing light. Yet the Title Principle provides a consistent principle of epistemic merit that can be accepted and permanently endorsed by philosophers, including Hume himself, who could not proceed to investigate the passions and morals in Books 2 and 3 of the *Treatise* without it. His endorsement of the Title Principle, once made, is never renounced; and it is for this reason that he never has occasion to return in the *Treatise* to the causes of perplexity surveyed in "Conclusion of this book."[7]

Although Hume ultimately finds sufficient epistemic merit in his beliefs to sustain continued philosophical inquiry, at no point in the *Treatise* does he give any indication that the successive states of mind that he reports and expresses in "Conclusion of this book" – including philosophical melancholy and delirium – were not really his own, nor that they were meant merely to dramatize the unfortunate consequences of his opponents' principles. On the contrary, the perplexities, and his reactions to them, arise naturally from his own empirical discoveries concerning human cognition, and not from their conjunction with any separate principles that he had antecedently rejected. Indeed, the proper response to those perplexities requires, on Hume's view, that they *not* be avoided; for only in undergoing and surmounting them can the Title Principle be naturally elicited and established.

Skepticism and Naturalism

Skepticism

Commentators and others often refer to "Hume's skepticism" without indicating what they mean by this term. This is a cause of serious confusion, for instances of

skepticism may differ from one another in many different dimensions. As Robert Fogelin (1985, 1993) has observed, "skepticism" may refer to a psychological lack of assent, to a positive prescription to refrain from assent, or to a theoretical doctrine concerning the absence of rational justification for assent; we may call this dimension of difference the *character* of skepticism. Moreover, as we have seen, theoretical doctrines concerning the absence of rational justification for assent may themselves be distinguished (at least) into *rational support skepticism*, which concerns the absence of rational support, and *epistemic merit skepticism*, which concerns the absence of epistemic merit. Skepticism, whatever its specific character, may also differ with respect to the range of propositions to which it is applied; we may call this the *domain* of skepticism. Within a given domain, instances of skepticism may be mitigated to various degrees or entirely unmitigated; we may call this the *extent* of skepticism. Finally, a given degree of skepticism may vary or remain constant; this we may call the *persistence* of skepticism.

While Hume may be a constant and unmitigated rational support skeptic about some topics, such as the general doctrine of the existence of bodies, his acceptance of the Title Principle means that he is not an unmitigated epistemic merit skeptic about the results of reasoning generally, whether demonstrative or probable. Yet his newfound conviction that human cognition is subject to many infirmities – what he calls his "skeptical principles" – leave him with an awareness that our beliefs have *less* epistemic merit than one might have supposed. He is thus a mitigated epistemic merit skeptic about every domain of belief. While he considered, and may even have temporarily adopted, unmitigated epistemic merit skepticism in the midst of his philosophical melancholy and delirium, his final endorsement of mitigated epistemic merit skepticism is constant. His actual lack of psychological assent is, generally speaking, mitigated as well, amounting simply to what he characterizes as a lack of dogmatism and zealotry in belief; however, he allows that his psychological assent may temporarily diminish further when he considers intently the infirmities of human cognitive nature, just as it may temporarily increase to certainty when he considers an object of inquiry in a full and convincing light. Likewise, he *prescribes* – at least for those who are, in his phrase, "associates in these researches or auditors of these discoveries" (T 1.4.7.14; SBN 272) – a constant degree of the same mitigated lack of assent, so far as that is psychologically possible.

Naturalism

"Naturalism," too – a term that Hume does not use – can be understood to mean many different things. The doctrine that there is nothing outside of nature may be called *metaphysical naturalism*; and different versions of this doctrine result from different specifications of the kinds of things that would be outside of nature, if they were to exist. These may include, for example, a personal God; abstract, universal, or immaterial entities; irreducible and inexplicable representational or normative qualities; or causal influences that are not subject to ordinary causal

laws, such as miraculous interventions, the "agent" causation of libertarian free will, or occurrences of special intellectual insight. A corresponding program of trying to explain phenomena without appeal to anything outside of "nature," in one of its specifications, may be called a version of *explanatory naturalism*. The doctrine that all beliefs are produced by operations of nature may be called *doxastic naturalism*. Finally, the doctrine that beliefs can, in virtue of the ways in which they result from operations of nature, also have epistemic authority and be rightful objects of assent from the perspective of truth-conduciveness, may be called *epistemic naturalism*.

It is clear that Hume accepts at least some versions of metaphysical naturalism, understood in this way. He is skeptical about a personal God, includes no abstract or universal entities in his ontology, and rejects the question of whether there are immaterial substances. While he recognizes representational and normative qualities, he does not regard them as inexplicable. The former he explains through the functional roles played by representations; and the latter he explains in terms of the features (of agents, objects, and cognitions) that produce certain feelings (moral sentiments, aesthetic sentiments, and assent) in us and the role that those feelings play in our lives. He recognizes no miracles, "agent" causation or instances of special intellectual insight ungoverned by ordinary causal laws. Correspondingly, he is an explanatory naturalist in many respects as well; most importantly, his science of man seeks to explain the operations of the human mind in accordance with causal laws that have exactly the same status as the laws governing any other part of nature. One important outcome of this explanatory naturalism is his doxastic naturalism; for in developing such explanations, he appeals to basic aspects of human nature – in particular, features of the imagination – to explain why human beings do and must believe in bodies, accept the uniformity of nature, and in other respects interpret the world as they do. Finally, he is also a kind of epistemic naturalist; for he regards the faculties that provide our beliefs as having prima facie epistemic authority, and his final conclusions allow many of those faculties – when properly corrected and refined by reflective self-application – to retain at least a share of that authority.

A constant unmitigated lack of psychological assent would clearly undermine each of these types of naturalism, preventing any beliefs from arising and foreclosing the acceptance of any explanations. Unmitigated epistemic merit skepticism would also undermine Hume's naturalism, in all of these senses, by condemning, even if not fully preventing, all beliefs and proposed explanations as lacking in truth-conduciveness.

As we have seen, however, Hume is not an unmitigated epistemic merit skeptic. When lack of assent and epistemic merit skepticism are properly mitigated, they undermine only the *dogmatic* acceptance of Humean naturalisms without preventing or condemning those naturalisms themselves. The results of Hume's explanatory naturalism in "Conclusion of this book" serve to support his mitigated skepticism by uncovering the infirmities of human cognition; and Hume's miti-

gated skepticism constitutes, in light of those infirmities, the only viable attitude with which to continue the naturalism of the *Treatise* in Books 2 and 3.

See 5: BELIEF, PROBABILITY, NORMATIVITY; 6: CAUSATION; 7: IDENTITY, CONTINUED EXISTENCE, AND THE EXTERNAL WORLD; 8: PERSONAL IDENTITY AND THE SKEPTICAL SYSTEM OF PHILOSOPHY.

Notes

1 The final section of Book 3 of the *Treatise* is also entitled "Conclusion of this book." Throughout this essay, I will use the title to refer exclusively to the final section of Book 1.

2 A similar problem concerns the relation between the skeptical considerations discussed in the final section of *An Enquiry Concerning Human Understanding* and the positive character of the earlier sections of that work (and of Hume's subsequent writings). Although this essay will be largely confined to the question of interpreting the *Treatise*, many of the same considerations apply to the role of the final section of the *Enquiry* as well. For a comparison of *Treatise* 1.4.7 with the final section of the *Enquiry*, see Garrett 1997: ch. 10.

3 Hume also uses the term "reasonable" to describe such disparate entities as requests, anger, rules of painting, and prices – all items for which there are reasons. He sometimes uses the terms "reasonable" and "unreasonable" more narrowly (e.g., in T 2.3.3.6–7: SBN 416–17; and T 3.1.1.10; SBN 458), to denote a thing's approvability or rejectibility (respectively) by the inferential faculty in virtue of its capacity to be true or false (so that actions, for example, are incapable of being true or false and hence are neither reasonable nor unreasonable); but he does so only when the surrounding context makes it very explicit that this is his meaning.

4 The first substantial expression of skepticism in the entire *Treatise* occurs in *Treatise* 1.4.2, "Scepticism with regard to the senses," where Hume indicates that an accurate and sustained investigation (of the kind in which he is then engaged) of the operations leading to the two primary forms of belief in bodies as "continued and distinct existences" naturally produces a very temporary diminution of belief in such existences. Given his view that this kind of investigation naturally produces such feelings of doubt, it would be self-undermining *not* to express such feelings in the course of conducting the investigation. The question of the *final* assessment of the belief in bodies, however, is not settled until a decision is reached about how to regard our cognitive infirmities in "Conclusion of this book." Similar considerations apply to the concern he expresses in *Treatise* 1.4.4 ("Of the modern philosophy") about the "contradiction" he there discovers between the belief in bodies and a piece of standard causal reasoning.

5 For further discussion of each of these elements, see Garrett 1997: ch. 10.

6 These distinctions are also drawn in Garrett 2004, which contains further discussion of the precise character of Hume's skepticism.

7 He does return to review some of them (and add others) in the final section of *An Enquiry concerning Human Understanding*. In the *Enquiry*, however, their effects on

the mind, culminating in what is explicitly described, and praised, as "mitigated skepticism," are simply explained rather than narrated or enacted.

References

Baier, Annette (1991) *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge, MA: Harvard University Press.

Beauchamp, Tom L., and Alexander Rosenberg (1981) *Hume and the Problem of Causation*. New York: Oxford University Press.

Broughton, Janet (2003) "Hume's naturalism about cognitive norms," *Philosophical Topics* 31, pp. 1–19.

Fogelin, Robert J. (1985) *Hume's Skepticism in the* Treatise of Human Nature. London: Routledge.

——(1993) "Hume's skepticism," in David Fate Norton (ed.) *The Cambridge Companion to Hume* (pp. 90–116). Cambridge: Cambridge University Press.

——(1998) "Garrett on the consistency of Hume's philosophy," *Hume Studies* 24, pp. 161–9.

Garrett, Don (1997) *Cognition and Commitment in Hume's Philosophy*. New York: Oxford University Press.

——(1998) "Ideas, reason, and skepticism," *Hume Studies* 24, pp. 171–94.

——(2001) "Reply to my critics," *Philosophy and Phenomenological Research* 61, pp. 205—15.

——(2004) "'A small tincture of Pyrrhonism': skepticism and naturalism in Hume's science of man," in Walter Sinnott-Armstrong (ed.) *Pyrrhonian Scepticism* (pp. 68–99). Oxford: Oxford University Press.

Hume, David (1967) [1745] *A Letter from a Gentleman to his Friend in Edinburgh*, ed. E. C. Mossner and J. V. Price. Edinburgh: Edinburgh University Press.

Kemp Smith, Norman (1941) *The Philosophy of David Hume*. London: Macmillan.

Loeb, Louis E. (2002) *Stability and Justification in Hume's* Treatise. Oxford: Oxford University Press.

Meeker, Kevin (2000) "Hume's iterative probability argument: a pernicious reductio," *Journal of the History of Philosophy* 38, pp. 221–38.

Millican, Peter (2002) "Sceptical doubts concerning induction," in Peter Millican (ed.) *Reading Hume on Human Understanding* (pp. 107–73). Oxford: Oxford University Press.

Morris, William Edward (2000) "Hume's conclusion," *Philosophical Studies* 99, pp. 89–100.

Owen, David (1999) *Hume's Reason*. Oxford: Oxford University Press.

Passmore, John (1952) *Hume's Intentions*. Cambridge: Cambridge University Press.

Ridge, Michael (2003) "Epistemology moralized: David Hume's practical epistemology," *Hume Studies* 29, pp. 165–204.

Singer, Ira (1995) "Hume's extreme skepticism in *Treatise* I IV 7," *Canadian Journal of Philosophy* 25, pp. 595–622.

——(2000) "Nature breaks down: Hume's problematic naturalism in Treatise i iv," *Hume Studies* 26, pp. 225–43.

Stove, D. C. (1973) *Probability and Hume's Inductive Scepticism.* Oxford: Clarendon Press.

Waxman, Wayne (1994) *Hume's Theory of Consciousness.* Cambridge: Cambridge University Press.

Winters, Barbara (1979) "Hume on reason," *Hume Studies* 5, pp. 20–35.

Further reading

Mounce, M. O. (1999) *Hume's Naturalism.* London: Routledge.

Norton, David Fate (1982) *David Hume: Skeptical Metaphysician, Common-sense Moralist.* Princeton, NJ: Princeton University Press.

Popkin, Richard (1951) "David Hume: his Pyrrhonism and his critique of Pyrrhonism," *Philosophical Quarterly* 1, pp. 385–407.

Stroud, Barry (1977) *Hume.* London: Routledge.

——(1991) "Hume's scepticism: natural instincts and philosophical reflection," *Philosophical Topics* 19(1), pp. 271–91.

Part III

The Passions

10

The Powers and Mechanisms of the Passions

Lilli Alanen

Introductory Remarks

A skeptic may ask whether Hume's toolbox of impressions and ideas, so useful in his deconstruction of rationalist metaphysics, can also serve as the basis for a positive moral psychology and phenomenology. Can Hume offer a coherent account of human emotions without additional presuppositions which, by his own empiricist standards, do not stand up to scrutiny? Even granting him all his assumptions about human nature and propensities in general, and about the endowments of the mind or self in particular (they could, after all, be seen as revisable empirical hypotheses), questions remain about the consistency of his system as a whole and its interest from the point of view of philosophy of mind and cognitive psychology. What account can he give, for instance, of representation and intentionality beyond a purely causal theory of reference? How is his famous copy principle to be understood? Does his mechanistic and associationist psychology leave any space for judgment? Can the human mind or reason be said to contribute anything in the forming, for instance, of evaluative judgments, as opposed to being a mere passive recipient of impressions and beliefs bestowed upon it through the action of external circumstances and events?

This chapter is not directly concerned with these questions, but the examination of Hume's account of the passions in Book 2, which is an integral part of his theory of perception, may contribute to their answering. The first critical book of the *Treatise*, which has received so much attention on other grounds, paves the way for a new kind of a genuinely naturalist moral psychology, which is developed in Books 2 and 3, where the traditional picture of reason as an independent governor of the passions is reversed. Hume's theory recognizes, instead, the central role and value of passions for moral agency and reflection and suggests means to their mastery that do not appeal to anything over and above the resources our human nature – animal *and* social – can offer. Hume's complex view of the nature and role of the passions has often been oversimplified and unjustly classified

either as an emotivist theory or, at the other extreme, as a purely cognitive theory where the affective component is reduced to a mere epiphenomenon. Such readings no longer seem widely viable, and if one were to apply current labels to Hume's broadly naturalist account of the passions, it would fall in between the usual dichotomies, since it recognizes that emotions are cognitively informed and have their own affective intentionality.

To show this, it will help to consider Hume's theory in the larger context of influences in which it is embedded, including not merely British empiricist epistemology and Scottish moral naturalism, but also the tradition of Cartesian philosophy of mind and moral psychology, variously developed by Malebranche, Spinoza, and Locke. Since this Cartesian background is the least explored in this context, I will look briefly at the terminology and concepts Hume adopts from it before turning to his controversial account of the passions themselves and its interpretation.

The Cartesian Background

Descartes transformed the study of human passions in proposing to examine them not as a moral philosopher but rather, as he explains in a letter prefacing his *Les Passions de l'âme*, "en physicien," as a scientist of nature. It is important not to misinterpret this claim. The point is not to deny the role and moral significance of the passions, defined as a subclass of thought, or to replace current discourse about passions by neuro-physiology, as much as to understand and describe their bodily causes and effects and to treat them as phenomena on a par with other natural changes and events. This is also how those of his followers and critics who took their inspiration from Descartes's account of passions understood it. The human mind or soul, although it also follows laws of its own is, qua embodied, subject to the same laws governing the rest of nature. Among the new ideas in the Cartesian treatise on the passions adopted by his followers are: (1) the mechanistic and mathematical ideal of explanation for physical bodies and nature in general; (2) the extension of a mechanistic account to all the functions of living and animal bodies; (3) the treatment of the human mind or soul as embodied or substantially united to the body (this took different forms in different authors); (4) the recognition of the impact on the embodied mind and its trains of thought of the movements in the body and the laws governing them; (5) the adoption of a strict methodological dualism, in virtue of which the mind and its modes are irreducible to the body and cannot, therefore, be fully accounted for in the same terms as the changes in the body; (6) the ensuing account of sensations, emotions, and passions as complex psycho-physical states, where the associations of motions in the nerves and the brain with thoughts or perceptions have been set up to serve the ends of the mind–body union by what the Cartesians referred to as the "Institution of Nature"; and finally, (7) the recognition that passions or

emotions, in addition to being caused by and accompanied by various motions and physiological changes in the body, also have their "natural," instant, and spontaneous expressions in its countenance and behavioral reactions.

The full significance of these ideas for moral psychology, as well as the tensions between some of them, became apparent only in the work of those of Descartes's successors who were committed to naturalism and who also took the examination of human passions to be important for the understanding and regulation of human conduct. Spinoza and Hume are cases in point. Unlike Hobbes, who may have developed his theory independently, they both resisted the identification of thoughts with motions in the brain, but they took the idea of mechanization further, developing a quasi-mechanical account not only of modes of thinking like imagination and the passions, which in the Cartesian framework are thoughts depending on corporeal motions, but of reasoning itself. Spinoza, relentless in his criticism of Descartes's substance-dualism, interaction theory, and concept of a free will, gives him credit for at least attempting to explain the passions. But he thought of himself as completing that project. He recognized only one substance, at once thinking and extended, and could hence argue in the *Ethics* that the passions follow the "same necessity and force of nature as all other particular things" and can be explained as any other natural phenomena. So he proposed an investigation that would "consider human emotions just as if it were an investigation into lines, planes or bodies" (Spinoza 1985: 492). The idea was still alive a century later. Hume, who cared little about metaphysical essences and distinctions, had no problem in including mental and moral phenomena within the framework of his broad naturalism. He ends his little "Dissertation of the passions," published in 1757, by stating that he wanted to show "that, in the production and conduct of the passions, there is a certain regular mechanism, which is susceptible of as accurate a disquisition, as the laws of motion, optics, hydrostatics, or any part of natural philosophy" (Hume 1992). What Hume takes over from the ideology of the mechanistic philosophy of nature is the idea of causal regularities and constant conjunctions between thoughts in accordance with laws of association, which in his famous analogy are supposed to operate in the mental world just like the Newtonian laws of attraction work in the physical world. As an empiricist, he contents himself with observing the phenomena, mental or material, and with describing observed conjunctions between them, without worrying about their metaphysical nature or how these connections come about. (T Introduction 10; SBN xviii–xix; T 1.1. 4.6; SBN 12–13) Not unlike his Cartesian predecessors, he also contents himself with appealing to natural endowments and original qualities of the mind when his own principles are insufficient to explain the regularities he observes.

Passions, for Hume and his Cartesian predecessors, are a species of mental states: they are modes of "thought" or "perception" (these are synonyms in the wide Cartesian sense of the terms adopted also by Hume), which are immediately caused and accompanied by bodily motions. In Part 3 of his *Ethics*, Spinoza defines

the passions that he calls "affects" as a subclass of affections, which – as he explains in Part 2 – are changes caused in the body by impact of other bodies (1985: 448, 460–78, 493). Hume defines them as a subclass of impressions, which are lively perceptions arising, directly or indirectly, by the actions of other bodies on our body and our bodily organs. Hume's impressions, like Spinoza's affections, are sensations or feelings of how external bodies or their ideas affect our body: they are the mental counterparts of bodily affections – although Hume, as we will see, does not really care about their metaphysical status and is indifferent as to whether they should be called states of the mind or of the body.

In the Cartesian framework, impressions or sense perceptions can be more or less *clear* (in the sense of vivid, forceful, present or manifest to consciousness), but they cannot be *distinct* in Descartes's (and Spinoza's) sense of the term: they are characteristically clear but confused perceptions. A paradigm example of a perception which is very clear without being distinct is pain; an example of a perception which is both clear and distinct would be a simple self-evident mathematical proposition, such as "2 + 2 = 4." Hume does not distinguish impressions and ideas on such epistemological grounds; he divides perceptions using the purely phenomenological characteristics of vivacity and liveliness. His force and vivacity, however, meet the same criteria as Cartesian "clearness": a perception is forceful and vivid which is present and manifest to consciousness – like the sound of the motor-saw you are actually hearing, or the pain you feel in cutting yourself. It is good to keep these similarities between Hume and his Cartesian predecessors in mind when turning to the new distinctions and definitions Hume introduces in his own analysis of the passions.

Hume differs from his Cartesian predecessors in many ways, not least through his commitment to anti-metaphysical naturalism and empiricism. His naturalism is broad enough to include human morality, action, and thought among its observable phenomena (T Introduction 10; SBN xix). As shown in other chapters of this book, he introduces a radically new empiricist account of causation and belief-formation, where contingent regularities of nature and associational patterns of thinking are given a central role. The first two books of the *Treatise*, "Of the understanding," and "Of the passions," were published together, separately from the third book, "Of morals," and make up a "compleat chain of reasoning by themselves" (T Advertisment; SBN xii). The account of ideas and association in Book 1 is as important to the understanding of Hume's account of passions, as the latter –with which we are solely concerned here – is to understanding his novel views of human reason and the nature of self or personhood.

Impressions and Ideas

Hume divides thoughts or perceptions into two kinds, impressions and ideas. The former differ from the latter only through the greater "degrees of force and liveli-

ness, with which they . . . make their way into thought or consciousness" (T 1.1.1.1; SBN 1–2). The ideas are but fainter copies or replicas of the first. Sometimes Hume describes them in terms of representation, sometimes he speaks of reflection: "The one seems to be in a manner a reflection of the other; so that all the perceptions of the mind are double, and appear both as impressions and ideas." The important point is that there is an exact correspondence between them: there is nothing in the idea that was not present in the impression (T 1.1.1.3; SBN 7). Whatever content the idea has was there already in the impression it reflects or copies (T 2.1.11.7; SBN 317). Hume wants to dissociate his use of the term "idea" from that of Locke who had "perverted it" in equating idea with perception. Impressions are *not* the felt manner in which lively perceptions come to mind, but are genuine perceptions themselves (T 1.1.1.1; SBN 2 n. 2). All our ideas are preceded by impressions that they exactly correspond to and represent.[1]

Hume also distinguishes between simple and complex impressions and ideas. Simple ideas, supposedly, cannot be further analyzed, whereas complex ones can be divided into parts. The impression of red is simple, and that of a street in Paris is complex. (See also 3: IMPRESSIONS AND IDEAS.) The exact correspondence holds only between simple impressions and simple ideas, with regard to their content. Sense impressions are of external things presently acting on the senses, ideas are that by which we think of what has been thus directly felt or perceived. Actual impressions can be retained in memory and reappear in consciousness as remembered or imagined ideas. Remembered perceptions are something in between impressions and ideas: they retain some of the force and vivacity of the former while sharing some of the bleakness or faintness of the latter (T 1.1.3.1; SBN 8–9). In recognizing, for instance, the taste of a sweet lemon cookie dipped in tea, the original taste, since long forgotten, may be re-actualized, bringing to mind many other pleasant or painful impressions experienced at the same time, in their original order of appearance.

Memory is tied to the order of the original perceptions. Imagination on the contrary is free to rearrange and transpose its ideas. One can think of the taste of those cookies without thinking of the taste of the tea, or of the smell of the room, or the pleasure of the company in which it was originally savored. One can imagine all of this without experiencing the pleasure or anxiety associated in one's memory with those impressions. One cannot remember a recent accident without remembering also its disastrous consequences and the fear and horror one felt in witnessing it. However, once its original vividness fades, one may imagine the same events and is now at liberty to transpose and change the details of their course as one pleases. One may ignore, and then forget, what a coward one was when it actually happened and imagine oneself instead as having taken heroic actions to advert it or to save the injured.

Once a perception has lost all of the force and vivacity of the original impression it turns into "a perfect idea" (T 1.1.3.1; SBN 8). Perfect ideas are cool and

faint copies of the lively perceptions that were once actually experienced, and all such distinctly perceived ideas can be separated from each other and recombined by imagination with others. Unless connected to fresh or remembered impressions, such ideas, like that of the number of hairs on your head, or of "2 + 2 = 4," which of themselves are perfectly indifferent, do not affect us one way or another.

The first and most important relation that Hume notes between perceptions is the exact correspondence between impressions and ideas (T 1.1.1.6; SBN 4) – let us call this the "copy relation" – while the second is that of their order of appearance and causal relation: ideas, as it turns out, are always first caused by the impressions they exactly represent (T 1.1.1.8; SBN 4). It is sometimes assumed in the Hume literature that the "copy relation" is the only kind of non-mechanical, intentional relation Hume accepts, but this is an oversimplification. Ideas are associated and combined in accordance with the three different laws of associations Hume introduces, not only by spatiotemporal contiguity and causality, but also by resemblance, which hardly counts as a mechanical principle. As will be seen, Hume grants in addition other natural connections and transitions between perceptions which cannot be explained by the laws of association solely, but are accounted for by unexplained "original qualities" and "organs" of the mind.

Passions as Reflective Impressions

Passions are a subset of impressions, differing from sense impressions in arising not directly from stimulation of sense organs by external objects or by the disposition of the body or from the movements of animal spirits in the neural system, but from other impressions or ideas. Hume calls them secondary impressions, or impressions of reflection, by contrast to the first, which are said to arise in the soul "originally, from unknown causes" (T 2.1.1.1; SBN 275). The order in which the so-called secondary impressions, which depend on our ideas, are produced is as follows:

> 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. . . . so that the impressions of reflexion are not only antecedent to their correspondent ideas; but posterior to those of sensation, and deriv'd from them. (T 1.1.2.1; SBN 7–8)

True to his empiricism and methodological dualism, Hume does not dwell on the physical causes of the impressions. The examination of sense impressions themselves belongs to "anatomists and natural philosophers." Their nature and true origin, Hume thinks, are ultimately inexplicable (T 1.3.5.2; SBN 84).

Bodily pains and pleasures are the source of many passions, both when felt and consider'd by the mind; but arise *originally in the soul, or in the body, whichever you please to call it, without any preceding thought or perception*. A fit of gout produces a long train of passions, as grief, hope, fear; but is not deriv'd immediately from any affection or idea. (T 2.1.1.2; SBN 275–6; emphasis mine)

Desire and aversion, hope and fear, arise immediately from ideas of pleasant or painful impressions, themselves inexplicable in terms of prior perceptions. That they make irruption in and affect our minds – our train of thought – is a matter of experience, which is the object of Hume's "anatomy" of the mind. By the pleasure or pain they cause, the impressions, or their idea-copies, affect our desire: we naturally seek pleasant impressions and avoid the painful ones.

Contrary to the order one might expect, given his observations about the causal priority of impressions, Hume starts by examining not the impressions but the ideas and their associations (T 1.1.2; SBN 8). Because passions are caused by and depend on ideas, the regularities that count in the explanation of the passions are not mechanical laws governing the body, but psychological laws of association between ideas (cf. Hume 1992: 133) which Hume discusses with detail and examples in the first two books of the *Treatise*, and of which the principle of causality is the most important.

That the passions are causally related to their antecedent ideas as well as to those that unfailingly succeed them, and that Hume's concept of causality is ultimately contingent, tend to obscure the fact that there are systematic relationships between the ideas that are constitutive of the passions but that cannot be explained through mere accidental brute associations. Some of the things Hume himself says about the passions when contrasting them with reason contribute to obscuring this fact. Consider the following often quoted passage:

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. (T 2.3.3.5; SBN 415)

This and other pronouncements are evidence for those who see Hume's view as paradigmatically emotivist. Passions, on this reading, are purely affective physiological states, comparable to shudderings, shiverings, and heat-flushes, without any cognitive or intentional content. But as many Hume scholars also note, even a brief glance at Book 2 of the *Treatise* shows that such a reading and the above claim itself, taken literally, are "wildly implausible" (Penelhum 1993: 128; see also Baier 1991: 163).

The very fact that the passions are caused by antecedent perceptions and not directly by physiological stimulation puts this reading into question. A passion, as a secondary impression, differs from a primary impression like that of being burnt,

or feeling hungry or sexually aroused, in that it is not produced by physiological changes or sensory stimulation directly, but requires specific interposed perceptions of its object and its qualities. Thus prior impressions of the beauty and riches, or of the kindness and trustworthiness of a person may cause one to love her, to wish her well and to care for her good, whereas repeated painful experiences of, say, her scorn, betrayal, and constant scheming make one hate her. Passions are natural affective reactions to prior perceptions, which can be actually experienced live impressions or recently experienced fresh impressions – for instance, ideas brought up by memory, or clusters of lively ideas that Hume calls beliefs. Beliefs, in Hume's account, are nothing but ideas enlivened by association to current, more lively impressions (T 1.1.1.1; SBN 12). Indeed forceful and lively impressions are judgments of existence. (See T 1.3.5.7; SBN 86; and Owen forthcoming.) Passions too are complex, intentionally structured states: they are always of, or directed to, some object or event, attuned to the contents of the perceptions causing them and pregnant with those which they in their turn unfailingly bring to mind. Love, for instance, always makes one think well of the person loved and comes with a desire to benefit her; hate, on the contrary, comes with a desire to hurt or avoid the person whom one hates.

Direct and Indirect Passions

Passions are lively reactive perceptions based on impressions of pain and pleasure. Hume divides them into direct and indirect passions (T 2.3.9.1; SBN 438). Direct passions (e.g. desire, aversion, grief, joy, hope, fear, despair, security, and volition) arise directly out of impressions of sensation (original impressions), which are caused by stimulation of bodily organs and nerves. More specifically, they arise directly from sensory impressions that are experienced as pleasant or painful, and alert us of some good or evil (T 2.1.4.1; SNB 276). This holds for indirect passions too, whose formation requires, in addition, a set of associative principles and ideas related according to these.

The latter, by which Hume starts his examination, are "pride, humility, ambition, vanity, love, hatred, envy, pity, malice, generosity, with their dependents." The first on his list are pride and humility, which play a central role in the forming of one's self-image and personal identity, and are crucial for morality too.

In explaining pride and humility Hume notes two general features in the "passions themselves" which are of equal importance in all indirect passions. The first is related to their causes and objects, which are determined by nature. Pride and humility, "by an original and natural instinct" always have the self as their "peculiar" object. Hume finds

> 'tis absolutely impossible, from the primary constitution of the mind, that these passions shou'd ever look beyond self, or that individual person, of whose actions and

sentiments each of us is intimately conscious. Here at last the view always rests, when we are actuated by either of these passions; *nor can we, in that situation of mind, ever lose sight of this object*. For this I pretend not to give any reason; but consider such a peculiar direction of the thought as an original quality. (T 2.1.5.3; SBN 286; my emphasis)

Not only are thoughts (in Hume's terms, ideas, images, beliefs) about their objects part of our indirect passions, but their objects always are some sentient, thinking thing. Pride and humility are always directed at this self that is at all times intimately present in one's thoughts; as for love and hatred, they are "always directed to some sensible being external to us" (T 2.2.1.2; SNB 329). Hume goes on to explain that this holds not only for their *object* but also for their *causes*, which are always related to some "thinking being" (T 2.2.1.6; SBN 331).

The second feature to which he draws attention is the sensation or impression of the passion itself, also an "original" (read: unexplained) quality:

The second quality, which I discover in these passions, and which I likewise consider an original quality, is their sensations, or the peculiar emotions they excite in the soul, and which constitute their very being and essence. Thus pride is a pleasant sensation, and humility a painful; and upon the removal of the pleasure and pain, there is in reality no pride nor humility. Of this our very feeling convinces us; and beyond our feeling, 'tis here in vain to reason or dispute. (T 2.1.5.4; SBN 286)

This second element, which Hume calls the "very being or essence" of a passion is the one that those who read Hume as an "emotivist," taking him to reduce emotions to mere sensations or feelings, have retained and focused exclusively on (e.g., Kenny 1963; Alston 1968; Deigh 1994). Hume of course helps them in emphasizing that these peculiar emotions constitute the very essence of passions. But it is rather peculiar that he should talk about essences or "the very being" of anything at all, given his general commitment not to pronounce himself on the inner natures or essences of things. Now if we take Hume seriously in this anti-metaphysical commitment, we should not give too much weight to his use of this term. "Essence" here certainly cannot mean necessary *and* sufficient conditions of the passion. All Hume needs and, as I understand it, does, is to emphasize that the affective impression or feeling is a necessary component in the complex of thoughts or ideas constituting a passion, its other necessary constituents being those causing it and those it brings to mind. The question now is why is this element – the affective impression – singled out as essential here? To understand this we have to consider how the indirect passions, which as shown by McIntyre (2000) are also importantly at work in the genesis of direct passions, are formed.

Pride and humility differ from other indirect passions in having always the same object, namely "the self, or that succession of related ideas and impressions, of which we have an intimate memory and consciousness." What distinguishes them

are the opposite ways in which they make one think of oneself, "as elated with pride," or "dejected with humility," and this again depends on their causes, which cannot be the same as their object, since they are contrary, and a man cannot at the same be proud and humble for the same reason. The *cause* is the idea which excites the passion, the *object* is that to which, once excited, "they direct their view"; and the secondary impression, in which the passion strictly speaking is said to consist, is the transitional element between the two (T 2.1.2.1–4; SBN 277–8).

The causes of pride, for instance, can be any valuable quality of the mind (e.g., wit, good sense, integrity), or of the body (strength, beauty, address in riding or dancing, skill in some particular technique or handicraft), or of things somehow related to us, like our country, children, relations, pets, dresses, and so on, any of which may cause either pride or humility (T 2.1.2.4–5; SNB 279). How these various things related to oneself happen to affect one at various times depends on their qualities and how they are perceived. For Hume makes a further distinction with respect to the cause, namely between "that *quality*, which operates, and the *subject*, *on which it is plac'd*" (T 2.1.2.6; SBN 279). I would not be proud of my house were it not for some of its properties – say its neat location, its tidyness, or pleasant façade, any of which I may find independently praiseworthy or admirable, and the contemplation of which actually affects me with a pleasant impression. On the other hand, those properties by themselves, no matter how great a pleasure they may give me, would not also make me proud were they not properties of something I could relate to myself, to a house I myself happen to inhabit, to possess, or to have fixed up. The shiny fur of the dog may please you very much but it would not make you feel proud unless you were its caretaker. Nor would you feel humiliated by the dog's misbehavior if you were not responsible for its proper training (T 2.1.2.5; SBN 279 and T 2.1.5.3; SBN 285).

