

THE *VERA CAUSA* PRINCIPLE IN 18TH CENTURY MORAL PSYCHOLOGY¹

FERNANDO MORETT

David Hume has been largely read as a philosopher but not as a scientist. In this article I discuss his work exclusively as a case of science; in particular as a case of early modern science. I compare the moral psychology of self-interest, sympathy and sentiments of humanity he argues for with the moral psychology of universal self-interest from Bernard Mandeville, presenting the controversy between the two as a case of theory choice.

David Hume, Adam Smith and Francis Hutcheson regarded the psychology of self-interest advanced by Thomas Hobbes and Bernard Mandeville as the rival theory to be defeated; it was a theory making ‘so much noise in the world’, Smith reports². Hume positively praises Mandeville’s theory in the first pages of the *Treatise*. This volume pays more attention to self-interest than *An Enquiry Concerning the Principles of Morals*, in which Hume discusses the disinterested passions of humanity and benevolence in more detail. Both Hume and Smith were highly critical of Mandeville’s theory, which they considered to be ‘malignant’ and ‘wholly pernicious’ because it leaves no grounds for ‘feelings sympathy and humanity’. They actually allude to Mandeville with epithets such as ‘sportive sceptic’ and ‘superficial reasoner’, whose reasoning is ‘ingenious sophistry’.³

Hume strongly criticises the self-interested individuals described by Mandeville who are ‘monsters’ ‘unconcerned, either for the public good of a community or the private utility of others’⁴; they are replicas of Ebenezer Scrooge who even at Christmas shows no humanity, no concern for others. In contrast, Hume describes a polite, sympathetic and utilitarian individual who, despite being self-interested, is capable of performing acts of disinterested benevolence and humanity, that is, a Scrooge who is morally reformed by secularised Christian values. Hume not only claims that true disinterested charity and beneficence exists grounded on natural sentiments of humanity, but he also claims that these sentiments can ‘overpower’ and ‘over-balance’ self-interest.⁵

The debate between the psychology of universal self-interest and the psychology of self-interested sympathetic individuals is about *true causes* or *true motives* for moral behaviour. This debate is concerned with the ‘metaphysical’ part of human psychology, that is to say, with unobservable entities and mental processes whose study Mandeville described as analogous to the study of the inner anatomy of the human body in medicine.⁶ Using the same naturalistic analogy from Mandeville, Hume explains that the moralist is a painter who is concerned with the beauty of moral behaviour portraying it

¹ Forthcoming in *Causes and Consequences*, 2015, Rom Harré and Fathali Moghaddam (eds), Springer.

² A. Smith (1790), p. 313.

³ *Ibid.*, pp. 308, 312; D. Hume (1772a), pp. 35, 48, 54, 95.

⁴ D. Hume (1772a), p. 48.

⁵ *Ibid.*, p. 77; D. Hume (1739-40), p. 313.

⁶ B. Mandeville (1732a), p. iii.

with ‘the most graceful and engaging airs’, whereas the moral anatomist is concerned with ‘the most hideous and disagreeable’ parts analogous to the ‘the inward structure of the human body, the position of the muscles, the fabric of the bones, and the use and figure of every part or organ.’⁷ Hume did not consider himself to be a moralist but a moral anatomist, he did not write any substantive normative moral argument, this is why the moral philosopher Francis Hutcheson criticised the *Treatise* because of its lack of ‘Warmth in the Cause of Virtue’.⁸

THE METHOD OF MORAL PSYCHOLOGY

Because the different passions and sentiments and their mutual operations cannot be observed, such a moral anatomy becomes metaphysics in search for the ‘hidden truths’ and ‘the secret springs and principles’ of the inward parts of human nature, which can only be discovered by ‘painful’ and ‘abstruse’ enquiry.⁹ Methodologically, the production of this new science represented a great challenge because of the difficulties of producing accurate and reliable knowledge of unobservable entities and processes in the mind by relying on observable behaviour. Indeed, Hume wanted to build this new ‘science of man’ as an ‘accurate’ and ‘true metaphysics’ of human nature based on ‘experience and observation.’¹⁰

In the introduction to the *Treatise* and the opening section of the *Enquiry*, Hume states his commitment to the observational and experimental method with an explicit reference to Francis Bacon, and also by quoting Isaac Newton. He considers the introduction of the experimental method as a key methodological innovation in the study of morality, which would allow one to treat ‘passions, motives, volitions, and thoughts’ as ‘matters of fact’ existing ‘in the mind’ just as it is done in physics with properties such as ‘sounds, colours, heat, and cold’, so that ‘discovery in morals, like that other in physics, is to be regarded as a considerable advancement of the speculative sciences’.¹¹ In line with this methodological commitment, he explicitly appeals to three epistemic criteria in his attempt to refute Mandeville’s psychological theory, namely inductive support, *experimentum crucis* and simplicity. All this seems to provide enough evidence for evaluating Hume’s work as a case of early modern science. Therefore, the moral psychology he was erecting can be evaluated by looking into its epistemic merits and methodological grounds.

⁷ D. Hume (1772b), p. 8; D. Hume (1739-40), p. 395.

⁸ Greig, J. Y. T. (1932), p. 32, Letter 13, 17th September 1739.

