Hume on Justice as a Virtue

Alessio Vaccari

St. Peter's College, University of Oxford

A thesis submitted for the degree of D.Phil. in Philosophy

Trinity 2023

Abstract

This thesis examines Hume's theory of justice as a virtue by addressing some of its textual problems and by highlighting its connection with Adam Smith's and John Stuart Mill's conceptions of justice. The thesis consists of five chapters. The first discusses the complexities surrounding the well-known 'Circle Argument' and argues that justice cannot be considered as a mere habit of action. The second examines the methodological features of Hume's account of justice as an artificial virtue and concludes that it can be seen as a form of vindicatory genealogy. The third examines the role of the 'dark passion' of resentment in the process of moralising justice. The fourth discusses the differences between Smith's and Hume's theory of justice with regard to the roles of resentment and utility. Finally, the fifth essay examines the lines of continuity between Hume's and Mill's genealogies and the points of contact in their conceptions of the virtue of justice.

Word Count: 73123 words

Supervisor: Peter Kail (St. Peter's College)

Second Supervisor: Roger Crisp (St. Anne's College)

Contents

Introduction

Hume considers justice an important part of morality, allowing the selfish and individuals limited in their benevolence to live together in society and to flourish individually and as a community. The analysis of justice is also the most complex part of Hume's moral theory. Indeed, he sees it not only as a set of rules that enable individuals to cooperate, but also as a virtuous trait that is approved because it is useful. However, he does not clearly explain what element of our psychology this trait is identified with, or what exactly its connection with usefulness is. Through a textual analysis of the *Treatise of Human Nature* and the *Enquiry Concerning the Principles of Morals*, the main aim of this work is to explore the various problems surrounding justice as a virtue and, in some cases, to suggest solutions to them.

The interpretation of the textual problems is carried out by favouring a systematic interpretation of the terms used by Hume, that is, by deriving their meanings from their various occurrences in the texts under consideration. In some cases, however, the interpretation also makes use of philosophical tools borrowed from contemporary ethics. In the first two chapters, for example, Hume's theory of justice is illustrated by concepts borrowed from Bernard Williams's conception of genealogy and contemporary virtue ethics. This allows aspects of Hume's treatment of justice that would otherwise remain hidden to be highlighted. The purpose is to provide a richer understanding of the problems and issues, but not to place Hume within these philosophical traditions as their precursor or advocate.

My research focuses on Hume's theory of justice. However, it also includes Adam Smith's and J. S. Mill's conceptions of justice, which are presented in the last two

chapters of the thesis. The purpose of this inclusion is twofold. On the one hand, it aims to highlight the importance and explanatory fertility of some of Hume's ideas, precisely because they are present in the theory of justice of authors who, despite having a different overall conception of morality from Hume, have taken up these ideas precisely because of their ability to solve a number of philosophical problems surrounding the notion of justice. On the other hand, beyond the differences between these authors, the operation also allows us to delineate the contours of a strand of modern reflection on justice that, centred on the positive role of human passions - not only the social ones but also the asocial one of resentment, on sympathy and the notion of general utility - is an alternative to the Hobbesian one centred on a contract.

In Chapter 1, I examine the problems that arise when we consider the motive of justice as an original trait of human nature, in the same way as the natural virtues of love of offspring or compassion. The work in this chapter has several aims. First, through an examination of the well-known 'Circle Argument', I illustrate the main elements of Hume's theory of virtue and explain why a virtuous trait is not reducible to a mere disposition to act in certain ways under certain circumstances, but consists of dispositions to have certain feelings and beliefs that are specific to each virtue. In the same context, by examining the notion of temper, I also argue that the possession of a virtue does not require that it be habitually expressed in behaviour, but is compatible with the fact that it is expressed infrequently. Second, on the basis of an examination of the conclusion of the Circle Argument, I reject the thesis, advanced by some, that justice is not a genuine virtue because, unlike the other natural virtues, it is merely a disposition to act. Third, following a line already taken by some interpreters, I argue that justice is not identifiable with the virtue of prudence or redirected self-interest.

The difficulties of conceiving of the virtue of justice as natural lead Hume to illustrate how the notion of justice arose in human culture. Chapter 2 examines the

characteristics of this account and argue that it can be regarded as a specific form of genealogical explanation. The chapter defends two claims. The first is that Hume does not regard the state of nature as a fiction that serves merely to illustrate the point of the practice of justice. On the contrary, it aims to provide a plausible explanation of how justice has evolved through various stages in human culture, based both on a dynamic conception of human nature and on some general facts about the circumstances in which human beings live. In particular, I examine the account of the origin of the virtuous duty motive of justice, thereby illustrating the origin of our common-sense judgments about justice as a virtue, which are examined in the Circle Argument. Following a general line introduced by Cohon, I argue that Hume's account resolves the problems raised by the Circle Argument regarding the status of justice as a virtuous trait. This renders implausible the thesis that justice is simply a habit of action rather than a virtue. The second thesis of the chapter is that this form of explanation has an important vindictive function. Far from weakening respect for justice, Hume's account provides those who already respect it with additional reasons to continue to do so. The point of the chapter is to show not only that Hume's account of the emergence of justice has a vindictive function, but that it can have it only because of its explanatory function.

Chapter 3 explores the explanatory role of the passions in explaining the emergence of justice, illustrating the role of the 'dark passion' of resentment. The chapter has several aims. The first is to show that Hume has an articulate conception of resentment. The second, beginning with a discussion of contemporary theories of anger and resentment, is to show that this conception is plausible. The third and final aim is to illustrate the role that resentment plays in the final stage of Hume's account of justice, which concerns the transition from the natural obligation of justice to the moral obligation. Here I argue that resentment is not only the crucial explanatory factor that accounts for the emergence of moral blame for injustice, but also, once moral blame is embedded in

our practices, an important secondary motive that can reinforce it. This runs counter to a common interpretation that attributes to Smith, and perhaps to Butler, but not to Hume, the idea that resentment is an explanatory factor for justice. However, *pace* Schwarze, I argue that claiming that resentment underlies moral blame does not mean that it is identical with sympathetic resentment.

Chapter 4 examines Smith's theory of justice as set out in *The Theory of Moral Sentiments*, with particular emphasis on the role played by the notions of resentment, sympathy, and utility. With regard to resentment, I argue that Smith extends the explanatory role of this passion in at least two important directions. The first is that it intervenes not only in the process of moralising justice, but also at a more fundamental and antecedent level that concerns the very emergence of this notion. Unlike Hume, who traces the convention that leads to the invention of rules of justice back to the passion of redirected self-interest, Smith explains both the possibility of recognising behaviour as just or unjust and the general abstention from injustice at work in natural resentment. Second, the contribution of this passion to the moralisation of justice is markedly different from that of Hume. Here, too, the role of resentment becomes predominant. Before civil law, stable respect for justice can be explained only by an evolutionary variant of natural resentment, sympathetic indignation, which Smith elucidates in terms of the relationship between sympathy and the desire for approval. I also argue in the chapter that Smith's conception of justice reserves a non-secondary and crucial role for utility, which is often obscured by the space his treatment reserves for resentment. On this point, too, Smith echoes Hume's treatment. I argue that utility is the notion that allows the scope of justice to be extended to classes of actions whose effects cause no present harm, even though they are contrary to the future interests of the society with which we can sympathise.

Chapter 5 is devoted to Mill's theory of justice in *Utilitarianism*. Although Mill's content of justice is much richer than Hume's, I argue that Mill explains the emergence

of this concept by following Hume's methodological approach. He starts with a problem at the level of the phenomenology of our common experience and our first-order judgments about justice, which he solves through a staged, naturalistic genealogical explanation of the emergence of this notion. Mill's genealogy is based on a conception of human nature which, like Hume's, is centred on sympathy, extended human interests and resentment. Moreover, like Hume, Mill justifies justice on the basis that it promotes the most important part of the general welfare of human beings.