Association and the Individuation of Passions

In themselves the passions of pride or humility are "simple and uniform impressions," and so cannot be defined, but since "they are the most common, of any, every one, of himself, will be able to form a just idea of them." They can be described (only) by the enumeration of the specific circumstances in which they are embedded (T 2.1.2.1; SBN 277). Among these are three general "properties of human nature, which tho' they have a mighty influence on every operation both of understanding and passions, are not commonly much insisted on by philosophers" (T 2.1.4. 2; SBN 283). Here is where Hume brings in his novel system of associations.

The first is the associations of ideas explained in Book 1. Our span of attention is very short and our thoughts are therefore in constant flux, though "however changeable they may be," their changes are not "without rule and method." The

rules governing them are the three laws of association. Thus they pass "from one object to what is resembling, contiguous to, or produc'd by it" (T 2.1.4.2; SBN 283). So, for instance, the idea of terrorism may bring on that of the US war on Iraq, followed by those of torture, humiliation, and disgrace.

The second property is a corresponding association of impressions, but the three principles are here reduced to one, namely association by resemblance:

> Grief and disappointment give rise to anger, anger to envy, envy to malice, and malice to grief again, till the whole circle be compleated. In like manner our temper, when elevated with joy, naturally throws itself into love, generosity, pity, courage, pride, and the other resembling affections. (T 2.1.4.3; SBN 283)

As soon as one passion is "actuated," others follow in train, according to the temper and the then prevailing prior affections, which set the tone of the variations to follow. Painful impressions attract, by resemblance, more of the same, and so do pleasant and joyful ones, as will be seen shortly more in detail.

The third property noted by Hume is that these two kinds of association, between ideas on the one hand, and impressions on the other, "very much assist and forward each other, and that the transition is more easily made when they both concur in the same object." Suppose you are in low spirits, and someone irritates or insults you. You will then "be apt to find a hundred subjects of discontent, impatience, fear and other uneasy passions"; and the more so, if you can ascribe them to the person you consider the original cause of your sad state.

> Those principles, which forward the transition of ideas, here concur with those, which operate on the passions; and both uniting in one action, bestow on the mind a double impulse. The new passion, therefore, must arise with so much greater violence, and the transition to it must be render'd so much more easy and natural. (T 2.1.4.4; SBN 284)

Carried away by enthusiasm Hume goes on to cite, from "an elegant author," this additional evidence about how the different senses assist each other by mutually strengthening their respective perceptions:

> any continued sound, as the musick of birds, or a fall of waters, awakens every moment the mind of the beholder, and makes him more attentive to the several beauties of the place, that lye before him. Thus if there arises a fragrancy of smells or perfumes, they heighten the pleasure of the imagination, and make even the colours and verdure of the landskip appear more agreeable; for the ideas of both senses recommend each other, and are pleasanter together than when they enter the mind separately . . . (Addison quoted by Hume, T 2.1.4.6; SBN 368)

Applying these general observations in his analysis of particular passions like pride and humility, Hume finds that their formation depends on two interrelated

sets of associated perceptions operating together: the so-called double relation principle. Pride is produced when the idea of its subject, say the house, is associated with the idea of its object, namely myself as its owner, but only if at the same time some idea of the qualities of the house gives me an additional pleasant impression, which in its turn by association of resemblance calls forth another pleasant impression related to the idea of myself as its owner. The latter is the aforementioned "simple and uniform impression of pride" (cf. T 2.1.2.1; SNB 277). But as should now be obvious, its identification and individuation as the passion of pride requires a wider context with a double set of interrelated perceptions: one pair of associated ideas – that of the subject causing it, and that of the object to which it directs the view, working in concert with one pair of resembling impressions causally related to the ideas of each – the pleasure that the sight of the color the house gives me, strengthening and forcing a resembling impression of pleasure caused by the thought of myself as having painted it.

This yields an answer to the question raised in the previous section about the importance of the affective element. The secondary impression, pride, aroused and strengthened by the original impression of its subject, is crucial in fixing the attention so pleasantly on the idea of its object, namely, the self. Without this affective enforcement, the same qualities of the subject of my pride, the house, would leave me entirely indifferent and contribute nothing to my self-esteem. I could still register its beauty and entertain a "perfect idea" of it without being moved by it at all, that is, without experiencing any pleasant impression in its train. Or it may come connected with a painful memory of some past traumatic event, so that as soon as the house or its bright color comes to my mind I feel sorrow and humiliation instead of being elated with pride.

The secondary impressions should not be thought of as discrete atoms of experience coming with their own "feels" or "qualia" that could be recognized apart from other perceptions accompanying them. Their character and effects – pleasant or painful – depend wholly on the network of ideas (beliefs) and impressions within which they are perceived. Our ideas and impressions, including that of our self (that bundle of current and past impressions and ideas), are in constant flux and interaction, wheeling us about in accordance with the force and vivacity with which they strike us, and the strength of the associations into which they merge. Any fresh impression of pain or pleasure is immediately *transfused* into some corresponding direct or indirect passion containing similar impressions. Our current train of ideas and their associations will work on us in accordance with the degree of liveliness or force of the passions involved. The stronger my already existing indirect passion of love for a person, the stronger is my direct passion (the desire to benefit or please her or possess her) aroused by a fresh sensation of her kindness or beauty. Likewise, the more vivacious the original impression of pain or pleasure, the more violent any other reactive passions thus triggered, as Hume's somewhat prosaic example illustrates:

> 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. Again, when these cloaths are consider'd as belonging to ourself, the double relation conveys to us the sentiment of pride, which is an indirect passion; and the pleasure, which attends that passion, returns back to the direct affections, and gives new force to our desire or volition, joy or hope. (T 2.3.9.4; SBN 439)

Similarly, I presume, if the suit belongs to my friend, it will renew my love and admiration for him or her.

Like Spinoza, Hume stresses the central role of the more or less mechanically governed imagination in causing and sustaining the passions: the two are so closely united that nothing can affect the imagination while remaining "entirely indifferent" to the affections (T, 2.3.6.1; SBN 422). Again these transitions of perceptions are explained only by invoking natural propensities of the mind and of the passions themselves.

> those two faculties of the mind, the imagination and passions, assist each other in their operations when their propensities are similar, and when they act upon the same object. The mind has always a propensity to pass from a passion to any other related to it; and this propensity is forwarded when the object of the one passion is related to that of the other. The two impulses concur with each other, and render the whole transition more smooth and easy. (T 2.2.2.16; SBN 339)

Hume provides a wealth of thought-experiments to illustrate how his double relation principle works, now in transfusing love and hatred of another into pride and humility for oneself, now in strengthening and enforcing existing passions, now in transferring them between objects which are already somehow related in our thought. Thoughts pass from the idea of one object to another according to their laws of association, determined now by contiguity, now by resemblance, now by causality, or, when none of these suffice, by original propensities. The secondary impressions, reflecting our affective reactions to the first idea, follow along, but, as we saw, they associate themselves only according to one of these principles, namely resemblance. Affective impressions that are not supported by concurrent associations of ideas will not develop into passions proper, nor are the ideas alone sufficient, for where our ideas of the cause of the affection fail to connect in a sustained way with an idea of its object, no passion properly occurs. The double relation of ideas and impressions enforcing each other is essential in the formation of the passions (T 2.2.2.4; SBN 333 and T 2.2.2.28; SBN 346). It is essential as well in the transitions of passions, so that "'tis not the present sensation alone or momentary pain or pleasure, which determines the character of any passion, but the whole bent or tendency of it from the beginning to the end" (T 2.2.9.2; SBN 381).

Some differences in how these principles work in different passions can be noted. Because "we are at all times intimately conscious of ourselves, our sentiments and passions," any new impressions or ideas of ourselves and our passions

"strike upon us with greater vivacity than the ideas of the sentiments and passions of any other person." And since anything that strikes us with great vivacity will catch our attention, it will also act on fixing it with the same force, keeping it "from wandering to other objects, however strong may be their relation to our first object" (T 2.2.2.13; SBN 339). This explains why, once our thought is set on ourselves, it does not pass easily on to others, no matter how closely they are related to us. So while love and hatred for relatives or other persons closely connected to us easily get transfused into pride and humility, this does not normally work the other way around, with this one exception: pride transfuses into love only when the loved one happens to be its cause by approving of us, of our action or belongings. For nothing "more readily produces kindness and affection to any person, than his approbation of our conduct and character: As on the other hand, nothing inspires us with a stronger hatred, than his blame or contempt" (T 2.2.2.27; SBN 346).

Normally, unlike other indirect emotions, pride and humility are unaccompanied by any desire or tendency to action. Hume therefore calls them "pure sensations." Pride is a pure sensation of pleasure because it remains fixed on oneself as long as it lasts, and does not prompt one to any action – the contemplation of self is its own fulfillment. Love and hatred, to the contrary, come with their proper desire to benefit the person loved, or to harm whom one hates, and this desire unfailingly feeds new emotions in its turn. Thus the analysis of the passion of love reveals that it can consist of three interrelated "impressions": (1) pleasure from beauty; (2) a bodily appetite for generation; and (3) a generous kindness or good will. Each of these impressions has a distinct object, and they produce each other through a set of three interrelated ideas. In all the passions, the sensation or felt emotion and the idea of its object are, as it were, preordained to occur together.

> Here then is the situation of the mind, as I have already describ'd it. It has certain organs naturally fitted to produce a passion; that passion, when produc'd, naturally turns the view to a certain object. But this not being sufficient to produce the passion, there is requir'd some other emotion, which by a double relation of impressions and ideas may set these principles in action, and bestow on them their first impulse. (T 2.2.11.6; SBN 396)

It is as if we had special affective organs attuned to certain objects, which work only when excited with sensations of pleasure and pain (T 2.1.5.7; SBN 287). This brings us to yet another distinction which may help shed some light on Hume's puzzling remark about passions not having "any representational qualities."

Perception and Perceiving

The metaphor of organs of the mind suggests a distinction between the very act or operating of the mind (or its mental "organ") and the object perceived or

sensed, which, although these never occur (are experienced) separately, nevertheless can be thought of apart, as two different aspects of the same occurrence or phenomenon.

Here again Hume may be seen as following his Cartesian predecessors in using the term "perception" broadly enough to cover both the act and the content of a perception. He notes that "all the acts of seeing, hearing, judging, loving, hating and thinking, fall under the general denomination of perception." He also writes:

> The mind can never exert itself in any action, which we may not comprehend under the term of perception; and consequently that term is no less applicable to those judgments, by which we distinguish moral good and evil, than to every other operation of the mind. To approve of one character, to condemn another, are only so many different perceptions. (T 3.1.1.2; SBN 456)

These actions too come divided into original or primary, and reflective or secondary. Judging (about moral good and evil), like loving, hating, sympathizing, being inflated with pride, or torn by fear or jealousy, are secondary or reflective perceptions, caused by prior acts of perceiving depending on the actual stimulation of the external or internal organs.

Applying the traditional framework of kinds of distinction to the complex phenomenon of perception, we can now separate, by thought or rational distinction, these two aspects, which are inseparable *in re*, that is, are always experienced together: (1) the mental act or operation and (2) its object or content. In other words, we can now distinguish the psychological fact of being actuated or having a perception on the one hand, and the object or content of the particular perception so acting on us on the other. In the same way as there must be some object seen or heard and some stimulation of the sensory organ, for an act of sense perception to occur, so there must be a prior perception of an object thus-and-so-qualified for a particular passion to be actuated. Hume, of course, runs over such distinctions with his rough system of impressions, ideas, and associations, but then he has to import specific "original qualities" or "organs of the mind" to account for why a certain kind of object unfailingly gets associated with a certain kind of passion (e.g., T 2.1.5.6; SBN 260, 287–8). When calling the passions original existences, Hume's focus, I surmise, is on the act-aspect of the secondary impression, and the manner of liveliness in which it is had. This act-aspect, considered separately, does not, qua psycho-physical fact or mode, contain any representative quality, yet it may be described as a bodily (physiological-cum-behavioral) state, abstracting from its other features (causes and objects), just as one can give a description of thirst in terms of a dryness in the throat and a lack of liquid in the body. But such a description, to pick out the state of the body so described as one of thirstiness, presupposes that the unpleasant sensation of dryness in the throat was associated with some idea of its quenching when first perceived, and thus that it has already been identified as such. So to the provocative statement, where Hume writes that being angry is being "possest with the passion, and having in

that emotion no more a reference to any other object," than when being thirsty, or sick, or of a certain tallness, one could agree: no more, perhaps, but also no less. Certainly there is no copy relation involved, so if representative relation and reference means just copy relation there is no problem. Being sick and five foot high presuppose standards of health and measurement and other relations that are not merely causal. If thirst comes with the idea of an object quenching it, anger likewise comes with a perception of some wrong being done to one and some idea of retaliation. Being an impression, anger is a lively perception, and as such it is something more than merely a felt manner of perceiving – it must also be of something or other. Because of its in-between position it must be "of" other ideas or impressions, though not in the sense of copying them, but in that of affecting the perception of whatever object they are reactions to. For something to be a perception it has to "strike upon the mind," and enter "into thought or consciousness" (T 1.1.1.1; SBN 1), that is, it has to merge into the set of ideas and patterns of associations currently dominating the mind, from which it gets its impulse or affective force and which it in its turn influences.

Passions and Moral Sentiments

The distinction between passions and moral judgements in Hume is very thin and depends, basically, on their objects and the point of view from which they are considered. Just as the thought of the sorry state of my house, my faults of character, or of your disapproval of what I thought of as my successes, causes me pain and humility, sustaining my painful awareness of myself and my shortcomings, so the contemplation of the characters, actions, and passions of other people, when "considered in general, without reference to our particular interest" causes us pain and pleasure, which makes us call them good or evil (T 3.1.2.4; SBN 472). All of this, again, comes about by "a primary constitution of nature."

> To approve of a character is to feel an original delight upon its appearance. To disapprove of it is to be sensible of an uneasiness. The pain and pleasure, therefore, being the primary causes of vice and virtue, must also be the causes of all their effects, and consequently of pride and humility, which are the unavoidable attendants of that distinction. (T 2.1.7.5; SBN 269)

Hume again rests his case on experience, which shows that noble and generous characters never fail to give us delight, whereas cruelty and treachery always displease us "by their very nature." It is not possible, Hume argues, that we should ever reconcile ourselves to such qualities either in our selves or others (T 2.1.7.6; SBN 269). Hume does not deny the importance of self-interest, but his naturalist phenomenology recognizes other sources of moral motivation as well. He believes that human beings are possessed with a number of natural endowments that work together with self-love but which cannot be reduced to or derived from it.

Sympathy, that natural propensity of ours "to receive by communication" the inclinations and sentiments of others (T 2.1.11.2; SBN 316), is one of them. It is explained through the liveliness of the ever-present idea or impression of ourselves, working in concert with the general resemblance between human beings to make us "enter into" and "embrace," as if they were our own, the sentiments of those we recognize as like ourselves (T 2.1.11.5; SBN 318). This, remarkably, happens by the sheer mechanism of passions: my ideas of your sorrow, or of the pains suffered by distant victims of war and terror, when vivid enough, are simply "converted into the very impressions they represent" (T 2.1.11.8; SBN 319). Among our natural endowments is benevolence, which Hume sometimes describes as fellow-feeling or humanity. That we can be motivated by altruism and benevolence is also a fact of experience, something we can observe, although, again, we cannot explain it. "It is needless to push our enquiries so far as to ask why we have humanity or fellow-feeling with each others. It is sufficient that this is experienced to be a principle in human nature." Far from deriving moral sentiments from self-interest, Hume explicitly excludes self-love and passions composed of it from his theory of the origins of morals, not because they are too weak, but simply because they do not have the right direction (EPM 9.1.5. Cf. T 3.1.2.4; SBN 472).

Humans are both social and moral animals. It is a natural fact that we take interest in the well-being of other beings resembling us, and react to wrongs done to them as well as to ourselves. Our affective reactions reveal what we care about, and among the things we care about are moral vice and virtue. We not only care about how we are viewed by other people, and we read their attitudes directly off the bodily and behavioral expressions of their affective reactions to what we say and do and feel. Passions and moral sentiments are propagated by direct imitation, by the impressions their expressions make on us. Human minds "are mirrors to one another" and "reflect each others' emotions." Not only passions and sentiments are reflected and often "reverberated," but so are our opinions. Thus the pleasure that the owner of the estate receives from his possession may cause pleasure and esteem in his admirers or dependants, "which sentiments again, being perceiv'd a'd sympathiz'd with, encrease the pleasure of the possessor," and so, by sympathy, our passions and opinions mutually reinforce and support each other (T 2.2.5.21; SBN 365). We "fancy ourselves more happy, as well as more virtuous and beautiful, when we appear so to others" (T 2.1.6.6; SBN 292). Other causes of pride, no matter how great, have little influence "when not seconded by the opinions and sentiments of others" (T 2.1.11.1; SBN 316–17).

The examination of passions is the starting point of Hume's moral philosophy and the ground from which he opposes rationalists seeking eternal moral values and laws independently of human interests and natural reactions in some objectively given intelligible order of things that only reason and rational intuition could detect. He starts from the observation of facts, of how, for instance, we feel bad about breaking promises or betraying friends, or in being blamed and shunned,

whereas the slightest mark of approval of our deeds or character never fails to delight us. He also observes that we all grant that morality influences our actions: we care about duties and rights and take offense when they are harmed. If we did not have these natural moral reactions we could not begin to make moral judgments, and the moral judgments we make are expressed in our moral sentiments and feelings. The vivacity and great influence of moral judgments on our conduct tell us that they are not, essentially, made up of ideas, but of (secondary) impressions. That they are lively, felt impressions, does not mean, as should now be clear, that they are mere qualitative raw feels having no cognitive content or import. It means only that whatever their content or object is, it is perceived in a lively and vivacious manner, which reflects how those objects matter to us. Their vivacity marks them out from ideas and from rational judgment about relations of ideas, even when they can be so gentle that we mistake them for ideas:

> Our decisions concerning moral rectitude and depravity are evidently perceptions; and as all perceptions are either impressions or ideas, the exclusion of the one is a convincing argument for the other. Morality, therefore, is more properly felt than jud'g of; tho' this feeling or sentiment is commonly so soft and gentle, that we are apt to confound it with an idea . . . (T 3.1.2.1; SBN 470)

To have moral sense, that is, a sense of virtue, is having a propensity to feel a particular kind of satisfaction from the contemplation of a character, and it is this feeling which *constitutes* the moral judgment of praise or blame. It is important, again, to stress that the judgment with its content is included in the feeling itself, not deduced from it or formed on its basis:

> We do not infer a character to be virtuous, because it pleases: But in feeling that it pleases after such a particular manner, we in effect feel that it is virtuous. The case is the same as in our judgments concerning all kinds of beauty, and tastes, and sensations. Our approbation is imply'd in the immediate pleasure they convey us. (T 3.1.2.3; SBN 471)

What makes an affective reaction into a moral reaction is not its "feel" by itself, but its content, that it is a vivid perception *of* a particular kind of object or action – willful murder for instance, child abuse, torture, manipulation, or seducing your neighbor's spouse. For the sentiments thus caused to count as properly moral sentiments, the particular object causing them must be considered in their general aspect – not just as related to our self-interest. Because morals excite our passions and influence our actions, they cannot be derived from reason, "which is utterly impotent in this particular" (T 3.1.1.6; SBN 475). Reason, moreover, has no influence on the passions, "but is and ought to be their slave" (T 2.3.3.4; SBN 414–15). The passions that reason ought to be governed by are the moral senti-

ments or feelings, and the indirect passions that they cause. Given the wide range of cognitive attitudes and postures included under Hume's heading of "passions," and the very narrow scope given to reason as a truth-detector and inference-maker, this should not give one pause. Hume's positive views about human nature and its original endowments, among which are our great sensitivity to the opinions of others, make him confident that the means of mastering and correcting the passions are found in that nature itself, provided it is properly assisted and checked by others of our kind. We may be wheeled around by the mechanisms of our passions and imagination, but we are all in the same boat, at the mercy both of our circumstances, and of how we fare under the reflective assessment of those of our fellow human beings, whose approval we cannot help caring about and whose interaction and support we need to stay afloat a little longer. When excessive, disorderly and contradictory emotions threaten to submerge us, their disapproval will help bring consistency into our emotional life and their approval sustain our moral, governing sentiments.

Notes

1   I am grateful to David Owen for kindly commenting on a draft of this paper and for emphasizing that what ideas represent *are* impressions. See the discussion in Owen forthcoming.

I wish to thank David Owen, Lisa Shapiro, Fred Stoutland, and the editor of this volume for helpful comments on earlier versions of this paper.

References

Alston, W. P. (1967) "Emotion and feeling," in *The Encyclopedia of Philosophy*, vol. 2, ed. P. Edwards. New York: Macmillan.

——(1968) "Moral attitudes and moral judgements," *Noûs* 2, pp. 1–23.

Deigh, John (1994) "Cognitivism in the theory of emotions," *Ethics* 104, pp. 822–54.

Descartes, René (1964–7) [1649] *Les Passions de l'âme*, vol. XI in *Oeuvres de Descartes*, 12 vols., ed. Charles Adam and Paul Tannery, rev. edn. Paris: Vrin/CRNS.

Hume, David (1992) [1757] "Dissertation of the passions," in *Four Dissertations*. South Bend, IN: Saint Augustine's Press.

Kenny, Anthony (1963) *Action, Emotion and Will*. London: Routledge & Kegan Paul.

Malebranche, Nicolas (1946) [1674–8] *De la recherche de la vérité*, 3 vols., ed. G. Lewis. Paris: Librairie Philosophique J. Vrin.

McIntyre, Jane (2000) "Hume's passions: direct and indirect," *Hume Studies* 26, pp. 77–86.

Owen, David (forthcoming) "Mechanics of mind: impressions and ideas," in David Fate Norton (ed.) *The Cambridge Companion to Hume*, 2nd edn. Cambridge: Cambridge University Press.

Spinoza, Baruch (1985) [1677] *Ethics*, in *The Collected Works of Spinoza*, ed. and trans. Edwin Curley. Princeton, NJ: Princeton University Press.

Further reading

Alanen, Lilli (2004) "Reflection and ideas in Hume's account of the passions," in J. Jenkins, J. Whiting, and C. Williams (eds.) *Persons and Passions*, Notre Dame, IN: University of Notre Dame Press.

Baier, Annette (1978) "Hume's analysis of pride," *Journal of Philosophy* 75, pp. 27–40.

——(1980) "Master passions," in A. Oksenburg Rorty (ed.) *Explaining Emotions* (pp. 403–23). Berkeley: University of California Press.

——(1991) *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge, MA: Harvard University Press.

Davidson, Donald (1980) "Hume's cognitive theory of pride," in *Essays on Actions and Events* (pp. 277–90). Oxford: Clarendon Press.

Owen, David (1999) *Hume's Reason*. Oxford and New York: Oxford University Press.

Penelhum, Terence (1993) "Hume's moral psychology," in David F. Norton (ed.) *The Cambridge Companion to Hume* (pp. 117–47). Cambridge: Cambridge University Press.

Wiggins, David (1998) "Natural and artificial virtues. a vindication of Hume's scheme," in Roger Crisp (ed.) *How Should One Live? Essays on the Virtues*. Oxford: Oxford University Press.

11

Hume's "New And Extraordinary" Account of the Passions

Jane L. McIntyre

Introduction

In the *Abstract* of *A Treatise of Human Nature* Hume introduced his discussion of Book 2 by saying, "'Tis of more easy comprehension than the first; but contains opinions, that are altogether as new and extraordinary" (A 30; SBN 659). But what, if anything, is "new and extraordinary" in Hume's treatment of the passions? The relationship of Hume's epistemological and ethical theories to those of his predecessors and contemporaries is acknowledged to be vital to understanding the issues he addressed, as well as his analyses of them. This would appear to be a sufficient reason to investigate the relationship of Book 2 of the *Treatise* to other works on the passions. A further reason is provided, however, by the elusive quality of Book 2 itself. Book 2 is filled with detailed descriptions of the causes of various passions, but reading them is like playing "Jeopardy": we know Hume's answers, but what were the questions? What did theories of the passions attempt to explain? What were the philosophical issues? Even when we can give a partial answer to these questions – for example, that Book 2 is concerned with the relationship of reason and the passions – our understanding of Hume's position may be hampered by a lack of knowledge of other commonly discussed positions and arguments.

In this chapter, I look to the literature of works on the passions in the seventeenth and early eighteenth centuries to shed some light on what is new and extraordinary in Book 2 of the *Treatise*. I begin by identifying the problems most often addressed in these works, and describing the solutions they proposed. In this context, I offer an analysis of the nature of Hume's disagreement with

traditional accounts of the passions. I argue also that Hume recognized an emerging incoherence within the commonly accepted account of the relation of reason to the passions, and attacked it. Finally, I argue that the *Treatise* does provide a genuine (and radical) set of solutions to the philosophical problems addressed in more traditional works, particularly through his novel introduction of the category of the indirect passions.

Background

First, what are passions? Although there is no one definition universally adhered to, many theorists in the seventeenth and eighteenth centuries agreed on certain central features. Passions are modifications of the mind (or soul) that occur on the apprehension of something good, evil, or novel. Passions often include tendencies to action, but it was recognized that some passions, most notably wonder, do not. Although it was not uncommon to argue that there could be desires and aversions in disembodied minds,[1] and that human passions could also be calm, typically passions involve bodily feelings. Nearly all writers treat human passions as a concomitant of the embodiment of the soul, and most of the philosophical problems addressed in works on the passions are rooted in the question of the nature of the human mind and its relationship to body.

In the seventeenth century it was common to identify three perspectives on the study of the passions. As a part of natural philosophy, the causes and effects of the passions were accounted for, often in physiological terms. As a part of moral philosophy, treatises on the passions discussed how the passions become good or evil. Theories of the passions were also incorporated into accounts of rhetoric. Both Descartes and his English contemporary Edward Reynolds explicitly invoke this threefold view of the passions, and the titles and prefaces of a number of other works also reflect it.[2] Authors rarely focused on the passions from only one of these points of view, however. Descartes, for example, stated that he would treat the passions as a natural philosopher, but he also discussed the control of the passions in relation to the knowledge of good and evil (Descartes 1985: 347–8). Works dealing with the morality of the passions often included, if only incidentally, assertions about the physiology of the passions (Watts 1729: 7; Grove 1749: 277). Malebranche was as concerned with the epistemological problem of the passions as a source of error, as he was with metaphysical issues. Nevertheless, this classification reflects the dominant areas of philosophical interest in the passions.

Philosophical treatises on the passions of all three types generally identified some primary or primitive passions and discussed the origin of specific passions as combinations of the more primitive ones. While there was no unanimity about which passions were primary, many works identified love as *the* primary passion.[3] Longer lists usually included love and hatred, joy and sorrow, desire and aversion, and wonder. Descartes and others explicitly rejected the scholastic division of the

passions into the concupiscible and the irascible, that is, those based on desire and those in which the passions moved in opposition to something, as they do in anger. The lists of the derivative passions appear to have been limited only by an author's love of detail: Watts, for example, includes accounts of passions such as "vexation."[4] Formulas for a number of compound passions, for example that envy is joy or sorrow mixed with malice, tend to be repeated by many authors (e.g. Watts 1729: 38; Grove 1749: 398).

Central Philosophical Issues in Works on the Passions

Mind and body

Because of their apparently compound nature, the passions exemplify the problem of the relationship of mind and body inherent in dualism. This connection is quite evident in Descartes. *The Passions of the Soul* (1649), composed in response to questions from Princess Elizabeth of Bohemia concerning how the soul and the body act on each other, explains the occurrence of the passions by appealing to the motions of the animal spirits in the brain. (Interestingly, the publication of another work, Reynolds's *A Treatise of the Passions* (1640) was also prompted by correspondence with Princess Elizabeth, as the dedication and the introduction to that work indicate.) In *The Search After Truth* (1674) Malebranche held a view superficially similar to Descartes's, but Malebranche's analysis of causality entailed that the animal spirits were only the natural (or occasional) but not the real causes of the passions. Malebranche's reasons here are interesting. He argued,

> I cannot understand how certain people imagine that there is an absolutely necessary relation between the movements of the spirits and blood and the emotions of the soul. A few tiny particles of bile are rather violently stirred up in the brain – therefore the soul must be excited by some passion, and the passion must be anger rather than love. What relation can be conceived between the idea of an enemy's faults, or a passion of contempt or hatred, on the one hand, and the corporeal movement of the blood's parts striking against certain parts of the brain on the other? How can they convince themselves that the one depends on the other, and that the union or connection of two things so remote and incompatible as mind and matter could be caused and maintained in any way other than by the continuous and all powerful will of the Author of nature? (Malebranche 1980: 338–9)

According to Malebranche, physiological accounts of the passions fail to capture necessary connections between the body and the mind; they fail to make the relation between the passions and their corporeal causes intelligible. Malebranche supplied the lacking connection through the will of God (1980: 448). Other philosophers developed accounts of the passions that retained causal explanations, but

within categories revealed through introspection. In a work wonderfully titled *An Essay upon the Art of Love, containing an Exact Anatomy of Love and all the other Passions which Attend It* (1702), Silvester Jenks aimed to set forth "the Anatomy of a living Soul." Jenks wrote:

> And thus it is in the Anatomy, I now speak of; in which our Moralists have great Advantage over our Physicians. The Soul, which they anatomize, is Living: All the Passions of it, are in actual motion . . . And when our Mind reflects attentively upon itself, the Main Springs which continually move us, are in open view . . . We cannot but discern, what thoughts pass in our Mind; and what Impressions they make there; at least those great Impressions, which Philosophers call Passions. (1702: 13–14)

Philosophers such as Jenks described the causes and composition of the passions in terms derived from experience, either instead of, or in addition to, the physiological explanations of writers such as Descartes and Malebranche.

The government of the passions

The government of the passions was the focus of the large number of works that considered the passions from a moral perspective.[5] Depicting the ideal state of the passions, the defects of persons who are slaves to their passions (a common metaphor)[6] and offering practical guidance on the control of the passions, these were not works that could be accused of wanting "Warmth in the Cause of Virtue." Further, they exhibited considerable agreement in their assessment of the passions.

First, a wide variety of authors agreed with the view expressed by Senault in *The Use of the Passions*, that

> there is no passion which is not serviceable to vertue, when they are governed by reason, and those who have so cried them down make us see they never knew their use, nor worth. (1671: 8)

In an account of the passions that preceded his *Discourses of the Love of God* (1729), Isaac Watts connects the usefulness of the passions to the fact that we are embodied, as did other writers:

> The *Passions* may be made also of considerable Service to the *Mind*, while it dwells in this embodied State: For tho' they do not inform us what is *Good* for the Soul and what is *Evil*; yet when Reason has judg'd what is *Evil* and what is *Good*, the Passions are ready to lend their vigorous Assistance to avoid or pursue. They fix all the natural Spirits and the Thoughts of the Mind strongly upon those Objects which excite them, and with a sudden call they awaken and excite all the Powers of Nature to act agreeably to them. (Watts 1729: 48)

The passions are useful because they focus our attention and move us to action. A stronger position, that the passions were not only useful inducements to action but necessary supplements to our rationality, was also not uncommon, as will be discussed below. The attitude that the passions ought to be rooted out, correctly or incorrectly identified with Stoicism, was often explicitly rejected. In *The Government of the Passions* (1700) William Ayloffe wrote, "utterly to destroy our Passions were foolish to attempt, since impossible to effect" (1700: 5).

As the quotations from Senault and Watts make clear, the passions were taken to be useful *when governed by reason*. There was considerable agreement, too, about what this meant: the passions are reasonable when they are directed at proper objects, and in a degree appropriate to those objects. Jenks stated concisely the position expounded at great length by others:

> This *Art of Love* comprehends the *Whole Duty of Man*. Let him *Love* what he ought according to the Rules and Measures of Right Reason; neither more, nor less than he ought. (1702: 10)

Francis Bragge, author of *A Practical Treatise of the Regulation of the Passions* (1708), provides another example of this:

> and the Exorbitancy [of the passions] consists, either in *Misplacing* 'em upon *Undue Objects*, Loving what we should Hate, and Hating what we should Love, etc. Or, if the *Object* be Right, in suffering them to be Excessive in *Degree*. (1708: 4)[7]

On this view, the proper objects of love are God, other persons, and oneself. The passions are properly ordered, and reasonable, when their objects and degrees reflect essential relationships among the objects themselves, particularly those of relative superiority and perfection. The status of self-love was a subject of some disagreement, but it was not unusual for degrees of self-love to be acknowledged as virtuous.[8]

There was consensus, too, about the characteristics of the person whose passions are not properly governed. The person who is a slave to the passions overvalues what the passions are attracted to; and undervalues what is hated;[9] fails to deliberate before action; and generally lacks the tranquillity necessary for the operation of reason.[10] Watts's view here is representative:

> 'Tis the Nature of Passion to fix our Minds only on those Properties, Qualities, and Circumstances that first raised it, and to make them appear more considerable; and indeed it suffers us not to attend with Impartiality to any thing else. Passion generally tends to make us blind and deaf to all Circumstances and Reasonings but those which confirm itself. (1729: 52)

This picture of the proper functioning of the passions, with its emphasis on reason guiding the passions, may be precisely what Hume scholars expect to find

when they look at rationalist treatises. But there is another surprising feature of these works that significantly complicates this idealized view. Many of these very same authors who argued that reason discovers the proper objects and degrees of the passions also acknowledged, explicitly and implicitly, that reason could not actually control the passions. Reason, on this view, should be our guide, but it is weak and inactive. Understanding this apparently paradoxical aspect of these traditional accounts of the passions will illuminate Hume's deep disagreements with them.

The Weakness of Reason

Despite that fact that Senault argued that the passions were virtuous when governed by reason, he did not actually hold that reason by itself has the power to control the passions. Rather, reason requires the assistance of grace.