⁹ D. Hume (1772b), pp. 6, 11.

¹⁰ *Ibid.*, pp. 6, 10; D. Hume (1739-40), p. 4.

¹¹ D. Hume (1739-40), pp. 301-302.

Mandeville himself explains the method he uses for producing his own psychological theory, he calls it the ‘Method of reasoning from Facts *à posteriori*’¹², which also consists of reasoning from observation and experience. He recognises reasoning from experimentation in chemistry also as an *a posteriori* method, although he explains that experiments cannot be performed on the brain. Alluding to René Descartes, he rejects *a priori* reasoning by claiming that ‘all our knowledge comes *à posteriori*, it is imprudent to reason otherwise than from Facts.’¹³ He also makes an important distinction between ‘conjecture’ and ‘knowledge’, explaining that the latter provides certainty while the former only provides a degree of probability.¹⁴

In this case, ‘certainty’ becomes a decisive epistemic criterion because a psychological theory consists of knowledge of unobservable processes and entities taking place inside the mind. Certainty has been a long-standing problem in psychology and across the social sciences, where an important distinction is made between the inside and the outside of any individual or collective action. This distinction has produced a methodological divide in the social sciences between methodological naturalism and hermeneutics, while in psychology a similar distinction created a divide between behaviouristic and cognitive psychology. Hume and Mandeville fall into the category of those scientists following naturalistic methods to learn about the inside of human action, that is to say, about mental entities and processes as causes for behaviour. Mandeville explicates his own *method of inference* to get knowledge of the mind by using an analogy with the knowledge of the inner parts and functioning of a spring-watch:

I don’t believe there is a Man in the World of that Sagacity, if he was wholly unacquainted with the Nature of a Spring-Watch, that he would ever find out by dint of Penetration the Cause of its Motion, if he was never to see the Inside: But every middling Capacity may be certain, by seeing only the Outside, that its pointing at the Hour, and keeping to Time, proceed from the Exactness of some curious Workmanship that’s hid; and that the Motion of the Hands, what number of Resorts soever it is communicated by, is originally owing to something else that first moves within. In the same manner we are sure that, as the Effects of Thought upon the Body are palpable, several Motions are produced by it, by contact, and consequently mechanically.¹⁵

Note that according to Mandeville only a ‘middling certainty’ can be attained on the knowledge about the mind. He explains why full certainty cannot be achieved by pointing to two important methodological constraints related to the anatomy of the brain. First, he explains that the anatomist can only have access to the brain when it is already dead, so the main ‘spring of life’ is gone, and therefore a full understanding of the

¹² B. Mandeville (1729), p. 175, see also p. 164.

¹³ *Ibid.*, p. 261, see also p. 186.

¹⁴ *Ibid.*, pp. 163, 167, 263.

¹⁵ *Ibid.*, p. 164.

its functioning is not possible. The second constraint is set by the limited scope of macroscopic observation, which is constrained to large parts and organs such as nerves, blood vessels, folds and windings, while millions of small cells remain unobserved. Because of these two methodological constraints, the psychologist cannot gather the information needed for producing a theory with a higher degree of certainty. Therefore, the ‘best Naturalist must acknowledge that he can only ‘give any tolerable Guesses’, or actually admit that in some cases ‘as to the mysterious Structure of the Brain itself, and the Oeconomy of it, that he knows nothing.’¹⁶

With those limitations, the inference to inner live processes and entities proceeds exclusively from observable behaviour. He is aware the epistemic challenge this implies by explaining that ‘it is impossible to judge of a Man’s Performance, unless we are thoroughly acquainted with the Principle and Motive from which he acts.’¹⁷ Therefore, by becoming thoroughly acquainted with the information available on the subjects, the scientist can reach a ‘middling’ or reasonable degree of certainty.

I present the controversy between the moral psychologies from Hume and Mandeville as a problem of theory choice. Current epistemic criteria used for theory choice such as novel predictions, falsifiability or ontological heterogeneity can be inappropriate for a choice between moral psychologies of the eighteenth century. Therefore, I use the three criteria used by Hume himself, namely inductive support, *experimentum crucis* and simplicity adding those from the *vera causa* principle, which in the late eighteenth and early nineteenth century became the systematised expression of the rules Isaac Newton advanced in the *Principia*. The *vera causa* principle addresses epistemic concerns analogous to those of Hume and Mandeville concerning the knowledge of ‘passions, motives, volitions and thoughts.’¹⁸ By using these four criteria, the controversy between the psychology of universal natural self-interest from Mandeville, and the psychology of natural self-interest and sympathy from Hume, becomes a standard case of theory choice.

This early modern controversy between these two psychologies not only has historical value because the debate about self-interest and unselfish behaviour or altruism continues. In recent years, social and public choice theories as well as neoclassical and welfare economics have been a main battle ground for this controversy. The controversy is also relevant today because some of the criteria used such as inductive support and simplicity are currently used in theory choice. Today, the inference to the best explanation and the realist argument on unobservable entities and mechanisms addresses the same problems the *vera causa* principle was trying to solve.¹⁹

¹⁶ *Ibid.*, p. 165-166.

¹⁷ B. Mandeville (1732a), p. 56.