In addition to illustrating these similarities, the chapter also aims to defend three claims. The first is that Mill, like Hume, identifies the virtue of justice as the motive that leads us to obey its rules not instrumentally but as an end in itself. But, as with Hume, this is compatible with the fact that the person who possesses the virtue of justice is reflexively able to grasp the connection it has with the stable interests of human beings. This does not mean, however, that it is this awareness that typically moves the virtuous to observe the practices of justice. Second, unlike Hume, Mill recognises a further role for virtue, which concerns the ability to resolve various conflicts, both within the principles of justice and between them and other moral obligations. These are conflicts Hume seems uninterested in describing. Finally, as the third thesis of the chapter, I argue that the two authors have a very similar position on whether the general inflexibility of the rules of justice is compatible with their violation in particular cases. In this way, the inviolability of justice is compatible with the fact that the virtue of justice requires that these rules be not applied in certain cases.

Chapter 1

The Circle Argument and the Moral Motive of Justice

Hume has a virtue-centred conception of morality that sees the evaluation of actions as depending on the value of the motivational dispositions that cause them. When we ask how we should evaluate motives, Hume's answer is strongly influenced by utilitarian concerns and focuses on the motives' effects in terms of immediate pleasure and/or their overall usefulness to human beings. Justice is a clear example of this conception. Hume sees justice not only as a set of rules that foster cooperation between individuals, but above all as a character trait that is commended for its utility. Justice, however, is also the virtue that raises the most problems for Hume since he never clearly investigated the motivational disposition that constitutes it, nor is it clear what the connection is between justice and utility.

This chapter examines this problem from the well-known Circle Argument, which highlights the difficulties surrounding the nature of this virtuous motive.

1.1 Background

Hume begins his discussion of morality in the *Treatise of Human Nature*[1] Book 3, Part I, arguing that our capacity to make moral distinctions does not depend solely on reason but depends crucially on our sentiments.[2]

Our experience 'must convince us' at every moment (*T* 3.1.2.2/SBN 470) that we find 'the impression arising from virtue' is 'agreeable', 'and that proceeding from vice' is 'uneasy' (Ibid.). This feeling is not the premise from which we infer that something has moral value, but we become aware of virtue and vice precisely by experiencing these pleasant or painful feelings.[3] As Hume famously states, 'To have a sense of virtue, is nothing but to feel a satisfaction of a particular kind from the contemplation of a character'. And he adds 'The very feeling constitutes our praise or admiration' (*T* 3.1.2.3/SBN 471). He refines his conception of these feelings by examining their phenomenological qualities, causes and effects in order to distinguish different forms of approval from moral approval.[4] Through psychological introspection, for example, we know directly that the feeling of approval is constituted by 'particular pleasures' (*T* 3.1.2.3/SBN 471), which are distinguishable from those related to the approval of the aesthetic qualities of a musical symphony or that of an excellent bottle of wine. The

[1] References to Hume's *Treatise of Human Nature* is cited with notations of the form *T j .k.m.n* / SBN *pqr*, the lower-case letters here standing for arabic numerals. Numerals immediately following *T* indicate book, part, section, and paragraph in David Hume 2000. *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton. Oxford: Clarendon Press; numerals following 'SBN' indicate page no. in David Hume 1978 (2nd edn.). *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press.

[2] If it depended on reason alone, this capacity would consist in one of the two activities in which reasoning takes place, i.e. the comparison of ideas or the acquisition of ideas concerning factual data. In the first case, we should be able to make moral judgements even when the relations between ideas apply to inanimate objects (*T* 3.1.1.24/SBN 466-7). In the second, the immorality of an action, such as premeditated murder, would only be ascertainable through ideas, such as those concerning the evil intentions of a criminal planning a murder (*T* 3.1.1.26). Hume argues that both these circumstances are contrary to our experience and concludes that discrimination between moral properties requires the intervention of an 'impression or feeling' (*T* 3.1.2.1/SBN 470).

[3] Cohon, who has characterised Book 3, Part 1, as an epistemology of value, has described this position by arguing that for Hume the basic awareness of vice and virtue is 'a direct apprehension by feeling', emphasising the partial analogy that exists between moral evaluation and sense perception. See Cohon 2008: 103.

[4] On this point, see G. Sayre-McCord 2016: 436.

contrast here does not seem confined to inanimate objects: we might approve of a physical quality, such as the dexterity of a tightrope walker, because of the proportionate elegance of his movements. But even here we should be able to distinguish the pleasurable feeling aroused by the beauty of acrobatic gestures from that involved in a genuine moral judgement of the tightrope walker. A further refinement concerns the distinction between moral approval and approval of what is advantageous. Moral feelings are only aroused when we consider 'a character in general, without reference to our particular interest' (*T* 3.1.2.4/SBN 472), i.e. without considering whether its effects satisfy our personal interests. Hume thus distinguishes moral approval from self-centred feelings aroused by the characters and actions of others by resorting not only to the phenomenological qualities of the pleasures involved, but also through reference to the general perspective from which the character is considered.

Book 3, Part 3 adds a further crucial element to his explanation of moral evaluation by illustrating how his sentiment-based view makes it possible to explain the fact that morality, in accordance with common sense, is a system of judgements that tend to be stable in the life of each individual and to be shared among those who constitute a community. He introduces this treatment as a theoretical problem that is, however, solved through an explanation of the development of morality and its point of view on the world.[5]

The theoretical problem is that the account of moral evaluation in Book 3, Part 1, is unable to explain the intersubjective character of common-sense morality because, as Hume explains from Book 3, Part 3, moral sentiments are typically caused by sympathy, which is an extremely variable mechanism of emotion transmission. As he illustrated in *Treatise* Book 2, sympathy is 'the propensity' that human beings have 'to receive by communication [the] inclinations and sentiments [of others], however different from, or

[5] On this aspect, see Crisp 2019: p. 151. See also Mackie 1980: chap. 7.

even contrary to, our own' (*T* 2.1.11.2/SBN 316). This transmission is influenced by the principles of association of ideas and impressions that Hume carefully described in Book 1. If the general similarity in physical and psychological characteristics between all human beings enables each person to sympathise with his or her fellow human beings, the greater similarity between some people (in tastes, age, nationality) as well as their spatio-temporal contiguity facilitates this process, making the vicarious sympathetic emotions of those closest and most similar to us proportionately more intense. How does the variability of sympathy affect the ability of morality to be stable and shared?

In *T* 3.3.1 Hume explains the causal relation between sympathy and moral sentiments in this way. Character traits that are the object of moral approval can be categorised on the basis of four criteria[6]: they are either advantageous or immediately agreeable to their possessor or to others. More precisely, a virtue is such if it possesses at least one of the following four qualities: (a) it leads to public utility, such as the social virtues of benevolence, charity, generosity and justice (*T* 3.3.1.11/SBN 578; see also *EPM* 2-3); (b) it is useful or advantageous to its possessor, such as discretion, industry and frugality (*T* 3.3.1.24/SBN 587; see also *EPM* 6); (c) it is immediately agreeable to its possessor, such as greatness of mind, courage and benevolence (*T* 3.3.1.28/SBN 590; see also *EPM* 7); (d) it is immediately agreeable to others, such as good manners, sound reasoning, cleanliness and wit (*T* 3.3.1.27/SBN 589-90; see also *EPM* 8).[7] These qualities

[6]It is worth bearing in mind the existence of a certain ambiguity on Hume's part in his use of the terms vice and virtue. In some instances, Hume uses them as if referring to notions that can be described in morally neutral terms. The benevolence of someone is in this sense ascertainable solely by considering his psychology and the effects this disposition has on other people irrespective of the feelings of approval it arouses. As R. Cohon has noted, this position does not entail any particular ontological commitment and is in fact compatible with both moral realism and moral anti-realism. What these two positions diverge on relates to a moral use of the terms vice and virtue, i.e. that which refers, for instance, to the property of benevolence as virtuous. Anti-realists argue that these properties depend on psychological reactions to the perception of the morally-neutral characteristics of benevolence, while realists argue that these properties are independent of our reactions. For a discussion of the meta-ethical issues of moral ontology related to this point, see Cohon 2008: 100-101.