> As her forces are but small, to achieve what is good, she hath yet smaller power to rule her Passions; and though she approve not their disorders, she knows not how to remedy them. (1671: 62)

Senault continued:

> If prophane *Philosophers* object unto us, that Reason was in vain allowed us to moderate our Passions, if she have no power over them; and that nature is a useless guide, if she herself have need of a Conductor, we must satisfie them by experience, and teach them without the holy Scripture, that there are disorders in man which Reason alone cannot regulate, and that we are subject unto maladies, which nature without grace cannot cure. (1671: 65)

Senault viewed the demonstration of the weakness of reason as preliminary to the acceptance of faith; in this respect his argument is similar to those of the fideists. Assertions of the weakness of reason play a role in many treatises on the passions, but they are not always accompanied with Senault's emphasis on the role of faith. The acceptance of the view that reason is relatively powerless is important enough to illustrate with several representative passages. Here is Reynolds:

> Herein consists a great part of Mans infelicitie, by the Fall; that albeit his Understanding itselfe be blinded, and therefore not able to reach forth any perfect *Good* to the inferior parts; yet that small portion of Light, which it yet retaineth for the government of our Actions, is become uneffectuall, as being able only to convince, but not to reforme. (1640: 44)

Samuel Clarke, in his sermon *The Government of Passion* (1711) connected this view to the grounds for holding that the passions were useful:

Men . . . are of a middle Nature, between these two States, between perfect Reason and mere irrational Appetites: Being indued with Appetites and Passions, to excite and stir them up to Action, where their bare abstract Understanding would leave them too remiss. (1711: 17)

Isaac Watts maintained:

the Passions were given us to assist the feeble influences of our reason in the practice of duty . . . Reason is too slow and weak, to excite a sudden and vigorous activity in many cases. (1811: 82)

Malebranche also held that

when the volition that commands is a volition of pure reason unsustained by some passion, it is so weak and languid that our ideas are more like phantasms that we can barely make out and that quickly disappear. (1980: 386)

If "bare abstract Understanding" leaves us remiss, and the influences of our reason are "feeble, slow and weak," the passions are not merely useful, but also necessary, to virtuous action. This view is acknowledged with regret by Butler in his sermon "On compassion":

Reason alone, whatever any one may wish, is not in Reality a sufficient Motive of Virtue in such a creature as man. (1726: 86)

If we look at the most practical parts of these works, at the guidance they offer for controlling the passions, the weakness of reason is reflected in another way. It might be expected that those who defend the position of reason as the guide of the passions would give the perception of truths a central role to play in that endeavor. The basic rules for governing the passions, found in most of the works on this topic in the hundred years before the publication of the *Treatise* can be summarized as follows:

1 Learn the causes of the passions (through self-study or by reading guides like these).[11]
2 Avoid the situations that you know, from past experience, cause those passions that you want to avoid.[12]
3 Correct or modify your opinions, because your opinions influence your passions. This strategy has many interesting subdivisions. Senault, Coeffeteau and others remind their readers that the judgments we form that are based on the senses tend to be mistaken in various ways.[13] If we remember the ways in which our opinions can be false, they are less likely to move the passions. Bragge recommends that we not "nourish fanciful aversions" and that we take the words and actions of others "by the best handle." Coeffeteau counsels, in a

similar vein, that we have many reasons to distrust the feeling that we have been injured, and that "we must not so suddenly give credit to this opinion." Reynolds goes further and suggests that we "give injuries a new name."[14]

4   Try to combat the first stirring of a passion, because it is weaker at that point.[15]

5   At the first signs of a passion you want to suppress, redirect your thoughts.[16] Although many authors suggest this, Watts (1729: 55) and Malebranche both noted explicitly that the passions are not under the control of the reason or the will, but may be raised by the perceptions of the mind, or the imagination. According to Malebranche,

> when we wish to stop some motion of the spirits that has been excited in us, it is not enough to will it to stop, for that alone is not always capable of stopping it, but cleverness must be used by representing to ourselves things contrary to those exciting and sustaining the motion . . . (1980: 388)

There was actually only one strategy for controlling the passions that was controversial: the method (later championed by Hume) of using one passion to oppose another. Although everyone acknowledged that this method of controlling the passions was effective, Senault, Bragge, and Clarke thought that it was dangerous, since it tends to strengthen other passions,[17] while Reynolds, Charleton, Malebranche, Watts, and Grove endorsed it.[18]

What is most remarkable in this list is the fact that the mechanisms for controlling the passions, in spite of being described as the rules by which reason could govern the passions, all rely on causal knowledge, imagination, belief, and the passions themselves. Nothing specifically tied to the intellect, or the perception of truth, is involved. This is consistent with the explicit acknowledgment (discussed above) of the weakness of reason. Yet these traditional theories continued to maintain that reason, weak or ineffectual as it was, should still act as a guide to our passions, even though it couldn't actually control them.

On these accounts of the passions, reason's role as a guide, its normative function, did not depend on reason's power to *produce* passions or actions. Hume, of course, explicitly argues against this in *Treatise* 2.3.3, "Of the influencing motives of the will." A central argument of that section is:

> Since reason alone can never produce any action, or give rise to volition, I infer, that the same faculty is as incapable of preventing volition, or of disputing the preference with any passion or emotion. (T 2.3.4.4; SBN 414)

Hume's own theory of the passions shows that the same forces appealed to by the popular rules for regulating the passions – the imagination and the passions themselves – can continue to do their work when stripped of any reference to reason. The causes of the violent passions, on Hume's account, and the factors

that increase and diminish the passions, would be familiar to readers of the practical guides: custom (T 2.3.5); the imagination (T 2.3.6); contiguity or distance in space or time (T 2.3.7–8). What would be very *un*familiar was Hume's insistence that reason's already diminished role could be written out of the story.

"Reason *Directs* and the Affections Execute"[19]

The view that reason can direct what it cannot execute may seem paradoxical, especially in the light of Hume's later arguments, but there are reasons why it might not have appeared to be so to its adherents. As was described in the section "The government of the passions" (above) it was common to hold that the passions were reasonable when they were directed at their proper objects in the correct degree. Reason acts as a guide to the passions by discovering essential relationships among things in the world – the natural inferiority of ourselves to God, for example – and the propriety (or impropriety) of the passions to objects of different types. If there are relationships like this to be discovered, reason can serve as a guide merely by providing the external standard by which passions could be evaluated. By itself, this would be a very attenuated role for reason to play. More commonly, this view was supported by assumptions about the place of reason in a hierarchy of faculties within the soul. Changes in this underlying theory of the soul, however, left many theories of the passions in a precarious situation, and vulnerable to Hume's attack.

The scholastic account of the soul had separated the intellect, the rational soul, from the sensitive soul which was the seat of the passions. The soul, like the world itself, was taken to be a structured hierarchy, in which reason and passion had distinct roles to play. Reason's function of guiding the passions was derived not merely from its ability to perceive the truth, but from its place in the structure of the soul. The sensitive soul was inferior to the rational soul; although it could get out of hand, its nature was to be led. The scholastic account of the soul was losing ground throughout the seventeenth century,[20] however, and its gradual demise is reflected in the works on the passions discussed so far in this chapter. Coeffeteau (1621) and Reynolds (1640) accepted it, apparently without question; Senault (1671) and Descartes (1649) explicitly argued against it, defending instead the view that the soul is simple and unified; Charleton (1674) discusses Descartes's argument, but continued to maintain the scholastic view. In the early eighteenth century however, the Cartesian view of the soul prevailed in the works of Clarke (1711), Watts (1729, 1811), and Grove (1740).

On the scholastic view of the soul, it was possible to hold that reason guided the passions even though it was powerless to produce them, for the sensitive soul was naturally inclined to act in accordance with reason. The normative force of reason was based on the hierarchical structure of the soul, not reason's power to produce passions or actions. On the "Cartesian" view of the soul, that is, the view that the

soul is simple, and is not composed of separate faculties, although it was still possible to talk of reason guiding the passions, this talk had lost the most important part of its metaphysical foundation. If "bare abstract reason" is incapable of causing passions, then there is nothing to fall back on to explain its operation as a guide. Without the scholastic account of the soul, or something equivalent to it, at most reason could function as a guide to the passions only in the attenuated sense of providing an external standard, not as a motivating force. As I will argue below, Hume's analysis of the passions undercuts this minimalist rationalist view.

The tension in the rationalist position is particularly evident in the work of Clarke. Clarke wanted to ground moral knowledge in relations of ideas, but he recognized that the perception of truth, by itself, is incapable of generating the passions that cause actions. As if to attempt to get around this problem, the moral relations that Clarke took to be perceived by abstract reason were relations of fitness that mimic the structured relationship in the scholastic view of the soul. It is *fit*, according to Clarke, that creatures like ourselves who have reason, should follow reason. But the question remains: why will the perception of this truth, if it is one, move us more than any other?

Hume's Connection to the Earlier Literature

One point that should now be evident is that Hume is not beginning his account of the passions *ex nihilo*. His project is not merely one of phenomenological description. The particular passions and the circumstances surrounding them that Hume discusses in Book 2 are drawn from a body of popular and philosophical literature, although he re-casts them in a significantly different form. In some cases, Hume makes explicit references to the explanations others have given of individual passions. For example, after explaining the role of sympathy in pity, and asserting that women and children are most subject to pity (!) Hume says,

> Those philosophers, who derive this passion from I know not what subtile reflections on the instability of fortune, and our being liable to the same miseries we behold, will find this observation contrary to them among a great many others, which it were easy to produce. (T 2.2.7.4; SBN 370)

The explanation Hume rejects is given by Hobbes, and repeated by Charleton. The observation that children are prone to pity also occurs in Descartes.[21] In other cases, Hume's arguments appear to build on earlier ones. Malebranche, for example, had argued that passions cannot be identified by their objects, because the same object may cause different passions if its relationship to us changes, and different objects may cause the same passion (1980: 391. See also Descartes 1985: 349). The discussion of the objects and causes of pride given in *Treatise* 2.1.3 is remarkably similar. Finally, some observations about the passions just were, or had become,

common ones. To illustrate, I will give two examples, from Coeffeteau and Reynolds. Related passages in Hume are noted in the endnotes. Coeffeteau wrote:

> resemblance breedeth *Love*, and unites the affections of men. The truth hereof appears, for that men do commonly *love* those, that are allayed unto them in neernesse of blood, so as kinsmen doe commonly love one another: or by some conformity of humours and complexions, which maketh melancholy men *love* the company of their like, and Joviall spirits delight in the company of them that are pleasant: Or by some commerce of profession, which maketh Philosophers to love Philosophers; and Painters delight in Painters . . . But notwithstanding that which we have said . . . this conformity of professions may ingender hatred and envy, that is to say, when they of one trade and profession, living of the art and labor, hinder one another. (1621: 141–2. Compare with T 2.2.4.6–7; SBN 354 and T 2.2.8.15; SBN 378.)

In a similar vein, Reynolds observed,

> as we *love* our *selves*, for the unitie which we have in ourselves; so, wheresoever we find any similitude to our *selves*, or character of our *selves*, either in *Nature* or *Habits*, upon that also doe the beames of this Affection extend . . . in some cases, similitude is the cause of Envie. (1640: 85, 89)

A unique feature of Hume's account of the passions is that he constructs a unified explanatory framework that makes sense of many of these accumulated observations. Hume's appeal to the double association of impressions and ideas, and to sympathy, although in the tradition of earlier natural histories of the passions, goes far beyond the causal explanations found in other works in both scope and generality. But although Hume assimilates quite a lot from previous accounts of the passions, his intention is to subvert them, not to reform them.

Central Philosophical Issues regarding the Passions: Hume's Alternative Analyses

The proper objects of the passions

Part 3 of Book 2, "Of the will and the direct passions," has commanded more attention than the rest of Book 2, and there is at least one explanation for this. In Part 3 Hume gives clear arguments against opposing positions. This raises the question of what Hume is doing in Parts 1 and 2. These are, ostensibly, on another topic, i.e. the indirect passions, and they contain little argument against other theories. Yet I believe that here, as in Part 3, Hume's account of the passions is fundamentally different from that of his rationalist opponents. Indeed, the very creation of the category of the indirect passions constitutes an important deviation from other accounts, as I will argue below.

Hume called pride and love the *indirect passions*: they are passions resulting from the association of the idea of the cause of a pleasurable sensation with the idea of oneself or another person. In the indirect passions Hume distinguished between the cause of the passion and its object. Consider the indirect passion of love: anything that is a cause of pleasure may be a cause of love, if it is related to another person. Beauty, a fine character, fine possessions – all are causes of love. The person related to these causes of pleasure is the object of love. Traditional accounts of the passions emphasized, in addition, a conformity between the passions and their proper objects: love could be misdirected, or too weak, or too intense, for a given object. Passions and their objects were not merely causally related, but their individual natures were also suited (or unsuited) to each other. Hume's argument in "Of the influencing motives of the will," that a passion is an "original existence, and contains not any representative quality" (T 2.3.3.5; SBN 415) constitutes a direct attack on this point, but the foundation of his argument is the detailed discussion of the causes and the objects of the indirect passions in Parts 1 and 2 of Book 2. These causal conditions of the indirect passions can be exhaustively given, but over and above these there is nothing in the nature of the passion itself, no "representative quality," that relates it to one object rather than to another. Whether it is proper to love or hate an object cannot be determined from the passion itself. It is relevant to note that anger, the passion Hume uses to illustrate his claim, was often condemned as unreasonable or unlawful, except when directed at its proper object, i.e. sin (Clarke 1711: 20; Watts 1729: 84).

It might be objected that Hume does say that pride is connected to its object, the self, by an *original quality*. Hume states:

> Now these qualities, which we must consider as original, are such as are most inseparable from the soul, and can be resolv'd into no other. (T 2.1.3.3; SBN 280)

But original connections are still contingent, as we can see from Hume's discussion of another example:

> benevolence and anger are passions different from love and hatred, and only conjoin'd with them, by the original constitution of the mind . . . This order of things, abstractedly consider'd, is not necessary. Love and hatred might have been unattended with any such desires, or their particular connexion might have been entirely revers'd. (T 2.2.6.6; SBN 368)

The original connection between pride and the self, therefore, should not be construed as an a priori connection between a passion and its proper object.

Hume did not reject the claim that the passions could be morally evaluated. As components of character, they are, indeed, the focal point of Hume's moral theory. But passions do not become virtuous or vicious by being directed at proper or improper objects. There is no hierarchy of objects against which the passions can

be measured, and no inherent conformity or disconformity between passions and their objects. Hume argued:

> we always consider the *natural* and *usual* force of the passions, when we determine concerning vice and virtue; and if the passions depart very much from the common measures on either side, they are always disapprov'd as vicious. A man naturally loves his children better than his nephews, his nephews better than his cousins, his cousins better than strangers, where every thing else is equal. Hence arise our common measures of duty, in preferring the one to the other. Our sense of duty always follows the common and natural course of our passions. (T 3.2.1.18; SBN 483; see also T 3.2.2.8; SBN 488)

The varying degrees of the "love of relations" are explained, in Hume's theory, by association. Given the principles of association, we love what is closely related to ourselves more than what is distant (T 2.2.4.2; SBN 352). This natural operation of the passions provides the context in which the passions are evaluated, taking over the role attributed, by earlier theories, to reason's discovery of a metaphysical order.

Hume's rejection of the metaphysical view underlying rationalist accounts of the passions is also reflected in his innovative introduction of the category of the indirect passions. Most traditional views had one dominant passion, love, with different degrees of love suitable to different objects, according to their natures. Hume's account of the passions, however, identifies two person-related passions, pride and love, with parallel structures. The causes of love and the causes of pride are exactly analogous. The object of pride is oneself, the object of love is another person, but there is no rank ordering of these objects, nor of the passions that relate to them. This is a fundamental difference between Hume and his predecessors.

Mind and body

I argued above that the rejection of the scholastic view of the soul eroded the foundation of traditional accounts of the passions. Hume's development of a successful alternative required an alternative theory of the mind as well, and the *Treatise* provides this. On Hume's analysis, the mind is a compound entity, but it is not composed of independent faculties, as in the scholastic account. The components of the mind are perceptions, unified by relations of resemblance, causation, and the operation of sympathy.[22] Although some perceptions are extended and some are unextended (T 1.4.5.10; SBN 235–6), Hume's account of causation enables him to avoid traditional problems about interaction (the problems that bothered Princess Elizabeth of Bohemia). Hume argued:

> All our perceptions are not susceptible of a local union, either with what is extended or unextended; there being some of them of the one kind, and some of the other: And as the constant conjunction of objects constitutes the very essence of cause and

effect, matter and motion may often be regarded as the causes of thought, as far as we have any notion of that relation. (T 1.4.5.33; SBN 250) [23]

Indeed, Hume is almost indifferent about "whether we consider the body as a part of ourselves" (T 2.1.8.1; SBN 298; see also T 2.1.9.1; SBN 303). The body is the natural extension of the Humean mind, itself a compound of entities of different types. The passions, together with their complex causes and effects in ideas, impressions, and actions, are just a part of this larger causal unity. For Hume, there is no further problem about the relationship of mind and body.

In contrast with the scholastic view of the soul (and with the rationalist views that borrow, somewhat inconsistently, its authority) Hume does not construe the mind as a natural hierarchy. Hume is not an "inverted rationalist." This claim requires some defense, because it appears to conflict with Hume's famous pronouncement that "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them" (T 2.3.3.4; SBN 415).

First, the causal structure of the mind, in which the passions conflict, interact, and are influenced by forces such as the imagination and belief, does not give the passions a position of clear superiority. Hume describes the mind as a causal system (T 1.4.6.19; SBN 261) and the passions are a part of that system. Second, the passions themselves, as a group, are not organized hierarchically. As I argued above, pride and love, the indirect passions, have parallel structures, and the direct passions are described as working in conjunction with them (T 2.3.9.4; SBN 439). But third, and, I think, most important, the passions are not autonomous rulers within individual persons because the passions are social. We are influenced by the passions of others through sympathy, and this is the most distinctive feature of Hume's account of the government of the passions.

The government of the passions

Traditional treatments of the government of the passions combined depictions of the ideal state of the passions with guidance on how to achieve it. I have argued that Hume rejected the view that there is an ordering of objects, from highest to lowest, which reason discovers and that our passions must adjust to. Hume's alternative standard is derived from what is natural and usual in human nature. A certain degree of selfishness is excused because it is common (T 3.3.1.17; SBN 583); lack of affection for one's own children is blamed because it is uncommon (T 3.2.1.5; SBN 478). Evaluations of the passions, however, are not arrived at exclusively through individual reflection, even in this enlarged context of the knowledge of human nature. The moral sentiment of one individual is not analogous to the rationalist's individual reason. On Hume's view,

> the minds of men are mirrors to one another, not only because they reflect each others emotions, but also because those rays of passions, sentiments and opinions may be often reverberated . . . (T 2.2.5.21; SBN 365)

Every pleasure is increased by being perceived by others, and sympathized with.

> Hatred, resentment, esteem, love, courage, mirth and melancholy; all these passions I feel more from communication that from my own natural temper and disposition. (T 2.1.11.2; SBN 317)

Because of this shared sentiment and reverberation, our passions, though pleasurable to ourselves, may be opposed by the reflected uneasiness they cause in others (T 2.1.11.11; SBN 321). The "merit or demerit" of our passions, therefore, is revealed not by individual reflection, but through sympathy.[24] And when we feel the sentiments of others through sympathy,

> they operate upon us, by opposing and encreasing our passions, in the very same manner, as if they had been originally deriv'd from our own temper and disposition. (T 3.3.2.3; SBN 593)

Through the operation of sympathy, therefore, the government of the passions takes on an important social component.

Conclusion

Hume proposed a bold alternative to the accounts of the passions found in many earlier works. He recognized that the role played by reason in controlling the passions was already much diminished, and he constructed a theory that required only the recognized forces of belief, the imagination, and the passions themselves. At the core of Hume's new theory of the passions was his controversial account of the self as a complex causal unity. On Hume's account, neither this self, nor the objects of its passions, display the hierarchical organization presumed by the rationalists. Only in the context of human nature can the passions be distinguished as virtuous and vicious. And, finally, the passions are not governed by autonomous reason, but by the social force of sympathy. In writing the *Abstract* Hume may have given in to the "wonderful partiality" we all have for ourselves, but his judgment does not seem exaggerated. In his account of the passions, as in the rest of the *Treatise*, there is much that is new and extraordinary.

Notes

1   Malebranche 1980: 337–8; Hutcheson 1728: 29; Isaac Watts 1729: 3.
2   Descartes 1985, vol.1: 327; Reynolds 1640: 42; Charleton 1674; Bragge 1708.
3   For example, Reynolds 1640: 74; Senault 1671:. 26; Ayloffe 1700: 15–16; Cureau de la Chambre 1650: 20; Jenks 1702: 10; to some extent Coeffeteau 1621, see for example p. 155.

4 Watts 1729: 38: "Vexation and Fretfulness is an active, busy and galling Sort of Sorrow, that hangs about the Spirit, and teases it, and makes it restless, and is generally join'd with Anger against ourselves or others."

5 See, for example, Coeffeteau 1621; Reynolds 1640, Senault 1671; Charleton 1674; Ayloffe 1700; Jenks 1702; Bragge 1708; and Watts 1729; and the following sermons: Burghope 1701; Clarke 1711; Grove 1740.

6 See, for example, Senault 1671: 94; Descartes 1985: 347; Ayloffe 1700: 37; Burghope 1701: 4; Bragge 1708: 17; Clarke 1711: 22.

7 See also Charleton 1674: Epistle Prefatory, pages un-numbered; Malebranche 1980: 364; Burghope 1701: 4; Watts 1729: 7; Grove 1740: 240.

8 See especially Jenks 1702: 28–9, and Bragge 1708: 69ff. Bragge includes an extended discussion on the nature of the self. On self love see also Coeffeteau 1621: 142–3; Reynolds 1640: 85; Charleton 1674: 93–4. However, Senault argued that "there is no disorder in the world which doth not acknowledge [self-love] for it's originall" (1671: 216).

9 Senault 1671: 61–2. Malebranche 1980: 370; Burghope 1701: 3; Watts 1729: 52.

10 Bragge 1708: 92; Clarke 1711: 19; Watts 1729: 51–2.

11 Senault 1671: 108; Charleton 1674, Epistle Prefatory, pages unnumbered: Burghope 1701: 1–12; all the works in this self-help genre implicitly endorse this.

12 Coeffeteau 1621: 622; Reynolds 1640: 321; Senault 1671: 107; Grove 1740:. 272, 287.

13 Senault 1671: 105; Coeffeteau 1621: 557; Malebranche 1980: 369–72; Burghope 1701: 3.

14 Bragge 1708: 153, 176; Coeffeteau 1621: 624; Reynolds 1640: 322.

15 Coeffeteau 1621: 614; Bragge 1708: 176; Burghope 1701: 12; Grove 1740: 275.

16 Coeffeteau 1621: 618; Charleton 1674; Grove 1740: 265.

17 Senault 1671: 116–17; Bragge 1708: 17; Clarke 1711: 7.

18 Reynolds 1640: 52–3; Malebranche 1980: 388; Watts 1729: 97; Grove 1640: 278.

19 Clarke 1711: 18.

20 For further discussion see Levi 1964, especially chs. 11 and 12; and Fiering 1981, especially ch. 3.

21 Hobbes 1994: 53; Charleton 1674: 129; Descartes 1985: 375.

22 For a full discussion of this, see McIntyre 1989.

23 For a fuller discussion of this, see McIntyre 1994.

24 See T 3.3.2, "Of greatness of mind," and the discussion of virtuous pride in Baier 1991: ch. 9.

References and further reading

Ayloffe, W. (1700) *The Government of the Passions, according to the Rules of Reason and Religion*. London: J. Knapton.

Baier, A. (1991) *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge, MA: Harvard University Press.

Bragge, F. (1708) *A Practical Treatise of the Regulation of the Passions*. London: John Wyat.

Burghope, M. (1701) *The Government of the Passions: a Sermon*. London: John Wyat.

Butler, J. (1726) *Fifteen Sermons preached at the Rolls Chapel*. London: James and John Knapton.

Charleton, W. (1674) *Natural History of the Passions*. [London] In the Savoy: James Magnes.

Clarke, S. (1711) *The Government of Passion*. London: James Knapton.

Coeffeteau, N. (1621) *A Table of Humane Passions*, trans. E. Grimeston. London: N. Okes.

Cureau del la Chambre, M. (1650) *The Characters of the Passions, translated into English*. London: John Holden.

Descartes, R. (1985) [1649] *The Passions of the Soul*, in *The Philosophical Writings of Descartes*, vol. 1, trans. J. Cottingham, R. Stoofthoff, D. Murdoch. Cambridge: Cambridge University Press.

Fiering, N. (1981) *Moral Philosophy at Seventeenth Century Harvard*. Chapel Hill: University of North Carolina Press.

Grove, H. (1740) "On the government of the passions," in *Sermons and Tracts*, vol. 4. London.

——(1749) *A System of Moral Philosophy*. London.

Hobbes, T. (1994) [1650] *Human Nature*, ed. J. C. A. Gaskin. Oxford: Oxford University Press.

Hutcheson, F. (1728) *An Essay on the Nature and Conduct of the Passions*. London: Garland facsimile edition, 1971.

James, S. (1997) *Passion and Action: the Emotions in Seventeenth Century Philosophy*. Oxford: Clarendon Press.

Jenks, S. (1702) *An Essay upon the Art of Love, containing an exact anatomy of love and all the other passions which attend it*. London[?].

Levi, A. (1964) *French Moralists*. Oxford: Clarendon Press.

Malebranche, N. (1980) [1674] *The Search after Truth*, trans. T. Lennon and P. Olscamp. Columbus, OH: Ohio State University Press.

McIntyre, J. (1989) "Personal identity and the passions," *Journal of the History of Philosophy* 27, pp. 545–57.

——(1994) "Hume: second Newton of the moral sciences," *Hume Studies* 20, pp. 3–18.

Reynolds, E. (1640) *A Treatise of the Passions and Faculties of the Soule of Man*. London: R. Bostock.

Senault, J. (1671) [1641] *The Use of the Passions*, trans. Henry, Earl of Monmouth. London.

Watts, I. (1729) *Discourses of the Love of God and the Use and Abuse of the Passions in Religion*. London: J. Clark and R. Hett.

——(1811) [1739] *The Doctrine of the Passions, explained and improved*, 3rd edn. Trenton.

12

Liberty, Necessity, and the Will

Tony Pitson

The question of whether free will ("liberty") can be reconciled with determinism ("necessity") arises unavoidably for Hume in light of his expressed intention to establish a science of man applying the "experimental philosophy" associated with the natural sciences (T Introduction 6–7; SBN xvi). For in attempting thus to "anatomize human nature" (A 1; SBN 646) Hume is assuming that it is possible to provide causal explanations of the occurrence of mental states, or "perceptions," and thereby exhibit the principles governing our mental lives. In what sense can we be considered free and responsible agents if the states of mind from which our actions proceed occur in accordance with the laws of human nature which Hume is seeking in the *Treatise* to identify?

Hume addresses the issues involved here in a discussion which is generally regarded as one of the classic contributions to the problem of free will and determinism.[1] Yet it is probably fair to say that until the late twentieth century, at least, there has been little serious examination of Hume's position in the secondary literature.[2] In what follows I shall look in some detail at what Hume has to say about both liberty and necessity in *Treatise* 2.3.1–2, and I shall also consider the relation of these sections to other parts of the *Treatise*. Among the critical issues with which I shall be concerned is the adequacy of Hume's account of the nature of liberty and of related notions such as that of moral responsibility. Finally, I shall address the important interpretive issue of the validity of the standard view of Hume's position as a form of compatibilism (or "soft determinism").

The discussion of liberty and necessity occurs within the wider context of Hume's treatment in Book 2 of the *Treatise* of the passions. In brief, Hume classifies the passions as instances of one of the two types of impression distinguished at the beginning of the *Treatise*, i.e., impressions of reflection as opposed to impressions of sensation (T 1.1.2; SBN 7–8). To the extent that they depend upon the original impressions of sensation, impressions of reflection are "secondary" (T 2.1.1; SBN 275–7). Hume introduces further divisions within the category of impressions of reflection considered as passions: between those which are calm

and those which are violent (and which comprise the emotions as we would ordinarily think of them); and in the case of the latter, between the "direct" (those which arise immediately from pain and pleasure) and the "indirect" (those which in general depend upon the prior occurrence of ideas). The will is classified as an impression of reflection; and although not "comprehended among the passions," it shares with the direct passions the characteristic of being an immediate effect of pain and pleasure. It also resembles indirect passions like pride and humility, or love and hatred, in being indefinable – presumably because, as in the case of these indirect passions, it consists in a simple and unanalysable impression (T 2.3.1.2; SBN 399).

In T 2.3.1–2 Hume puts forward two sets of arguments which may be classified as, respectively, the necessity and the liberty arguments. Hume introduces the former by characterizing the will as "*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.*"[3] Now, according to Hume, it is "universally acknowledg'd" that the behavior of material bodies is governed by necessity so that one such "body or action" may be considered "the infallible cause of another." He refers to the results of his earlier investigation of the idea of necessary connection in T 1.3.14: in particular, that we are unable to discover the "ultimate connection" of bodies or objects and are acquainted only with their "uniform and constant conjunction," as a result of which we regard them as being causally related. The idea of causal necessity itself is derived from the "determination of the mind to pass from one object to its usual attendant, and infer the existence of one from that of the other" (T 2.3.1.4; SBN 400). On this account of the idea of necessary connection it involves two essential components: the constant union of the objects involved and the resulting inference from one to the other. The question to be considered in the present context is whether the "actions of the mind" exhibit the same sort of necessity. According to Hume, we do in fact find in human beings generally (regardless of sex, age, education, etc.) "the same uniformity and regular operation of natural principles"; in our case, just as in the case of natural objects, like causes produce like effects.[4]

Hume recognizes that we may resist this argument for ascribing necessity to our actions by denying that they display genuine uniformity, for example, by appealing to the fact that people may act out of character. But he is able to counter this by pointing out that there are varying degrees to which material objects provide us with experience of uniformity; and, in any case, even where such objects fail to behave in accordance with our expectations we still conclude that this is to be ascribed to "contrary and conceal'd causes" so that their operations are always equally necessary. Likewise, any apparent irregularity in human behavior should not lead us to deny that our actions are governed by a similar necessity.

The next stage in the necessity argument is to establish not only that there is a uniformity in human actions, just as in the case of the operations of natural objects, but also that it determines us to make inferences from the causes and

effects involved (i.e., motives and actions in the case of human beings). The conclusions we draw about people's actions from a consideration of their motives, character, and circumstances fall under the heading of "moral evidence." This provides the basis for our knowledge of historical facts and the kind of reasoning involved is employed also in disciplines such as politics and economics. But it also enters into our ordinary life at every point, since our actions are based on assumptions about the attitudes and responses of others. The reasoning in which we habitually engage reflects the belief that the "actions of the will" arise from necessity. The inferences involved are typical of causal inferences in resting on experience and observation of the constant union of the causes and effects in question. As Hume had earlier observed, "there is but one kind of necessity" and it is a mistake to suppose that there is a distinction to be drawn between moral and physical necessity (T 1.3.14.33; SBN 171). This is borne out by the fact that the same piece of probable argument may appeal to both moral and natural evidence (T 2.3.1.17; SBN 406–7; cf. EHU 8.19; SBN 90–1). If anyone disputes whether human actions exhibit necessity this can only be a result of ascribing a different meaning to this notion from that claimed by Hume. The doctrine of liberty, insofar as it disputes that our actions are governed by necessity, effectively denies that they may be causally explained and so reduces them to chance events.

In T 2.3.1 Hume has been principally concerned to establish that our voluntary actions provide instances of necessity and that in denying this the doctrine of liberty is to be rejected. In the next section he identifies several reasons for the prevalence of this doctrine, notwithstanding its absurdity or unintelligibility. The first of these is the difficulty of persuading ourselves that any action we have performed was governed by necessity, so that we could not have done otherwise. Hume introduces here an important distinction between liberty of *spontaneity* ("which is oppos'd to violence") and liberty of *indifference* ("which means a negation of necessity and causes").[5] It is only the former kind of liberty, according to Hume, that we are concerned to preserve, yet it is "almost universally confounded" with liberty of indifference, with the result that we do not suppose that our voluntary actions are necessitated (T 2.3.2.1; SBN 407–8). A second source of the doctrine of liberty is a "*false sensation* or *experience*" of indifference. Voluntary actions seem to occur independently of any determining cause: whatever we do we can readily imagine having willed or chosen to act quite differently, and this appears to show that "the will itself is subject to nothing." But this conclusion cannot be established by any such exercise of the imagination. We should compare our perspective as agents with that of the spectator who is able commonly to infer our actions from our motive and character, and would be able to do so in all cases were he acquainted with "the most secret springs" of our dispositions (T 2.3.2.2; SBN 408–9).

Hume also ascribes acceptance of the doctrine of liberty to the influence of religion (T 2.3.2.3; SBN 409). Liberty of indifference is opposed to necessity, and the latter doctrine is thought to be harmful to religion and morality. According

to Hume, however, necessity is essential to both religion and morality to the extent that the institutions of reward and punishment, whether human or divine, presume that our actions, both good and evil, are the effects of motives with which they are customarily conjoined. But the doctrine of liberty of indifference, which essentially reduces our actions to chance events, prevents us from being responsible for any of our actions, and not just those which are casual or accidental. For it destroys any real connection between actions and the agents who perform them in spite of the fact that we are praised or blamed for our actions only insofar as they proceed from some cause in our character or disposition. This is why virtue and vice actually depend upon the principles of necessity, even if the "common opinion" supposes otherwise, and it explains why we do not blame people for harmful actions which are performed ignorantly, hastily, or unpremeditatedly (T 2.3.2.5–6; SBN 410–11). Actions are vicious or criminal only insofar as they proceed from vicious or criminal passions, but the doctrine of liberty of indifference prevents us from establishing any connection of this kind.

Hume refers to his earlier discussion of the idea of necessary connection (in T 1.3.14) and his definitions of "cause" as a philosophical relation and as a natural relation. In the former case, necessity is a matter of "the constant union and conjunction of like objects," and in the latter of "the inference of the mind from the one to the other." Hume claims that necessity, in both these senses, is commonly allowed to belong to the will, so that no one denies that we are able to make inferences about human actions based on "the experience'd union of like actions with like motives and circumstances." Perhaps one reason for resisting this is that there is thought to be *more* to necessity as it involves the operations of matter. If this is the received view, however, then it is rejected by Hume: in this respect, he changes the "receiv'd systems" with regard to material objects rather than with regard to the will (T 2.3.2.4; SBN 409–10).

This, in broad outline, is what Hume has to say in T 2.3.1–2 about the dispute concerning liberty and necessity. It gives rise to a number of questions, including the following. Does Hume successfully make out a case for his claimed parallel between moral and natural necessity? How should we understand Hume's account of liberty of spontaneity? Is liberty of spontaneity essentially distinct and separable from liberty of indifference, as Hume claims? There is the further question of whether Hume's views about liberty and necessity allow for an adequate treatment of moral responsibility. We will begin by looking more closely at what Hume has to say about the doctrine of necessity in relation to our voluntary actions.