¹⁸ D. Hume (1739-40), p. 301.

¹⁹ See R. Harré (1960), P. R. Thagard (1978); N. Cartwright (1983), pp. 6, 87-99, and P. Lipton (2004).

To get true and accurate knowledge of unobserved entities and processes causally responsible for observable effects became a major challenge in the eighteenth century. Hume's main concern, and indeed the main problem for Newton and other natural scientists at the time, was about the criteria required for accepting any explanation based on unobserved entities. The rules of reasoning advanced by Newton constitute a response to this concern. Mandeville's methodological analogy with the inference to the inner pieces and functioning of a spring-watch also reflects the same concern and awareness of the problem. The epistemic justification of the existence and causal efficacy of gravitation, self-interest and sympathy was a main scientific challenge. In the late eighteenth century this led to the development of the *vera causa* principle by Thomas Reid, John Hershel and Charles Lyell.²⁰ This method was later used by Charles Darwin in his defence of genetic variation across very long periods of time as the true cause for the origin of new species, and migration as the true cause for the existence of colonies of the same species found in distant places.²¹

The first rule of natural philosophy as stated by Newton dictates that '*No more causes of natural things should be admitted than are both true and sufficient to explain their phenomena.* As the philosophers say: Nature does nothing in vain, and more causes are in vain when fewer suffice. For nature is simple and does not indulge in the luxury of superfluous causes.'²² By appealing to this methodological rule, Newton was trying to prove the existence of gravitation as the *vera causa* of the attraction between celestial bodies against the vortex theory advanced by Descartes which, by multiplying causes unnecessarily, depicted nature as superfluous and idly complex.²³ Therefore, theoretical simplicity was not an instrumental principle but a realist one justifying the choice for theories with fewer unobservable entities.

Like Newton, Thomas Reid also defines the *vera causa* principle by using the two criteria of truth and causal sufficiency; he writes that 'when men pretend to account for any of the operations of nature, the causes assigned by them ought, as Sir Isaac Newton taught us, to have two conditions, otherwise they are good for nothing. *First*, they ought to be true, to have a real existence, and not to be barely conjectured to exist without proof. *Secondly*, they ought to be sufficient to produce the effect.'²⁴ John Herschel explained that 'Newton has applied the term *verae causae*; that is, causes recognized as having real existence in nature, and not being mere hypotheses or figments of the

²⁰ See M. Ruse (1976); and R. Laudan (1982).

²¹ C. Darwin (1859), pp. 352, 482,

²² I. Newton (1726), p. 794.

²³ See I. Newton (1726), pp. 786-790, 939-944; R. Descartes (1644), pp. 106-109, 188-191. Besides not being able to account for a number of data, Descartes's vortex theory was more complex, it postulated three different kinds of unobservable matter and stratified bands in each vortex, where heavenly and earthly bodies move down and around due to centrifugal forces.

²⁴ T. Reid (1785), p. 80.

mind.²⁵ As we know, Mandeville draws a similar distinction between ‘conjectures’ and ‘knowledge’, arguing that only ‘middling certainty’ can be attained in the knowledge of the passions and the mind.

Just like the controversy between Newton and Descartes was about the true causes of the same set of phenomena, namely the motion of the planets, the controversy between Hume and Mandeville was about the true causes of the same domain of human behaviour. Newton introduced the first rule to prove that his theory had only used the sufficient number of causes, whereas Descartes used more than a sufficient number of them. In spite of stating his commitment to Newton’s method for the creation of moral psychology as a new science, Hume does not mention nor discuss Newton’s first rule. Nonetheless, I believe the use of this rule for evaluating his moral psychology is both justified and adequate. To explain the same domain of human behaviour, Mandeville uses one cause or motive only, i.e. self-interest, whereas Hume uses two, i.e. self-interest and the sentiment of humanity and benevolence.

The actions from Greek and Roman characters such as Pericles, Marcus Tullius Cicero, Publius Thrasea Paetus and King Henry IV of France are presented as examples of unselfish acts of patriotism, statesmanship and friendship motivated by sentiments of humanity and benevolence. The main problem with them is the lack of consideration Hume gives to the possible existence of self-interested motivations, which he could then refute. In contrast, Mandeville considers and refutes possible unselfish motivations for precisely the same kinds of actions Hume is using in support of his own theory.

Because of this refutation Mandeville’s theory must be chosen as the simpler one, while Hume’s theory loses the contest standing as a theory ‘indulging in the luxury of superfluous causes’ just like Descartes’s vortex theory did against Newton’s theory of gravitation. This is shown in the paragraphs below where the cases of Pericles, Marcus Tullius Cicero, Publius Thrasea Paetus and King Henry IV are discussed.

INDUCTIVE SUPPORT

Newton’s rule number four explains the epistemic power of induction as follows: ‘*In experimental philosophy, propositions gathered from phenomena by induction should be considered either exactly or very nearly true notwithstanding any contrary hypotheses, until yet other phenomena make such propositions either more exact or liable to exceptions.*’ This rule should be followed so that arguments based on induction may not be nullified by hypotheses.²⁶ The refutation of Mandeville’s psychological theory relies on the existence of acts of disinterested benevolence and humanity motivated by sympathy, which Hume claims is a natural

²⁵ J. Herschel (1830), p. 144.