[7] Against a disjunctive reading of this criterion of virtue see Hursthouse 1999: 67-70. She has proposed a conjunctive interpretation of the four criteria that places Hume's theory of virtue in continuity with that of Aristotle. According to others, the test of virtue is a relational one and should be elaborated from the final remark of the section 'Of Goodness and Benevolence' where Hume writes: 'And 'tis a most certain rule,

are all hedonistically constructed, since Hume, as Roger Crisp rightly pointed out, 'is merely distinguishing between qualities which produce pleasure indirectly, and those which do so directly, as soon as they are confronted'.[8] The relation between moral evaluation and sympathy is thus as follows. Every moral spectator is able to detect the causal connection between a quality of a particular person's mind and the direct or vicarious pleasures it produces. It receives through sympathy these pleasures and its vicarious pleasures, in turn, become the approval of that character trait.

Let us return to the objection. Since sympathy is variable in its operations, if our judgements actually depended on it, they would be characterised by that same variability, and we would therefore evaluate a trait that has positive effects on people we know as more virtuous than if its effects were felt by strangers living in a distant country (*T* 3.3.1.14).

As is well known, Hume addresses this problem through the indication of privileged points of view, which he calls a 'common point of view' (*T* 3.3.1.30/SBN 591) To put it very schematically, Hume argues that only the moral sentiments we experience when we consider the effects of a character trait on its possessor or on those who have connexion with him can become the criterion for our correct approval and disapproval and thus the basis for our stable and shared moral judgements.

Through the illustration of the common view, Hume explains how his view of morality is equipped to solve the problems of conflict and communication that arise from a sentiment-based conception of moral judgement.[9] The moral point of view is not that of our partial and idiosyncratic feelings, but is the expression of our impartial sympathy, which creates a shared language about vices and virtue and allows for moral agreement

---

that if there be no relation of life, in which I cou'd not wish to stand to a particular person, his character must so far be allowed to be perfect. If he be as little wanting to himself as to others, his character is entirely perfect. This is the ultimate test of merit and virtue' (*T* 3.3.9/SBN 606). See Abramson 2008: 249-254.

[8] See Crisp 2019: 147.

[9] See Crisp 2019: 151. See also Cohon 2008: 150-158.

and discussion. In relation to this, as we will see more clearly in the following chapters, the common point of view provides, the psychological resources to access solutions to conflicts that arise from the partiality of our natural affections in resource-poor contexts.

The above provides the general framework within which Hume's theory of justice is set. Hume proposes a sentiment-based moral epistemology according to which we become aware of moral qualities through specific agreeable and painful feelings. He proposes a virtue-based conception of the moral value of actions according to which right actions are those produced by virtuous motivational dispositions. Finally, if we ask what they are, Hume's answer is that they are those traits that, when considered from an impartial point of view, tend to produce pleasure in people either immediately or in the long run.

In the next section I will examine in more detail the relationship between the value of motives and that of actions set out in the argument that is known as the Circle Argument and examine how its conclusion influences the status of the virtue of justice.

1.2 The Circle Argument

The Circle argument is one of the best known and most discussed arguments in Book 3, Part 2, of the *Treatise*. Through its two premises and conclusion it addresses three important issues of Hume's moral philosophy, namely the source of the moral merit of actions, the relation between virtue and motivation, and the relation between virtuous motive and duty. The first premise states that the moral merit of an action depends on the moral quality of the motive that caused it. The second premise further specifies the nature of the moral quality referred to in the first premise: to assess the morality of an action or a motive is to determine whether it is vicious or virtuous. The second premise, in

accordance with the first, argues that the viciousness or virtuousness of an action depends on the corresponding viciousness or virtuousness of a motive that cause it. From these two premises, Hume derives what is known in the literature as the 'First Virtuous Motive Principle'[10], i. e. for every type of virtuous action there must be a first virtuous motive that is different from 'the sense of morality or duty' (*T* 3.2.1.8/SBN 479) or the 'regard to the virtue of the action' (*T* 3.2.1.4/SBN 478).[11] Indeed, to claim that the first virtuous motive is the sense of duty or the regard to the virtue of the action would be to fall into circular reasoning. These two types of motivation can become motives for virtuous action only after it is identifiable as virtuous from a non-moral motive different from these two.

After formulating the argument, Hume applies it to a typical case of just behaviuor and concludes that none of its possible motivations satisfy the conditions posed by the Circle Argument, and therefore that justice is not a virtue. As is well known, the argument is the basis for introducing a broader conception of virtue than that stated in the Circle Argument, on the basis of which Hume argues that justice is an artificial virtue.

On the basis of this argument, some commentators have expressed doubts as to whether the notion of virtue plays a crucial explanatory role in Hume's theory of justice. James Harris famously argued that this 'cannot be given a virtue-theoretic construal', but rather should be placed in the tradition that originated in Grotius, centred on the respect of perfect rights concerning omissions rather than actions.[12] This kind of criticism seems to be based on a view of virtue elaborated within the tradition of virtue ethics that draws its inspiration mainly from Aristotle and that does not take much account of the way the language of virtue is elaborated in Hume's moral philosophy. To this end, the first part of this chapter will be devoted to an examination of the Circle Argument, where Hume

[10] See Garrett 2007: 258.
[11] I will explain later the difference between a 'sense of duty' and a 'sense of the virtue of an action'.
[12] See Harris 2010: 25. Harris follows a line that considers Grotius' jurisprudence as fundamental to understanding Hume's theory of justice. This position has been supported, for instance, by Forbes 1975: chap 1-2; Haakonssen 1981: chap 1; Buckle: chap. 5; Stewart 1992: chap. 1.

establishes some crucial theses of his theory of virtue. This examination is preliminary and necessary in order to verify whether, against Harris, justice is not only artificial but, in a crucial sense, also an artificial virtue.

1.2.1 The First Premise of the Circle Argument

Hume begins his formulation of the Circle Argument with a premise that I will call the Motive-centred Claim. This states that the moral merit of an action depends neither on its internal characteristics, e.g. whether or not it respects certain principles or meets formal requirements, nor on its consequences but depends instead on (the moral qualities of) the motives that gives rise to the action. Hume claims:

> 'Tis evident, that when we praise any actions, we *regard only the motives* that produced them, and consider the actions as *signs* or indications of certain *principles in the mind* and *temper*. The external performance has no merit. We must look within to find the moral quality. This we cannot do directly; and therefore fix our attention on actions, as on external signs. But these actions are still considered as signs; and the ultimate object of our praise and approbation is the motive, that produc'd them. (*T* 3.2.1.2/SBN 477, my italics)

What is the status of this thesis? Hume is not explicit on this point. It is certainly a meta-ethical thesis, concerning the general virtue-centred view of morality. However, it is also a thesis that Hume derives from our first-order judgements, from the common-sense morality he describes. Indeed, Hume writes that when we blame someone's

behaviour, we retract our moral reproach if we discover that the seemingly blameworthy action conceals a virtuous motive (*T* 3.2.1.3/SBN 477-78).

The thesis of the primacy of motives over actions has a long history in Western philosophical thought and played a crucial role in Kantian ethics, which attributed the value of behaviour to the intentions of the agent. In eighteenth-century Scottish philosophy, Francis Hutcheson, immediately before Hume, had subscribed to this thesis, arguing that the moral value of actions depends on a single type of motivation, traceable to benevolent desires that are evaluated in proportion to the magnitude of the good they intend to promote. The Motive-centred Claim fits well into this canon, but it also expresses a peculiar point: in order to have moral value, actions need not necessarily be caused by a motive that is permanently manifested in a person's life. This is what follows from Hume's claim that the value of actions derives not only from the 'principles in the mind' but also from what he calls 'temper'. Far from weakening his meta-ethical view, this clarification aligns it more closely with his view of first-order ethics.

The Motive-centred Claim holds that motives have a constitutive relation to and/or are produced by the 'principles in the mind' and 'temper'. This follows from the two theses presented in the first paragraph. If in fact (i) approval of actions depends on approval of motives, and (ii) approved actions are so because they are considered signs of the principles of mind and temper, therefore (iii) these motives are or are produced by the 'principles in the mind' and 'temper' that confer moral value on the actions. This opens up two questions. The first concerns the meaning of the expressions 'principles in the mind' and 'temper'. Are they synonyms or do they have different meanings? If they are different, how do they differ? The second concerns the nature of the link between these psychological elements and motivation. Are 'principles in the mind' and 'temper' species of motives or do they only motivate indirectly by causing other motives?