The Doctrine of Necessity

Two issues, at least, arise here. One has to do with Hume's attempt to establish that human actions are, to the same extent as natural events, subject to necessity. The other has to do with Hume's view of our beliefs about this aspect of human

behavior. In regard to the former issue, the question is whether Hume's claims about the necessity to be found in our actions are established by the kinds of consideration to which he refers. On the compatibilist interpretation, this is equivalent to the question of whether Hume succeeds in establishing the truth of determinism as an account of human behavior. This might be seen as part of the more general problem of Hume's commitment to universal causation. Hume regards the causal maxim – i.e., that *whatever begins to exist, must have a cause of existence* (T 1.3.3.1; SBN 78) – as something which is commonly taken for granted. Yet it is neither intuitively nor demonstratively certain, for we are concerned here with a matter of fact rather than an "infallible" relation of ideas. The causal maxim seems in fact to be a product of our experience of causally related events as ones which are constantly conjoined, so that we acquire the habit of regarding all events as ones which belong to pairs of events so conjoined even if this has not been confirmed by experience.

How, then, does this bear on the necessity argument of T 2.3.1? If determinism, or necessity, cannot be shown to hold in the natural world, then it is unlikely that it may be shown to apply in the moral world. But this suggests that we should see Hume's argument as an attempt to establish only that the doctrine of necessity applies to human actions in just the same way, or to just the same extent, that it applies to natural objects or events. In referring to the discussion of T 2.3.1–2 in the *Abstract*, Hume claims to be putting "the whole controversy" concerning liberty and necessity "in a new light, by giving a new definition of necessity"; and the claim is that the elements which, according to this definition, are essential to necessity – i.e., union and inference – are features of human actions as well as the actions of matter (A 34; SBN 661). The obvious objection, that human behavior exhibits a degree of inconstancy or irregularity, may be answered by pointing out that the same is true of material effects in so far as their causes are often "variable and uncertain" (A 32; SBN 661). It seems, then, to be a feature of Hume's "new definition" that the application of the doctrine of necessity does not presuppose that we are dealing with entirely uniform regularities.[6] What matters is that the uniformities we do experience should be sufficient to enable us to make the kinds of causal inference to which Hume refers (Garrett 1997: 126). Hume is, however, writing as a philosopher who ascribes "the contrariety of events" to "the secret operation of contrary causes" – in contrast to the "vulgar," who are prepared to ascribe this contrariety to the absence of necessitating causes (T 1.3.12.5; SBN 132). Hume's claim "that all actions of the will have particular causes" (T 2.3.2.8; SBN 412) reflects the assumption that volitions as mental occurrences will be subject to the principle of universal causation in the same way as any other kind of event, recognizing that the principle itself can be established neither empirically nor in any other way.

Even if Hume is granted this claim about actions of the will, we might wonder why they must have specifically psychological causes in the form of motives (Botterill 2002: 286). The answer, in short, appears to be that voluntary actions are in general a response to pleasure and pain, and are therefore associated with

the presence of motives in the form of the "direct" passions. Hume further claims that the same motives always produce the same actions (EHU 8.7; SBN 83; cf. T 2.3.1.5; SBN 401); in doing so, he is attempting to apply the fourth of his "rules" for judging cause and effect (T 1.3.15.6; SBN 173) to the moral as well as to the natural realm and hence to claim that the same kind of uniformity may be involved in each case. But is Hume entitled to claim that motives as the presumed causes of our voluntary actions operate in this uniform kind of way? He responds to the objection that our actions may on occasion appear to be irregular and extraordinary not only by claiming that the same kind of apparent irregularity is to be observed in the natural world, but also by suggesting that the internal principles which govern our voluntary actions may after all operate in a uniform manner, just as in the case of natural causes and effects (EHU 8.12; SBN 88). But the *evidence* for this claimed uniformity of motive and action is not compelling. Moral sciences such as history may indicate the pervasiveness of certain kinds of motive in the actions of human beings, but scarcely confirm the uniform connection between motive and action claimed by Hume.

Some of these points bear on the question of the nature of our *beliefs* about the doctrine of necessity as it applies to human actions. On the face of it, this aspect of Hume's discussion appears to involve him in an inconsistency. Having explained that necessity is to be defined by reference to the two definitions of cause "of which it makes an essential part," i.e., in terms of the constant union of like objects or the inference from one to the other, Hume goes on to say that

> necessity, in both these senses, has universally, tho' tacitly, in the schools, in the pulpit, and in common life, been allow'd to belong to the will of man, and no one has ever pretended to deny, that we can draw inferences concerning human actions, and that those inferences are founded on the experience'd union of like actions with like motives and circumstances. (T 2.3.2.4; SBN 409)

The prevalence of the doctrine of liberty, however, is explained by supposing that we "imagine we feel that the will itself is subject to nothing" (T 2.3.2.2; SBN 408; cf. EHU 8.23n; SBN 94n). How can we both acknowledge the regularity of the conjunction between volitions and motives while also imagining that the will itself may be subject to nothing? In the corresponding discussion in the first *Enquiry* 8.6–8, it is said that "all mankind have ever agreed in the doctrine of necessity" that it is "universally acknowledged that there is a great uniformity among the actions of men," and that "we readily acknowledge a uniformity in human actions and motives." On the other hand, Hume goes on to ask why mankind exhibits a reluctance to acknowledge the doctrine of necessity and, indeed, to profess the contrary opinion.

Is Hume, then, guilty of providing an ambivalent or inconsistent account of our beliefs about the doctrine of necessity? It seems, in fact, that any ambivalence or inconsistency is to be charged, rather, against those who hold these beliefs. Thus Hume refers to "Our way of thinking" (in regard to liberty and necessity)

being "absolutely inconsistent" (T 2.3.1.13; SBN 404). This inconsistency is a product of "confus'd ideas" and "undefined terms": hence the need for the clarification of the notions of liberty and necessity which Hume attempts to provide. One of the points to emerge from this clarification is that there is a sense in which we may be seen to agree with the doctrine of necessity, even if we are apparently reluctant to acknowledge the truth of the doctrine in so many words. Once it is recognized that all that can be meant by "necessity" in this context is that there is a constant conjunction of actions and motives (with the consequent inference from one to the other), then it is evident that this is reflected in our ordinary reasoning about actions and their causes. In this sense, the belief in the doctrine of necessity might be seen as an implicit feature of the kinds of explanation which we ordinarily provide of human actions. Hume is, in effect, *reminding* us in T 2.3.1 of the nature of necessity as a feature of our voluntary actions, which is something we are liable to forget, partly because we may imagine feeling that the will itself is subject to nothing. We may also forget because we mistakenly suppose that the idea of necessity implies something like force or constraint, in spite of the fact that we have no awareness of any such thing when we engage in voluntary action (T 2.3.2.1; SBN 407). Hence Hume's remark

> I do not ascribe to the will that unintelligible necessity, which is suppos'd to lie in matter. But I ascribe to matter, that intelligible quality, call it necessity or not, which the most rigorous orthodoxy does or must allow to belong to the will. (T 2.3.2.4; SBN 410)

There is a discrepancy between our commitment to the doctrine of necessity as reflected in our ordinarily reasoning about voluntary actions and their causes, and our reluctance explicitly to acknowledge the truth of this doctrine. As we might expect, there is an important distinction to be drawn in this respect between the situation of the "vulgar" and that of philosophers. Philosophers are provided with evidence of the truth of the doctrine of necessity in the form of history, politics, and economics, all of which testify to the use of "moral evidence" in arriving at conclusions about the causes of people's actions, including their motives (T 2.3.1.15; SBN 404–5; cf. EHU 8.7; SBN 83). This enables philosophers to recognize that the apparent uncertainty of events may be ascribed to "the secret operation of contrary causes" rather than to any uncertainty in the causes themselves, in accordance with the maxim that "the connexion between all causes and effects is equally necessary" (EHU 8.13; SBN 86–7).

The Doctrine of Liberty

We have seen that in T 2.3.1–2 Hume appears to identify the doctrine of liberty with liberty of indifference.[7] In this latter sense liberty is equated with chance, and

is thus incompatible with necessity. Hume is therefore committed to rejecting the notion of free will where this is understood to involve the claim that the will itself is undetermined (T 2.3.1.18; SBN 407). Even if Hume's position on the debate between liberty and necessity is a form of compatibilism, we should not suppose that he represents necessity as being compatible with freedom of the will itself. The relevant question for Hume is the nature of "the motivating influences of the will" (T 2.3.3; SBN 413–18): that it *is* invariably subject to such influences is a central feature of the doctrine of necessity to which Hume subscribes.

Even in the *Treatise*, however, Hume clearly means to allow for our possession of a kind of liberty which is compatible with necessity: what he classifies as liberty of spontaneity. This gives rise to a number of questions. How, for example, does this notion of liberty relate to the notion of liberty of indifference, bearing in mind Hume's rejection of the latter? And does the notion of liberty of spontaneity provide an adequate basis for the ascription of moral freedom and responsibility? Before addressing these questions we should clarify the alternative notion of liberty which is rejected by Hume. While liberty of indifference might be understood as the view that our *actions* are uncaused or undetermined by antecedent conditions, it seems clear that Hume is concerned with the view that our *willings*, considered as the immediate psychological causes of our actions, are themselves uncaused or undetermined. (This is reflected in his reference to the "*false sensation* or *experience*" of the liberty of indifference which leads us to suppose that "the will itself is subject to nothing"; see p. 218 above.) So far as the meaning of "liberty" is concerned, what is at issue is its application to our voluntary actions (EHU 8.23; SBN 95). Hence Hume's claim that "*liberty . . . can only mean a power of acting or not acting, according to the determinations of the will*": this is the liberty of spontaneity which, apart from providing the most common sense of the word "liberty," is the only species of liberty we are concerned to preserve (T 2.3.2.1; SBN 407–8).[8] Liberty of spontaneity, so understood, involves two sorts of condition: a negative one, consisting in the absence of constraint, and a positive one which requires that our actions are caused by our volitions. The positive condition appears to provide a rather minimal notion of agency, as reflected by the fact that on Hume's account non-human animals are also capable of voluntary behavior (T 2.3.9.32; SBN 448). A more serious worry, perhaps, is that liberty of spontaneity does not, by itself, imply that agents have any power over the determination of their wills; but that they should do so might seem an important condition for moral responsibility. We should not assume, however, that Hume himself considers liberty of spontaneity to provide more than a necessary condition for the ascription of moral responsibility. His account of the conditions for responsible action appears to ascribe to us a certain kind of *moral* liberty in so far as our actions are not just the product of volitions, but are also caused by our characters or mental dispositions (T 2.3.2.6–7; SBN 410–12; EHU 8.29–30; SBN 98–9). We are distinguished from animals, for example, by our ability to act in accordance with the calm passions, as Hume classifies them, i.e., we are able to form a conception of

our longer-term interest and thereby exhibit what Hume calls strength of mind (T 2.3.3.10; SBN 418). This might help to explain Hume's apparent reluctance to regard animals as moral agents (T 3.1.1.25; SBN 467), while doing so in a way that is consistent with his own account of the basis for the distinction between virtue and vice (Pitson 2003).

While Hume himself evidently sees it as crucial to distinguish the different notions of liberty of indifference and liberty of spontaneity, it might be questioned how far they are truly independent of each other. It has been suggested, for example, that liberty of spontaneity entails a capacity of the agent which is part of liberty of indifference. The thought here is that we can be described as possessing a power of acting or not acting according to the determinations of the will (in accordance with Hume's definition of liberty of spontaneity – or "hypothetical" liberty, EHU 8.23; SBN 95), only if our choices are themselves undetermined, so that having chosen to act in a particular way we could have chosen to do otherwise, in accordance with liberty of indifference (Kenny 1975: 123). This view reflects a certain conception of the nature of volition: i.e., that it involves the rational consideration of an action which enables the agent to exercise this two-way power of choice. It is clear, however, that Hume himself regards volitions as the immediate product of passion rather than reason (see T 2.3.3; SBN 413–18). So far as Hume's conception of "hypothetical" liberty is concerned, we are free just to the extent that we are able not only to act on the choice we do make, but also to act on the alternative which is not chosen. This is why liberty is to be opposed to constraint, as in the case of the person who is a prisoner and in chains. Even supposing that, for whatever reason, the prisoner is content to remain in his cell, he is deprived of liberty to the extent that it would be impossible for him to leave should he wish to do so.[9]

But what, then, of the view that liberty of spontaneity is inseparable from liberty of indifference as consisting in the power not only to act on the choice we do not make, but also to make that alternative choice? It appears to be a mistake to suppose that someone cannot correctly be described as performing a voluntary action, or one that is the product of a genuine choice, if the action is in fact necessitated. In cases of overdetermination, for example, it is simply not true that the agent could have done otherwise even though he is doing what he chooses to do (Frankfurt 1969: 835–6). But more importantly, it seems just wrong to suppose that we would no longer hold an agent responsible even when there is a reasonably clear sense in which he could not have done otherwise – given, e.g., the nature of his moral commitments (Dennett 1984: 133). What cases of this latter kind appear to establish is that in determining someone's responsibility for an action we are concerned with the relation of that action to features of the agent which fall under Hume's reference to the person's character and disposition, and *not* with his or her possession of liberty of indifference as Hume defines it.[10]

A fundamental issue concerning Hume's endorsement of liberty of spontaneity is whether this provides the basis for distinguishing between cases of genuinely

voluntary action and action which might appear to be compelled and thus not genuinely free and responsible. The kind of example often invoked here is that of the kleptomaniac whose acts of stealing are supposed to be the product of internal compulsion (in contrast to the actions of the discriminating shoplifter). The problem which is thought to arise for Hume in this case is that the kleptomaniac's actions appear to satisfy the negative condition for liberty of spontaneity to the extent that no one prevents the acts of stealing from taking place. It is less clear, however, that they satisfy the positive condition of resulting from the choice or volition of the person concerned (Árdal 1966: 88). Indeed, it appears that the actions are performed regardless of what the person wills or chooses, so that even if he had chosen *not* to steal he would have failed to act accordingly. It is just this kind of factor that, arguably, distinguishes this from the ordinary case of stealing. In any case, there must be doubt as to whether in a case like that of kleptomania the conditions for the kind of moral liberty which Hume apparently recognizes are satisfied. For it seems clear that the kleptomaniac cannot be acting on the basis of a calm passion, as Hume understands this notion: on the contrary, the klepto-maniac is apparently unable to act in accordance with his recognition of what is in his own longer-term interests.[11]

Before we leave the question of the adequacy or otherwise of Hume's account of what it is for an action to be free and responsible, there is a further matter to be considered. On Hume's view of necessity it appears that our choices or volitions must themselves have a causal explanation, and one that might be expected to appeal to internal psychological causes which, like the choices or volitions them-selves, represent a response to pleasure or pain or to the prospect of their occur-rence. The candidates will be, in Hume's terms, the direct or indirect passions that might typically be cited as motives for the resulting actions. Motives them-selves reflect other features of the mind of the agent: in particular, the agent's character, which typically manifests itself in the disposition to experience certain kinds of passion and to act accordingly. As Hume also indicates, the character of the agent may itself be ascribed to certain causes, both "moral" and "physical" (Hume 1987: 198). Even if factors of the former kind – which "work on the mind as motives or reasons" – predominate, they are still ones which, like the purely physical causes, do not themselves fall within the control of the agent. This last point is liable to generate a familiar worry about the attempt to combine an accep-tance of necessity (or "determinism") in regard to the causes of our voluntary actions with a commitment to the possibility of ascribing responsibility for those actions to the agent himself. Thus it might be argued that Hume has failed to establish that moral responsibility is consistent with an unrestricted determinism which applies to motives as the effects of prior causes. In fact, it has been suggested that Hume encounters here a problem which is akin to the theological problem he himself identifies for those who believe that "The ultimate Author of all our volitions is the Creator of the world" (EHU 8.32; SBN 99–100). If this belief results in the conclusion that God – as the ultimate cause of our volitions – is the

only responsible agent, it seems possible that a secularized version of Hume's argument might persuade us that we are not responsible agents at all (Stroud 1977: 152).

Now, this line of thought may be associated with a demand for the liberty of indifference which Hume justifiably resists to the extent that it apparently prevents us from being motivated by the prospect of reward or punishment (T 2.3.2.5; SBN 410; EHU 8.28; SBN 97–8). But, in any case, Hume might respond here as he does to the theological version of the doctrine of necessity, i.e., that such "remote and uncertain speculations" will fail to influence the sentiments of approbation or blame which "arise from the natural and immediate view" of each other's characters, dispositions, and actions (EHU 8.35; SBN 102). On this naturalistic understanding of the notion of responsibility, it seems evident that Hume would find no obstacle here to his view that moral responsibility is not only compatible with but actually required by the doctrine of necessity.[12]

Hume's Compatibilism

While it seems evident that Hume's position with regard to the apparent conflict between liberty and necessity is a form of compatibilism, there is scope for debate about the precise nature of his compatibilism. What appears beyond question is that Hume is neither a "hard" determinist for whom the possibility of free and responsible action is an illusion, nor on the other hand a libertarian for whom the will is free in the sense of being uncaused or undetermined. The question at issue is whether Hume's compatibilism may be classified as "standard" in the sense that his attempt to combine acceptance of the doctrine of necessity with a commitment to liberty of spontaneity (or "hypothetical" liberty) depends on his account of the meaning of the concepts involved. The tendency so to classify Hume's position is reflected in the common view of this position as one which exemplifies an approach originated, perhaps, by Hobbes, and continued by such empiricist-minded philosophers as Mill, Russell, and Ayer. This understanding of Hume's compatibilism is apparently supported by Hume's own reference to the liberty/necessity dispute as a verbal one which might be resolved by a "few intelligible definitions" (EHU 8.2–3; SBN 81; cf. T 2.3.1.18; SBN 407). It is this that Hume appears to have in mind in referring to his "reconciling project with regard to the question of liberty and necessity" (EHU 8.23; SBN 95).

There are, however, various reasons why the classification of Hume's position as a "standard" form of compatibilism may be resisted. One of these has to do with an argument which appears to play a central role in the standard form of compatibilism: what has been dubbed the *contrastive* argument (Botterill 2002: 279). In essence, the argument claims that freedom or liberty is to be contrasted with *constraint* rather than with *causation*, so that an action's being caused or necessitated is compatible with its being freely performed (Ayer 1982: 19). While

some commentators have suggested that this argument forms the principal basis of Hume's compatibilism (Penelhum 1975: 120), it may be disputed whether Hume ever advances the contrastive argument, especially since he avoids identifying freedom with the absence of constraint or compulsion (Botterill 2002: 296–8). It is true that in accounting for the prevalence of the doctrine of liberty (i.e., of indifference) Hume does refer to the assumption that the idea of necessity seems "to imply something of force, and violence, and constraint" of which we are not aware in our ordinary voluntary actions (T 2.3.2.1; SBN 407). But it appears that Hume is responding here to a possible objection to the doctrine of necessity – i.e., that we have no awareness of internal compulsion when we act – rather than arguing for the compatibility of this doctrine with liberty of spontaneity. While the standard interpretation assumes that Hume's reconciling project depends upon his account of liberty, it seems clear that the necessity argument represents Hume's primary contribution to the "long disputed question concerning liberty and necessity" (EHU 8.2; SBN 81; cf. A 34; SBN 661). This, in turn, suggests that Hume's liberty arguments – i.e., the antilibertarian argument and the spontaneity argument – are liable to be misconstrued if they are seen as providing the basis for Hume's compatibilism (Russell 1995: 21). How, then, should we see the relation between these arguments and Hume's necessity argument? The principal emphasis of the latter is to call into question a view about the nature of causal necessity as it operates in the natural realm, i.e., that there is *more* to this relationship among objects than merely constant conjunction and inference (as reflected in Hume's two definitions of "cause"). Hume claims that his own account of necessity as "something, that exists in the mind, not in objects" (T 1.3.14.22; SBN 165) also applies to the relation between volition and bodily movements, contrary to those who claim that we are provided here with an experience of energy or power (T 1.3.14.12; SBN 632–3; cf. EHU 7.9–20; SBN 64–9). This enables Hume to respond to the libertarian whose position reflects the assumption that it is possible to perceive something like a necessary connection between natural causes and effects but not between volitions and actions, so that relations of the latter kind fail to provide genuine instances of causal necessity (T 2.3.2.1; SBN 407; EHU 8.21; SBN 92). This is to be contrasted with the standard interpretation, according to which the purpose of the necessity argument is to provide an account of causation which distinguishes it from compulsion or constraint.

Another objection to the standard interpretation of Hume's compatibilism is that the necessity argument, which denies that there is more to causation than union and inference, actually undermines Hume's liberty arguments (Russell 1995: 53–5). These arguments allegedly presuppose that we are responsible for our actions to the extent that we are able to produce or bring them about through the exercise of causal power, while the necessity argument calls into question this way of understanding the causal relationship. Why, however, should we accept that Hume's liberty arguments commit him to any such view of the conditions for responsible agency? In the case of Hume's anti-libertarian argument, what is at

issue is our accountability for our actions, and Hume suggests that this depends on their proceeding from some cause in the character and disposition of the agent. But in characterizing this relation Hume refers to a *uniformity* in the actions that flow from character; and he adds that "this uniformity forms the very essence of necessity" (T 2.3.1.10; SBN 403). So far as liberty of spontaneity is concerned, apart from the absence of constraint or compulsion the essential condition is that our actions should proceed from our volitions. The relation between volitions and actions is, as Hume insists, something that can be known only from experience of their conjunction and not from awareness of any power by which they are connected (EHU 7.10; SBN 64–5). Hume does not appear to see this as posing any difficulty for the way in which we ordinarily hold people responsible for their actions. In fact, the standard interpretation appears justified to the extent that Hume's "hypothetical" liberty is considered to be necessary for moral responsibility (liberty when so defined is "essential to morality," EHU 8.31; SBN 99).

The standard interpretation might be faulted for missing the true significance of Hume's necessity argument in this context, but it does not seem to be seriously at fault in its view of the relation of this argument to Hume's liberty arguments. Its real shortcoming is the failure to recognize the relationship of Hume's notion of liberty of spontaneity, e.g., to his *naturalistic* account of attributions of responsibility – one which makes reference to the moral sentiments that arise when we draw inferences from people's actions to their mental causes (Russell 1995: 64). Since these sentiments arise only as a result of our ability to make such inferences, this evidently presupposes both the doctrine of necessity, as Hume understands it, as well as liberty of spontaneity. While it is important that we should recognize this naturalistic dimension to Hume's use of both the necessity and liberty arguments, this is not in itself inconsistent with the central thrust of the standard interpretation, i.e., that Hume in T 2.3.1–2 is providing an account of the nature of the notions of liberty and necessity which allows us to accept that our voluntary actions may be free and responsible while also exhibiting the uniformity which is at the heart of Hume's "new definition" of necessity.

See also 3: IMPRESSIONS AND IDEAS; 6: CAUSATION; 9: HUME'S CONCLUSIONS IN "CONCLUSION OF THIS BOOK"; 10: THE POWERS AND MECHANISMS OF THE PASSIONS; 11: HUME'S "NEW AND EXTRAORDINARY" ACCOUNT OF THE PASSIONS; 13: REASON, PASSION AND THE INFLUENCING MOTIVES OF THE WILL; 14: HUME'S ARTIFICIAL AND NATURAL VIRTUES; 15: VIRTUE AND THE EVALUATION OF CHARACTER.

Notes

1   Hume returns to the issues involved in Section 8 of the first *Enquiry*. The principal difference between the two discussions lies with Hume's more conciliatory approach

in the latter to the apparent conflict between liberty and necessity – to the extent that he is able to refer there to his "reconciling project with regard to the question of liberty and necessity" (EHU 8.24; SBN 95). This in turn reflects the fact that in the *Enquiry* Hume explicitly identifies a sense of "liberty" which may be seen to apply to human actions without dispute, while in the *Treatise* "the doctrine of liberty" is described as being "absurd . . . in one sense, and unintelligible in any other" (T 2.3.2.1; SBN 407). As various commentators have maintained, however, the differences between the two discussions are more apparent than real.

2 Important exceptions are provided by Garrett 1997; Penelhum 2000; and – especially – Russell 1995. A closer examination of the secondary literature reveals interesting differences concerning not only the nature of Hume's position on the problem of liberty and necessity but also the philosophical merits of that position.

3 This should be contrasted with the view of the will as a faculty: a view which is bound up with the notion of "free will."

4 This is a presupposition of Hume's program of the science of man, as referred to above.

5 As Hume indicates, the distinction is a scholastic one; see the discussion in Kenny 1975.

6 See EHU 8.13; SBN 86: "All causes are not conjoined to their usual effects with like uniformity."

7 The doctrine of liberty in the *Treatise* is essentially equivalent to libertarianism – i.e., the denial that our volitions are subject to necessity – as a form of incompatibilism.

8 This way of formulating the notion of liberty of spontaneity could give rise to misunderstanding. Hume is not saying that having chosen to act, the agent might equally have chosen *not* to act and in that way exhibit the power of acting or not acting according to the determinations of his will. Rather, whatever choice the agent *does* make, he is free to the extent that he is able to act (or refrain from acting) accordingly.

9 One is reminded here of Locke's example of the man who wakes up in a room which is, unknown to him, locked but where he happens to prefer staying to going away (*Essay Concerning Human Understanding*, 2.21.10). Locke's verdict is that while this man's stay is voluntary (presumably, because he does not try to leave), he is not free or at liberty in this respect.

10 As Dennett points out, if responsibility really depended on establishing that in the circumstances in question the agent could have done otherwise, then it is unlikely that we should ever be able to resolve this issue (1984: 135). This, indeed, is indicated by Hume's own treatment of our attempt to persuade ourselves that the will is subject to nothing.

11 Hence, it might be suggested that even if the kleptomaniac *is* free to the extent that he is not prevented from stealing, he would fail to satisfy Hume's conditions for responsibility insofar as his actions are not a true reflection of character and so not liable to give rise to disapproval and blame (see Garrett 1997: 136).

12 This is not to deny that Hume's account of the conditions for responsible action is scarcely adequate as it stands; but it is another matter whether that account can reasonably be criticized by reference to Hume's view of the doctrine of necessity as applying to voluntary human actions.

References

Árdal, P. (1966) "Liberty," in *Passion and Value in Hume's* Treatise (pp. 80–92). Edinburgh: Edinburgh University Press.

Ayer, A. J. (1982) "Freedom and necessity," in G. Watson (ed.) *Free Will* (pp. 15–23). Oxford: Oxford University Press.

Botterill, G. (2002) "Hume on liberty and necessity," in P. Millican (ed.) *Reading Hume on Human Understanding* (pp. 277–300). Oxford: Oxford University Press.

Dennett, D. (1984) *Elbow Room: the Varieties of Free Will Worth Having*. Oxford: Clarendon Press.

Frankfurt, H. (1969) "Alternate possibilities and moral responsibility," *Journal of Philosophy* 65, pp. 829–33.

Garrett, D. (1997) "Liberty and necessity," in *Cognition and Commitment in Hume's Philosophy* (pp. 118–36). Oxford: Oxford University Press.

Hume, D. (1987) *Essays: Moral, Political, and Literary*, ed. Eugene Miller. Indianapolis: Liberty Press.

Kenny, A. (1975) "Spontaneity, indifference, and ability," in *Will, Freedom and Power* (pp. 122–44). Oxford: Blackwell.

Penelhum, T. (1975) "The will," in *Hume* (pp. 111–30). London: Macmillan.

——(2000) "Hume and the freedom of the will," in *Themes in Hume: the Self, the Will, Religion* (pp. 156–76). Oxford: Oxford University Press.

Pitson, A. E. (2003) "Hume on morals and animals," *British Journal for the History of Philosophy* 11, pp. 639–55.

Russell, P. (1995) *Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility*. Oxford: Oxford University Press.

Stroud, B. (1977) "Action, reason and passion," in *Hume* (pp. 141–70). London: Routledge.

Further reading

Bricke, J. (1984) "Hume's volitions," in V. Hope (ed.) *Philosophers of the Scottish Enlightenment* (pp. 70–90). Edinburgh: Edinburgh University Press.

Clarke, R. (2003) "Freedom of the will," in S. Stich and T. Warfield (eds.) *The Blackwell Guide to Philosophy of Mind* (pp. 369–404). Oxford: Blackwell.

Davidson, D. (1980) "Freedom to act," in *Essays on Actions and Events* (pp. 63–81). Oxford: Oxford University Press.

Harris, J. (2005) "Hume's reconciling project," in *Of Liberty and Necessity* (pp. 64–87). Oxford: Oxford University Press.

Hobbes, T. (1991) "Of the liberty of subjects," in Richard Tuck (ed.) *Leviathan* (pp. 145–54). Cambridge: Cambridge University Press.

Jenkins, J. (1992) "Are our choices free?," in *Understanding Hume* (pp. 91–107). Edinburgh: Edinburgh University Press.

Kemp Smith, N. (1949) [1941] "Liberty and necessity," in *The Philosophy of David Hume* (pp. 433–41). London: Macmillan.

Mill, J. S. (1906) "Of liberty and necessity," in *System of Logic*, 8th edn. (pp. 547–52). London: Longmans.

Penelhum, T. (1998) Critical notice of P. Russell, *Freedom and Moral Sentiment*, *Canadian Journal of Philosophy* 28, pp. 81–94.

Russell, B. (1957) "The doctrine of free-will," in *Why I Am Not a Christian* (pp. 28–33). London: Allen & Unwin.

Strawson, P. (1974) "Freedom and resentment," in *Freedom and Resentment and Other Essays* (pp. 1–25). London: Methuen.

A useful selection of further readings on topics associated with the free will/determinism debate may be found at the Determinism and Freedom Philosophical Website http://www.ucl.ac.uk/~uctytho/dfwIntroIndex.htm edited by Ted Honderich. This site contains excerpts from Hume together with other historical and contemporary writings.

Part IV
Morals

13

Reason, Passion, and the Influencing Motives of the Will

Mikael M. Karlsson

The interplay of cognition and desire in motivation is a much-debated subject in contemporary philosophy, and like most philosophical subjects it is not new. David Hume's *Treatise of Human Nature* is widely considered to contain the quintessential formulation of a certain position on this subject, a position which is therefore styled "Humean." This view is *sentimentalist*, emphasizing the key role of the *passions* – in particular the sentiments of desire and aversion – in motivation. It opposes *rationalist* or *cognitivist* motivational accounts, according to which we may act upon considerations delivered by reason alone, unassisted by sentiment. Reason is here conceived as a power of cognition or understanding, disjoint from sentiment.

Motivational rationalists do not normally claim that actions *in general* are the offspring of purely rational motives, only that *certain sorts* of actions are. Specifically, among the kinds of actions that are commonly said by motivational rationalists to derive from reason alone are those that we call *moral* actions. When we act from motives of *duty* or *morality*, we act from purely rational motives, these thinkers claim. Moral rationalists – those who would ground moral judgments, moral motives, and moral action in human rationality – tend to consider sentiment untrustworthy as a source of morality, since our sentiments are all too easily swayed by adventitious factors, not in good accord with morality. Indeed, doing our moral duty seems to be often *opposed* by sentiment: what we are obligated to do is all too often not what we would *like* to do or would *enjoy* doing.

What Hume is thought to have offered in the *Treatise* was not so much a novel motivational account – for the basic view surely dates back to antiquity – but new kinds of arguments in its favor and against opposing views. The immediate targets of Hume's critical arguments were the views of the moral rationalists of his time,

including William Wollaston (1659–1724), Samuel Clarke (1675–1729), and John Balguy (1686–1748). More importantly, however, Hume's *Treatise* set a challenge to his younger contemporary, Immanuel Kant (1724–1804). In his *Grundlegung zur Metaphysik der Sitten* (1785), and other works, Kant attempted to show that Hume's arguments could be resisted and that the motivational account of the *Treatise*, as Kant understood it, could not be correct for *moral* action – if such action truly existed, which Kant was professedly unable to prove. Contemporary philosophers concerned with motivation, and especially those concerned with moral motivation, divide for the most part into Humeans and Kantians, although the influence of Aristotle (384–322 BC) is also widely felt.

Our excursion in this chapter will begin with an examination of the account of motivation for which Hume is renowned. After articulating this theory, we will consider for contrast a recent rationalist, or perhaps more properly, cognitivist, account due to Thomas Nagel (1937–) and John McDowell (1942–) and will consider some of the criticisms of the "Humean" view which this cognitivist account is supposed to meet. We will then consider an account that we will maintain is Hume's own and which is interestingly different from the "Humean" account. Hume's genuine account is not one that has been well appreciated in the philosophical literature on motivation, but it appears to have a number of important advantages over the theories which have received better press. In this connection, as in many others, Hume may be treated as a living philosopher with something fresh and vital to contribute to current debates.

I

The motivational theory for which Hume is famous goes something like this: human agents have various wants or desires. These passions turn the things or states of affairs which are the objects of our desire into objects to be sought. In connection with these desires, we also form judgments as to how the things desired may be obtained, or the wanted states of affairs realized, by means of actions that we might perform. Our repertoire of judgments may include some which indicate that the circumstances are presently ripe for acting so as to achieve a desired goal. The combination of the above-mentioned desires and cognitions constitutes a motive for an agent to act.

Now, Hume describes *reason* as "the discovery of truth and falsehood" (T 3.1.1.9; SBN 458); more properly, it is the *faculty* that we apply to the discovery of truth and falsehood. In the Humean account, reason, so understood, plays an important role in motivation. Among other things, "it discovers the connexion of causes and effects, so as to afford us means of exerting any passion" (T 3.1.1.12; SBN 459). However, Hume maintains that the deliverances of reason alone are not sufficient to move us to action. Reason, he declares, "is perfectly inert and can never either prevent or produce any action" (T 3.1.1.8; SBN 458). That is

because what reason delivers is *judgments concerning matters of fact* or *relations of ideas*; these are the sorts of judgments which have reference to the external or internal world and which are capable of truth and falsity:

> Truth or falsehood consists in an agreement or disagreement either to *real* relations of ideas or to *real* existence and matter of fact. What . . . is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. (T 3.1.1.9; SBN 458; cf. T 2.3.3.2; SBN 413)

But such a judgment as to what is or is not the case (or as to what could be made to be the case through action) cannot itself move us to act, for we might not want, desire, or care in any way about the objects or situations (actual or prospective) with which the judgment is concerned. If I judge that there is a warm, flakey, buttery *pain au chocolat* before me on the table, I am not thereby moved to action (let us say, to the action of eating this delicacy), for I may not be hungry or may not especially care for *pain au chocolat*. The wanting, desiring, or caring is an *additional* element, which is the *locus* of conation or motivation. Despite the fact that it works together with the deliverances of reason in the context of motivation, the passionate element is *distinct* from the cognitive elements, falling clearly outside of the domain of reason; for, as Hume puts it:

> A passion is an original existence . . . and contains not any representative quality, which renders it a copy of any other existence . . . (T 2.3.3.5; SBN 415)

> ’Tis impossible, therefore, [that our passions] can be pronounced either true or false, and be either contrary or conformable to reason. (T 3.1.1.9; SBN 458)

It is an important aspect of the motivational theory for which Hume is famous that the cognitive and sentimental components of our motives are distinct in this way, and that the cognitive elements – the deliverances of reason – are, by themselves, incapable of moving us to action: "Reason is wholly inactive," Hume says (T 3.1.1.10; SBN 458), and "alone can never produce any action" (T 2.3.3.4; SBN 414). Thus does Hume forestall the claims of motivational rationalism.