²⁶ I. Newton (1726), p. 796.

component of the human mind. Because Mandeville's theory is universal, one case of disinterested benevolence is enough for refuting it. Hume, however, wants to do more than that because he is seeking to reduce the scope of self-interest by enlarging the scope of sympathy producing as many cases as possible of disinterested benevolence and humanity.

The psychological capacity humans have for sympathising with others is the main foundation for any benevolent and humanitarian action. Hume defines sympathy as the 'communication' of the 'inclinations and sentiments' of others 'however different from, or even contrary to our own', so 'hatred, resentment, esteem, love, courage, mirth and melancholy; all these passions I feel more from communication than from my own natural temper and disposition', 'and 'tis certain that we may feel sickness and pain from the mere act of imagination, and make a malady real by often thinking of it.'²⁷ Within the different passions and inclinations sympathy may elicit, Hume selects only those related to humanitarian and benevolent actions.

There is hardly any case of disinterested benevolence and humanity discussed in the *Treatise*, so specific cases can only be found in the *Enquiry*. There, Hume quotes statesmanship, patriotism, motherly love, friendship and love relationships as strong evidence of disinterested actions, which can prove the existence of humanitarian motives.

The first particular case he discusses is that of Pericles, the 'great Athenian statesman and general', who in his death-bed stopped his friends from paying public tribute to him by citing all his great achievements as a statesman, which he described as 'vulgar advantages' in comparison to the 'the most eminent' of his accomplishments, namely that 'no citizen has ever yet worn mourning on my account.' Further cases include Marcus Barea Soranus, roman proconsul in Asia, and Publius Thrasea Paetus, roman senator and republican leader, who rebelled against the emperor Nero; they were 'intrepid in their fate, and only moved by the melting sorrows of their friends and kindred. What sympathy then touches every human heart!' Then, he quotes the case of the mother who 'loses her health by assiduous attendance on her sick child, and afterwards languishes, and dies of grief, when freed, by its death, from the slavery of that attendance'. And also friendship and love relationships when people love and care for others even at the expense of being hurt, like King Henry IV of France whose 'amours and attachments' 'during the civil wars of the league, frequently hurt his interest and his cause.'²⁸ Hume argues that in all these cases, a sympathetic sentiment prompts humanitarian and unselfish actions, which benefit the citizens within a country, children within a family and friends and lovers within a close circle. No self-interested motivation is considered, self-interest plays no role.

Unlike Hume, Mandeville does consider if statesmen, patriots, mothers, friends and lovers act because of self-interested motives such as being flattered, adored and

²⁷ D. Hume (1739-40), pp. 206-207.

²⁸ D. Hume (1772a), Pericles, p.8; Marcus and Publius, pp. 40-41; mother, p. 93; Henry IV, pp.64-65.

glorified. He argues for universal self-interest as the true explanation of these and other kinds of behaviour. In order to prove this, he discusses different social groups giving special weight to those where true humanitarian and unselfish motives are expected such as cardinals, nuns, friars, mendicant orders, mothers, soldiers, kings, ministers and members of the court.

Mandeville explains that cardinals and other clergymen enjoy large fees, housing, food and other comforts for their service. Moreover, the College of Cardinals in Rome is described as ‘the best School to learn the Art of Calling’, where ‘each Member, besides the Gratification of his own Passions, has nothing at Heart but the Interest of this Party.’²⁹ When priests and nuns provide emotional help and comfort to the poor and the rich, they also act motivated by self-interest for they want to ensure a place in heaven and veneration on earth. No ‘seraphic love’ exist among nuns and friars living in convents and monasteries because ‘none of them sacrifice their darling Lusts’, they breach the vow of celibacy by having children, which they then abort or hide after being born. Mendicant orders ‘differ in nothing but their Habits from other sturdy Beggars, who deceive people with a pitiful voice’ to get the money they will later spend indulging in their own ‘Appetites, and enjoy one another’.³⁰

Similar conclusions extend to members of the Royal Courts, who despite being named and employed to serve the public interest, they ‘rob the Publick’ instead and are dominated by ‘excess of Vanity and hurtful Ambition unknown among the poor’. Furthermore, ‘Envy, Detraction and the Spirit of Revenge, are more ranging and mischievous in Courts that they are in Cottages’.³¹ Government ‘Offices of the greatest Trust are bought and sold’, ministers affect the public interest by engaging in bribery and corruption in order to advance their own self-interest or that of their own party.³² The King sends their subjects to death and impoverishes his own country by ‘obstinately maintaining A War that has almost utterly destroy’d his Kingdom’, or because of excessive spending caused by his love for ‘Pomp and Luxury’.³³ Soldiers do not risk their lives for others, they act motivated by their own self-interest trying to avoid public shame if they hide or run away, while at the same time they also seek personal glory and immortality.³⁴

Mothers love their children, however they love themselves and their own preferences even more. The same woman who can neglect, give away or even kill ‘her Bastard in the most execrable manner’ because of public shame or burden ‘may take care of, cherish and feel all the tenderness for her Infant’ if it is born in proper marriage. By giving excessive care and protection to their offspring mothers care more about their

²⁹ B. Mandeville (1729), pp. 54-55.