In order to answer the latter question, it is necessary not only to understand what 'principles in the mind' and 'temper' are, but also what 'motives' are. This is a very difficult task since the range of 'motive' is quite broad in Hume's philosophy encompassing not only, as Don Garrett noted, 'character traits, abilities, dispositions, and recurring passions as well as occurrent desires'[13], but also, as Geoffrey Sayre McCord has pointed out, the objects that cause those psychological factors.[14] For the purposes of this paper, however, it is not necessary to address this second question, and here I will assume that Hume's argument does not depend on the strong thesis that the principles of mind or temper cannot be contingently motivating.[15]

The question I shall therefore deal with in this section is as follows: what does Hume mean by 'principles in the mind' and 'temper' within this context? To answer it, a brief digression on parts of Book II of the *Treatise* is necessary.

The expression 'principles in the mind' generally refers to those general axioms, not further explainable, that form the basis of Hume's science of the mind. In some cases, it may also indicate psychological elements that are further explicable, as when 'principle' is used as a name for sympathy (*T* 3.3.2.2/SBN 592), comparison (*T* 3.3.2.5/SBN 594-5) or authority (*T* 2.1.11.9/SBN 320-1). In accordance with his naturalist framework, principles play different roles which we can schematically divide into two functions.

First, they account for the origin and modes of association of the passions. Principles are, for example, the double relation of impressions and ideas that accounts for the origin of indirect passions (*T* 2.1.4.4/SBN 284; *T* 2.1.5.1/SBN 285), or the fact that direct passions, such as desire, joy, fear 'arise immediately ... from pain or pleasure' and do not need any additional idea to make their entry into the mind (*T* 2.1.2.4/SBN 277).

[13] See Garrett 2007: 257.
[14] See Sayre-McCord 2015: 436-8.
[15] The argument is too strong since Hume's list of virtues is very broad and includes natural virtues such as and wit (*T* 3.3.1.27/SBN 589-90). While the idea that benevolence is constitutively motivating is obviously true, it not clear whether this property also applies to a trait such as wit.

Moreover, principles identify the operating rules of the imagination that facilitate or hinder the transition from one passion to another (*T* 2.2.2.16/SBN 339-40). Principles do not all have the same force: when they conflict with each other, some of them tend to prevail, as when passions mix with each other in contrast to the resemblance of their impressions, but in accordance with the 'whole bent' or 'tendency of action' that characterizes them (*T* 2.2.9.2/SBN 381).

Second, principles denote the passions that exercise control over the will and produce actions. This use of the term principle, which is mostly found in Book 2, Part 3, of the *Treatise* where Hume discusses the direct passions, has different semantic nuances. Sometimes, 'principle' is used to highlight a very precise feature of the passions, namely, the fact that they have a constant union with both the actions they produce and the circumstances in which the actions take place (*T* 2.3.1.4/SBN 401; *T* 2.3.1.17/SBN 406-7). At other times Hume uses 'principles' to highlight the very fact that certain motivating passions, such as instincts, express the uniformity and regularity of the operations of human nature, regardless of 'the difference of sexes, ages, governments, conditions, or methods of education' (*T* 2.3.1.5/SBN 401). This is, for example, the case with sexual desire and the care of relatives for their children, which are the basis of the origin of any social formation (*T* 2.3.1.8/SBN 401-2).

Following this line of reasoning, Hume refers not just to these instincts as 'principles', but to the various direct passions that, divided into calm and violent passions, vie for the domination of the will. He claims that the direction of the will is entrusted to 'two principles' that he identifies with 'calm passions' and 'violent passions' (*T* 2.3.4.10/SBN 418). The former includes either certain natural instincts ('such as benevolence and resentment, the love of life, and kindness to children') or other inclinations ('such the general appetite to good, and aversion to evil, consider'd merely as such') (*T* 2.3.4.8/SBN 417). They typically produce little emotional turmoil in the mind

and are known more for their effects on conduct than for their immediate feelings. The latter, on the other hand, are violent passions that often lead us to act against our own interests, as when we desire the harm of someone who has wronged us, without considering the real advantage we might gain from it (*T* 2.3.4.9/SBN 417-418).

To sum up, the expression 'principles of the mind' that appears in the Circle argument refers to motivating passions which are considered as types and not as tokens of passions. Hume asserts that these principles typically operate in human beings in a way that cuts across the processes of civilisation and culture. Moreover, they have a necessary connection to the circumstances and actions of human beings and thus ground predictions about human behaviour that are extremely useful for our associated lives.

Let us now look at the other notion of which morally worthy actions are signs. Hume does not offer a systematic analysis of 'temper' and so the many nuances of its meaning must be gleaned from the scattered observations in Book 2. Hume first mentions 'temper' when describing a passion that prevails in the individual mind and imparts a direction to its natural variability. He claims:

> All resembling impressions are connected together, and no sooner one arises than the rest immediately follow. 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. 'Tis difficult for the mind, when actuated by any passion, to confine itself to that passion alone, without any change or variation. Human nature is too inconstant to admit of any such regularity. Changeableness is essential to it. And to what can it so naturally change as to affections or emotions, which are *suitable to the temper*, and *agree with that set of passions, which then prevail* ? (*T* 2.1.4.3/SBN 283)

'Temper' describes a set of passions, connected to each other by relations of impressions and/or ideas, which tends to prevail over other desires and passions (*T* 2.1.4.3/SBN 283), making a person inclined to have a certain set of sense impressions, beliefs, and to behave in a certain way.

Note that the above passage is compatible with two different interpretations of the nature of temper. The first is that temper is identified with a set of passions while the second is that temper is a dispositional property that causes that set of passions. The first interpretation can, in turn, take two forms.[16] We can argue that the temper is identical with the impressions a subject has had and the actions he has actually performed that are related to the passions that constitute that specific temper. A joyful temper, for example, will consist of a person actually taking pleasure in his or her thoughts and actions up to that point. But we can instead claim that temper is rather about a person's counterfactual behaviour and thoughts. Thus, the attribution of jealousy to my brother is only true if - and due to the fact that - in the case that his wife went out with a friend, or he did not hear from her by phone several times a day, he would begin to doubt her fidelity and lose his peace of mind. In the latter case, temper is identical with a certain set of mental states and behaviours that would be present in a person's mind should certain circumstances arise.

Unlike non-dispositional interpretations, the dispositional one holds that temper is not reducible to a set of passions, desires and beliefs that the agent actually had or could have under certain conditions, but is identified with the causal powers that produce and explain the occurrence of these mental states. In this manner temper is a real property of the mind whose existence does not depend on being known to anyone, including its possessor. Which of these interpretations is more plausible for Hume?

[16] For this interpretation I follow Miller 2018: 9 ff. Miller in fact distinguishes a dispositionalist interpretation of character traits from a non-dispositionalist interpretation, which in turn distinguishes between a summary view and a conditional view, which I illustrate below in the text.

In a later passage, Hume seems to narrow down the range of possible interpretations. Our temper tends to focus our attention towards mental contents that are related to the passionate episode we are experiencing. This makes us inclined to form certain beliefs that, in turn, determine the mind to establish a double connection, of impressions and ideas, with other passions, making them particularly violent and therefore able to control more easily the will and behaviour. Hume claims:

> [...] a man, who, by any injury from another, is very much discompos'd and ruffled in his *temper*, *is apt to find a hundred subjects of discontent, impatience, fear, and other uneasy passions*; especially if he can discover these subjects in or near the person, who was the cause of his first passion. 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)

The claim that the possession of a certain temper makes us inclined to form certain beliefs rather than others seems to point towards the last two interpretations suggested above. The passage is indeed compatible with both the dispositional and the non-dispositional counterfactual views. There is however a general problem with the idea that Humean philosophy can accommodate the idea of a dispositional property and thus that temper can be dispositional in character. Dispositional language presupposes the possibility of distinguishing the power to experience certain thoughts and desires from the fact of actually experiencing them as occurring thoughts and desires. Although Hume sometimes uses "disposition" precisely to describe certain types of passions, he had, as is well known, serious doubts about the distinction between power and its exercise. In Book 1, for instance, he writes:

> The distinction, which we often make betwixt *power* and the *exercise* of it, is equally without foundation. (*T* 1.3.14.34/SBN 171)

In view of this fact, I will not explore this issue further, and what I say about temper from now on will be compatible with both the dispositional and the non-dispositional counterfactual views.