The account of motivation just rehearsed can be further deployed in opposition to the moral rationalism which claims that "virtue is nothing but a conformity to reason . . . [and that it is] possible, from reason alone, to distinguish betwixt moral good and evil" (T 3.1.1.4; SBN 456–7; cf. T 2.3.3.1; SBN 413). This cannot be correct, Hume argues, because:

> morality is always . . . supposed to influence our passions and actions, and to go beyond the calm and indolent judgments of the understanding. And this is confirm'd by common experience, which informs us that men are often govern'd by their duties, and are deter'd from some actions by the opinion of injustice, and impell'd to others by that of obligation. (T 3.1.1.5; SBN 457)

And:

> Since morals . . . have an influence on actions and affections, it follows [given the account of motivation], that they cannot be derived from reason . . . because reason alone . . . can never have such an influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. (T 3.1.1.6; SBN 457; cf. T 3.1.1.12; SBN 459)

II

Although, as we have said several times, Hume is famous for the account of motivation described in the previous section, a number of authors have argued that in certain important respects this account is not really Hume's (Kydd 1946; Árdal 1966; Norton 1982; Baier 1991; Radcliffe 1999; Karlsson 2000, 2001). I have thus referred to it, and will continue to refer to it, as a "Humean" account of motivation, indicating by the scare quotes that, although widely attributed to Hume, the account may actually diverge from his own.

As the passages we have quoted show, Hume *does* maintain that judgments, which are creatures of reason or the understanding, and passions, which are creatures of sentiment, are distinct sorts of entities, and that judgments alone, that is, without the influence of passion or desire, cannot move us to action. And on the basis of that, Hume *does* maintain that morality cannot be founded on reason alone. Morality incites to action, but moral motives cannot be purely rational, for there are no purely rational motives, or so Hume claims.

What is arguably not Humean in the "Humean" account is its representation of a desire for some object or state of affairs as the starting point in the motivational process, or at least as an essential constituent of that starting point. So far, the account seems to say that the motive for an action is made up out of two elements: a desire for some object or state of affairs (call it G) and a judgment or belief to the effect that by acting in a certain way (by doing A, a certain sort of action), G may be realized or achieved. A diagram may make this clearer; see Figure 1 (p. 239).

The boxed items in the diagram are the desire and judgment mentioned above. When these are conjoined, they induce an agent (here called S) to act. The downward fat arrow with the dots indicates a defeasible causal connection: S does not necessarily do A given the conjunction of the items in the box, for various other factors (e.g. that S has desires for conflicting goals) may intervene, or some partial element may be missing (e.g. the perception that the time is just now ripe for achieving G by doing A). But the idea is that the items in the box together dispose S to do A, and we may understand this as a kind of causal influence. Since the items in the box do have at least this kind of causal efficacy, they together constitute a motive for S to do A, for a *motive* (let us agree) is a subjective state which disposes an agent to a certain sort of action.

Figure 1   The "Humean" model: first approximation

> COGNITION
> *S* judges that by doing *A* she could realize or achieve *G*
> +
> DESIRE
> *S* desires to realize or achieve *G*

:
⇓

ACTION
S does *A*
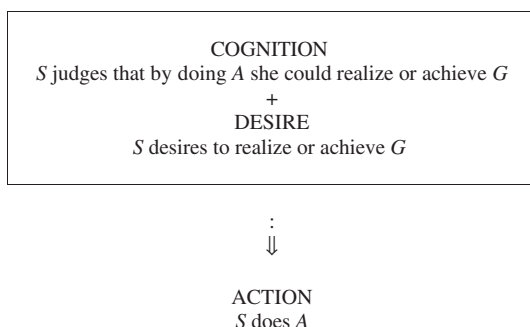
There is a certain difficulty with this "Humean" model which we must address before continuing: A "Humean" theorist is surely bound to think that *S*'s doing *A* will be causally connected with *S's wanting or desiring to do A*, but this desire is so far nowhere accommodated in the model. There would seem to be two ways of giving it an explicit place. One way would be to include it within the box, as deriving from *S*'s desire to realize or achieve *G* by what Thomas Nagel has called the "transmission of motivational influence across the relation between means and ends" (Nagel 1970: 33); see Figure 2.

Figure 2   The "Humean" model: first alternative

> COGNITION
> *S* judges that by doing *A* she could realize or achieve *G*
> +
> DESIRE
> *S* desires to realize or achieve *G* and thus desires to do *A*

:
⇓

ACTION
S does *A*
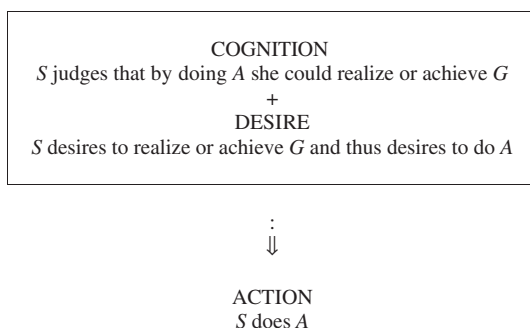
The exact character of this "transmission" is not clear, but it will not be further investigated here. Including *S*'s desire to do *A* within the box in this way will be understood to indicate that this desire is simply a specification of *S*'s desire to realize or achieve *G* in the light of her judgment that by doing *A* she could realize or achieve *G*. "*S* desires to do *A*" might be seen as a redescription of *S*'s desire to realize or achieve *G*, licensed by the judgment that *A* is a means to *G*. The upshot of this is that *S*'s motive for doing *A* is understood to be the same as in the first "Humean" diagram: it consists of the cognition and the passion (desire) originally specified.

Figure 3 The "Humean" model: second alternative

```
┌──────────────────────────────────────────────────┐
│                     COGNITION                      │
│ S judges that by doing A she could realize or achieve G │
│                         +                          │
│                      DESIRE                        │
│         S desires to realize or achieve G          │
└──────────────────────────────────────────────────┘

                          ↓

                       DESIRE
                   S desires to do A

                          ⋮
                          ⇓

                       ACTION
                      S does A
```
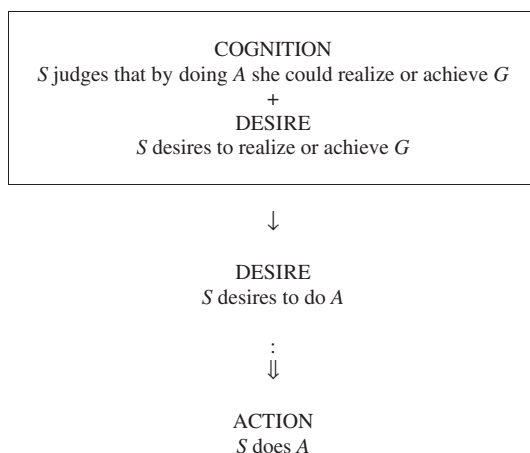
A second way of explicitly accommodating S's desire to do A within the "Humean" model removes it from the box; see Figure 3 above.

The relationship indicated by the thin downward arrow is left unspecified. That arrow indicates merely that *S*'s desire to do *A* derives in some way from the boxed items, but whether logically (and then, whether monotonically or non-monotonically), causally (and then, defeasibly or non-defeasibly), or otherwise we do not say here. The significance of exporting *S*'s desire to do *A* from the box is that of indicating that this desire, so described, is in some significant way distinct from *S*'s desire to realize or achieve *G* and is dependent upon the latter in the motivational process (or at least in the description of this process) as represented by this "Humean" model.

III

When *S*'s desire to do *A* is detached from the boxed elements, as in the latest diagram, that desire in itself appears as a motive for *S* to do *A*, and this raises some pertinent questions: first, are we still to understand the boxed items as constituting *S*'s motive to do *A*, even if she now has a distinct motive lying beyond the box and seemingly in closer causal proximity to her action? And also, given that *S*'s desire to do *A* is a more proximate impulse to her action than are the boxed items, why do we need to refer to the latter at all in our motivational story? Why, in other words, would a sentimentalist who subscribes to the second version of the "Humean" model (Figure 2) not be satisfied merely with the *minimalist* account, diagrammed in Figure 4 (p. 241)?

Figure 4 Minimalist model

```
┌─────────────────────────────┐
│           DESIRE            │
│       S desires to do A      │
└─────────────────────────────┘

              ⋮
              ⇓

             ACTION
            S does A
```
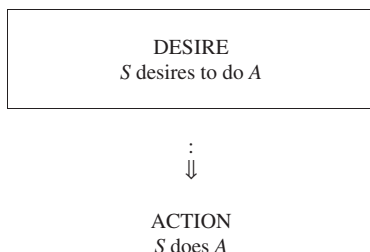
This last question presses upon us particularly in the light of the idea, to which many sentimentalists profess to subscribe, that desires are *unaccountable*, in the sense that no reasons can be given for or against them. They "simply assail us," as Thomas Nagel puts it (1970: 29). This idea of the unaccountability of desires is, indeed, commonly attributed to Hume on the basis of his denial, examined above, that a passion (such as a desire) could be "either contrary or conformable to reason" (T 3.1.1.9; SBN 457). "['T]is only in two senses that a passion can be call'd unreasonable," Hume says:

> First, when a passion . . . is founded on the supposition of the existence of objects, which do not really 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; SBN 416)

"And even then," Hume maintains, "'tis not the passion, properly speaking, which is unreasonable, but the judgment" (T 2.3.3.6; SBN 416). "Where a passion is neither founded upon false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it" (T 2.3.3.6; SBN 416).

We shall presently take a closer look at Hume's views on these matters. But for now let us notice that the "Humean" view is, at any rate, that *S's desire to do A*, so described, is accounted for, and needs to be accounted for, in terms of *S's* reasons, at least in the cases meant to be covered by the model. In those cases – which "Humean" theorists evidently suppose to be the widest range of cases to which it is interesting to apply a motivational theory – *S's* desire to do *A*, so described, is not unaccountable – not the sort of desire that "simply assails" the agent – but derives, in the account, from another desire (or at least a desire otherwise described), which is the desire to achieve a certain goal *G*, together with a judgment to the effect that doing *A* would be a means to that goal. And we should notice that these items together (the boxed items in Figure 3) constitute a reason that *S* has for desiring to do *A* (see the following paragraphs). The "Humean" theorist *may* be of the opinion that *S's desire to realize or achieve G*, so described, is unaccountable in terms of reasons; but this unaccountability does not pertain

to *S's desire to do A*, so described, for here doing *A* is desired as a means to realizing or achieving *G*.

But we may ask: Even if the "Humean" theorist thinks that *S*'s desire to do *A* is accountable in terms of reasons in the manner just described, why should he be concerned to *give* that account, given that *S*'s desire to do *A* may be understood *in itself* to be a motive for *S* to do *A* (particularly if he favors the second "Humean" diagram, Figure 3)? What the "Humean" motivational theorist evidently recognizes (if only implicitly) is that a motivational model rich enough to be of interest for the purposes of moral philosophy – or of practical philosophy more generally – needs to connect up an agent's motive for doing an action *A* with her *reason* for doing, or being disposed to do, *A*. It has often been urged that an account of reasons for action must show their connection with motivation (e.g. Nagel 1970), for, it is claimed, nothing can properly be called a reason for action that does not dispose an agent to act. But we may likewise insist that an account of motives for action – at any rate, for human action – must show the way in which those motives connect up an agent's reasons. A human action, as opposed to just any "act of a man" (to quote St Thomas), is something that an agent does for a reason. This is an understanding which I believe Hume shared and which most of us probably share. As motivational theorists, "Humean" or not, we are looking for an account of *being moved to act for a reason*.

Now if all we have to say in characterizing someone's motive for doing or attempting *A* is that she was disposed to do *A*, as in the minimalist model, we have perhaps named, but not illuminated her motive; for we have not described her motive in such a way as to reveal its connection with her reason for acting. When an agent is moved to do something for a reason, that reason will evidently illuminate her motive as well as her action. It will make plain on what grounds she is so moved. If *S* were merely assailed by a desire to do *A*, her only reason for doing, or attempting to do, *A* would be that "she just wanted to." This is not a very robust reason. It might even be said that she had *no* reason but rather "just wanted to," with no further explanation of why. I myself am very doubtful about there being very many real examples of things which we do for no other reason than that we just feel like doing them. And if there are such examples, where there is *really* nothing more to be said than that we wanted to, I think we might question whether the things thus done ought to be categorized as actions (i.e. human, or intentional, actions). But however this may be, it seems to me clear that the "Humean" motivational theorist seeks to catch within his net only those cases where there *is* something more to be said: something which connects up an agent's *motive* for acting with her *reason* for acting; and it is the boxed items in our diagrams which are indicative of such reasons.

"Humean" motivational theorists have tended not to unpack the "first approximation" model in the ways represented by the two "alternative" models, and have therefore tended to think of the boxed items in the "first approximation" model as both the agent's motive to act and the agent's reason to act; and they have thus

often neglected to distinguish analytically between an agent's reasons and her motives. If the "Humean" model is unpacked as in the first alternative, then the agent's motive and the agent's reason for acting are taken to be the same (even if analytically distinguished). If the model is unpacked as in the second alternative, then, as we have seen, the agent may be more easily represented as having a motive to act that is distinct from her reason for acting. But the boxed items will still constitute a motive to act, though in some sense distinct from the motive which is S's desire to do A, so described. As we saw, the latter motive, so described, is not very plausibly considered to be S's reason for doing A. The specification of S's reason for doing A, if it is thought to be given at all in this version of the model, is provided by the boxed items.

A final question is pertinent at this point: Why does *S's desire to realize or achieve G*, so described, not serve by itself as a motive to act (namely, as a motive to realize or achieve G)? It may surely be thought of as such, but it is an inchoate motive since it cannot produce action unless the agent knows (or thinks that she knows) *how*, that is, by means of what sort of behavior, G may be realized or achieved. It might be best to say, from the "Humean" point of view, that S's desire to realize or achieve G is the *motor* of her action, but not yet the *motive*. It is in combination with a judgment concerning means that the motor yields up a motive.

IV

The "Humean" account of motivation has been criticized by numerous authors, from contemporaries of Hume, such as Thomas Reid (1710–96), to our own contemporaries, such as Thomas Nagel and John McDowell. As we noted earlier, Immanuel Kant went to great lengths to try to circumvent it, and his influence persists in current discussions. What has particularly exercised the authors just mentioned (and many others) is the application of the "Humean" account to *moral* motivation – to the account that is to be given of the motives that underlie moral action, which are said to be motives of duty or morality. None of these authors has denied that the "Humean" model gives the right picture of a certain (and perhaps very wide) range of cases which lie outside of the moral sphere.

It is not our purpose here to consider the criticisms of the "Humean" model in any detail. Nevertheless, a limited consideration of the dissatisfaction of critics with the "Humean" model, along with a look at what sort of thing has been proposed as an alternative, will help us to understand the underlying issues and ultimately Hume's own views.

We will briefly canvass John McDowell's account, which builds upon some key ideas of Thomas Nagel and, more remotely, of Aristotle and Kant (cf. also Foot 1972a, 1972b). McDowell devotes his attention to the explanation of an agent's actions in terms of her *reasons*. But since McDowell believes that an agent's reasons must dispose her to action – specifically, to the action for which they are reasons

– to count as reasons at all, he makes little distinction between an agent's *reason* for acting and her *motive* for acting. Moreover, since the kind of explanation that McDowell evidently has in mind is *causal* explanation, he considers an agent's reasons, and her motives, as causes of her actions.

Now McDowell says:

> When we explain an action in terms of the agent's reasons, we credit him with psychological states given which we can see how doing what he did, or attempted, would have appeared to him in some favourable light. A full specification of a reason must make clear how the reason was capable of motivating; it must contain enough to reveal the favorable light in which the agent saw his projected action. (1978: 14–15)

In McDowell's view, the "Humean" model may in some cases – which McDowell (1982) calls the "Humean" cases – reveal this favorable light. For instance, if a man takes an umbrella with him when he leaves his home in the morning, we may explain this in terms of his reasons by saying that he judges that by taking his umbrella he can keep himself dry (let us grant that it is threatening to rain) and that he wants to keep himself dry. McDowell is willing to grant that this last-mentioned want or desire is "independently intelligible": a desire for something whose appeal is plain (even to those who might not happen to share it), thus revealing the favorable light in which the agent sees his action.

But McDowell argues that in the moral case (and also in prudential action), there is generally no such independently intelligible desire to appeal to and therefore no satisfactory account of reasons-motivated action that accords with the "Humean" model. Suppose, to take an example of my own (Karlsson 2001: 44–5) a certain person, John, judges that he could spare an old woman a lot of discomfort by giving up his seat on the bus. And let us suppose that, having noticed this, he does give up his seat on that account. The "Humean" model explains his action (and accounts for his reasons and his motive) by attributing to him, in addition to the judgment just mentioned, a desire to spare the old woman a lot of discomfort. Now McDowell's view is evidently that this desire is not patent. It is not a desire for something whose appeal is inherent, or plain to everyone, and appeal to it does not, therefore, in itself reveal the favorable light in which John sees his action. The desire is not "independently intelligible" and needs *itself* to be explained. To be sure, John is motivated to give up his seat and sees this action in a favorable light; but this is because John is *considerate* and sensitive to the needs of others. When he judges, therefore, that he can spare someone a lot of discomfort by acting in a particular way, he is motivated to act in that way. Being considerate involves seeing actions of this kind in a favorable light, finding them appealing just because they meet the needs of other people. McDowell maintains that seeing an action in this way is a *cognition*, and one that is not shared by inconsiderate persons. An inconsiderate person, Jane, may agree that by giving up

her seat, she could spare the old woman a lot of discomfort; but as an inconsiderate person, Jane does not thereby see (or perceive, or cognize) the action of giving up her seat as appealing; it does not appear to her in a favorable light. Thus, McDowell maintains, there is a *cognitive* difference between John and Jane; the difference between them is not merely one of sentiment.

John's *reason* for giving up his seat is his perception that by doing so he can spare the old woman a lot of discomfort; and this is also his *motive*, since he is inclined to give up his seat in the light of having that reason (that cognition). Following Thomas Nagel, McDowell would admit that since John is motivated to give up his seat by the cognition that by so doing he could spare the old woman a lot of discomfort we may attribute to him a desire to spare the old woman discomfort. However, this desire, though "present," is not efficacious in the motivational process. Rather, it is itself explained by the same factors that explain John's action (or at least his inclination to act in a particular way); it is, in Nagel's language, a "motivated" desire (Nagel 1970: 29–32; McDowell 1978: 14–16; Karlsson 2001: 32–5).

The view maintained by Nagel and McDowell concerning the nature and role of desires in this context is obscure in a number of respects (Karlsson 2001). But what they seem to be saying is that the kind of desire in question here can be attributed to John (and not to Jane) on the basis of the fact that John's cognition – that by giving up his seat he could spare the old woman a lot of discomfort – motivates him to give up his seat, whereas a nominally similar cognition does not move Jane. Probably the fairest account of their position is that the desire-attribution is little (if anything) more than a redescription of the aforementioned fact. It is thus incapable of explaining that fact. We cannot consider the desire thus attributed as an independently intelligible element which is causally efficacious in producing John's action or in moving him to act. The desire attribution is intelligible only in the light of the fact that certain considerations (cognitions) move John to act. For the moral case, at any rate, the "favorable light in which the agent saw his projected action" is therefore not revealed by reference to his desire to realize or achieve a certain goal; and it may be said that cognitivist motivational theorists like McDowell are in fact *generally* unimpressed with explanations of why an agent S was disposed to perform an action A which cite his desire to realize or achieve an objective G along with his thought that doing A would be a way of achieving G. We shall return to this point later.

McDowell's account of moral motivation is represented by what we may call the "cognitive model"; see Figure 5 (p. 246).

The boxed element in this figure is the S's reason for doing A (or for being so motivated) and is likewise S's motive; McDowell identifies reasons and motives. This cognitive motive *is* a motive in virtue of a certain condition of the agent: one which makes the cognition to be a motive and which also affects its character as a cognition in such a way as to put the action with which the cognition is concerned (doing A) in a favorable light. *Virtues* (like being considerate) are

Figure 5   The cognitive model
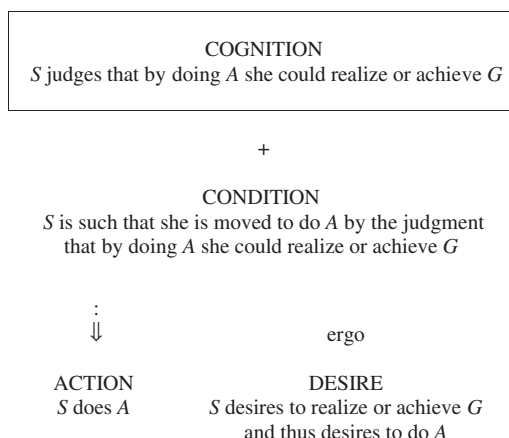
conditions of this kind, as McDowell emphasizes. Given such a condition, the boxed cognition comes to be both a cause and a rationalization for doing A; this is what is represented on the left side of the diagram. Absent such a condition, it would be neither. On the right side of the diagram we have the attributed desire to do A, which is not a part of the causal (motivational) or justificatory process and whose relation to the elements above it is (as here represented) a logical (one is tempted to say a "merely" logical) one (Nagel 1970: 29–30).

V

Let us now turn to what I believe is Hume's actual account of motivation and to a look at the way in which cognition and sentiment are connected therein. Confessedly, the texts are not transparent, and Hume's rhetoric is a serious hindrance to understanding what I believe to be the ultimate import of his theory. But a studied reading of the relevant portions of the *Treatise* (esp. 2.3.3; 2.3.9; and 3.1.1) appear to me to give a reasonably univocal picture.

   Hume's account of motivation cannot be disentangled from his idea about the good, and about the goods aimed at in our actions. For Hume, *good* and *evil* – perhaps we should rather say *the good* and *the bad* – are equated with *pleasure* and *pain* (T 2.1.1.4; SBN 276; T 3.1.1.1; SBN 399; T 2.3.9.1; SBN 438, T 2.3.9.8; SBN 439). A good is something that offers pleasure in prospect; I have elsewhere spoken of things having *pleasure in store* for a person (Karlsson 2000). Now Hume evidently maintains that a person forms a desire or aversion for something when, and only when, she judges that it has (or is likely to have) pleasure or pain in store for her (T 2.3.3.3; SBN 414; T 2.3.3.7; SBN 416–17, T 2.3.9.1; SBN 438; T

2.3.9.7; SBN 439; T 3.1.1.12; SBN 459), and this desire or aversion moves her to pursue or avoid the thing in question:

> 'Tis obvious, that when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity, and are carry'd to avoid or embrace what will give us this uneasiness or satisfaction. 'Tis also obvious, that this emotion rests not here, but making us cast our view on every side, comprehends whatever objects are connected with its original one by the relation of cause and effect. Here then reasoning takes place to discover this relation; and according as our reasoning varies, our actions receive a consequent variation . . . 'Tis from the prospect of pain and pleasure that the aversion or propensity arises toward any object: And these emotions extend themselves to the causes and effects of that object, as they are pointed out to us by reason and experience. (T 2.3.3.3; SBN 414)

Pain or pleasure in prospect are evidently the only basis upon which such desires or aversions are formed. I may suppose that something has pleasure in store for me and may therefore form the desire to "embrace" this (i.e. to realize the pleasure that this object has in store):

> [B]ut whenever you convince me of my mistake, my longing ceases. I may will the performance of certain actions as means of obtaining any desir'd good; but my willing of these actions is only secondary, and founded on the supposition, that they are causes of the propos'd effect; as soon as I discover the falsehood of that supposition, they must become indifferent to me. (T 2.3.3.7; SBN 417)

This view has been referred to as Hume's "psychological hedonism" (e.g. in Árdal 1966: 69–79; cf. Karlsson 2000: 16–21). Schematically, Hume's account may be represented as in Figure 6 (p. 248).

In that figure O is the "object" that is thought to have pleasure in store for the agent S. Let us note that this "object" may itself be an activity, such as hunting, gaming, or philosophizing rather than a non-activity with which we may enter into a relationship through the medium of action. Where the "object" is itself an activity, the relationship into which we enter with it for the sake of realizing pleasure would simply be to engage in it (T 2.3.10.7–10; SBN 451–2).

We may notice a certain complication in this latest diagram, which is that the second of the judgments included in the upper box presupposes the first; and this in turn complicates the relation of these elements to the sentimental elements in the lower box. In what follows, we will skirt the difficulties raised by this aspect of Hume's account, as our space here is insufficient to deal with them. Incidentally, just as with the second alternative "Humean" model, the relationship indicated by the thin downward arrow (here within the lower box) is left unspecified. That arrow indicates merely that S's desire to do A derives in some way from her desire to realize or achieve the pleasure that O has in store for her, but whether logically, causally, or otherwise we do not say here.

Figure 6 Hume's model

```
┌─────────────────────────────────────────────────┐
│                    COGNITION                       │
│    S judges that O has pleasure in store for her   │
│                         +                          │
│  S judges that by doing A she could realize the    │
│      pleasure that O has in store for her          │
└─────────────────────────────────────────────────┘

                         ⋮
                         ⇓

┌─────────────────────────────────────────────────┐
│                     DESIRE                         │
│  S desires to realize the pleasure that O has in   │
│                 store for her                      │
│                         ↓                          │
│                 S desires to do A                  │
└─────────────────────────────────────────────────┘

                         ⋮
                         ⇓

                      ACTION
                      S does A
```
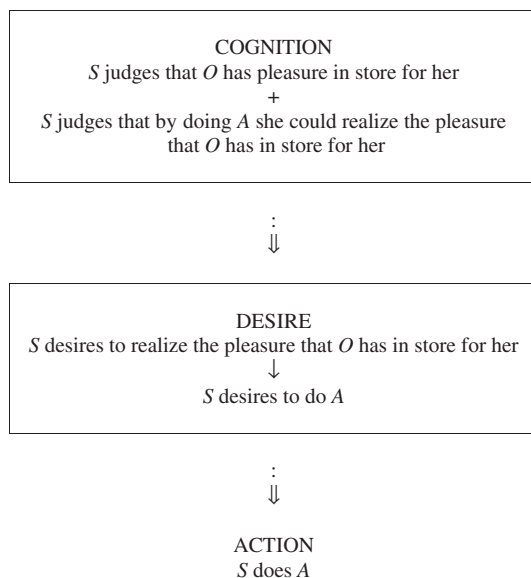
Now, in this account, the starting point of the motivational process is a *cognition*, or rather a pair of cognitions: one to the effect that something has pleasure in store and the other to the effect that by acting in a certain way, this pleasure might be realized. These cognitions are, in Hume's terms, creatures of reason, capable of truth or falsity:

> A person may be affected with passion by supposing a pain or pleasure to lie in an object, which has no tendency to produce either of these sensations, or which produces the contrary to what is imagin'd. A person may take false measures for the attaining of his end, and may retard, by his foolish conduct, instead of forwarding the execution of any project. These false judgments . . . extend not beyond a matter of *fact* . . . (T 3.1.1.12; SBN 459; italics in original)

As creatures of reason, the cognitions in question are presumably "inert." However, Hume's view is evidently that these cognitions *cause* us to form a desire to act in some way that we suppose will enable us to obtain the pleasure that an object has in store. It is thus very misleading of Hume to insist so emphatically upon the inertia of reason. A large part of his point seems to be that it is a contingent fact that, in sentimental beings such as ourselves, the cognitions in question lead to the formation of certain passions. If we were purely rational beings, we could have these same cognitions without forming any motives whatever: the deliverances of reason are not *inherently* (but only, in the context of human nature,

contingently) motivating. It is on that account that he emphasizes that "reason *alone* can never produce any action" (T 2.3.3.4; SBN 414; emphasis added).

Motives are creatures of our sentimental nature, which Hume seems to think of as *responding* to certain deliverances of the understanding. Motives are not produced by, nor do they reside within, our reason or understanding. For, as Hume has set things out, that which is within the domain of the understanding is that which can be true or false, and such is not the case with the passions. Thus it is within the domain of sentiment that motives reside, and it is sentiment that is, in that sense, the source of action. The passions are inherently, not merely contingently, motivating; and in that sense sentiment is active, while reason is inert.

In insisting upon this, Hume evidently sees himself as forestalling all talk of "the combat of passion and reason" (T 2.3.3.1; SBN 413) which, he supposes, rests upon the presumption of there being *two* active powers, reason and passion, which could issue contradictory motivational impulses. Hume's argument is that there is only one *independent* active power, the power of sentiment, and any motivational influence exerted by the understanding must work through the medium of sentiment. And that being the case, Hume thinks, we cannot imagine that reason could ever "oppose passion in the direction of the will" (T 2.3.3.1; SBN 413).

However, Hume's exaggerated emphasis on this point obscures what is arguably a much more important aspect of his own account. For although it may be a merely contingent fact that humans are sentimental beings, with sentiments that respond to the deliverances of reason, once this fact is taken into account, cognitions are seen as *causes* of motives and (at a slightly further remove) actions, and are in that important sense *active*. It is thus not well said by Hume that "[r]eason is, and ought only to be the slave of the passions" (T 2.3.3.4; SBN 415). Even while asserting that reason is inert and inactive, Hume details its causal *influence*:

> [R]eason . . . can have an influence on our conduct . . . after two ways: Either when it *excites a passion* by informing us of the existence of something which is a proper object of it; or when it discovers the connection of causes and effects, so as to afford us means of exerting any passion. (T 3.1.1.12; SBN 459; emphasis added)

It is in fact just the two sorts of cognitions mentioned here by Hume which ground the motivational process in his account, as shown above in the schematic picture of Hume's model (the items in the upper box, Figure 6).

Is Hume then, after all, a motivational rationalist? Is his a *cognitive* account of motivation? Strictly speaking, no, since Hume assigns desires a necessary role in motivation: cognition without passion cannot move us to action. However, like the motivational cognitivists, and in contrast to "Humean" accounts, Hume grounds the motivational process in cognition rather than desire. According to his

account, certain kinds of cognitions dispose an agent to action of a particular sort; and we earlier agreed that motives are subjective states which dispose an agent to a certain sort of action. Even so, Hume would decline to call these cognitions *motives*, for their causal role must be mediated by sentiment. What he would identify as the motives for action are the passions – specifically, the desires – "excited" by those cognitions and which inherently dispose the agent to act. In the schematic picture of Hume's account given above this passionate element is indicated by the lower box.

If the lower box in Figure 6 is indicative of the agent's *motive*, what are we to say of the cognitions in the upper box? I think that in Hume's picture the items in the upper box constitute the agent's *reason* for acting, although he does not quite say this. But if we understand matters in that way, then in contrast to McDowell's cognitive motivation model, and in contrast to the first approximation "Humean" model (and to the general tenor of "Humean" motivational theory), Hume's account distinguishes between the reasons and motives for an action, and shows at the same time how these reasons and motives are connected up.

Strikingly, Hume's account meets the motivational cognitivist's complaint that reference to an agent's desire to realize or achieve an objective G fails to reveal the "favorable light" in which she saw her projected action, and hence to explain her disposition to perform such an action (wherefore the cognitivist denies that this desire could be, or found, her reason for acting). For motivational cognitivists (or at least Nagel and McDowell) maintain, as we saw, that the agent's desire to realize or achieve the objective G is as much in need of explanation as her disposition to do A (at least in the case of moral action); and they maintain moreover, that the explanation of the desire and the disposition are the *same*, deriving from the agent's reason for acting (or being disposed to act), which reason, they maintain, is a certain cognition. Now, on the account here attributed to Hume, he agrees with all this. If it be asked, "Why (i.e. for what reason) did S do A?" Hume's answer would *not* be "in order to achieve G, which she desired, in the belief that she could achieve G by doing A." Rather Hume would answer, "in order to realize the pleasure that something (as she thought) had in store for her, in the belief that by doing A she might realize that pleasure." Such an explanation reveals (though not in the way imagined by McDowell!) the favorable light in which the agent saw her action: she saw it as a way of obtaining pleasure.

If we are to appreciate the power of Hume's account, we must remind ourselves at this point that Hume maintains that *pleasure* and *good* (and likewise *pain* and *evil*) are identical (T 2.1.1.4; SBN 276; T 3.1.1.1; SBN 399; T 2.3.9.1; SBN 438; T 2.3.9.8; SBN 439). Just what this thesis comes to is unclear, and we cannot belabor it here; but despite its unclarity, it is of great importance in the present context. For, in the light of this thesis, Hume's account of reasons and motives represents the agent as acting in order to realize some *good* (or to avoid some evil). And this means that an agent's reasons for acting *justify* (or would, if true, justify) her action: putting the action in a favorable light is seeing it as a means to

obtaining pleasure, but this is to see it as a means to obtaining some good. On Hume's account, the fact that *S* desires to realize or achieve some objective *G* cannot be represented (as in the "Humean" account) as her reason for being disposed to do *A*, which she supposes would be a way of achieving *G*; for this desire does not justify doing *A*, since no connection has yet been made with her thereby achieving some good. To put it another way, *S*'s desire to realize or achieve *G* remains itself to be justified as a desire for the achievement of some good. Referring to Figure 6, representing Hume's account, we can say that (assuming the identity of *pleasure* and *good*) the items in the upper box (the agent's *reason*) provide a justification of the items in the lower box (the agent's *motive*), as well as justifying the action thus motivated.

But the account makes a further important claim, namely that (*contingently*, in the human animal) the cognitions which provide justifying reasons for certain passionate motives and for the actions thus motivated, *cause* (or as Hume puts it, "excite") us to form those motives and thereby bring about those actions. Contingently, in other words, the so-called "justifying reasons" for our actions are likewise "motivating reasons" (which Hume would, however, decline to call "motives"). It has usually been thought that, with his tirade on the "inertia" and "inactivity" of reason, Hume excludes this possibility. But we maintain here that such is not the case. Hume is only saying that a creature might be such as to be unmoved by the judgment that a certain sort of action would bring it some good, even while understanding that that action would thereby be justified. Human nature happens to be such that we *are* moved by such considerations.