³⁰ B. Mandeville (1732a), pp. 153-154.

³¹ B. Mandeville (1729), p. 60.

³² B. Mandeville (1732a), p. 115.

³³ *Ibid.*, p. 166.

³⁴ *Ibid.*, pp. 63-65, 275; (1729), p. 304.

own preferences even by ruining their own offspring, who then becomes overspoiled and overdependent.³⁵

Mandeville extends his theory of self-interest to physicians, lawyers, tradesmen and beggars among other groups. More importantly, he extends his theory to charity. He argues that the rich and famous help the poor and needy to pride themselves on charitable behaviour, while they are also eager for flattery, honour and the public tribute. He explains that charity is given to hospitals, orphanages and universities because ‘they are the best Markets to buy Immortality at with little Merit.’³⁶ Charity is a highly relevant case in moral psychology and moral philosophy because it was, and still is considered by some, as definitive evidence on the existence of true sympathetic unselfish behaviour, so any successful refutation of such evidence would have important consequences for the respective moral philosophy and the moral psychology supporting it.

Unlike Mandeville, Hume does not consider if statesmen, mothers, friends and lovers act expecting to be flattered, adored and glorified. This is important because it would allow Hume to advance an argument against any self-interested motivation. His method instead is that of simple enumerative induction, which was criticised by Francis Bacon, who explains that ‘induction proceeding by simple enumeration is a childish affair, its conclusions are unsafe, it opens itself to the threat of the contradictory instance, and generally bases its verdict on facts fewer than necessary, and among these only the ones which are readily available.’³⁷

Both in the *Treatise* and the *Enquiry*, Hume highly praised the work of Newton and Bacon, whose methods he claimed to be following. The lack of consideration to ‘contradictory instances’, that is, to self-interested motives is an important flaw in Hume’s method, who takes the obvious as true making no further enquiry. An expectation for glory and public tribute might be the motivation for Pericles, Marcius and Publius, whereas overindulgence in sexual passions and fun might be the motivation for Henry IV, even at the expense of losing a war. Similar considerations apply to the devoted mother, who might be motivated by the veneration she gets from her child. Because Hume does not consider opposite motivations, the cases he presents provide poor support to his theory leaving it vulnerable to refutation.

Another important weakness of Hume’s argument is the very small number of cases he presents of disinterested actions in both the *Treatise* and the *Enquiry*. He was aware of this because he finishes his defence of disinterested benevolence by claiming that ‘these and a thousand of other instances are marks of general benevolence in human nature’,³⁸ which is a poor justification for not providing further evidence. The cases of Pericles, Marcius, Publius, Henry IV and devoted mothers are all cases of ‘particular benevolence’ delivered to individuals we have a close connection with. In contrast,

³⁵ B. Mandeville (1732a), pp. 75-76.

³⁶ *Ibid.*, p. 265.

³⁷ F. Bacon (1620), p. 163.

³⁸ D. Hume (1772a), pp. 93-94.

‘general benevolence’ refers to all those individuals outside this close group, for instance those living in other countries and any distant place. Hume gives no example of this type of general benevolence; in a footnote he only writes: ‘I assume it is as real, from general experience, without any other proof.’³⁹

Hume’s inductive evidence of disinterested benevolent and humanitarian actions is clearly small, and it is subject to easy dismissal because he does not consider the existence of self-interested motivations and how they could be contested. In contrast, the cases discussed by Mandeville are numerous and diverse, ranging from cardinals and nuns to kings, mothers and charity acts. He also explains suicide through self-interest and extends his theory to other societies such as Spain and Holland.⁴⁰ Besides the two volumes of *The Fable*, there is also *An Enquiry into the Origin of Honour and the Usefulness of Christianity in War* as well as several other essays and books, where Mandeville refined his psychological theory and enlarged the number of cases supporting it. The amount of evidence he provides largely surpasses the evidence supplied by Hume in the *Treatise* and the *Enquiry*. In consequence, Hume’s psychological theory is poorly supported and it therefore is hardly convincing, while Mandeville’s theory is better supported and it therefore is more compelling.

EXPERIMENTUM CRUCIS

If inductive support is considered insufficient to decide the controversy between Hume and Mandeville, it could perhaps be settled by presenting a successful *limiting case* in the critical region of the domain. The *experimentum crucis* could do this. Hume explains that ‘it is easy to attain what natural philosophers, after Lord Bacon, have affected to call the *experimentum crucis*, or that experiment, which points out the right way in any doubt or ambiguity.’⁴¹ The experiment under consideration is a case of *benevolence to enemies*. Because of its exceptional features within the domain of benevolent actions, it becomes a limiting case, even though it actually is not an experiment but rather a case of ‘cautious observation’ and ‘experience’ from records in history.⁴²

Hume describes how Demosthenes, a prominent Greek politician of the fourth century B.C., helped his long-standing enemy Aeschines, who was leaving the city after being sent out to political exile. Demosthenes ‘secretly followed, offering him money for

³⁹ *Ibid.*, p. 92.

⁴⁰ B. Mandeville (1729), pp. 135-137; (1732a), pp. 185-191, 194-196, 209-210.

⁴¹ D. Hume (1772a), p. 37.