Regardless of the nature of the temper, the question we should consider is: how often must the thoughts, motivating passions and actions that are the expression of a particular temper operate in an individual mind for that particular temper to be attributed to it? Quite frequently or, at the limit, a very few times? Hume does not explicitly address this question and, where he alludes to an answer, the answer is not clear-cut. Schematically, at least two different scenarios are possible.

The first is the one in which a certain 'set of passions' tends to become predominant in an individual's life, as when a person tends to be governed by calm passions rather than violent ones or vice versa (*T* 2.3.10/SBN 418). In that case, temper has a 'constant union and connections' with the actions of an agent and is identified with his 'natural temper' (*T* 2.1.11.2/SBN 316-17; *T* 2.2.4.7/SBN 354) and his 'general character' (*T* 2.3.3.10/SBN 418). In addition to the discussion of calm and violent passions, this use of 'temper' takes account of some important aspects of our social life. Hume claims, for example, that temper can influence the sympathetic process by affecting the readiness with which the idea of the passion that is the object of the sympathetic episode is transformed into the corresponding impression. Indeed, this conversion tends to take place more easily if 'our natural temper gives us a propensity to the same impression, which we observe in others, and makes it arise upon any slight occasion' (*T* 2.2.4.7/SBN 454). Moreover, temper also explains the particular twist that our social and

emotional interactions can take. Hume points out that although human beings, through sympathy, find the company of other people pleasant, regardless of the degree of relationship they have with them, they tend to associate with each other 'according to their particular tempers and dispositions, and that men of gay tempers naturally love the gay; as the serious bear an affection to the serious' (*T* 2.2.4.6/SBN 354).

As these examples show, temper can denote a set of passions which are typical of a person, i.e. which manifest themselves permanently in the course of his/her life, and thus are part of his/her character. This reading is not at odds with what Hume tells us about the nature of temper, and is indeed compatible with the dispositional and non-dispositional counterfactual views of it.

The second use of temper, however, refers to a disposition to have a particular set of 'motives' that although it is "a constant cause in the mind" of a person '*operates only at intervals*' and therefore 'does *not infect the whole character*' (*T* 2.3.2.7/SBN 411-2, my italics). In this second sense, and unlike the first, temper does not manifest itself steadily in a person's life. Its characteristic thoughts, desires, and actions pass from being dispositional or conditional to occurring only at a distance of time or rarely. This means that in this second sense, temper is not identical with individual character, understood as a stable causal psychological factor of the mind.

Hume uses this notion of temper when he refers to the 'hasty temper' that sporadically causes us to perform evil actions that are done 'without forethought' (*T* 2.3.2.7/SBN 411-12). Or, referring to his mood of despondency resulting from the radical scepticism in which he was trapped in Book I, he yearned for something that would serve to 'compose my character from that spleen, and invigorate it from that indolence, which sometimes prevails in me' (*T* 1.4.7.14/SBN 272-3). Moreover, in line with this second use, Hume clarifies how temper contrasts with the set of passions that express the general character of the agent. While discussing the conflict between violent and calm passions,

Hume argues that although strength of mind implies the 'prevalence of the calm passions above the violent' (*T* 2.3.4.10/SBN 418), no one possesses this virtue so consistently that he or she never yields to violent passion. Indeed, Hume describes this possible variation as a 'variation in temper', suggesting that a particular combination of violent passions (violent temper) may take the place of the '*general character*' of the agent dominated by calm passions (*T* 2.3.4.10/SBN 418).

Having established that 'principles in the mind' and 'temper' have a slightly different meaning we can go back to the original problem posed at the beginning of this chapter. Why does Hume say in the first paragraph of the Motive-centred Claim that actions have merit when they are signs of motives traceable to 'certain principles in the mind and temper'?

One plausible answer is that Hume wants to emphasise that an action is meritorious not only when it is produced by a motive that is a sign of a certain stable character trait, but also when it is a sign of a temper that occurs only rarely or at times. More precisely, the first paragraph of the Motive-centred Claim holds that an action is approved when it is produced by a motivation which is the sign of a certain 'principle in the mind', i.e. a certain type of motivating passion (e.g. benevolence, care of children, etc.), and which indicates the presence of a certain temper, i.e. a disposition which might not coincide with a trait of character but operate only occasionally or very rarely.

This interpretation of the Motive-centred Claim has one advantage two advantages. First, it aligns Hume's ethics with our common-sense morality, which underlies the conception of the right of actions conducted in Book 3, Part 2. It shows that upholding the moral priority of motives does not imply that one cannot morally evaluate an action that is performed for a motive that is somehow out of character. Holding this view would imply arguing that if a person who generally does not care for others in their daily interactions, proves capable of helping them when they are in dramatic

circumstances, those actions, as they are rarely performed, could not be considered morally worthy. The Motive-centred Claim rejects this conclusion. As I have argued, its main thesis about the moral priority of motive over action does not imply any constraint on whether the temper of which the motive is a sign must manifest itself habitually or constantly and not occasionally.[17]

Secondly, this interpretation contributes to the unity of the *Treatise* by adding a further similarity, in addition to those already indicated by Hume, between the causes of moral sentiments and those of indirect passions. Consider his position on the causes of hatred. Hume argues that although a harmful action, considered as such, is too brief to be associated with the idea of its author and thus to arouse the passion of hatred towards him, an intentional harmful action is capable of arousing this passion. The intention survives the action, which is by its nature perishable, and can therefore be placed in imaginative connection with certain qualities associated with the object of hatred, that is, the idea of the aggressor. This relation of ideas, together with the relation of resemblance between the painful impression of harm and that of hatred generates this passion. Hume claims:

> But here we must make a distinction. If that *quality* in another, which pleases or displeases, *be constant and inherent in his person and character*, it will cause love or hatred independent of the intention: But otherwise a knowledge and design is requisite, in order to give rise to these passions. One that is disagreeable by his deformity or folly is the object of our aversion, tho' nothing be more certain, than that he has not the least intention of displeasing us by these qualities. But if the uneasiness proceed not from a quality, but an action, which is produc'd and annihilated in a moment, 'tis necessary, in order to produce some relation, and connect this action sufficiently with the person, that it be

[17]As we shall see later by examining the second premise of the Circle Argument, this is not all that Hume has to say about temper making action morally worthy. In particular, every single episode of temper must be an expression of a token that has certain effects on its possessor or on other people.

> deriv'd from a particular fore-thought and design. 'Tis not enough, that the action arise from the person, and have him for its immediate cause and author. This relation alone is too feeble and inconstant to be a foundation for these passions. It reaches not the sensible and thinking part, and neither proceeds from any thing *durable* in him, nor leaves any thing behind it; but passes in a moment, and is as if it had never been. On the other hand, an intention *shews certain qualities*, which remaining after the action is perform'd, connect it with the person, and facilitate the transition of ideas from one to the other. We can never think of him without reflecting on these qualities; unless repentance and a change of life have produc'd an alteration in that respect: In which case the passion is likewise alter'd. This therefore is one reason, why an intention is requisite to excite either love or hatred. (*T* 2.2.3.4/SBN 348-349, my italics)

Hume argues that although a harmful action that is disconnected from the motive that caused it cannot arouse hatred, an intentional action, although out of character, is able to arouse that passion. Why? Because intentional actions typically indicate certain mental qualities. What are these qualities of the agent's mind that sustain an intention and are not 'constant and inherent' in its character? The most plausible answer is that these qualities are signs of the agent's temper that, although it is imaginatively connected with his mind, does not necessarily have to occur so permanently as to be identified with the character. And this is exactly what happens with morally reprehensible actions which are signs of a certain temperament but not of character.

1.2.2 The Second Premise of the Circle Argument

The second premise of the Circle Argument specifies the nature of praiseworthy action mentioned in the Motive-centred Claim: Hume is interested in explaining our moral approval of virtuous action.