One might want to resist Hume's identification of *good* with *pleasure*. But even if one rejected that identification, one might nevertheless admire what I call the "Hume-ish master model," of which Hume's own model is an instantiation; see Figure 7 (p. 252).

We noted above the extent to which Hume's account was in agreement with the cognitivist account of reasons and motives offered by Nagel and McDowell. But his account also diverges from theirs in important respects, and I find Hume's account superior in several respects. This may be easier to appreciate by looking at the Hume-ish master model than by looking at Hume's own model.

Nagel and McDowell maintain that what explains (in terms of her reasons) both an agent's disposition to do *A* and her desire to realize or achieve a certain objective *G* is her judgment that by doing *A* she could realize or achieve *G*. A cognition falling under this latter description will, however, only serve as a reason for an agent who is in a suitable condition (e.g. who has certain virtues). For an agent in such a condition, this cognition puts doing *A* in a favorable light. For other agents, a cognition falling under the same description will not be the same cognition and will very likely not put doing *A* in a favorable light. Now, on this way of putting things, the justifying function of the agent's reason is not well brought out, for there is no explicit connection made between the agent's reason and acting for the good. This is just what *is* perfectly explicit in the Hume-ish model

Figure 7   Hume-ish master model

```
┌─────────────────────────────────────────────┐
│                  COGNITION                    │
│      S judges that O has good in store for her│
│                      +                        │
│      S judges that by doing A she could realize│
│        the good that O has in store for her   │
└─────────────────────────────────────────────┘

                       ⋮
                       ⇓

┌─────────────────────────────────────────────┐
│                   DESIRE                      │
│  S desires to realize the good that O has in store for her │
│                      ↓                        │
│              S desires to do A                │
└─────────────────────────────────────────────┘

                       ⋮
                       ⇓

                    ACTION
                    S does A
```
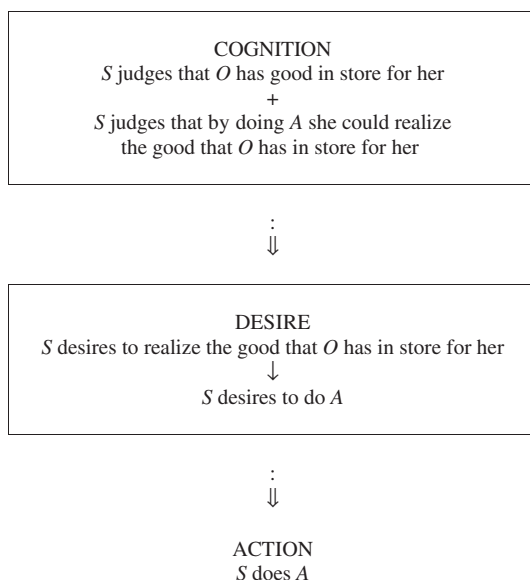
(Figure 7); for the agent's reason for performing an action A (and for being moved to do this) is that she judges some G to be a good, *together* with the judgment that by doing A she could realize or achieve G. In the account given by Nagel and McDowell, the latter judgment is the agent's whole reason. In the Hume-ish model, "S judges G to be a good" seems to me to represent precisely the element that is missing from the cognitive model, and from the "Humean" model. In the end, McDowell seems to me to maintain that the favorable light in which the agent saw her prospective action is not revealed by citing the agent's reason according to the McDowell formula, unless the addressee of such a reasons-explanation shares the virtues or other relevant characteristics of the agent herself. It is true that if an explanation is given according to the Hume-ish master model, the addressee of that explanation will not appreciate the favorable light unless he grasps why the objective G which the agent judged to be a good was so adjudged. It therefore matters what an admirer of a Hume-ish account has to say about the good. Hume himself assumes that the appeal of *pleasure* is transparent, and this is an assumption that he shares with many. That fact notwithstanding, a critic might find reason to reject Hume's equation of the good with pleasure while continuing to admire the virtues of the Hume-ish master model.

VI

Let us complete our excursion by summing up the ways in which Hume's account of motivation seems to come out better than either the "Humean" or McDowell-

type cognitive accounts. First, it distinguishes clearly, as any such account ought to do, between *reasons* and *motives* for action. An agent's reason justifies her action in the sense of making clear why the action is worth doing. An agent's motive is the impulse which brings an action about, or at least disposes the agent to act in a certain way. These items may be ontologically the same, but they must at least be distinguished analytically. In Hume's account, they are ontologically distinct, which – whether or not it be thought correct – ensures the analytical distinction.

Second, Hume's account of an agent's reason for acting is arguably better than what can be derived from either the "Humean" or the cognitivist accounts. The "Humean" account suggests that an agent's reason for acting consists in her wanting something and seeing a way to obtain it through action. Critics like McDowell complain that in certain cases, this does not make clear why the action is worth doing; for surely one's merely wanting something does not make it worth pursuing. On the other hand, what McDowell offers fails equally to show how the agent's reason connects up with anything that she considers to be good. In the first lines of his *Nicomachean Ethics*, Aristotle makes the fundamental point that every human action or pursuit aims at some good (or something thought to be good) and that this is what justifies any action. And that is what Hume's account tells us: an agent's reason for acting is that she thinks that the action will realize some good.

Third, Hume offers us a clear picture of the relationship between an agent's reason for acting and her motive. Her reason relates to her motive both through practical rationality, showing why the motive is worth having and the action worth doing, and also causally: the reason "excites" the motive. If anyone thinks that this is the wrong account, they must argue for an alternative which is at least as clear.

Fourth, in contrast with Kantian approaches, which (using this label loosely) include the cognitive account offered by Nagel and McDowell, Hume offers a single account of motivation, which covers moral motivation as a special case. For Kant and his descendents, moral motivation works differently from motivation in other domains and requires a separate account. Thus we find McDowell maintaining that we must distinguish between "Humean" and "non-Humean" domains of action. McDowell argues that motivation in moral action does not submit to a "Humean" analysis but requires a cognitive account. However, McDowell is content to apply the "Humean" model to much of the remainder of human action. Hume himself evidently denies that the "Humean" analysis applies anywhere!

Moreover, although moral action comes in for special treatment by Hume, it falls nevertheless under the motivational model that applies to every kind of action. The key point about this is evidently that in all cases, according to Hume, an agent's reason for action would have to be that by acting in a certain way she believed that she could realize some good. Hume's way of explaining what is distinctive about moral motivation involves the implication of the moral sentiments in the motivational process (see 14: HUME'S ARTIFICIAL AND

NATURAL VIRTUES, pp. 257–8) for a discussion of the moral sentiments). That is too long a story to include here; but it is surely an advantage of Hume's account that moral motivation is represented as structurally similar to motivation in other domains.

In sum, Hume's account of the interplay between reason and passion in motivation constitutes a philosophically interesting alternative to the accounts which have received more discussion. It is an alternative that warrants a great deal of further attention and exploration.

References

Árdal, P. (1966) *Passion and Value in Hume's* Treatise. Edinburgh: Edinburgh University Press. Rev. edn. 1989.

Aristotle (1908) *Nicomachean Ethics*, trans. W. D. Ross. Oxford: Clarendon Press.

Baier, A. C. (1991) *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge, MA: Harvard University Press.

Foot, P. (1972a) "Morality as a system of hypothetical imperatives," *Philosophical Review* 81(3), pp. 305–16.

——(1972b) "Reasons for action and desires," *Proceedings of the Aristotelian Society*, Suppl. 46, pp. 203–10.

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten* [Groundlaying for the metaphysics of morals], vol. IV. Riga: Akademie-Ausgabe.

Karlsson, M. (2000) "Rational ends: Humean and non-Humean considerations," *Sats – Nordic Journal of Philosophy* 1(2), pp. 15–47.

——(2001) "Cognition, desire and motivation: 'Humean' and 'non-Humean' considerations," *Sats – Nordic Journal of Philosophy* 2(2), pp. 30–58.

McDowell, J. (1978). "Are moral requirements hypothetical imperatives?," *Proceedings of the Aristotelian Society*, Suppl. 52, pp. 13–29.

——(1982) "Reason and action – III," *Philosophical Investigations* 5(4), pp. 301–5.

Kydd, R. (1946) *Reason and Conduct in Hume's* Treatise. London: Oxford University Press.

Nagel, T. (1970) *The Possibility of Altruism*. Oxford: Oxford University Press.

Norton, D. F. (1982) *David Hume: Common-sense Moralist, Skeptical Metaphysician*. Princeton, NJ: Princeton University Press.

Radcliffe, E. S. (1999) "Why beliefs alone never motivate: Hume on the generation of motives," *Hume Studies* 25(1–2), pp. 101–22.

Further reading

Brown, C. (1988) "Is Hume an internalist?," *Journal of the History of Philosophy* 26, pp. 69–87.

Cohon, R. (1993) "Internalism about reasons for action," *Pacific Philosophical Quarterly* 74, pp. 256–88.

——(1997) "Is Hume a noncognitivist in the motivation argument?," *Philosophical Studies* 85, pp. 251–66.

——(2000) "The roots of reasons," *Philosophical Review* 109(1), pp. 63–85.

——(ed.) (2001) *Hume: Moral and Political Philosophy*. Aldershot: Ashgate/Dartmouth.

Darwall, S. (1995) *The British Moralists and the Internal 'Ought' 1640–1740*. Cambridge: Cambridge University Press.

Garrett, D. (1997) *Cognition and Commitment in Hume's Philosophy*. Oxford: Oxford University Press.

Hampton, J. (1995) "Does Hume have an instrumental conception of practical reason?," *Hume Studies* 21(1), pp. 57–74.

Kemp Smith, N. (1941) *The Philosophy of David Hume*. London: Macmillan.

Mackie, J. L. (1980) *Hume's Moral Theory*. London: Routledge.

Magri, T. (1996) "Natural obligation and normative motivation in Hume's *Treatise*," *Hume Studies* 22(2), pp. 231–53.

McDowell, J. (1979) "Virtue and reason," *The Monist* 62(3), pp. 331–50.

Millgram, E. (1995) "Was Hume a Humean?," *Hume Studies* 21(1), pp. 75–93.

Norton, D. F. (1984) "Hume's moral ontology," *Hume Studies* (special issue), pp. 189–214.

Persson, I. (1997) "Hume – not a 'Humean' about motivation," *History of Philosophy Quarterly* 14(2), pp. 189–206.

Radcliffe, E. S. (1994a) "Hume on motivating sentiments, the general point of view, and the inculcation of 'morality'," *Hume Studies* 20(1), pp. 37–58.

——(1994b) "Hume on passion, pleasure, and the reasonableness of ends," *Southwest Philosophy Review* 10, pp. 1–11.

——(1996) "How does the Humean sense of duty motivate?," *Journal of the History of Philosophy* 34(3), pp. 383–407.

——(1997) "Kantian tunes on a Humean instrument: why Hume is not *really* a skeptic about practical reasoning," *Canadian Journal of Philosophy* 27(2), pp. 247–69.

——(forthcoming) "The Humean theory of motivation and its critics," in E. S. Radcliffe (ed.) *A Companion to Hume*. Oxford: Blackwell.

Smith, M. (1994) *The Moral Problem*. Oxford: Blackwell.

Snare, F. (1991) *Morals, Motivation and Convention*. Cambridge: Cambridge University Press.

Stroud, B. (1977) *Hume*. London: Routledge.

——(1993) "Gilding or 'staining' the world with 'sentiments' and 'phantasms'," *Hume Studies* 19(2), pp. 253–72.

Sturgeon, N. (2001) "Moral skepticism and moral naturalism in Hume's *Treatise*," *Hume Studies* 21(1), pp. 3–83.

——(forthcoming) "Hume on reason and passion," in D. Ainslie (ed.) *Critical Essays on Hume's* Treatise. Cambridge: Cambridge University Press.

Svavarsdóttir, S. (1999) "Moral cognitivism and motivation," *Philosophical Review* 108(2), pp. 161–219.

14

Hume's Artificial and Natural Virtues

Rachel Cohon

Hume's analysis of the virtues and vices, which occupies most of Book 3 of the *Treatise*, is original and provocative. He believes he has shown in 3.1.1–2 that

"An action, or sentiment, or character is virtuous or vicious; why? because its view causes a pleasure or uneasiness of a peculiar kind."

He infers, therefore, that

"In giving a reason . . . for the pleasure or uneasiness, we sufficiently explain the vice or virtue." (3.1.2.3; SBN 471)

Since feelings of approval (the peculiar pleasure) are how we, as observers, know a character is laudable, uncovering whatever it is that causes a trait to evoke these sentiments both proves that the trait is laudable and explains why it is. Thus Hume apparently believes he can offer, for each admirable character trait, an argument that is both causal and normative at once. That is, he gives an argument showing a trait has such and such features that cause near-universal human approval, and thereby he both explains the mechanism by which we approve it and demonstrates that – since we approve it – it is indeed a virtue. His analyses contain some surprises. The main themes of his theory of the virtues in the *Treatise* are (1) his distinction between natural and artificial virtues, and (2) the role of sympathy in explaining why traits count as virtues and vices.

Hume prefers the virtue-oriented approach to ethics of the ancient Greek and Hellenistic philosophers to the duty- and rule-oriented approach of moralists of his own day and more recent past. Even by present-day standards Hume is a virtue ethicist. For example, he endorses a version of the view that the primary object of moral evaluation is a trait of character, and he understands a good action as what the virtuous agent would do. But there is a striking difference between Hume's

virtue theory and those of Aristotle and present-day neo-Aristotelians. Aristotelians define the virtues as those character traits that play a constitutive role in the flourishing human life of those who possess them; this is their criterion of virtue. Hume, by contrast, defines a virtue as any quality of the mind that evokes the feeling of approval in an observer when it is contemplated in an unbiased way (T 3.1.2.3–4; SBN 471–2; T 3.3.1.3; SBN 574–5; T 3.3.1.15–17; SBN 581–3). Hume gives no explicit account of the good or flourishing human life, and whatever implicit account he may have (for example at T 3.3.6.6; SBN 620–1) does not serve as a criterion for identifying the virtues. The question whether the virtuous person is happy is, for him, a matter for empirical investigation.

In his examination of the various virtues Hume takes a new position on a question that divides his predecessors, whether morality is natural or conventional ("artificial"). To say a virtue or duty is *natural* is usually to say two things: that the character trait or action itself is not something socially invented but is a (perhaps refined) manifestation of a familiar feature of human nature, and that the trait or action's goodness (its being a virtue or a duty) is not the result of inculcated social custom but rather is solely a consequence of something in the nature of things (whether in the rational order of the universe, the divine will, or human nature). Hobbes and Mandeville see morals as conventional, while Hutcheson, Locke, and others see them as natural. Hume is the only one to take an intermediate position: some important virtues are natural, but others, equally important, are socially invented. He mocks Mandeville's contention (Mandeville 1988: 41–57) that the very concepts of vice and virtue are foisted on the public by scheming politicians for the purpose of managing us more easily. If there were nothing in our experience to give rise to the concept of virtue, Hume says, no lavish praise of heroes could generate it (T 3.2.2.25; SBN 500). But while natural features of the human mind account for some of the traits we have and their classification as virtues, they cannot account entirely for others.

The Moral Sentiments

At the point at which Hume begins his investigation of the virtues and vices, he has argued that all our moral evaluations of persons and their character traits arise from our sentiments. The virtues and vices are exactly those traits that produce approval and disapproval, respectively, in whoever contemplates the trait – whether its possessor or another – in a disinterested fashion, without regard to self-interest. Approval (approbation) is a pleasure and disapproval (disapprobation) is a pain or uneasiness. Approval and disapproval are "passions" or "affections" (emotions, as we would say today) that feel different from other psychic pleasures and pains, and they are caused only by the contemplation of someone's quality of mind.

For Hume the basic object of moral evaluation is a quality of mind, a character trait. Thus the typical moral judgment is that some trait, such as a particular

person's benevolence or laziness, is a virtue or vice. When we observe someone's *action* we take it as a sign of the inner quality that must have produced it, and derive our moral evaluation of the action from our assessment of that inner quality. For example, if we disapprove a man's neglect of his children, we condemn it because we take it to be the manifestation of a "character" (trait) of parental indifference or self-indulgence; and we change our evaluation of the act if we learn that the man was devoted to his children but was prevented from caring for them by some obstacle (T 3.2.1.3, 5; SBN 477, 478). A character trait, for Hume, is a psychological disposition to feel a certain sentiment or combination of sentiments, ones that are typically motivating. Therefore, given Hume's sentiment-based account of moral judgment, a moral judgment is an observer's feeling of moral approval or disapproval elicited by the observed person's disposition to have certain motivating sentiments.

Sympathy

In Book 2 Hume describes a psychological mechanism he calls sympathy, which proves crucial to his accounts of virtue and approbation. Sympathy is the propensity we have "to receive by communication [the] inclinations and sentiments [of others], however different from, or even contrary to, our own" (T 2.1.11.2; SBN 316). People's emotions are contagious. Hume explains this phenomenon in terms of his theory of the mind, whose contents he divides into *impressions* (vivid, lively perceptions such as sensory impressions and such feelings as love or anger) and *ideas* (the faint, much less lively copies of impressions which are used in thought). Sympathy operates as follows. First, when I observe the outward expressions of a person's passion in his "countenance and conversation," this brings into my mind the idea of his passion. For example, seeing him clench his fists and shout causes me to believe he is angry. Observing the typical cause of a passion also conveys its idea into the mind: if we see the instruments laid out for someone's surgery, they evoke ideas in us of fear and pain. Hume claims that at all times each of us possesses a maximally vivid and forceful impression of himself. According to Hume's associationism (developed in Book 1 and used throughout the *Treatise*), the liveliness of one perception will be automatically transferred to those others in the mind that are related to it by resemblance, contiguity, or cause and effect. Now, all human beings, regardless of their many differences, are similar in basic bodily features and functions, and in a parallel way are similar in their psychic features, including their passions. So any person I contemplate resembles me in having a human body and mind subject to the same characteristic responses as my own. And some persons may further resemble me in some more specific feature such as age or nationality. Because of the resemblance between us and my nearness to the observed person, some of the great liveliness of the impression of myself is transmitted to my idea of his sentiment. The sole difference between an idea of a

passion and an actual passion is the degree of liveliness or vivacity each possesses. So great is this newly-acquired liveliness that the idea of his passion in my mind (the mere thought of it) becomes an impression, and I actually experience the passion.

The Distinction between Natural and Artificial Virtues

Hume observes that "our sense of every kind of virtue is not natural; but . . . there are some virtues, that produce pleasure and approbation by means of an artifice or contrivance, which arises from the circumstances and necessities of mankind" (T 3.2.1.1; SBN 477). He will argue not only that our approval of certain characters is socially engineered, but also that the traits themselves could not exist without cooperative human invention. He divides all virtues into the natural and the artificial, and gives separate accounts of the two kinds.

Hume apparently supposes that there are certain characteristic sentiments (including impulses, desires, and capacities for love and hatred) that are simply part of human nature, just as it is in the nature of animals of various other species to be timid or ferocious, loyal or indifferent, according to their kind. If there were human beings who belonged to no society but only cooperated in small family groups, they would have these natural sentiments. (However, Hume does not claim a priori or revealed knowledge of which sentiments these are; he gives arguments to show that some of the sentiments familiar to us that may *seem* part of the original equipment of the species could not be so; T. 3.2.1.9; SBN 470–80; T 3.2.5.2–5; SBN 516–18; T 3.2.6.9; SBN 531–3.) The natural virtues are more refined and completed forms of those natural human sentiments. They include the virtues of attachment and devotion to particular individuals, and so tend to exhibit partiality (T 3.2.2.8; SBN 488–9). (It is a virtue, for example, to be devoted to one's child; it is no virtue to be indifferent as between one's own child and all other children.) The artificial virtues are, by contrast, traits we need for successful *im*personal cooperation. Hume argues that our natural sentiments are too partial to give rise to these on their own, without intervention. There are just a few, carefully described artificial virtues: honesty with respect to property (often called equity or "justice," though this is a strangely narrow use of that term), fidelity to promises (sometimes also listed under "justice"), the (less stringent) sort of international justice appropriate to heads of state, allegiance to one's government, chastity (refraining from non-marital sex), and modesty (both primarily for women and girls), and good manners. A great number of individual character traits are listed as natural virtues, but the main types discussed in detail are greatness of mind ("a hearty pride, or self-esteem, if well-concealed and well-founded," T 3.3.2.11; SBN 598), goodness or benevolence (an umbrella category covering generosity, gratitude, friendship, and more), and such natural

abilities as prudence and wit, which, Hume argues, have a good claim to be ranked as moral virtues though traditionally they are not.

A further difference Hume observes between the natural and artificial virtues is that the particular actions manifesting the natural virtues directly benefit someone on each occasion when they are performed, whereas a particular action manifesting one of the artificial virtues may not benefit anyone, although it contributes to a systematic practice that is highly beneficial to all. Each generous act, for example, benefits its recipient; but a particular act of repaying a loan (an act of justice) might impoverish the debtor while restoring to a miser wealth he will never enjoy. Hume thinks that nonetheless we need people who are disposed to strict, exceptionless adherence to the rules of property and promise if we are to maintain any sort of society; and without society it would be very difficult to survive. (In his later *Enquiry concerning the Principles of Morals*, however, Hume drops the distinction between artificial and natural virtues.)

The analysis of each of the artificial virtues has the same basic structure. Hume first shows that a particular familiar virtue is artificial and not natural, either by means of an argument (where this is a controversial view of the virtue in question) or by pointing out the obviousness of this fact (where it is widely accepted or has been well defended by others). Then he gives an account of why human beings would want to invent the virtue in question (what problem it solves for human beings), and how the relevant convention might have come into being. The two virtues of honesty with regard to property ("justice") and fidelity to promises and contracts have only rarely been thought artificial; even Hobbes, who regards property as artificial, gives credence to the idea that in the state of nature human beings have a duty to keep their covenants (e.g., *Leviathan*, 1991: 102–3). Consequently it is in some detail that Hume argues that these two virtues are artificial, not natural. (By contrast, he thinks it too obvious to require defense that the feminine tendency to recoil from improper sexual advances is not a naturally-occurring trait.) Hume argues for the artificiality of honesty and fidelity by means of a *reductio* in each case, showing that when we suppose that the character trait exists and wins our approval without help from a collaborative social arrangement, insoluble paradoxes arise; hence we must conclude that the supposition is false, and although we did not previously suspect it, some sort of social engineering has occurred. He then goes on to explain the point of the relevant convention and to sketch a plausible story of how it might have formed.

Each sketch of the genesis of an artificial virtue is an account of how free, unsubordinated, un-coerced individuals could come to develop the phenomena that he describes: a sense of honor about property and promise-keeping, respect for international justice, allegiance to rulers, a disposition to chastity, deferential manners. Even in the invention of government, though subordination of some people to the authority of others is the result, the process is carried out by individuals who act freely and initially have no masters – indeed, among whom authority is unknown. Hume does not argue that there never is any

coercion involved in these developments, and historically perhaps there was. He simply describes entirely non-coercive mechanisms that could in principle produce these results. (This is true for the adults. Children, who are inculcated with new feelings of pride and shame that uphold the new conventions, are not offered a choice.) These voluntary steps ultimately lead to profound psychological changes in the members of the societies formed by means of the conventions.

Below we take up Hume's accounts of two of the artificial virtues, honesty and fidelity.

Honesty with Respect to Property

The circle

Hume gives a cryptic argument to show that our approval of material honesty must be the product of convention. We have seen that all virtuous actions derive their goodness solely from virtuous motives – from motivating passions in the agent's "mind and temper" of which we independently approve. For example, if someone cooks a meal for a sick friend, I approve the action because I believe it was motivated by his desire for his friend's good. (If he did it solely to curry favor I would judge the action quite differently.) It follows from this that the motive that originally "bestows a merit on any action" can never itself be moral approval of the action (awareness of its virtue), but must be a non-moral feeling or impulse, one distinct from "the regard to [the action's] virtue" (T 3.2.1.4; SBN 478). For if the virtue-imparting motive of the action *were* the agent's sense of the action's virtue – if that were why he did it, and why we approved it – then we would be reasoning in a vicious definitional circle. It would be impossible to specify what we approve (what makes the action virtuous), or what the agent's motive is. We know that any approval of the action is approval of the agent's motive. And this motive (by hypothesis) would be approval of the action. But such approval itself can only be approval of the action's motive. So the agent is moved to act by his own approval of what moves him to act. But what is that?

For every virtue, therefore, there must be in human nature some non-moral motive (some motive other than the moral sentiment) that characteristically motivates actions expressive of that virtue; and that motive, by eliciting our approval, makes actions so motivated virtuous.

However, there is no morally approved (and so virtue-bestowing) non-moral motive of honest action. The only approved, reliable motive that we can find that manifests the virtue of equity is a moral one, the sense of virtue or "regard to the honesty" of actions. The honest individual repays a loan not (merely) out of self-interest, nor (merely) out of concern for the well-being of the lender (who may be a "profligate debauchee" who will reap only harm from his possessions), nor from concern for the public good (which we rarely think of when we pay our

bills), but from a "regard to justice, and abhorrence of villainy and knavery" (T 3.2.1.9, 13; SBN 479–80, 482). This appeal to the motives of moral approval and disapproval, however, is "evident sophistry and reasoning in a circle," as we saw above. Now, nature cannot have "establish'd a sophistry, and render'd it necessary and unavoidable." Therefore "the sense of justice and injustice is not deriv'd from nature, but arises artificially . . . from education, and human conventions" (T 3.2.1.17; SBN 483). The logic of this part of the argument is not perfectly clear, but Hume's intent is to show that if we imagine honesty to be a natural virtue we commit ourselves to a sophistry, and therefore honesty is instead man-made.

Hume's account of the genesis of the social convention that on his view creates honesty with respect to property is meant to cope with the circularity. We will see how it may do so.

The origin of material honesty

Hume poses two questions about the rules of ownership of property and the associated virtue of material honesty: what is the artifice by which human beings create them, and why do we attribute moral goodness and evil to the observance and neglect of these rules?

By nature human beings have many desires but are individually ill-equipped with natural strength, bodily weapons, or instinctive skills to satisfy them. We can overcome these deficiencies by social cooperation: combining our strength with others, using division of labor to acquire skills, and helping one another in times of weakness. People obtain the idea of cooperation from their experience with the small family groups into which they are born; seeing how advantageous it is for family members to work together, they try to implement this with strangers, to achieve greater material well-being. However, with non-intimates, factors in nature (including human nature) cause people to come into conflict. Both natural and produced goods are moderately scarce relative to what people want, and the goods we want are easy to take from us. We are selfish and have only "confined generosity" (generosity to those dear to us in preference to others). So we will quarrel over who gets what, and collaboration will disintegrate. No remedy for this is to be found in "our natural uncultivated ideas of morality" (T 3.2.2.8; SBN 489), for our natural ideas of good and evil endorse loyalty to loved ones over strangers. An invention is needed.

Hume argues that we create the rules of ownership of property originally in order to satisfy our avidity for possessions for ourselves and our loved ones, by linking material goods securely to individuals to avoid conflict and maintain our lucrative cooperation. Within small groups of cooperators, individuals signal to one another a willingness to conform to a simple rule: to refrain from the material goods that others obtain by labor or good fortune provided those others will

observe the same restraint toward them. (This rule in time requires more detail: specific rules determining who may enjoy which goods initially and how goods may be transferred by consent.) This signaling is not a promise, but an expression of conditional intention: I will not interfere with objects in your possession if you will not interfere with those in mine. The usefulness of such a custom is so obvious, once proposed, that others will soon catch on and express a similar intention, and the rest will fall in line. The convention develops tacitly, as do conventions of language and money. When someone in such a small society violates this rule, for example by taking what is designated as another's, the rest know it and exclude the offender from their cooperative activities. Once the convention is in place, justice (of this sort) is defined as conformity with the convention, injustice as violation of it. Indeed, the convention defines property rights, ownership, financial obligation, theft, and related concepts, which had no application before the convention was formed. It is so easy to devise and appreciate this invention that human beings would not live for long in isolated family groups; their ingenuity would quickly enable them to invent property so as to reap the substantial economic benefits of cooperation with non-intimates. With reliable possession of the products of cooperation, all could better satisfy their natural greed.

Greed, and more broadly self-interest, is the motive for inventing property; but we need a further explanation why justice (adherence to the rules of ownership) is virtuous and injustice is vicious. Hume argues that once the convention is present, there are forces at work that moralize it. As a society grows larger, people may cease to see their own property violations as a threat to the continued existence of a stable economic community, and this reduces their motivation to conform. But when they consider violations by others, they partake by sympathy with the resulting uneasiness of their victims and all of society. Similarly they all feel a sympathetic pleasure on contemplating the public benefits of adherence. Such disinterested uneasiness and pleasure are instances of moral disapproval and approval. We extend this to our own behavior, because human sentiments operate in large part according to general rules. We cannot help but disapprove our own dishonest acts. This natural process of moralization, however, is "forwarded by the artifice of politicians" (T 3.2.2.25; SBN 500), who assist nature by encouraging esteem for justice and abhorrence of injustice in order to govern more easily. Parents and other educators contribute to this further artifice as well. Seeing that having the character of probity will benefit their children and make them useful to society, parents shape the "tender minds" of the young to regard the observance of the rules of equity as "worthy and honourable, and their violation as base and infamous" in each and every instance (T 3.2.2.26; SBN 500–1). Thus material honesty becomes a virtue.

Hume's genetic account of property makes no patriarchal assumptions about the nature of the family, and it explicitly denies that the creation of ownership depends on any promise or contract. Its concept of convention is of an informal

practice of mutual compromise for mutual advantage that arises incrementally and entirely informally, without the use of central authority or force. The process of moralization likewise proceeds in small, spontaneous steps. (On the originality of this account, see Baier 1988.)

The motive of material honesty

Does this account resolve the circularity problem? Does it specify any non-moral motive of honest action? Interpretations vary. The reading that seems to fit the text best (Cohon 1997 and forthcoming) is that while Hume does think there is a characteristic motive of honest action, a motive other than (even enlightened) greed, he relaxes the requirement that the approved motive of every virtuous type of action must be some natural impulse or feeling distinct from moral approval. Only natural virtues are subject to this requirement, because only for them is it impossible to identify good actions initially without recourse to a motive. In the development of the convention of property, instead of relying on an approved motive to identify the good actions, we first rely on the convention to do so. This permits us to escape the circle. Then, in the process by which the convention is moralized, parents and politicians create a motivating form of the moral sentiment itself. Approval of honesty and disapproval of "villainy and knavery" are enhanced by their trained associations with pride and shame so that they become motivating. Not all feelings of approval are strong enough to move us to act, but some are, as Hume argues in T 3.1.1; and among these is the approval of honesty that is inculcated into the young once the convention of property is in place. That process creates a trait of character. Observers in turn approve this (moral) motive to action (the "sense of honesty"), and we further approve actions derivatively because they are motivated by it. Thus honesty has an approved, characteristic motive and is a virtue. (For readings that differ in various ways from this one, see Penelhum 1975: 154–5; Mackie 1980: 79–80; Haakonssen 1981: ch. 2; Harrison 1981: esp. p. 8; Baron 1982; Gauthier 1992; and Darwall 1995: 309–11.)

Fidelity to Promises

Fidelity is the virtue of being disposed to keep promises and contracts. Hume has in mind not exchanges of favors between friends, but the sorts of promises involved in commerce and other self-interested transactions (especially those between strangers or mere acquaintances). As with material honesty, paradoxes arise if we attempt to analyze the virtue as a non-conventional (natural) one. The same circularity puzzle arises about the approved motive of fidelity that Hume tackles at length in connection with honesty; but with fidelity he pays more attention to a different conundrum. Unlike Hobbes and Locke, who help themselves to the

concept of a promise or contract in their imagined state of nature, Hume argues in the *Treatise* that the performative utterance "I promise" would be unintelligible in the absence of background social conventions, and that the moral obligation of a promise depends on such conventions as well.

Suppose the practice of giving and receiving promises did not depend on a socially defined convention. In that case, what could we mean by the utterances we use to make promises, and how does our saying these words obligate us to fulfill them? First, a hearer only understands that a promise is made if it is apparent that the speaker grasps the meaning of the words he uses, in particular as purporting to obligate him. If the speaker does not understand, or merely speaks in jest, or otherwise clearly does not use the words in a way that purports to reveal his will, we do not understand a promise as really being made. Thus for effective use there must be some act of the speaker's mind that he expresses by the special phrase "I promise" and its synonyms, and his moral obligation results from this act of the mind. (This seems to be Hobbes's assumption in *Leviathan* ch. xiv, 1991: 67, where the implicit signs of covenant – as distinct from the explicit ones – are clear signs of the person's will.) What mental act or state is expressed? It could not be either the desire or resolution to act, since it does not follow from our desiring or resolving to act that we are morally obligated to do so; we can resolve to do something and change our mind later without wrongdoing. Nor could it be the volition to *act*, since that only comes into being when the time comes to act, not ahead of time. And of course, one can promise successfully (incur obligation by promising) even though one has no intention to perform; a lying promise is still morally binding. So the mental act that incurs obligation is not the intention to do the promised action either. The only likely act of mind that might be expressed in a promise is a mental act of *willing to be obligated* to perform the promised action. This conforms to our common view that we bind ourselves by choosing to be bound.

But the thesis that one can actually bring an obligation into existence by willing to be obligated is absurd. What makes an action obligatory is that its omission is disapproved by unbiased observers. If a promisor were to make some action obligatory by an act of will within her mind, that inner volition would have to cause all observers (including the agent), who before the promise was made had no special feelings about the act or its omission, henceforth to feel moral disapproval of its omission by the agent. But no inner act of will can cause *that*; sentiments are not subject to such voluntary control. Thus, there is no such act of the mind. Hence one cannot express such an act in words, and others cannot understand one to do so. Even if people in their natural (pre-conventional) condition "cou'd perceive each other's thoughts by intuition," they could not understand one another to bind themselves by any act of promising, and could not be obligated thereby. Thus the necessary condition for a natural obligation of promises cannot be fulfilled. We may conclude that this obligation is instead the product of group invention to serve the interests of society.