⁴² Hume was aware of the impossibility of performing experiments in psychology the ‘new science’ as they were performed at the time in physics. In the introduction to the *Treatise* (p. 6), he explains that ‘we must therefore glean up our experiments in this science from a cautious observation of human life, and take them as they appear in the common course of the world, by men’s behaviour in company, in affairs, and in their pleasures.’

his support during the exile, and soothing him with topics of consolation in his misfortunes “Alas!” cries the banished statesman, “with what regret must I leave my friends in this city, where even my enemies are so generous!”⁴³ The case is presented as proving the existence of true generous sympathetic passions causing benevolent acts, which overpower selfish sentiments of hatred and revenge. Hume concludes that ‘compelled by these instances, we must renounce the theory, which accounts for every moral sentiment by the principle of self-love.’⁴⁴ According to him universal self-interest has thus been refuted and, because benevolence to enemies is a limiting case, the likelihood of finding disinterested benevolence in the middle regions increases.

Compared to the Mandevillian theory, Hume’s conclusions seem naïve and deceiving, or at least incomplete. Again, this is because he does not consider testing his own explanation against the existence of self-interested motivations by asking how much admiration, public tribute, reputation and power or personal relief from remorse Demosthenes would get by giving money to his enemy. An explanation based on a humanitarian act motivated by true sympathy can raise the degree of belief on a true concern for relieving Aeschines’s emotional pain and economic hardship, by showing how the rival explanation based on a self-interested passion could be dismissed. This omission undermines the confirmation value of the evidence Hume is presenting.

Hume should have considered at least as equally likely a self-interested motivation, so that apparent benevolence to enemies could be a calculated act of self-promotion. Indeed, because his theory holds that both self-interest and sympathy cause moral behaviour regardless of the external aspect of it, both of them should in principle be considered as equally likely. Then, a test should be performed or further evidence provided for choosing one cause over the other, which could then become an *experimentum crucis*. Methodologically and epistemologically, a theory with two or more causes is more challenging because the number of tests and the need for evidence as well as the overall uncertainty increases proportionally to the size of the causal set. These are the consequences of multiplying unnecessarily the number of unobserved causes, in this case self-interest and disinterested benevolence, which is a violation of the first rule from Newton discussed above. A further and potentially more damaging problem arises from Hume’s decision to stop at the most obvious explanations of moral behaviour which are accepted in ‘common life’, that is, folk psychology explanations. This problem is discussed below in the next section.

⁴³ D. Hume (1772a), p. 36.

⁴⁴ *Ibid.*, p. 37.

SIMPLICITY

Hume argues that a theory which holds both self-interest and disinterested benevolence is simpler than a theory based only on self-interest. His argument consists of two parts, the first one proceeds by a direct comparison between self-interested and disinterested actions; the second part relies on an analogy with secondary self-interested passions.

First part. Hume actually criticises simplicity as a criterion by expressing doubts on the prospects for accomplishing in moral psychology the ‘perfect simplicity’ observed in physics. In spite of these reservations, he still insists that a theory with two fundamental motives is simpler or, more precisely, imperfectly simpler.⁴⁵ The ‘selfish theory’, he claims, is more complex because it uses ‘very intricate and refined reflections’, where ‘metaphysical regards to a self-interest’ are ‘twisted and moulded, by a particular turn of the imagination’ of the scientist, so it can explain a ‘variety of appearances’.⁴⁶ Hume accepts that people may deceive themselves with respect to the ‘predominant motive or intention’, which is ‘indeed, frequently concealed from ourselves, when it is mingled and confounded with other motives, which the mind, from vanity or self-conceit, is desirous of supposing more prevalent.’⁴⁷ However, he thinks that theoreticians such as Mandeville have gone too far by inferring self-interested motivations. All this makes the selfish theory more complex in the theoretical or ‘imagined’ descriptions it provides of the inner workings of self-interest. Because of such descriptive complexity, Hume argues that the theory is ‘fallacious’⁴⁸, that is, false. By contrast, folk psychology—which he calls common life’ reflections or reasonings⁴⁹—offer simpler explanations, and because of this simplicity it should be considered as the probable true psychology of moral behaviour:

Many an hypothesis in nature, contrary to first appearances, has been found, on more accurate scrutiny, solid and satisfactory [...] that there is a general presumption for its arising from the causes, which are the least obvious and familiar. But the presumption always lies on the other side, in all enquiries concerning the origin of our passions, and of the internal operations of the human mind. The simplest and most obvious cause, which can there be assigned for any phenomenon, is probably the true one.⁵⁰

Hume recognises that physics has succeeded in going beyond first appearances. However, he believes that going beyond the most obvious causes is a methodological

⁴⁵ *Ibid.*, p. 92.

⁴⁶ *Ibid.*, pp. 91, 93.

⁴⁷ *Ibid.*, p. 93.

⁴⁸ *Ibid.*, p. 93.

⁴⁹ *Ibid.*, p. 45; (1772b), p. 121; (1739-40), p. 307; see also D. Livingston (1984) and A. Baier (1991), pp. 1-27.