> It appears, therefore, that all *virtuous* actions derive *their merit* only from virtuous motives, and are consider'd merely as signs of those motives. (*T* 3.2.1.4/SBN 478, my italics)

The meaning of the second premise is elaborated a little further on, where Hume introduces the case of a person mocking virtuous behaviour. He claims:

> But may not the sense of morality or duty produce an action, without any other motive? I answer, It may: But this is no objection to the present doctrine. When any virtuous motive or principle is common in human nature, a person, who feels *his heart devoid of that principle*, may hate himself upon that account, and may perform the action without the motive, from a certain sense of duty, in order to acquire by practice, that virtuous principle, or at least, to disguise to himself, as much as possible, his want of it. A man that really feels no gratitude in his temper, is still pleas'd to perform grateful actions, and thinks he has, by that means, fulfill'd his duty. (*T* 3.2.1.8/SBN 479, my italics)

The second premise, read in conjunction with the passage above, introduces an important restriction in the moral evaluation of actions. There is a moral difference between a genuinely virtuous action, (e.g. one motivated by gratitude) and an action that only appears to be so but in fact it is not because it is motivated only by a sense of duty or of the morality of an action.[18] Hume mentions here a well-known issue that goes back at least to Aristotle's treatment of the distinction between doing the virtuous thing and acting virtuously, which is typical of children taking their first steps in virtue (NE II.4.17-

[18] As I will argue in a moment, examining the conclusion of the Circle Argument, these two motives are not identical for Hume although in some cases it might appear that he believes they are.

33)[19]. Are the sense of the morality of an action and the sense of duty the same motive or are they two different motives? The sense of duty, as Hume mentions above, is that motive that impels us to perform an action because of the fear that not performing it would show others that we lack a virtuous motive. Duty is therefore the motive that impels us to perform a virtuous action to avoid the pain of others' moral disapproval.[20] Instead, the sense of the morality of an action seems to refer more generically to the motive that drives us to perform an action when, lacking the virtuous motive, we perform it not only to avoid the pain of others' disapproval but also merely to acquire the pleasures of approval. Rachel Cohon considers the two motives essentially identical, while Geoffrey Sayre-McCord argues instead that the identification is implausible because while it is certainly true that 'all actions that are a duty are virtuous' it is not equally true that 'every virtuous action will be a duty'.[21] The solution proposed by Sayre McCord seems more in line with Hume's extremely extensive list of virtues, which include not only qualities that promote the interest of society but also merely agreeable qualities, such as irony or eloquence, which it would seem odd to consider as qualities that, if lacking in a character, would constitute a source of moral blame. Hume does not seem interested in emphasising any distinction between the two motives here. However, as we shall see shortly, the distinction may be operative in the conclusion of the Circle Argument, where the motive of duty is no longer mentioned and only the reference to that of the 'sense of morality of an action' remains.[22]

Before examining the conclusion of the Circle argument, let us dwell again on the distinction between virtuous action and acting according to virtue. This sheds light on a further aspect of the notion of virtuous disposition that we only partly mentioned when

[19] The reference edition of *Nicomachean Ethics* is Aristotle 2000, *Nicomachean Ethics*, edited and translated by Roger Crisp. Cambridge: Cambridge University Press.
[20] On this motive see Sayre-McCord 2015: 438.
[21] Ibid.
[22]This position is also held by Sayre-McCord 2015: 438.

examining the Motive-centred Claim. As the example of gratitude shows, virtue typically consists of a motivational disposition that does not coincide with a tendency to act in certain ways given certain circumstances, but also includes the tendency to experience certain passions and beliefs that explain why the agent in those circumstances acted in a certain way rather than another.[23] If the virtuous motive of gratitude were merely an expression of a blind disposition to act, it would not be possible to distinguish between virtuous actions that express gratitude and non-virtuous actions that merely mimic gratitude.[24]

That said, one must proceed with some caution. How far this thesis can be generalised to all virtues is indeed a matter of debate. Annette Baier has, for instance, argued that the thesis does not apply to the humean virtues of frugality and diligence, which in fact would be nothing more than habits of action. [25] In the same vein, Rachel Cohon, questioned whether natural abilities such as wit or intelligence are made up of emotional dispositions.[26] Granting these cases, I believe that they remain exceptions to his general theory of virtue and that the mere disposition to repeat of a type of action is not a sufficient property for the possession of a virtue. Although Hume is not explicit on the point, this general thesis can be derived indirectly in several places, the clearest of which is section 7 of the *Enquiry Concerning the Principles of Morals*[27], where he explains how a trait can be approved from the sympathy spectators have for the agreeable psychological states of the agent that constitute the trait under evaluation. Note that the

[23]A contemporary Humean version of the elements that constitute a virtue can be found in the description of what Bernard Williams calls the 'S' or 'subjective motivational set'. See Williams 1981: 101-113.
[24] On this aspect, see Cohon 2020: 146. See also Sayre-McCord 2015: 437.
[25] See Baier 2010: 68
[26] See Cohon 2015: 146.
[27] Hume's *Enquiry concerning the Principles of Morals* is cited with notations of the form *EPM* followed by two numbers indicating the chapter and paragraph in David Hume, D. 1998 [1751]. *An Enquiry concerning the Principles of Morals*, ed. Tom L. Beauchamp, The Clarendon Edition of the Works of David Hume. Oxford: Oxford University Press, 1998. Numbers following 'SBN', indicate the corresponding page in David Hume 1975 [1751]. *Hume's Enquiries*, ed. L. A. Selby-Bigge and P. H. Nidditch, 3rd edn. Oxford: Oxford University Press.

traits assessed in this way constitute a very large group that includes most of the traits the *Treatise* classifies under the headings 'Of greatness of mind' (*T* 3.3.2) and 'Of goodness and benevolence' (*T* 3.3.3). Thus, benevolence is characterized not only by the actions it typically yields, but also by the 'very softness and tenderness of [its] sentiment' and 'its delicate attention' towards those to whom it is addressed (*EPM* 7.19/SBN 257).[28] Similarly, a constitutive element of the 'GREATNESS OF MIND' is the awareness of excellence in one's own abilities that manifests itself in the passion of 'noble pride' that characterizes this virtue (*EPM* 7.4/SBN 252). Again, the virtue of magnanimity is only described as a habit of action, but from 'serenity and contentment' in the face of adversity and in the 'care of preserving liberty' (*EPM* 7.17/SBN 256).

If this interpretation is correct, the two premises of the Circle Argument hold that actions that have moral value are virtuous actions, the value of which depends entirely on that of their motive. The motive that has moral value or virtuousness possesses two characteristics: (a) it is different from the sense of duty or morality of the action and (b) it is an expression of a virtuous disposition that is typically not identical with a mere disposition to feel certain desires. Virtue is a complex psychological state that typically includes dispositions to experience certain passions, beliefs and certain pleasant feelings as well as the ability to pay attention to certain aspects of the circumstances of action. Moreover, these elements do not appear to be identical for each virtue, they change according to the behavioural profile that is associated with each virtue.[29]

Let us now examine more precisely what is the moral status of the motives of duty and the sense of the morality of an action. Hume distinguishes them from genuinely

[28] R. Cohon 2008: 146.

[29] See on this point, Abramson 2002: 305. Abramson has rightly argued that this aspect of Hume's theory of virtue should lead Humean scholarship to look with suspicion on a simplistic conception of the division between natural and artificial virtues according to which natural virtues are simply original instinctive dispositions inscribed in human nature. She insisted that these dispositions need to be educated in order to be virtuous. In particular, the different cognitive and affective elements that make them up need to be educated. On this same subject, without the polemical overtones that characterise Abramson's position, see also, R. Cohon 2008: 162.

virtuous motives. The reason for this distinction is made clear in the conclusion of the Circle Argument that we are now going to examine.

1.2.3 The Conclusion of the Circle Argument. The Problem with the Virtue of Justice

The Circle Argument conclusion consists of two propositions which are in a relation of implication. The first claim holds that

> The first virtuous motive, which bestows a merit on any action, can never be a *regard to the virtue* of *an action*, but must be some other *natural* motive or principle. (*T* 3.2.1.4/SBN 478, my italics)

From this claim, Hume derives a further proposition known as the undoubted maxim

> In short, it may be established as an undoubted maxim that no action can be virtuous, or morally good, unless there be in human nature some motive to produce it, distinct from the *sense of its morality*. (*T* 3.2.1.7/SBN 479, my italics).