The origin of promises

Once property is invented, possession is stable and goods are exchanged by consent. But as the convention of property has been described so far, it only permits simultaneous swapping of visible commodities. Everyone would profit greatly if people could be trusted to provide goods or services later in exchange for benefits given now, and could exchange goods that are distant (a house a mile away) or described generically ("a bushel of corn"). But for people without the capacity to obligate themselves to future action, such exchanges could only happen if, once one party performed, the second were to perform in her turn out of gratitude alone, which would rarely occur.

Moralists and politicians cannot alter human selfishness and ingratitude, but they can help us learn a better way to satisfy our appetites "in an oblique and artificial manner" (T 3.2.5.9; SBN 521). First, people will easily see that additional kinds of exchanges would serve their interests. As soon as they express this interest to one another, everyone will be moved to invent and to keep agreements. To invent them, they first create a form of words to mark these new exchanges (and distinguish them from the generous reciprocal acts of friendship and gratitude). When someone utters this form of words, he is understood to express a resolution to do the action in question, and he "subjects himself to the penalty of never being trusted again in case of failure" (T 3.2.5.10; SBN 522). This penalty is made possible by the practice of the group, who simply refuse to make new contracts with those who have broken their word in the past. This "concert or convention" (ibid.) in turn transforms human motives to act. A person is moved by self-interest to give the promising sign (in order to obtain the other party's cooperation), and once she has given it, self-interest demands that she do what she promised to do so as to ensure that people will exchange promises with her in the future.

Some interpreters (Haakonssen 1981; Gauthier 1992) say that this enlightened self-interest remains the only motive for keeping one's promise. But Hume says the sentiment of morals comes to play the same role in promise-keeping that it does in the development of honesty with respect to property (T 3.2.5.12; SBN 523); so it seems the moral sentiment not only becomes "annex'd" to promise-keeping but further motivates it (Cohon 2004 and forthcoming). In larger, more anonymous communities, the fear of exclusion is no longer strong enough to motivate scrupulous adherence to contracts; one's past defections might not be known to all. A sentiment of moral approval of promise-keeping arises as the result of sympathy with all who benefit from the practice. This is aided by a second artifice, the well-meaning psychological manipulation of the people by parents and politicians, which yields a near-universal admiration of fidelity and shame at breaking one's word (T 3.2.2.25–6; SBN 500–1). This provides a moral motive for promise-keeping even in transactions that will not affect one's reputation.

The puzzle resolved

In reality, promises obligate not because they express some unique self-obligating act of the speaker's mind, but because the speaker makes them against the background of a convention, and everyone disapproves defection from the conventional rules. Obligatory action is simply action whose omission is disapproved by observers, and breaking a promise is certainly disapproved by observers who share in this convention. All that we understand the speaker to express when she says "I promise" is that she resolves to act. But the convention makes it possible for her to bind herself morally by these words.

The Natural Virtues and Sympathy

In addition to the various artificial virtues (only two of which have been discussed here), Hume identifies a great number of natural virtues, among them a due degree of well-concealed pride, generosity, compassion, gratitude, friendship, clemency, constancy, and industry. The list differs in different parts of the text. Nature is parsimonious, however, and Hume hints early in Book 3 that our approvals of all the many virtues are unlikely to stem from a multitude of separate instincts. In the *Treatise* Hume's principle interest in the natural virtues lies in explaining the root causes of our approval of them all and the corresponding disapproval of the natural vices. The mechanism of sympathy ultimately accounts for this. Sympathy also explains our approval of the artificial virtues, as we have seen. But in order to trace our moral judgments of the natural virtues and vices to an origin in sympathy in a way he finds satisfactory, Hume finds he needs to expand his theory of sympathy.

When I come to share in the affections of strangers and feel pleasure because they are pleased, as I do when I experience aesthetic enjoyment of a well-designed ship or fertile field that is not my own, my pleasure can only be caused by sympathy (T 3.3.2.7–8; SBN 596–7). Similarly, Hume argues, when we reflect on a mental quality knowing its tendency either to produce enjoyment or uneasiness for strangers, we come to feel enjoyment when the trait is beneficial or agreeable to them and uneasiness when it is harmful or disagreeable to them. This human response to the tendency of a character trait to affect the sentiments of people with whom we have no affectionate ties can only be explained by sympathy.

We greatly approve the artificial virtues, which have been shown to be inventions contrived solely for the interest of society. Interestingly, we approve them in all times and places, even where our own interest is not at stake. For example, we admire the justice and allegiance of figures in history and in distant lands. We do so only because we realize that these traits are likely causes of benefit to the whole society of that time or place. But approval of such causes of benefit to others can only be the product of sympathy. The moral approbation we feel toward these

traits of character, then, is the pleasure generated by sympathy with the people of their societies. We find the character traits – the causes – agreeable because they are means to an end (social well-being) that we find agreeable as a result of sympathy.

When it comes to traits we approve naturally (without any social contrivance), our approval of them and hence their status as virtues can also be explained as resulting from sympathy with the pleasure of those who receive benefit from them.

Hume never worries that different individuals within a society (or even across societies) will disagree in their moral assessments of such traits as justice or fidelity to promises. He apparently expects complete unanimity of judgment about the artificial virtues. The reason for this seems to be that approval of the artificial virtues is the effect of sympathy with the whole of society. In any given case, all observers occupy the same synoptic perspective when they sympathize, and so experience the same (degree of) pleasure or uneasiness in the same instances. There is of course the danger that an observer may confound her selfish interests with her sympathetic concern for society; but if she has attained sufficient discernment to distinguish these, her moral evaluation of any artificial virtue or vice will match that of any other discriminating individual. However, while some of the natural virtues (such as beneficence) benefit society generally, many of them do not improve the lot of society as a whole, so it cannot be sympathy with all of society that generates our approval of them. Many individual character traits, such as friendship or diligence, benefit only a few people: often only their possessor's close associates or even just their possessor himself. Yet we do approve them. So if our approval of these is to originate in sympathy, it must be sympathy with particular individuals. And this introduces a potential difficulty. For our moral evaluations of these virtues tend to be fairly consistent with one another, yet our emotional responses to individuals are quite varied.

The Common Point of View

The sympathetic transmission of sentiments between individuals can vary in effectiveness depending on how much the two people resemble one another and how close or far apart they are in space. I receive the sentiments of someone very much like me or very near me much more powerfully than I do those of someone unlike me or far away. Yet people's *moral* assessments do not vary depending upon whether the person they evaluate resembles them in language or temperament, or is close or remote. We do not regard a disloyal friend in our own town as more vicious than one in a distant land just because the former is contiguous with us. Indeed, our moral assessments of people remain stable even though we draw closer or move farther away from them over time. Any theory that traces the origin of moral judgment to subjective emotional reactions to persons or their actions must explain the fact that moral evaluations made by one individual at different times

and many individuals in a community tend to be fairly uniform. To do so, Hume claims that people do not make their moral judgments from their own individual points of view, but instead select "some common point of view, from which they might survey their object, and which might cause it to appear the same to all of them" (T 3.3.1.30; SBN 591). With respect to the natural virtues and vices, this common point of view is composed of the intimate perspectives of the individuals who have direct interactions with the person being judged. To make a moral evaluation I must sympathize with each of these persons in their dealings with her; the blame or admiration I feel as a result of this imaginative exercise is my genuine moral assessment of the subject's character. In that assessment I also overlook the small external accidents of fortune that might keep the trait from producing its usual beneficial or hurtful effect; instead I acquire by sympathy the pleasure or uneasiness that I imagine people *would* feel were the trait to operate as it ordinarily does. "Experience soon teaches us this method of correcting our sentiments, or at least, of correcting our language, where the sentiments are more stubborn and inalterable" (T 3.3.1.16; SBN 582).

Specific Natural Virtues

Hume's natural virtues fall into three main categories: the virtues of greatness of mind, which are outgrowths of the natural sentiment of pride; the virtues of goodness or benevolence, which are outgrowths of the natural sentiment of love (Hume uses "love" very broadly to include esteem, respect, and compassion as well as more intimate feelings); and the natural abilities, which are either true talents such as intelligence and humor or outgrowths of the natural sentiment of calm self-interest, such as application. Below we consider the first two types. In explaining our approval of virtues of the first type, Hume introduces further complexities into the sympathy-based account. His explanation of the second type of virtue works more smoothly.

Virtues based on pride and humility

Hume observes that different proportions and manifestations of the passions of pride and humility (or shame) are virtuous or vicious. An obvious and "overweaning conceit" is disapproved by any observer (is a vice), and "modesty," meaning an accurate assessment of our own weakness, is "esteem'd virtuous" (T 3.3.2.1; SBN 592). But it turns out that valuing oneself on the basis of one's genuine strengths is a virtue.

To explain these opposite assessments of closely related dispositions, Hume describes an interaction between sympathy and a distinct psychological mechanism he calls comparison. That mechanism compares a sympathetically communicated sentiment with the observer's inherent feeling in his own case, causing the observer

to feel a sentiment opposite to the one she observes in the other (pleasure if the other is suffering, pain if the other is pleased). For example, when I see a storm-tossed ship in the distance while I am safe on shore, sympathy brings me distress at the passengers' plight; but if that borrowed distress is not too great, I feel its comparison with my own sense of security, and this gives me a pleasure in my relative safety. When someone displays excessive pride (out of proportion to his true merit), others, via sympathy, come to feel some of his pleasant sentiment of pride (without believing in its objects), but this draws their attention to their own lesser feeling of pride, and they suffer uneasiness by comparison. This uneasiness is moral disapproval of the other's conceit, and thus conceit is a vice.

The operation of comparison depends on the observer's idea of the other's feeling being just vivid enough and not too vivid. If it is too faint, I will not feel any contrasting emotion by comparison. If it is extremely vivid, sympathy will eliminate comparison and I will forget myself in the other person's transmitted emotion. This is why pride based on a correct assessment of one's merits does not irritate others in the same way that unfounded pride does. Self-esteem founded on an accurate assessment of one's strengths is agreeable to oneself and indispensable for using one's capacities to good effect, and observers can share in these resulting pleasures via sympathy. The passion transmitted to the observer is more vivid because it is "accompanied with [the observer's] belief" in the person's merit (T 3.3.2.6; SBN 595), and so sympathy prevails and comparison does not take hold. Thus while Christians say they admire abject humility more than well-founded self-esteem, it follows from Hume's account that in practice they do not actually do so. (On Hume's defiance of the Christian view of pride, see Baier 1991: 205–7.) Although excessive pride is a natural vice and both modesty and self-esteem are natural virtues, in society people create the *artificial* virtue of good breeding (adherence to customs of mutual deference in accordance with social rank) to enable them each to conceal their own pride so that it does not shock the pride of others. We can never know whether our own pride is well-founded or excessive, so we hide it behind a superficially humble exterior. Genuine humility, though, is no part of our duty.

Courage and military heroism are also forms of pride. Though the student of history can see that military ambition has mostly been disadvantageous to human society, laying waste provinces, when we contemplate the "dazzling" character of the hero, immediate sympathy with his pride irresistibly leads us to approve it (T 3.3.2.15; SBN 601).

Virtues based on love

The virtues that are manifestations and refinements of the passion of love (or esteem or tenderness) are largely virtues of attentiveness toward intimates, although they can be extended to strangers whose plight touches us. Given natural human selfishness, we cannot expect most people's concerns to extend much beyond the

"narrow circle" of family, friends, and other close associates (T 3.3.2.2; SBN 602). But that leaves plenty of opportunity for them to manifest virtues. Our approval of those traits that may be grouped together under the heading of goodness and benevolence, such as generosity, humanity, compassion, and gratitude, arises from sympathy with the people close to the individual whose trait we judge. When we imaginatively adopt the common point of view, we correct for the local distortions of sympathy by entering into the feelings of those close to the person even if those intimate associates are remote from us. The joy a trait produces in them becomes our own approval of the trait, and moves us to admire or love the trait's possessor. The sufferings a trait produces have the analogous, reverse effect.

Conclusion: Sympathy and Justification

Hume concludes that the approbation of nearly all the virtues can be traced to sympathy. Sympathy alone accounts for our approval of honesty, fidelity to promises, and the other artificial virtues; and also for our approval of those natural traits that conduce to the benefit of society, of a person's intimates, and of the person herself. All these virtues show themselves to be causes of human advantage or pleasure, and all the corresponding vices show themselves to be causes of harm or pain. The happiness and sorrow of strangers affects us only by sympathy. Thus almost all moral judgment is caused by sympathy. (The sole small exception is our approval of traits immediately agreeable to the observer.) And if Hume means to make a normative argument as well, he has shown that sympathy is what *makes* traits virtues and vices. Sympathy is the foundation of morals.

Hume ends the *Treatise* by observing that if we reflect on the way that sympathy with the joys and sorrows of others gives rise to (nearly) all our moral judgments, it is clear that the phenomenon of moral evaluation arises out of an inborn concern we all have with human well-being. And this concern is admirable. The moral sentiments are derived from a noble source – one that evokes further approval. We approve the sense of virtue itself.

Hume has told us which traits we judge to be virtues and vices – which traits we do approve and disapprove. And he has explained what causes those specialized feelings of pleasure and pain in each case. But one must wonder whether he has also argued (and whether successfully, or not) that the traits we feel to be virtues really *are* virtues. Besides describing our moral evaluations and their sources, has he attempted to justify them? Perhaps he has; or perhaps he has in effect denied that any justification is possible beyond our favorable and unfavorable feelings. He leaves us with a tantalizing ambiguity: whether all evaluations or norms are on his view reducible to approval and disapproval, or whether instead there is some standard by which our approvals (our moral judgments themselves) can be justified or corrected. Besides what we do approve, is there anything further he can say about what we should approve?

There are signs in the text that Hume sees pleasure and pain as the ultimate foundation of norms. Since moral judgments simply are approvals and disapprovals evoked in a particular way, it seems that Hume acknowledges no further appeal to justification. The pleasure of approbation is what makes a trait a virtue, period. So whatever we approve is virtue. *Should* we approve as we do? The only sense that can be made of this question, it seems, is to construe it as asking whether we approve approving as we do. And perhaps this is why Hume points out at the end of Book 3 that when we reflect on our moral sentiments and the causes that trigger them, we approve the whole package. He realizes that one can wonder whether the various traits that trigger our approval should do so; and he finds that they should, in that the causal process he has described itself evokes approval. (See Baier 1991: 196–7, 277; Korsgaard 1996: 49–66.)

But if this is his view, then his theory allows for only the most minimal criticisms of current moral opinion, and no real criticism of moral judgments themselves. One can say, truly, that an observer mistakes her selfish feeling for a moral judgment, or that she fails to take up the common point of view and so responds in a parochial or local way rather than in a way that is moral. One could even say this of all the members of a social class or society. But these are cases not of making a mistaken moral judgment but of confusing a non-moral evaluation with a moral one. Hume could say of such an observer that she thinks she judges something a virtue when in reality she does not judge it a virtue but rather finds it personally advantageous or appealing from her own point of view; but he cannot say that she incorrectly judges it a virtue. Even a moral evaluation based on incorrect factual beliefs, such as a false but widespread belief that a neutral trait or practice is harmful to society, cannot be rejected as erroneous. The factual belief is itself an error, of course, but those who believe it and reflect on the trait or practice from the common point of view come to feel a genuinely moral sentiment of disapproval, since it has the right phenomenological quality and is caused by sympathy with those who (the observers suppose) are affected by the trait. If suitably generated sentiment is the sole standard of virtue and vice, then the trait or practice is indeed vicious. Hume would of course predict that were the false belief to be corrected, observers would come to approve the practice instead. But if the sole standard of virtue is (properly-triggered) sentiment, all he can say about this change is that what was a vice before ceases to be one after the factual error is corrected. On this way of understanding the foundation of all evaluation in feelings, one cannot say that a trait we all approve is not really a virtue, because, for example, it is sexist, class-biased, or frivolous. If we approve it, and our approval is generated by contemplation of it from the common point of view, then it is a virtue. This is disappointing to the reader who wishes to say that some widely held moral opinions within his own society are mistaken, such as a double standard of chastity for men and women, the more relaxed sense of justice thought suitable to princes but not to private citizens, or duplicity concerning one's own degree of self-esteem, all traits approved by Hume's contemporaries (or so he argues) and

approval of which is caused in just the right way. (Hume discusses each of these, though not all are mentioned above.) According to this interpretation, it would be simply incoherent to say that people's general disapproval of a woman's flirting, for example, is unjustified. (This interpretation and the following one are touched on in Cohon 1997.)

There is room, however, for an alternative interpretation. Hume also shows signs of thinking that it is possible for our approval of a trait or disposition to be warranted or unwarranted, and this may be his actual position. Approval of a trait is warranted provided the trait has some systematic causal connection to the social good, although that connection may be quite "oblique," as it is for the artificial virtues. (A trait may cause the good of society directly, by alleviating the misery of individuals within it; or it may contribute to a practice which, because it is generally followed, contributes to the social good, although a particular individual's trait may motivate acts that do not alleviate anyone's misery but even cause some. In the latter case our approval of the trait is nonetheless justified given the causal link between widespread possession of the trait and social stability.) If this is his view, then approvals based on false beliefs about the effects of a trait are unjustified, for the trait does not cause the social good that observers think it causes. So also are approvals of traits on the basis of too narrow a perspective – one limited to the observer's own gender, class, or race, for example – for if one is thinking of too few people one will misjudge the effects of a trait on society as a whole. One may, of course, still confuse moral with non-moral evaluations; but there is the further possibility of saying that the approval itself is moral but mistaken. There is, on this interpretation, the possibility of a divergence between what we judge to be a virtue or vice and what we should so judge. The trouble with this reading, however, is that it brings Hume close to abandoning his effort to make sentiment – even adjusted sentiment – the foundation of moral norms, and shifts him toward a proto-utilitarianism of character traits. What makes a trait really a virtue now, it seems, is not its emotional impact on the observer but its benefit to society. This is far from classical utilitarianism in many respects (it is still not a maximizing theory, and it is character-based rather than action-based), but nonetheless it is in tension with the explicit normative sentimentalism of the *Treatise*.

The text of the *Treatise* is ambiguous as between these two interpretations (though it somewhat favors the first). This ambiguity runs right through Hume's ethical writings. Which view is most consistent with all that he says will determine, for many, whether Hume's theory of the virtues and vices is ultimately acceptable.

References

Baier, A. (1988) "Hume's account of social artifice: its origins and originality," *Ethics* 98, pp. 757–78.

——(1991) *A Progress of Sentiments: Reflections on Hume's* Treatise. Cambridge, MA: Harvard University Press.

Baron, M. (1982) "Hume's noble lie: an account of his artificial values," *Canadian Journal of Philosophy* 12, pp. 539–55.

Cohon, R. (1997) "Hume's difficulty with the virtue of honesty," *Hume Studies* 23, pp. 91–112.

——(2004) "Hume: moral philosophy," in E. Zalta (ed.) *Stanford Encyclopedia of Philosophy* (Winter 2004 edition), http://plato.stanford.edu/archives/win2004/entries/hume-moral/

——(forthcoming) "Hume on promises and the peculiar act of the mind," *Journal of the History of Philosophy.*

Darwall, S. (1995) *British Moralists and the Internal 'Ought' 1640–1740.* Cambridge: Cambridge University Press.

Gauthier, D. (1992) "Artificial virtues and the sensible knave," *Hume Studies* 18, pp. 401–27.

Haakonssen, K. (1981) *Science of a Legislator: the Natural Jurisprudence of David Hume and Adam Smith.* Cambridge: Cambridge University Press.

Harrison, J. (1981) *Hume's Theory of Justice.* Oxford: Clarendon Press.

Hobbes, T. (1991) [1651] *Leviathan*, ed. R. Tuck. Cambridge: Cambridge University Press.

Hutcheson, F. (1971) [1728] *Illustrations on the Moral Sense*, ed. B. Peach. Cambridge: Cambridge University Press.

Korsgaard, C. (1996) *The Sources of Normativity.* Cambridge: Cambridge University Press.

Locke, J. (1986) [1690] *Second Treatise on Civil Government.* Buffalo, NY: Prometheus Books.

Mackie, J. L. (1980) *Hume's Moral Theory.* London: Routledge.

Mandeville, B. (1988) [1723] *The Fable of the Bees*, ed. F. B. Kaye. Indianapolis: Liberty Press.

Penelhum, T. (1975) *Hume.* London: Macmillan.

Further reading

Anscombe, G. E. M. (1981a) "On promising and its justice and whether it needs be respected in foro interno," *Collected Philosophical Papers*, vol. 3 (pp. 10–21). Minneapolis: University of Minnesota Press.

——(1981b) "Rules, rights and promises," in *Collected Philosophical Papers*, vol. 3 (pp. 97–103). Minneapolis: University of Minnesota Press.

Árdal, P. (1977a) "Another look at Hume's account of moral evaluation," *Journal of the History of Philosophy* 77, pp. 405–21.

——(1977b) "Convention and value," in G. P. Morice (ed.) *David Hume Bicentennial Papers* (pp. 51–68). Austin: University of Texas Press

——(1996) *Passion and Value in Hume's* Treatise. Edinburgh: Edinburgh University Press.

Cohon, R. (1997) "The common point of view in Hume's ethics," *Philosophy and Phenomenological Research* 57, pp. 827–50.

——(ed.) (2001) *Hume: Moral and Political Philosophy.* Burlington, VT: Ashgate.

Darwall, S. (1994) "Hume and the invention of utilitarianism," in M. A. Stewart and J. P. Wright (eds.) *Hume and Hume's Connexions* (pp. 58–82). Edinburgh: Edinburgh University Press.

Dees, R. H. (1997) "Hume on the characters of virtue," *Journal of the History of Philosophy* 35, pp. 45–65.

Gill, M. B. (2000) "Hume's progressive view of human nature," *Hume Studies* 26(1), pp. 87–108.

Hume, D. (1985) *Essays Moral, Political, and Literary,* ed. E. Miller. Indianapolis: Liberty Press.

Jensen, H. (1977) "Hume on moral agreement," *Mind* 86, pp. 497–513.

McIntyre, J. (1990) "Character: a Humean account," *History of Philosophy Quarterly* 7, pp. 193–206.

Miller, D. (1981) *Philosophy and Ideology in Hume's Political Thought.* Oxford: Clarendon Press.

Norton, D. F. (1982) *David Hume: Common-sense Moralist, Skeptical Metaphysician.* Princeton, NJ: Princeton University Press.

——(ed.) (1993) *The Cambridge Companion to Hume.* Cambridge: Cambridge University Press.

Raphael, D. D. (ed.) (1991) *British Moralists 1650–1800,* vol. 1: *Hobbes–Gay.* Indianapolis: Hackett.

Scanlon, T. M. (1990) "Promises and practices," *Philosophy and Public Affairs* 19, pp. 199–226.

Schneewind, J. B. (1990) "The misfortunes of virtue," *Ethics* 101, pp. 42–63.

Snare, F. (1991) *Morals, Motivation and Convention.* Cambridge: Cambridge University Press.

Whelan, F. (1981) *Order and Artifice in Hume's Political Philosophy.* Princeton, NJ: Princeton University Press.

15

Virtue and the Evaluation of Character

Jacqueline Taylor

Introduction

Hume's moral philosophy may plausibly be construed as a version of virtue ethics. Among the central moral concepts of his theory are character, virtue, and vice, rather than rules, duty, or obligation. The importance of character focuses theoretical attention on the emotions and other affective aspects of our moral psychology that are essential to our developing an appreciation of what is worthwhile and praiseworthy. Yet a conspicuous feature of Hume's ethical theory is an emphasis on moral sentiment as the means by which we *evaluate* character as virtuous or vicious. His theory thus contrasts with another version of virtue ethics, one often associated with Aristotle. Aristotelian virtue ethics focuses on the perspective of the virtuous agent, examining how one becomes virtuous, and the role of virtue in practical deliberation and living a good life. In contrast, Hume is much more interested in how we recognize and evaluate traits of character than in how we become virtuous, or in how virtue relates to deliberation and living well.

The difference between the Aristotelian and Humean versions of virtue ethics is not simply one of emphasis, that is, of seeing deliberation as more important than evaluation, or vice versa. Rather, the two theories have different views about moral knowledge: what moral knowledge is, how we acquire it, and what it is for. The Aristotelian view focuses on the virtuous agent's ethical capacities, especially moral perception and deliberation. For Aristotle, the fully virtuous agent has the wisdom needed to deliberate well, which is in turn required in order to lead a flourishing life. The virtuous agent sets the standard for moral knowledge, and is a model for others to imitate as they cultivate their own virtuous character. The Aristotelian's emphasis on practical reason thus grounds an agent-centered view of moral knowledge. In contrast, Hume's theory explains how we recognize character traits as virtues and vices, and how we form standards for our sentiment-based appraisals of these traits. The identification and valuation of character traits is a

social process, requiring conversation, and at times, negotiation and debate. On this view, moral knowledge about which characters are praiseworthy or blameworthy is a collectively established resource.

Seeing moral evaluation as a social practice distinguishes Hume from seventeenth- and eighteenth-century philosophers with whom, because they also focus on the moral sentiments, he is often associated. Certainly part of the explanation for Hume's emphasis on sentiment and evaluation lies in the historical context in which he was thinking about morality. Hume engages directly with an important debate of his day that emerged in response to Hobbes's "selfish" theory. Those who disagreed with the "selfish" theory of Hobbes and Mandeville divided into two general camps: moral intellectualists, such as Ralph Cudworth and Samuel Clarke, and moral sense theorists, such as Francis Hutcheson and Shaftesbury. In many respects, Hume agrees with the moral sense camp, and follows Hutcheson in setting out the inadequacies of the intellectualist position as well as criticizing the selfish theory. He agrees with Shaftesbury and Hutcheson that we have strong social motives, and that our approval of these motives reinforces both our appreciation of virtue and our motivation to act virtuously (see Shaftesbury 2000; Hutcheson 2002, 2004). It is the reflexive approval of social motives that makes the emphasis on evaluation important for the moral sense theorists. But rather than appealing to an innate moral sense that reflexively approves of benevolence, Hume argues that sympathy, our natural capacity to participate in one another's pains and pleasures, is the source for our moral approbation of a variety of virtues. Hume sees it as an evident benefit of his approach that grounding the moral sentiments in sympathy can accommodate a broad range of virtues: convention-based virtues of justice and etiquette, self-regarding virtues, even some talents and some immediately agreeable virtues such as wit or eloquence, as well as the social virtues recognized by his moral sense predecessors. The appeal to sympathy also means that Hume must explain how we correct and cultivate our moral "taste." His explanation of the cultivation of our moral taste yields an account of moral knowledge as the product of shared reflection, scrutiny, and conversation.

In this essay, I first examine Hume's understanding of character and its importance for us. I then turn to our natural responses to various character traits, and Hume's account of how sympathy with those responses is the source of the moral sentiments. Hume invokes the notion of a standard of virtue in order to undercut objections that sentiments with their source in sympathy will lack an objectivity that we think our moral evaluations should have. So next I look in some detail at the kind of errors to which sympathy makes us susceptible, and at how the strategies we use to correct these errors renders sympathy self-correcting. After then considering how our moral sentiments have authority for us, particularly when they are the verdicts of those who have cultivated an active habit of moral evaluation, I consider more fully why moral knowledge is, on the Humean view, a collectively-established social resource that is important for moral education and illustration, moral motivation, and our sense of moral identity.

The Importance of Character

Hume insists that character is the proper object of moral evaluation. We may sometimes consider actions independently of motive and character, looking instead, for example, at the consequences of a certain course of conduct. But Hume argues that what really makes a difference for us is a person's character. A person's reasons for acting typically reflect the characteristic ways in which she responds to situations. Her actions thus show us the kind of character she has. Suppose two people each present me with a gift of money, with the happy result that I now have enough wealth to start a business. Yet, if one gave me the money to curry favor further down the line whereas the other genuinely cared about my being successful in the new business venture, this difference in motive, and not just the gift of money, influences how I feel about the two people. The first one acts out of self-interest and perhaps manipulatively, while the second acts from beneficence. I am not indifferent to the fact that they did not act from the same motive, and I naturally feel more good will towards the beneficent person. As Hume notes, when "a good disposition is attended with good fortune, which renders it really beneficial to society, it gives stronger pleasure to the spectator, and is attended with a more lively sympathy" (T 3.3.1.21; SBN 585). But in spite of our warmer response to such fortunate virtue, we do not say that it is *more virtuous* than the benevolent disposition without the fortune. Hence character, the quality or motive from which someone acted, has an influence on our moral responses to the person independently of the consequences of her action.

Nevertheless, Hume indicates that careful judgment is needed to identify accurately traits of character. We are frequently partial towards those who benefit us, especially when it occurs on a regular basis. And just as we naturally love our benefactors, we tend to hate those who harm us. We thus may think well of our friends and allies who act with our interests in view, even if they are pernicious to others (as Hume writes, our ally's cruel conduct "is an evil inseparable from war"). And we may likewise deny a good character to our enemies, however much they benefit those in their charge; "we detest them under the character of cruel, perfidious, unjust and violent." This "method of thinking" with respect to our friends and foes is one that "runs thro' common life" (T 2.2.3.2; SBN 348). We also tend to find people disagreeable by their "deformity or folly," however unintended on their part (T 2.2.3.4; SBN 348). As Hume notes, "this is a clear proof, that, independent of the opinion of iniquity, any harm or uneasiness has a natural tendency to excite our hatred, and that afterwards we seek for reasons upon which we may justify and establish the passion" (T 2.2.3.9; SBN 351). Even accidental injuries tend to elicit a first, natural response of anger. To counter these common tendencies, we must distinguish between the pain or pleasure someone causes us and their "particular design and intention" (T 2.2.3.3; SBN 348). The intention connects the person's actions with something yet more "durable in him"

(T 2.2.3.4; SBN 348). Moreover, knowing whether someone has contempt for me or holds me in esteem influences my own love or hatred of him.

A person's attitudes, actions, and conduct show us what sort of character she has. Hume describes our actions and attitudes as "signs" of our character. Since we cannot directly "look within to find the moral quality" of an agent, we must "fix our attention on actions, as on external signs . . . but . . . the ultimate object of our praise and approbation is the motive, that produc'd them" (T 3.2.1.2–4; SBN 477–8). Not only someone's actions, but also her "words, or even wishes and sentiments," as well the responses of others to her actions and attitudes, allow us to make inferences about her character (T 3.3.1.5; SBN 575). Character consists of what Hume refers to as "durable mental qualities," settled habits of feeling and dispositions to respond and act in certain ways. Some of these are instinctive, and so will appear in some degree in most people. Hume identifies resentment, benevolence, and care of young children as among our instinctive propensities, ones we expect to move most people to respond in particular situations (T 2.3.3.8; SBN 417). He also characterizes some of the passions and what we think of as moods as traits of character. Such passions as kindness, anger, melancholy, hopefulness, or fear, can become established ways of responding to situations or other people. These become "durable principles of the mind, which extend over the whole conduct, and enter into the personal character" (T 3.3.1.4; SBN 575) (McIntyre 1990; Baier 1991).

Hume presents a fourfold classification of the virtues. A virtue is a character trait that is useful or agreeable, either to the person who possesses it or to others (T 3.3.1.30; SBN 590). The socially useful virtues include both artificial and natural virtues. The artificial virtues, which include honesty, loyalty, and fidelity, are primarily those we associate with the conventions of justice that establish our legal and civil obligations and rights. Hume also counts among the artificial or convention-dependent virtues the chastity and modesty required of women to help to preserve marital unions and families; gallantry on the part of men is likewise artificial. The natural social virtues include "meekness, beneficence, charity, generosity, clemency, moderation, equity" (T 3.3.1.11; SBN 578). Among the virtues characteristic of those who are "good and benevolent" we find "humanity, compassion, gratitude, friendship, fidelity, zeal, disinterestedness, liberality" (T 3.3.3.3; SBN 603). People who are "great" rather than good possess traits such as constancy, fortitude, and magnanimity, which suit them for a life of leadership and heroism (T 3.3.4.3; SBN 608). Traits useful for their possessor include "prudence, temperance, frugality, industry, assiduity, enterprise, dexterity" (T 3.3.1.24; SBN 587). The need for modesty in our dealings with others makes an internal "due degree of pride" perhaps the most important self-regarding virtue (T 3.3.2.8; SBN 596).

The inclusion of self-regarding as well as other-regarding traits makes Hume's catalogue of virtues broader than that of moral sense theorists such as Shaftesbury and Hutcheson. For the moral sense theorists, the moral sense is almost an instinctive approval of various forms of benevolence. By grounding the moral sentiments

in sympathy, Hume is able to include all the mental qualities that produce pleasure and pain for their possessor or others. Significantly, not all virtues tend to the public good. We also experience an immediate approval for such qualities as wit or an easy manner. These qualities are pleasing to others and we sympathize with that pleasure without considering whether wit or an easy manner promote the public good.

Two other features of Hume's catalogue are also noteworthy: the inclusion of traits some might label as talents or abilities rather than virtues, and the inclusion of traits that we may not be able to acquire through our voluntary efforts. Hume notes that some people might insist that the advantageous qualities such as industry, perseverance, patience, temperance, or frugality, required for success in the world, are natural abilities rather than virtues. But he responds that "in common life and conversation," we tend to praise whichever qualities please us and blame those that make us uneasy whether they are social virtues or self-regarding ones (T 3.3.4.4; SBN 609). Someone might object that our approval of talents or abilities is different in kind from that of our moral approval, but Hume responds that there are different kinds of approval even for some of the agreed-upon virtues. A second objection might be that talents are something bestowed by fortune rather than acquired. But people may not be able through their voluntary efforts to cultivate patience, or courage, or a number of other virtues, or for that matter, overcome a quick temper or some other vice. Regardless of whether it is voluntary, the quality produces pleasure or pain, and so earns our praise or blame.

Finally, notice that Hume suggests that there is broad agreement on which traits of character are virtues and which are vices. This agreement spans cultures and historical epochs; so, for example, "we give the same approbation to the same moral qualities in *China* as in *England*" (T 3.3.1.14; SBN 580). In his later works, in the second *Enquiry*, "A dialogue," and some of the essays, Hume changes his views about the universality of the recognition and valuation of character traits, and argues that there are culturally and historically specific virtues and vices. But even in the *Treatise*, Hume himself surely knew that his views about what we count as virtues would not garner widespread support. Including talents is controversial for theological or other views that place a value on free will. And some would surely find it frivolous to count qualities such as wit or eloquence as virtues. Nevertheless, Hume has given us a good explanation for why we can agree on the reasons for valuing or despising these different kinds of traits: we approve of useful or agreeable traits and disapprove of those that are harmful or disagreeable.