⁵⁰ D. Hume (1772a), pp. 92-93.

mistake in psychology, mainly because of the ‘abstruseness’ of the alleged motive and its functioning. He illustrates this by explaining how such ‘imagined’ functioning is false in the case of a rich man and patron who is grieving at the death of a poor man, who was also his friend: ‘how can we suppose, that his passionate tenderness arises from some metaphysical regards to a self-interest, which has no foundation in reality? We may well imagine, that minute wheels and springs, like those of a watch, give motion to a loaded wagon, as account for the origin of passion from such abstruse reflections.’⁵¹ Mandeville uses the analogy with the spring-watch in the inference to unobservables and, in the analogy he makes between anatomy and psychology, he insists in paying attention not to the obvious parts such as bones, muscles and nerves but to the minute parts, that is, the ‘small trifling Films and little Pipes that are either over-look’d, or else seem inconsiderable to Vulgar Eyes.’⁵²

In contrast, the ‘common life’ psychology pays attention to obvious causes when it explains that because the patron is rich, it is unlikely or false that he grieves the death of a poor friend because of self-interest. The same folk psychology also explains that if the rich man and patron dies, the poor man falsely ‘may flatter himself, that all his grief arises from generous sentiments, without any mixture of narrow or interested considerations.’⁵³ The epistemic standards of ‘common life’ psychology establish that the ‘simplest and most obvious cause’ is the *vera causa*, the true cause.

Again, following Mandeville’s method, the alternative hypothesis with non-obvious concealed motivations explaining the same behaviour must also be tested with the information available. For instance, it should be considered if the same false or self-deceptive flattery Hume places on the poor man grieving a rich friend and patron should also be considered in the case of the rich man, who may flatter himself for grieving the death of a man who is actually poor, dull and ignorant.

Hume is right suggesting that by taking a step into a deeper explanatory level the risk of failing increases. Nonetheless, physics had proved the success of taking this methodological step with the many scientific advances it made. Only by disregarding this success as Hume does, his own argument can gather some support. Why psychology should be different from physics? He would have to answer this question in order to gain more support for the division he makes by placing early modern moral psychology into a separate category. The philosophical debate on such a divide between social and natural sciences remains open, and the current advances made by both behaviouristic and cognitive psychology have not solved the problem yet.

With Mandeville, Hume shares a commitment to produce a metaphysics of human nature, that is, a psychology postulating unobservable entities and processes. Because of this commitment they both share the same epistemological and methodological challenges and risks. However, the argument Hume puts forward rejecting on the one

⁵¹ *Ibid.*, p. 93.

⁵² D. Mandeville (1732a), p. iii.

⁵³ D. Hume (1772a), p. 93.

hand the metaphysical descriptions of the selfish theory, and accepting the metaphysics of folk psychology on the other is weak and unpersuasive. Concealed self-interested motives are a main challenge to his theory, they are hardly considered and the few occasions when they actually are, they are quickly and unconvincingly dismissed. Those concealed motives were the main contribution from Thomas Hobbes and Bernard Mandeville to the explanation of moral and political behaviour.

Second part. Hume also presents an argument by analogy trying to prove the simplicity of his theory. He explains that ‘If we consider rightly the matter, we shall find, that the hypothesis, which allows of a disinterested benevolence, distinct from self-love, has really more *simplicity* in it, and is more conformable to the analogy of nature’.⁵⁴ The analogy is set against primary appetites like hunger and thirst, which have the acts of drinking and eating as their primary ends. When drinking and eating are independently performed without feeling hunger or thirst, both their motivation and pleasure are ‘secondary and interested’. Similarly, the desire for fame and power becomes independent and secondary, although it is derived from the primary passion of ‘self-love, and a desire of happiness.’ In the same manner, acts of benevolence and humanity are performed even when people are not in need, such acts are also secondary. By analogy, if the love of fame and power ultimately derives its pleasure from self-interest; generosity to the prosperous person must ultimately derive its pleasure from disinterested benevolence or, as Hume writes, ‘from the combined motives of benevolence and self-enjoyment’. Therefore, any secondary passion, interested or disinterested, cannot be explained unless the respective primary passions exist.

The comments and remarks I have made earlier also apply to this case, i.e. secondary passions also have to be tested against self-interest. Enumerative induction is too weak, and a single case is even weaker unless supplementary support is provided. Even if the analogy is accepted as it stands now, it is still weak like most analogies are. Further tests or evidence have to be provided, so that the analogy can gain epistemic strength. However, even if the analogy becomes fully warranted, it still remains difficult to accept that a theory with two fundamental motives is simpler just because it relies on the non-abstruse, obvious and easy metaphysics of folk psychology.

CONCLUSION

In sum, on all four criteria for theory choice, namely *vera causa*, inductive support, *experimentum crucis* and simplicity, the psychological theory from Hume achieves a lower score than the psychological theory from Mandeville. Consequently, there are better

⁵⁴ *Ibid.*, p. 94, all quotation in this paragraph are taken from this page.

grounds for believing in a psychology of universal self-interest than in a psychology combining both self-interest and sentiments of humanity and generosity.