The two claims make explicit what is the nature of the motive that bestow moral merit on actions. Hume calls it the 'first virtuous motive' to emphasise that he is examining a motive whose knowledge is a necessary condition for identifying a type of action as an expression of a particular virtue (brave, grateful, etc.). As noted in the previous paragraph, once actions are identifiable by means of this description, an agent can perform them even if he lacks this motive simply because of 'regard for the virtue of an action'.[30] The

[30] See Sayre-McCord 2015: 438.

conclusion argues that this second type of motivation can only be operative once the first, i.e. the first virtue motive, is present in human beings and that therefore the two motivations do not coincide.

Note that in the first paragraphs of his discussion of justice (*T* 3.2.1.4-8), Hume mentions both motivations 'sense of duty' and 'regard to the virtue of the action' (in its variants, 'sense of its morality' or 'regard to this merit'). In the two claims that constitute the conclusion of the Circle Argument, however, 'sense of duty' disappears. This may indicate that Hume wants his 'undoubted maxim' to have the widest possible scope so as to cover all the non-virtuous motives that through moral sentiments prompt us to perform an action that mimics virtuous ones.

Importantly, the conclusion characterises the 'first virtuous motive' as 'natural', without further specifying the meaning of this notion. As is well known, in *Treatise* Book 3, Part 1 Hume distinguishes five senses of 'natural' according to whether it is opposed 'to miracles', 'to what is rare and unusual', 'to artifice' i.e. the product of human ingenuity, 'to civil' or 'to moral' (*T* 3.1.2.7.9/SBN 474-5). In the context of his discussion of justice, on the other hand, particularly in fidelity to promises, the meaning of 'natural' is restricted to the first three oppositions and in particular to non-artificial. It is plausible then to hold that the meaning of 'natural' in the case of the 'first virtuous motive' must be derived from one of these contrasts. Since in this part of Book 3 Hume is about to characterise justice as an artificial virtue, one might think that the first virtuous motive is natural insofar as it is not artificial. [31]

[31] Some might argue that this interpretation seems to be confirmed by one circumstance. Immediately after the first claim of the Circle Argument conclusion, Hume exemplifies the first virtuous motive with the father's "natural affection" for his child (*T* 3.2.1.5/SBN 478). This is an instinctive propensity that human beings feel for their children that is not subordinate to what others do and is therefore independent of participating in a particular practice. One could then conclude that if Hume exemplifies the natural virtue-imparting motive with a non-artificial motive, then this is the meaning of 'natural' which is at stake in the conclusion. This argument is however not conclusive. The contrast between the naturalness of the motive of love for children and the motive of duty is not along the artificial/non-artificial but moral/non-moral line. The duty of care for children is in fact characterised as a motive that depends on the activity of moral

This hypothesis is not fully convincing. Indeed, at this stage of the argument Hume has not yet explained the distinction between natural and artificial virtues, which will be examined from *Treatise* 3.2.2 onwards. Therefore, it seems implausible that Hume is characterising the first virtuous motive using a contrast that he has only mentioned, but has not yet discussed. As Rachel Cohon has argued, this is not the only hypothesis on the table. Another possible interpretation is that the relevant meaning of 'natural' is not to be found in the opposition to artificial, but to moral. According to this hypothesis,-the first virtuous motive is natural because its existence does not depend on the activity of our moral sentiments, that is on approving the motive of an action from the common point of view.

Cohon's hypothesis is more consistent with the content of the conclusion, which is about the contrast between a type of motive, i.e. 'regard to the virtue of an action' or 'sense of its morality', that depends on the activity of moral sentiments and types of motive that instead cause those feelings, i.e. the first virtuous motive. With this meaning of natural in place, the conclusion of the Circle Argument holds that the first virtuous motive of a type of action, e.g. a compassionate action, is a non-moralised disposition such as solicitude towards the suffering of others, which must be identifiable by an impartial spectator before it can become, through impartial sympathy with its tendential effects, the object of moral approval and, as a result, a virtuous disposition. By shifting the focus from the spectator to the agent, the conclusion argues that the motive of solicitude towards the suffering of others must be operative and intelligible within our moral practices before someone can be motivated to perform a compassionate action out the concern for the morality of compassionate action. Hume indicates a powerful reason

feelings (avoid moral blame) whereas natural love for children does not. Therefore, as I will argue shortly, the example supports an interpretation of 'natural' as non-moral rather than non-artificial.

for accepting the undoubted maxim: if the contrary thesis were accepted, one would end up with a form of circular reasoning.

> To suppose, that the mere regard to the virtue of the action, may be the first motive, which produc'd the action, and render'd it virtuous, is to reason in a circle. (*T* 3.2.1.4/SBN 478)

Circularity depends on the fact that if we were to violate the indubitable maxim, we would be forced to define a virtuous action as that action which is produced by compliance with the virtuous action. As Sayre-McCord has argued, this kind of circularity is problematic for two reasons. First, the motive of duty would be devoid of content and therefore could not cause any particular action, let alone a virtuous action. Second, Hume could not provide any explanation for the virtuousness of the action and his virtue-based theory would be incapable of serving its intended explanatory purposes.[32]

The conclusion of the Circle Argument poses a serious threat to the thesis that justice is a virtue. This is because the motive of justice is apparently incapable of satisfying the undoubted maxim: there is no non-moral motive that can be a stable cause of the kind of actions that are in the class of just actions. Hume claims:

> From all this it follows, that we have *naturally* no real or universal motive for observing the laws of equity, *but the very equity and merit of that observance*; and as no action can be equitable or meritorious, where it cannot arise from some separate motive, there is here an evident sophistry and reasoning in a circle. (*T* 3.2.1.17/SBN483)

The only motive seems to be the motive of respect for the morality of justice. So, and this is the difficulty, the claim that justice is a virtue involves running into the circular

[32] See Sayre-McCord 2015: 440-441. For a similar argument, see also Cohon 2008: 170.

reasoning we have just indicated. The problem is exacerbated by the fact that Hume throughout Book 3, Part 2, never makes either of the two claims that could have really solved the problem. On the one hand, he never challenges the position of first-order ethics that justice is a character trait that elicits our moral approval and not simply a set of useful rules. On the other, he never distances himself from the thesis that the indubitable maxim applies to all virtues and not just the natural ones where justice has no place.

As is well known, Hume's so-called solution consists in stating that justice is an artificial virtue, i.e. it is a character trait that produces pleasure and is approved by some human artifice (*T* 3.2.1.1/SBN 477). However, as Rachel Cohon has pointed out, it is doubtful that this can be regarded as a solution to the problem posed by the Circle Argument: artificiality, in itself, does not explain why common sense believes that all virtues respect the undoubted maxim.

In Chapter 2 and partially at the end of this chapter, I will try to argue that one possible explanation for this is that from the point of view of our first-order judgments, respect for just actions is regarded as an end in itself and not as a means to the general benefit of society. In our daily practices, we do not act justly nor do we approve of our behaviour because we take the anatomist's point of view, which explains the emergence of justice as an effective solution to a fundamental problem of human coordination over scarce goods. That is, we do not believe that justice is the effect of a series of artifices, but treat it as a motive inscribed in human nature on a par with caring for one's offspring or gratitude to our benefactors.

Before proceeding further, however, it is appropriate to dwell on why Hume believes that no non-moral motive identifies the virtue of justice. Hume's explanation is not otiose, but highlights two important features of the motivational disposition that constitutes this virtue.

Hume considers a typical case of an action that common-sense morality would consider just, namely the 'restoring a loan' (*T* 3.2.1.9/SBN 479-80) and excludes the possibility that it can be identified with any of the non-moral motives that he considers possible candidates for this role, namely 'self-love', '*regard to public interest*' and '*private benevolence*' (*T* 3.2.1.10-16/SBN 480-483). Each of these motives, although it may cause just action in some circumstances, is not capable of causing what Don Garrett has called the 'full behavioural profile' of justice i.e. 'the whole set of actions required by the virtue of justice'.[33] Hume therefore does not exclude the possibility that in some circumstances both private and public benevolence or self-interest cause behaviour that complies with the rules of justice. This is not sufficient to identify those motives with the virtuous trait of justice. From what Hume says, those natural motives do not in fact make the just actions they cause virtuous, despite the fact that these actions supposedly respect the rules of justice. The virtue of justice, i.e. that trait that we admire from the common point of view and that confers moral merit on the actions it produces, is only that motivational disposition thanks to which we act justly in the many different circumstances that are provided for by what Hume will explain to us to be the rules of justice.