Sympathy, the Indirect Passions, and Moral Sentiment

Traits of character comprise one element of a larger set of relatively durable features of persons that we regard as having particular importance for us. The other features

include physical appearance and abilities, wealth, power and other external advantages, or the lack of these (see T 2.1.7–10). Their importance for us is shown by our strong response to them, feelings of love or hatred when we find them in others, and pride or humility when they concern ourselves. Hume describes the passions of love, hatred, pride, and humility as "indirect" passions. The indirect passions are not simple, direct responses to pleasure or pain, but rather are complex passions, always directed towards persons and signaling our recognition and valuation of one or more of these durable features that contribute to making the person who she is (Ainslie 1999). We might think of the indirect passions as evaluative attitudes, that is, belief-informed, affective responses that express our sense of someone's worth in virtue of one or more of these features. While we may love someone just because he is related to us, the indirect passions are typically produced in response to the person's possession of some feature such as good character, to which we independently attach value. So a person's virtue, beauty, or wealth makes her lovable to us, and is something in which she appropriately takes pride.

Hume's explanation of moral evaluation draws an intimate connection between moral sentiment and these four key indirect passions. First, he suggests that our most natural response to the characters of those people closest to us, including our family, friends, and colleagues, will be an indirect passion: some form of love towards those with a virtuous character, and some form of hatred towards those exhibiting vice. In Book 2, Hume describes virtue and vice as "the most obvious causes" of the indirect passions (T 2.1.7.2; SBN 295). In Book 3, in the section on the natural virtues and vices, he writes:

> Now since every quality in ourselves or others, which gives pleasure, always causes pride or love; as every one, that produces uneasiness, excites humility or hatred: It follows, that these two particulars are to be consider'd as equivalent, with regard to our mental qualities, *virtue* and the power of producing love or pride, *vice* and the power of producing humility or hatred. In every case, therefore, we must judge of the one by the other; and may pronounce any *quality* of the mind virtuous, which causes love or pride; and any one vicious, which causes hatred or humility. (T 3.3.1.3; SBN 574)

We love the virtuous among our friends, family, and associates, and likewise we take pride in our own virtue. Our natural response to vice in those with whom we have dealings is hatred, while our own vices cause us to feel humility.

Second, our perception of virtue and vice depends on our capacity to sympathize with these natural responses. Hume describes sympathy as a principle of the imagination that makes it possible for us to communicate our passions, sentiments, and even our opinions to one another. Sympathy can work in an immediate way, like a contagion, causing us to laugh, for example, just because someone else is laughing. More typically, though, sympathy allows us to interpret and respond to

the passions of others. The passions are perceptions of the mind, and since we do not have access to one another's minds, we must infer the passions of others, just as we infer character traits (some of which are passions), from their "signs." We read the passions and sentiments of others from their facial expressions, conversation, demeanor, and conduct. From a person's outward signs, and drawing on our own passionate experience, we form a belief about what she is feeling. As Hume observes, there is a "general resemblance" between ourselves and others, so that "we never remark any passion or principle in others, of which, in some degree or other, we may not find a parallel in ourselves" (T 2.1.11.5; SBN 318). The similarity between oneself and others explains why the force of sympathy goes beyond that of custom. We have a lively conception of ourselves, which "is always intimately present with us," so that we conceive "with a like vivacity of conception" whatever is related to us (T 2.1.11.4; SBN 317). Sympathy conveys the vivacity from the conception we have of ourselves to the idea we have of the person's passion, converting that idea into a passion.

Sympathizing with the passions or situations of someone else does not involve imagining oneself in place of the other. Indeed, our sympathetic participation in the passionate lives of others broadens our own passionate experience. For example, sympathizing with another "takes us so far out of ourselves, as to give us the same pleasure or uneasiness" that the person experiences, which in turn frequently moves us to respond to the situations of others (T 3.3.1.12; SBN 579). This point is important with respect to moral evaluation since our sympathetic appreciation of others' responses helps us to focus properly on traits of character instead of on consequences or our own interests. Sympathy may thus be thought of as having in part an epistemological function. It helps us move beyond our own concerns and gain a clearer view of what characters are like.

We should also note that sympathetic pains and pleasures can be disinterested. For example, when we sympathize with the pleasure the rich man gets from his wealth, our sympathetically produced pleasure, which in this case is an esteem or admiration, does not depend on our expecting to benefit from his wealth. Sympathy is the source of the different kinds of disinterested approval or disapproval we direct towards others. This single source for our different sentiments has what Hume sees as a distinct advantage with respect to the moral evaluation of character since the catalogue of virtues expands to include self-regarding as well as social traits, convention-based requirements for just conduct as well as good manners, and those mental abilities that are gifts of fortune rather than acquired traits.

Moreover, without this capacity for sympathy, we would typically respond to others when their conduct and character affected ourselves (or those so close to us that we regard their pains and pleasures as bearing directly on our own). Hume makes the point more clearly in the second *Enquiry*, arguing that if we lacked this capacity to sympathize with the pains and pleasures of others, we would likewise be indifferent to vice and virtue (EPM 5.40; SBN 226). In short, sympathy is the source of the moral sentiments.

Above, I said that there is an intimate connection between the indirect passions and the moral sentiments insofar as it is our sympathy with the former that produces the latter. It is worth noting that some commentators argue that Hume is advancing the view that the moral sentiments are themselves calm forms of love and hatred. The main textual evidence for this view includes the claim that virtue is the power of producing love (T 3.3.1.3; SBN 574); the claim that when we take up a shared moral perspective it makes possible a calm determination of the passions (T 3.3.1.18; SBN 583); and the claim that our moral approbation and blame "is nothing but a fainter and more imperceptible love or hatred" (T 3.3.5.1; SBN 614). Drawing on this evidence, Pall Árdal (1966) argues both that the indirect passions of love and hatred, on the one hand, and the moral sentiments on the other, have similar causes, namely, certain mental qualities, and that the moral sentiments are objective variants of the indirect passions. Philip Mercer (1972) endorses Árdal's argument, and adds that by taking up the general point of view (whereby we overlook our own interest and compensate for distance) we form a habit of objective judgment. Christine Korsgaard (1999) agrees with Árdal and Mercer that moral approbation is a species of love, and she also argues that Hume appeals to the shared moral perspective to establish a normative standard for love. She thinks it is possible to read Hume as claiming that virtue is *the* cause of love, and that while we do not always love the virtuous even from the common point of view, that evaluative perspective nevertheless shows us that we *ought* to love the virtuous since they are worthy of our love and entitled to our good will.

Nevertheless, the text leaves considerable room for disagreement with the interpretation of the moral sentiments as calm indirect passions. Annette Baier argues that while "moral sentiment is at least as complicated as any indirect passion," it is a "reversal" of the indirect passions scheme, since it takes characters rather than persons as its object (Baier 1991: 134). In contrast to our *loving* someone because she is witty, the moral sentiment is our *approval* of wit, whomever we find possessed of it (see also Norton 1982). Baier also argues that Hume proves more skeptical than we might have expected about the motivational influence of the moral sentiment since he claims it is sufficient merely to express our praise and blame (1991: 185–6). He does not *require* us either to love those we find praiseworthy, or to show good will to the virtuous, although sometimes we may.

Sympathy, Sentiment and Impartial Evaluation of Character

In Book 2, Hume introduces a distinction between the calm and violent impressions of reflection, writing: "of the first kind is the sense of beauty and deformity in action, composition, and external objects." The sense of beauty and deformity

in action is our sense of morality, or moral "taste." One reason for designating as calm the sentiments that arise from our sense of beauty is that they generally feel faint, even "imperceptible." "This division" of the passions is, however, "far from being exact" (T 2.1.1.3; SBN 275). Under various conditions, the violent passions may produce no felt emotion whereas the moral sentiments may be felt intensely. Indeed, our moral sentiments are sometimes among our strongest feelings: "there is no spectacle so fair and beautiful as a noble and generous action; nor any which gives us more abhorrence than one that is cruel and treacherous" (T 3.1.2.2; SBN 470). We find heroic characters "dazzling," and "cannot refuse" them our admiration (T 3.3.2.15; SBN 600). Cruelty is "the most detested of all vices," and arouses in us "a stronger hatred than we are sensible of on any other occasion" (T 3.3.3.8; SBN 605). We also tend to have stronger moral sentiments towards those who resemble the people with whom we usually interact. When a virtuous character is distant from us, but he and his acquaintances resemble those in our own community, their situation "interests us strongly by sympathy" so that we correspondingly feel a strong approval of his character. In such cases, "we approve of his character, and love his person, by a sympathy with the sentiments of those who have a more particular connexion with him" (T 3.3.3.2; SBN 602).

But our moral sentiments may be weak, and we may make a moral judgment about someone's character without feeling any sentiment at all. Variations in our sentiments of praise and blame occur because each of us remains confined, to some extent, to the perspective of our present situation. If confinement to our present perspective were entire, it would be impossible for us either to reach agreement with others or even to make consistent appraisals of our own actions over the course of our life. Fortunately, our capacity for reflection enables us to correct our sympathy and we can "in our thoughts place ourselves" in "some steady and general points of view" (T 3.3.1.15; SBN 581). From these perspectives we "form some general inalterable standard" to which our moral sentiments should conform (T 3.3.3.2; SBN 602). A second reason for regarding the moral sentiments as calm is that they arise from this process of reflection and correction, making our moral evaluations similar to the warranted judgments of the understanding.

From our present perspective, our evaluations of character vary, owing to a failure to regulate our sympathetic responses. Hume identifies three different causes of error, deriving from our natural partiality and the influence of the associative principles on sympathy. What I will designate the "remoteness" error occurs in cases where the agent is located at a distance from us so that our sympathy is too weak to produce the appropriate sentiment of praise or blame. The error of "countervailing interest" arises when our own interest or partiality is strong so that we confuse our interested love or hate of someone with moral approval or blame. The "consequentialist" error results from our evaluating the consequences of someone's having acted or not, rather than her character, so that we fail to separate the virtuous or vicious disposition from the accidental circumstances attending it.

The Errors of Remoteness and Countervailing Interest

Let us look at the remoteness error and the error of countervailing interest together, since Hume offers the same solution to them. We have seen that our moral sentiments arise when we sympathize with the pains and pleasures of those affected by the character of an agent. Although the general resemblance between others and ourselves makes sympathy possible, sympathy is influenced by the associative principles of resemblance, cause and effect, and contiguity. The influence of these principles makes sympathy "very variable." So we sympathize more easily with the passions and opinions of people who resemble us, and with those to whom we are related (Hume regards familial relationship as a "species" of causation). Others' nearness or distance, whether physical or temporal distance, or a more figurative distance based on culture, interests, or values, frequently affects our sympathy: "we sympathize more with persons contiguous to us, than with persons remote from us: With our acquaintance, than with strangers: With our countrymen, than with foreigners" (T 3.3.1.14; SBN 580). We naturally sympathize more easily with, and so feel a more lively praise or blame for an agent who is near to us than we do for someone who is physically or temporally remote from us. The problem is that when an agent and the people affected by her character are distant from us, "our sympathy is proportionably weaker, and our praise or blame fainter and more doubtful" (T 3.3.3.2; SBN 602). But it would be inconsistent to approve of a character trait in the agent close to us while not finding as admirable the same trait in a person who is distant from us. Hume acknowledges that we *judge* the two characters to be equally admirable, even if our sympathy is stronger and so produces a stronger sentiment in the case that is closer to hand. So he must explain how sympathy and sentiment can be the source of our moral admiration if our sympathy-based sentiments vary with distance.

A second failure in moral evaluation occurs when we fail to put aside a concern for our own interest or the interests of those to whom we are partial, and judge only from our own narrow perspective. As we saw in the section "The Importance of Character," failure to correct self-interest or partiality may prevent us from seeing the vices of our friends or acknowledging the virtues of our enemies. We generally think of courage as admirable whether we are reading about it in a historical narrative or witnessing it in the local heroes of our own community. But self-interest and prejudice lead us to find displeasing the bravery of our enemy and condemn it as blameworthy, even though others find it praiseworthy. In such cases, the passions of love and hate that arise from our interested perspective "are apt to be confounded" with moral sentiments, since these passions and sentiments "naturally run into one another." In other words, we mistake our *interested* hatred for *moral* blame: because the person is our enemy and her bravery works in opposition to our own interest, we tend, in our uneasiness, to think of her actions as

villainous. But it does not follow that the two different kinds of disfavor, interested and moral, are not distinct. The difficulty is sufficiently real, however, that it takes a person "of temper and judgment" to avoid "these illusions" (T 3.1.2.4; SBN 472).

Hume's emphasis on temper as well as judgment indicates the importance of both feeling appropriately and reasoning soundly. Responses of admiration or condemnation based on our partiality or prejudice do not count as moral evaluations. " 'Tis only when a character is considered in general, without reference to our particular interest, that it causes such a feeling or sentiment, as denominates it morally good or evil" (T 3.1.2.4; SBN 472). The solution to the remoteness error requires us to reconstruct imaginatively the agent's circle of associates, making the moral picture more vivid so that we sympathize more readily. Hume suggests that the corrective process for the remoteness error is as natural as correcting our judgment about the size of an object viewed from a distance in spite of the apparent sensible evidence. He notes, "such corrections are common with regard to all the senses; and indeed 'twere impossible we cou'd ever make use of language, or communicate our sentiments to one another, did we not correct the momentary appearances of things, and overlook our present situation" (T 3.3.1.16; SBN 582). In the case of countervailing interest, the moral evaluator remains focused on his own interest. When we find ourselves doing this, we must "loosen" ourselves from our "first station," that of self-interest, and make the moral picture more inclusive (T 3.3.1.18; SBN 583). Again, we sympathize with the effects of an agent's character on himself and his circle of acquaintances. And we

> consider not whether the persons, affected by the qualities, be our acquaintance or strangers, countrymen or foreigners. Nay, we over-look our own interest in those general judgments; and blame not a man for opposing us in any of our pretensions, when his own interest is particularly concerned. (T 3.3.1.17; SBN 582)

When we make the effort to broaden or reconstruct imaginatively the agent's circle of acquaintances, the immediate feelings that arise from our present perspective may nevertheless remain unchanged. We may continue to feel angry or humiliated, and so feel hatred for our enemy since her bravery opposes our interest. But by taking up this more general point of view, by discounting the strength of our own interest, and considering instead the effects of her bravery from a wider perspective than that of our own, we come to recognize that our immediate response is interested rather than moral. We can thus express verbally, if not through sincerely felt sentiment, praise for the merit of our enemy's bravery. In this way, we *judge* the two characters to be equally admirable, even if our sympathy is stronger and so produces a stronger sentiment in the case to which we are partial.

To correct both the remoteness error and the error of countervailing interest, we must adopt a shared perspective on the agent's character. Hume identifies two problems with this variability of sympathy that he thinks motivate us to correct it

by taking up a shared perspective. First, our own situation "is in continual fluctuation," so that an individual currently distant from us may in time become a familiar acquaintance. Second, our own situation frequently differs from that of others, leading to conflict with one another. Rather than relying on our own "peculiar" perspective, we need to find a common point of view with others if we want to "converse together on any reasonable terms" (T 3.3.1.15; SBN 581). We fix our view of what someone's character is like by adopting what Hume refers to as a "common point of view." Here is a succinct description of the process of taking up a common perspective.

> Every person's pleasure and interest being different, 'tis impossible men cou'd ever agree in their sentiments and judgments unless they chose some common point of view, from which they might survey their object, and which might cause it to appear the same to all of them. Now in judging of characters, the only interest or pleasure, which appears the same to every spectator, is that of the person himself, whose character is examin'd; or that of persons, who have a connexion with him. And tho' such interests and pleasures touch us more faintly than our own, yet being more constant and universal, they counter-ballance the latter even in practice, and are alone admitted in speculation as the standard of virtue and morality. They alone produce that particular feeling or sentiment, on which moral distinctions depend. (T 3.3.1.30; SBN 590)[1]

The common point of view helps us to move beyond our present and particular situation, and so to disregard the variations in sympathy and sentiment due to vivacity or distance. By sympathizing with the responses of those whose interests are affected by someone's character or with the person's own sense of pride or humility with respect to his character, we extend our concern to the perspective of those most familiar with the character. Their responses are "more constant and universal," so that they "counter-ballance" our own interests "even in practice." Taking up this shared perspective has a steadying or stabilizing effect on our judgment. The responses of the agent and her acquaintances "are alone admitted in speculation as the standard of virtue of morality," since they fix our view of what the agent's character is really like and thus help us to calibrate our sense of someone's praiseworthiness or blameworthiness. By following this method of evaluation, we gradually form a "general unalterable standard" to which our moral sentiments should conform, so that our moral evaluations become correspondingly "constant and establish'd" (T 3.3.3.2; SBN 602).

The Consequentialist Error

Lastly, Hume discusses cases where our sympathy is not activated at all. What I am calling the consequentialist error reflects our tendency to respond to the results of someone's action, rather than her character, so that we fail respond to an agent's

character when she is prevented from acting as she characteristically would. It is true that admirable or harmful character traits, when manifested in action, tend to produce valuable or harmful results, respectively. But it is the character trait towards which we direct our moral admiration or blame. So it is a mistake to regard the results of someone's action as harmful or beneficial independently of the character that produced it. Similarly, it is a mistake to disregard someone's character just because she was unable to act as she characteristically would, and there are no consequences – those effects comprising the responses of her acquaintances to her character – with which to sympathize. Now so far, we have focused on cases where a moral evaluator uses imagination to reconstruct the agent's circle of acquaintances, and sympathy to participate in the responses of all the members of the circle. But Hume's discussion of this third error shows that reconstructing and sympathizing with the whole of the agent's circle may not by itself be sufficient to produce the appropriate moral assessment, since in cases involving the consequentialist error there are few, if any, "effects" or "signs" of character in evidence for us to sympathize with.

The solution to the consequentialist error lies in adopting another kind of "steady and general" perspective. In Book 1, Hume argued that we distinguish the "superfluous" or "accidental" circumstances of an event from the properly efficacious causes by employing the general rules associated with the understanding and sound reasoning (T 1.3.13). We also rely on these general rules in the cases involving the consequentialist error, although in this case they produce a moral sentiment rather than a judgment of the understanding. From experience, we know the usual effects of benevolence and so are able to judge the benevolent disposition as if it were a "compleat" cause, even when it cannot be exercised to produce the action usual to such a disposition. Because the "character is . . . fitted to be beneficial to society, the imagination passes easily from the cause to the effect, without considering that there are still some circumstances wanting to render the cause a compleat one" (T 3.3.1.20; SBN 584). On reflection, we still approve of the benevolent person even though she is currently in reduced circumstances. As Hume puts it, "virtue in rags is still virtue" (T 3.3.1.19; SBN 584). One might object here that you cannot be benevolent without something to give, but that is not really right. If you are poor but benevolently disposed you would give if you could, and you can certainly express benevolent affections; in the same way, if no one stands in need of your beneficence that does not mean you do not have the trait, but rather that you have no reason to exercise it. It is on the other hand usually true that if you have something to give and you lack the benevolent trait, then beneficence will not result.

In cases involving the consequentialist error, Hume argues that the moral evaluator's sentiments are among a set of passions belonging to the imagination. In Book 2, Hume describes desires and inclinations "which go no farther than the imagination, and are rather the faint shadows and images of passions, rather

than any real affections" (T 2.3.10.5; SBN 450). For example, we will be pleased with the utility of a well-built city, arising from a sympathy with the pleasure of the inhabitants, even though we have no affection for them. In cases where the signs of character are not in evidence, the moral sentiments depend on envisioning in our imagination the typical efficacy of a particular character trait, an efficacy not actually in evidence, so that we are "mov'd by degrees of liveliness and strength, which are inferior to *belief*, and independent of the real existence of their objects (T 3.3.1.20; SBN 584). By imaginatively representing to ourselves the effects of benevolence, the sentiment aroused does not feel the same as those elicited by actual benevolent action, but it is "of a like species." The feeling of these two species of sentiment is sufficiently different, however, that they can be contrary, yet coexist without destroying each other: "These emotions are so different in their feeling, that they may often be contrary, without destroying each other." Hume characterizes these sentiments of the imagination as not forceful enough to "controul our passions" and actions, but as capable of influencing "our taste." Notice that because they arise from the corrective effects of general rules on the imagination, and are thereby rendered more stable and consistent, their influence on our taste is authoritative in a way the immediate and unreflective passionate response to the actual consequences is not; as Hume observes, "the imagination adheres to the general views of things, and distinguishes betwixt the feelings they produce, and those which arise from our particular and momentary situation" (T 3.3.1.23; SBN 586).

The Authority of the Moral Sentiments

In correcting each of these three kinds of error – remoteness, countervailing interest, and consequentialist – we often continue to experience those feelings arising from our particular and present perspective: whatever is near to us, affects our own interest, or is perceived vividly, and has an immediate influence on our passions. In our natural and immediate consideration of the consequences, we may, for example, feel disappointment about the ineffective virtuous agent; or we may feel strong love for a benevolent companion but not be moved by an equally benevolent person who is a stranger to us. Attention to the passages where Hume talks about correcting our responses from a general point of view shows him repeatedly making clear that it is not necessarily our passions or sentiments that get corrected. "The passions do not readily follow the determination of our judgment," and "the heart does not always take part" with our corrections, "or regulate its love and hatred by them" (T 3.3.1.18; SBN 583; T 3.3.3.2;SBN 603). Our present situation keeps us to some extent self-interested or partial, or focused on the immediacy of what happens before us. In "our thoughts," however, we adopt a more general perspective, and although "the passions do not always follow

our corrections . . . these corrections serve sufficiently to regulate our abstract notions, and are alone regarded, when we pronounce in general concerning the degrees of vice and virtue" (T 3.3.1.21; SBN 585).

Hume's phrase "abstract notions" is one of several that make moral judgment sound more like the verdict of reason than of moral taste or sentiment. For example, we disregard the variations in sympathy and sentiment "in our general decisions, but still apply the terms expressive of our liking or dislike." We learn to correct "our language, where the sentiments are more stubborn or inalterable." In attempting to "over-look our own interest in those general judgments; and blame not a man for opposing us in any of our pretensions," and making "allowance for a certain degree of selfishness in men," we arrive at "the general principle of blame or praise." If these reflections fail to correct our passions, we say "that reason requires such an impartial conduct" on our part. Or when a sympathetically produced sentiment is weaker than interest or partiality, it nevertheless is "equally conformable to our calm and general principles," and "'tis said to have an equal authority over our reason, and to command our judgment and opinion" (T 3.3.1.16–18; SBN 582–3).

But Hume insists, "the approbation of moral qualities most certainly is not deriv'd from reason, or any comparison of ideas; but proceeds entirely from a moral taste" (T 3.3.1.15; SBN 581. See also T 3.3.1.27; SBN 589). And although he uses the language of reason, he links our reflection-informed moral judgment with the calm passions and sympathy rather than reason. He reminds us of "what we formerly said concerning that *reason*, which is able to oppose our passion; and which we have found to be nothing but a general calm determination of the passions, founded on some distant view or reflection" (T 3.3.1.18; SBN 583). Our use of the terms expressing praise or blame indicates a judgment informed by reflection about how we *would* feel if we were affected by the agent's character. Sympathetic consideration of the love or hate of the person's acquaintances for him is for us a guide to his praiseworthiness or blameworthiness. Hume's reference to what was formerly said about improperly calling *reason* the calm determination of the passions is to his discussion, in the section on the origin of government, of the strategy we use to correct our direct passions (in this case self-interest) and determine "what is in itself preferable," or most choiceworthy (T 3.2.7.5; SBN 536). Our passions tend to "solicit" us and "plead in favor of whatever is near and contiguous," which, if we yield to them, leaves us satisfying a very short-term interest at the expense of what is best for us in the long run (T 3.2.7.2; SBN 534). When we consider the objects of our passions at a distance, from a more dispassionate perspective, those vagaries that make the short term good seem so tempting disappear, and we instead prefer what is most choiceworthy. Here, it is through our passionate response, in this case, a reflective form of interest, or long-term interest, that we determine what is really preferable. In the case of moral evaluation, Hume's claim that we *would* find the virtuous person lovable if we were an acquaintance in her circle, suggests that from the reflective, shared

point of view we *do* find her admirable and praiseworthy (even if we don't feel love for her).

Hume draws some initial parallels between the calm passions and the traditional conception of reason in the third part of Book 2. I want to survey some of those points briefly because they point to the possibility of constructing an account of good moral judgment as itself a virtue. Hume does not explicitly set out such an account in the *Treatise*, although he does in the second *Enquiry* and in the essay "Of the standard of taste" (Taylor 2002). We already noted in the section "Sympathy, the Indirect Passions, and Moral Sentiment," that some of our passions are naturally calm and so are mistaken for the conclusions of reason. As he wraps up his discussion of the will and the passions in T 2.3.8, Hume reviews the distinction between the calmness and violence of the passions, now explicitly associating the calm passions with how we use the term "reason" when we are talking about choosing good and avoiding evil.

> What we commonly understand by *passion*, is a violent and sensible emotion of the mind, when any good or evil is presented . . . By *reason* we mean affections of the very same kind with the former; but such as operate more calmly, and cause no disorder in the temper: Which tranquillity leads us into a mistake concerning them, and causes us to regard them as conclusions only of our intellectual faculties. (T 2.3.8.13; SBN 437)

Hume characterizes the calm passions here as like reason because they motivate us without producing emotional agitation. But he argues that some of our calm passions are also like reason both because they are *stronger* than, and hence able to oppose their violent counterparts, and because they *are reflection-informed* responses. These claims are critical to seeing how the reflective moral evaluations we make are strong enough "to controul our taste," if not our actions.

The account of the calm passions in Book 2 deals with the direct passions that move us to action. These passions do not influence the will "in proportion to their violence, or the disorder they occasion in the temper." Rather, through custom and its "own force," a passion can "become a settled principle of action" and "the predominant inclination of the soul," which no longer produces sensible agitation. A calm passion can thus be strong, while a violent one can be weak, a "momentary gust" that quickly subsides once it meets with opposition or uncertainty (T 2.3.4.1; SBN 418). Echoing Joseph Butler's distinction between active, practical habits and passive ones, Hume also argues that custom makes it easier and more pleasant to engage in active habits, in turn giving them "new force" (T 2.3.5.5; SBN 424). The imaginative *effort* involved in making moral judgments, either from the common point of view or by relying on general rules, and the *importance* of our moral judgments, together make it plausible to regard moral evaluation as an active habit.

Finally, some of the settled calm passions that may be mistaken for reflective reason comprise the virtue of strength of mind. Hume describes strength of mind

as "the prevalence of the calm passions above the violent," a prevalence which depends on the peculiar temper and disposition of the individual (T 2.3.3.10; SBN 418). Although "the violent passions have a more powerful influence on the will . . . the calm ones, when corroborated by reflection, and seconded by resolution, are able to controul them in their most furious movements" (T 2.3.8.13; SBN 437). The moral evaluator with strength of mind will find it easier to set aside her own interests, bring the distant characters closer, and disregard consequences in favor of character. Her moral praise or blame is informed by reflection on the vagaries of her passions and the imagination, on the commonality of interest among humankind, on the naturalness of being more concerned with the characters of those who affect one, and on our tendency to get caught up in the present and what is actually happening around us. Her resolution to stabilize her own internal judgments of others and to converse intelligibly with others will lead her to form an active habit of moral evaluation, and in turn cultivate a strong and authoritative moral taste.

Moral Knowledge as a Shared Resource

Let us return now to the point I made in the introduction to this essay about moral knowledge being a shared resource. The Humean appeal to moral sentiment is to a *cultivated* form of moral response, and signals a particular way of theorizing about the social development and character of our sense of morality. We might regard moral evaluation as a social practice in which, ideally, everyone participates. Because we direct our moral attention to what about a person's character, manifested in her attitudes or conduct, has a special relevance for how well or badly our lives go, it is important to us that others in our moral community endorse and agree with our judgments. The standards for appropriately praising or blaming, admiring or condemning, arise from what Hume calls an "intercourse of sentiments," that is, from conversation and debate among the members of the moral community (T 3.3.3.2; SBN 602). From the common point of view, we give a fair hearing to the views of others, thereby facilitating intelligible conversation. Moral evaluators also subject their attitudes to mutual scrutiny and attempt to reach agreement in outlook. The moral sentiments thus exhibit shared reflection on the value of a broad range of character traits and qualities, as well as social policies and practices, that make a difference to how well people live.

The agreement aimed for is not sheer consensus, however, but agreement in *judgment* about the worth of particular traits or kinds of character. Such judgments incorporate reflective beliefs about which character traits are virtues and which vices. Additionally, sympathetic concern for the effects of character traits on the interests, commitments, and lives of the agent herself or others introduces an affective quality, a felt sentiment of pleasure or uneasiness, to moral judgment. The common point of view and the practice of moral evaluation thus enable moral

evaluators to develop a moral taste, a *sense* of the value of particular traits or kinds of characters.

Hume has stressed the need for internal stability, which we achieve when moral evaluation becomes an active habit and when we are able to converse on intelligible terms with others. The point about conversing intelligibly relates to the social purpose served by our moral evaluations, and the importance of seeing moral knowledge as a shared resource. As Hume puts it, our exchange of mutually intelligible sentiments "in society and conversation, makes us form some general unalterable standard," and our moral evaluations are renderered "sufficient for discourse, and serve all our purposes in company, in the pulpit, on the theatre, and in the schools" (T 3.3.3.2; SBN 602). The moral sentiments attest to judgments about the worth of various character traits, and these judgments differ from our personal loves and hates. We are interested in educating our children, in our homes and in the schools, and in some cases from the pulpit. We want the kind of moral education, and the institutions fostering it, to shape children's characters and allow "the sentiment of honour . . . [to] take root in their tender minds," so that they can become useful and agreeable, to themselves and others (T 3.2.2.26; SBN 500). We want our public exemplars, for example, those on the stage and those with social or political power, to reflect our values and shared sense of what is good. Far from being merely subjective reactions, our moral evaluations have a real influence on what our society looks like. It is, as Hume recognizes, crucially important for us to make accurate identifications of character and to have a proper appreciation of virtue and an abhorrence of vice.

The moral sentiments have a productive rather than a representative function, and Hume famously describes them, in the second *Enquiry*, as "gilding and staining all natural objects" and raising "in a manner a new creation" (EPM Appendix 1.21; SBN 294). We direct these sentiments towards certain aspects of persons, their "mental qualities" or character traits, as these are exhibited in conduct, attitudes, or policies. We noted earlier that character traits have a special relevance for how well we attend to a range of needs and concerns important to our living together with some degree of success. Through our actions and attitudes we establish our character and earn a reputation with others. As the *Enquiry* claim about the moral sentiments gilding or staining characters suggests, the appraisal of our character and conduct by others not only contributes to our reputation, but also informs our sense of our own character. Those of us who are trustworthy and do our part in society will merit admiration, and may sometimes receive the love and good will of others. Hume suggests that "the most considerable effect that virtue and vice have upon the human mind" is, in the case of virtue, to elicit love and pride as well as admiration, and in the case of vice, to arouse hatred and humility in addition to moral disapproval (T 3.1.2.5; SBN 473).

In cultivating the proper appreciation for what is praiseworthy about someone's character, we can also turn the practice of moral appraisal on our own character. Virtuous character is a particularly durable source of pride. Our "peace and inward

satisfaction" depend on our cultivating a virtuous character, as well as moral taste, so that our mind can "bear its own survey" (T 3.3.6.6; SBN 620). Our appreciation of virtue and blame of vice may thus be deployed as practical attitudes that influence our own choices and conduct.

Note

1   See also the discussion at T 3.3.1.14–18; SBN 580–3. At T 3.3.3.2; SBN 602, Hume describes slightly differently the process of taking up a common point of view. He is here considering the qualities that make someone a "good" person. These are the qualities that make us useful and agreeable to those in our "narrow circle" of friends, family, neighbors, and colleagues. Hume notes that we expect most people to be concerned about those closest to them. When we sympathize with the narrow circle of an agent distant from us, our sympathy is less lively than the concern we feel for those in our own circle. Nonetheless, we extrapolate from our own situation or that of those with whom we easily sympathize. When we encounter cases resembling those with which we're familiar, we know from reflection that we would find someone equally as worthy or harmful were we a member of the distant agent's circle. By this strategy of extrapolation, we "arrive at a more constant and establish'd judgment" about the value of various characters. Again, our judgments may fail to regulate our sentiments and, yet the former are "sufficient for discourse, and serve all our purposes in company, in the pulpit, on the theatre, and in the schools."

References

Ainslie, Donald (1999) "Scepticism about persons in Book II of Hume's *Treatise*," *Journal of the History of Philosophy* 37, pp. 469–92.

Árdal, Pall (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.

Hutcheson, Francis (2002) [1728] *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*, ed. Aaron Garrett. Indianapolis: Liberty Fund.

——(2004) [1725] *An Inquiry into the Original of our Ideas of Beauty and Virtue*, ed. Wolfgang Leidhold. Indianapolis: Liberty Fund..

Korsgaard, Christine M. (1999) "The general point of view: love and moral approval in Hume's ethics," *Hume Studies* 25, pp. 3–41.

McIntyre, Jane L. (1990) "Character: a Humean account," *History of Philosophy Quarterly* 7, pp. 193–206.

Mercer, Philip (1972) *Sympathy and Ethics*. Oxford: Clarendon Press.

Norton, David Fate (1982) *David Hume: Common-sense Moralist, Skeptical Metaphysician*. Princeton, NJ: Princeton University Press.

Shaftesbury, Anthony Ashley Cooper (2000) [1711] *Characteristics of Men, Manners, Opinions, Times*, ed. Desmond M. Clarke. Cambridge: Cambridge University Press.

Taylor, Jacqueline (2002) "Hume on the standard of virtue," *Journal of Ethics* 6, pp. 43–62.

Further reading

Baier, Annette C. (1980) "Master passions," in A. O. Rorty (ed.) *Explaining Emotions* (pp. 403–23). Berkeley and Los Angeles: University of California Press.

Dees, Richard H. (1997) "Hume on the characters of virtue," *Journal of the History of Philosophy* 35, pp. 45–64.

Falk, W. D. (1975) "Hume on practical reason," *Philosophical Studies* 27, pp. 1–18.

Radcliffe, Elizabeth S. (1996) "How does the Humean sense of duty motivate?," *Journal of the History of Philosophy* 34, pp. 383–407.

Taylor, Jacqueline (1998) "Moral sense theories," in Edward Craig (ed.) *Routledge Encyclopedia of Philosophy*, vol. 6. London and New York: Routledge.

Index