Compared to Hume, Mandeville was epistemically more cautious, more rigorous and was more aware of the uncertainty involved in making inferences to unobserved entities and processes in the mind. For tackling the problem of rival explanations, Mandeville's own method has two important steps. The first one consists of a test and a deductive inference, that is, any claim on disinterested benevolence as the motivation for action must be tested against the rival hypothesis, namely self-interested motives. This test takes the form of a refutation. If the rival hypothesis becomes refuted, then via a disjunctive syllogism the alternative one significantly increases its chances of being true. The second step supplements the first one, first by adding a detailed, penetrating and sharp description of how self-interested motives may operate in the particular case under scrutiny, and second by adding numerous cases where selfish motivations are confirmed via a simple induction leading to a generalisation from other similar cases. In contrast, Hume relies on enumerative induction and the number of cases he presents is small. Moreover, the folk psychology he relies on had already been discredited, among others, by Thomas Hobbes⁵⁵ in Britain and François de La Rochefoucauld⁵⁶ in France.

⁵⁵ T. Hobbes (1642) and (1651).

⁵⁶ F. de La Rochefoucauld (1678).

REFERENCES

- Bacon, Francis (1620) *The Instauration Magna Part II: Novum Organum and Associated Texts Novum Organon*. Critical edition by Graham Rees, 2004, New York: Oxford University Press.
- Baier, Annette (1991) *A Progress of Sentiments: Reflections on Hume's Treatise*. Cambridge, MA: Harvard University Press.
- Cartwright, Nancy (1983) *How the Laws of Physics Lie*. New York: Oxford University Press.
- Darwin, Charles (1859) *On the Origin of Species*. Facsimile of the First Edition with an Introduction by Ernst Mayr, 1964, Cambridge, MA: Harvard University Press.
- Descartes, René (1644) *Principles of Philosophy*. Translation by V. R. Miller and R. P. Miller, 1983, Dordrecht: Kluwer Academic Publishers.
- Greig, J. Y. T. (1932) *The Letters of David Hume, vol. I*. Oxford: Oxford University Press.
- Harré, Rom (1960) 'Metaphor, Model and Mechanism', *Proceedings of the Aristotelian Society* 60: 101–122.
- Herschel, John (1830) *A Preliminary Discourse on the Study of Natural Philosophy*. Facsimile edition, 2009, New York: Cambridge University Press.
- Hobbes, Thomas (1642) *De Cive*. Critical edition by Howard Warrender, 1983, New York: Oxford University Press.
- _____ (1651) *Leviathan, or the Matter, Form and Power of Commonwealth, Ecclesiastical and Civil*. Critical edition by Noel Malcom, 2012, Oxford University Press.
- Hume, David (1739-40) *A Treatise of Human Nature: Being an Attempt To Introduce The Experimental Method of Reasoning Into Moral Subjects*. Critical edition by David F. Norton and Mary J. Norton, 2007, Oxford: Oxford University Press.
- _____ (1772a) *An Enquiry Concerning the Principles of Morals*. Critical edition by Tom L. Beauchamp, 2006, New York: Oxford University Press.
- _____ (1772b) *An Enquiry Concerning Human Understanding*. Critical edition by Tom L. Beauchamp, 2006, New York: Oxford University Press.
- La Rochefoucauld, François de (1678) *Collected Maxims and Other Reflections*. Critical edition by E.H. Blackmore, A.M. Blackmore, and F. Giguère, 2007, New York: Oxford University Press.
- Laudan, Rachel (1982) 'The Role of Methodology in Lyell's Science', *Studies in the History of Philosophy of Science* 13 (3):215-249.

- Lipton, Peter (2004) *Inference to the Best Explanation*. London: Routledge.
- Livingston, Donald (1984) *Hume's Philosophy of Common Life*. Chicago: The University of Chicago Press.
- Mandeville, Bernard (1729) *The Fable of Bees. Part II*. Facsimile edition, 1988, Indianapolis: Liberty Fund.
- _____ (1732a) *The Fable of Bees: or, Private Vices, Public Benefits. Part I. With an Essay on Charity and Charity-Schools. And a Search into the Nature of Society. And a Vindication of the Book*. With a Commentary Critical, Historical and Explanatory by F. B. Kaye, 1924. Facsimile edition, 1988, Indianapolis: Liberty Fund.
- _____ (1732b) *An Enquiry into the Origin of Honour and the Usefulness of Christianity in War*. Facsimile edition with an Introduction by M. M. Goldsmith, 1971, London: Frank Cass & Co. Ltd.
- Newton, Isaac (1726) *The Principia: Mathematical Principles of Natural Philosophy*. New translation by J. B. Cohen and A. Whitman with a Guide by I. B. Cohen, 1999, California: University of California Press.
- Reid, Thomas (1785) *Essays on the Intellectual Powers of Man*. Critical edition by D. Brookes, 2002, Edinburg: Edinburg University Press.
- Ruse, Michael (1976) 'Charles Lyell and the Philosophers of Science', *British Journal for the History of Science* 9 (2): 121-131.
- Smith, Adam (1790) *The Theory of Moral Sentiments*. Edited by D. D. Raphael and A. L. Macfie, 1976, Oxford University Press. Facsimile edition, 1982, Indianapolis: Liberty Fund.
- Thagard, Paul R. (1978) 'The Best Explanation: Criteria for Theory Choice', *The Journal of Philosophy* 75 (2): 76-92.

* * *