It is undoubtedly true - as Harris claims - that there is no single possible motive behind behaviour that complies with the rules of justice. However, this does not mean - *pace* Harris - that there is no single motivational disposition that is identified with the virtue of justice. This disposition is in fact the one that is capable of causing just actions in all the circumstances in which justice applies. It is precisely this trait that elicits moral approval from the common viewpoint and it is therefore this that identifies the virtue of justice.

[33] See Garrett 2007: 265.

The second characteristic that emerges from this examination concerns the particular force that characterises this motive. The case under discussion, which he identifies as a typical case of just action, is that of the repayment of a loan. Hume could of course have chosen a different case, such as abstaining from another's property. But this type of case would not have highlighted the fact that behaving justly is not always undemanding but entails costs in terms both personal and moral advantage.[34] Hume observes, for example, how repayment may be contrary not only to our advantage, but also to virtuous behaviour moved by benevolence or compassion. For example, repayment might force me not to provide for the material welfare of my family or friends at a time when they are in distress (*T* 3.2.1.13/SBN 482). Again, just action could be contrary to benevolence when I know that the sum of money, once returned, will do more harm than good to my creditor (*T* 3.2.1.13/SBN 482). In these conflicts the just action must never yield before other virtuous actions that are morally appreciable on the side of natural virtues. Indeed, justice entails inflexible rules (*T* 3.2.3.3/SBN 502-3; see also *T* 3.2.6.9/SBN 531-3; *T* 3.2.6.10/SBN 533). Of course, this does not mean that in order to possess the disposition of justice I must always act justly whenever circumstances require it and that an act of injustice proves that I do not possess this virtue. But it means that the possession of this virtue entails a motivational disposition deeply rooted in individual character that makes us feel such profound disgust towards injustice that it enables us to act contrary to what other virtues of our character would require us to do in that situation. It is precisely this deep-rooted trait that arouses our moral admiration and that we approve of from the common point of view.

Having made these clarifications, we need to examine the way in which Hume explains the emergence of the moral motive for justice and the controversial question of

[34] This point was also highlighted by Annette Baier who also reflected on the complicated cultural implications of this practice related to intolerance towards the Jewish people. See Baier 2010: 25-26.

whether having this disposition is actually beneficial for everyone. Before examining these issues, however, it is appropriate to briefly introduce what the distinction between natural and artificial virtues consists of.

1.3 Natural and Artificial Virtues

As we have already seen, in both the *Treatise* and the second *Enquiry* Hume distinguishes the virtues on the basis of the four criteria of the useful and the immediately agreeable to oneself and others, thus explaining the sources of moral approval. In the *Treatise*, but not in the second *Enquiry*, the virtues are also notoriously distinguished into natural and artificial. The distinction has been the subject of extensive debate in recent years, the extent of which, as Kate Abramson has correctly observed, depends mainly on the fact that Hume scholars 'have grossly exaggerated the "naturalness" of Hume's natural virtues'.[35] The overabundance of meanings attributed to 'natural' in the natural virtues has led to an increase in the levels of comparison between these virtues and the artificial ones. Indeed, Abramson distinguishes at least eight meanings of natural virtues that have been contrasted with as many meanings of artificial virtues. She claims:

> According to the traditional interpretation, natural virtues are: (1) dispositions whose objects, scope, force, and expression are dictated by the constitution of the human mind as such; (2) unalterable, or nearly so; (3) motivating dispositions in which the agent's conception of value does not play an essential role; (4) common; (5) culturally invariant; (6) traits consisting neither wholly nor partly of general rules; (7) traits whose approval does not necessarily refer to the general conformity of persons in a particular 'scheme or

[35] See Abramson 2015: 333.

system of action'; and (8) dispositions whose content can be specified without reference to a particular 'scheme or system of action'.[36]

As Abramson notes, the legitimacy of the first seven meanings is a matter of heated debate, and only the last appears to be uncontroversial. For the purposes of this chapter, I will leave aside the complexities associated with the controversial meanings of natural and assume that the distinction between natural and artificial virtues can be understood along the lines indicated by Abramson's eighth meaning of natural. Following this, I shall assume that the distinction under consideration basically concerns two points. The first is whether the positive effects produced by individual virtuous actions are available irrespective of the number of people performing such actions.[37] The second is whether the practical content of a virtuous disposition, what it typically prompts one to do, is identifiable independently of the rules that emerge from the patterns of actions on which we all agree. Natural virtues are motivational dispositions whose characterisation entails a positive response to both points. By contrast, the characterisation of artificial virtues entails a twofold negative response. The overall utility produced by these virtues is only available when they are shared by a sufficiently large number of agents; similarly,

[36] K. Abramson (2015): 333-334.

[37] Hume claims: 'The only difference betwixt the natural virtues and justice lies in this, that the good, which results from the former, arises from every single act, and is the object of some natural passion: Whereas a single act of justice, consider'd in itself, may often be contrary to the public good; and 'tis only the concurrence of mankind, in a general scheme or system of action, which is advantageous. When I relieve persons in distress, my natural humanity is my motive; and so far as my succour extends, so far have I promoted the happiness of my fellow-creatures. But if we examine all the questions, that come before any tribunal of justice, we shall find, that, considering each case apart, it wou'd as often be an instance of humanity to decide contrary to the laws of justice as conformable to them. Judges take from a poor man to give to a rich; they bestow on the dissolute the labour of the industrious; and put into the hands of the vicious the means of harming both themselves and others. The whole scheme, however, of law and justice is advantageous to the society and to every individual; and 'twas with a view to this advantage, that men, by their voluntary conventions, establish'd it'. (T 3.3.1.12; SBN 579; see also *T* 3.2.2.22; SBN 497). On this point, see Árdal 1966: 117; Baron 1982: 541–42; Baier1991: 231; Haakonssen 1993: 186; Darwall 1995: 292–93; Magri 1996: 233–34; Abramson 2015: 353.

these motivational dispositions would be devoid of practical content if there were no pattern of action on which actors converge.[38]

The list of natural virtues is extremely broad and includes both qualities that are typically approved because they are useful and advantageous to others or to those who possess them and because their mere appearance, regardless 'of reflections on their tendency to the happiness of mankind', arouses immediate approval. To the first group belong what he calls 'social virtues', i.e. qualities such as 'Meekness, beneficence, charity, generosity, clemency, moderation, equity' (*T* 3.3.1.11/SBN 578-9), which are in fact approved of because they have the tendency to promote the good of others. It also includes qualities such as '*prudence, temperance, frugality, industry, assiduity, enterprize, dexterity*' (*T* 3.3.1.24/SBN 587), which make us useful to ourselves and enable us to promote our own interest. To the second group belong, on the other hand, both qualities such as 'wit, and a certain easy and disengag'd behaviour', which are immediately pleasing to others, whose approval does not seem to depend on sympathy but on 'by particular original principles of human nature' (*T* 3.3.1.27/SBN 590) and qualities immediately pleasing to the person who possesses them, which, as we have seen in the previous sections, constitute a heterogeneous class that is also approved on the basis of their utility.

The artificial virtues, on the other hand, constitute a less broad group that includes justice, i.e. the motivational dispositions to respect property and promises, allegiance to the sovereign, chastity and modesty. They have in common that their motives and approval depend on the existence of various artifices, the nature and staged explanation of which I shall examine in detail in Chapter 2. Although Hume is explicit about which character traits identify the artificial virtues, he is not as clear about which traits identify

[38] See also EPM App. 3.1–2 (SBN 303). On this point, see Cohon 1997: 93-95; Abramson (2015): 353. For a comprehensive account of the various positions, see also Fieser 1997: 373–388.

justice. Sayre-McCord argued that since all the artificial virtues are discussed in Book 3, Part 2, entitled 'Of justice and injustice' they all constitute different aspects of the virtue of justice.[39] This broad interpretation, however, can be challenged on the basis of several considerations. Firstly, those made available by common respect for justice are fundamental human goods since they constitute necessary circumstances for the constitution of any human society. However, as Annette Baier has observed, it seems doubtful that chastity and modesty, whose only good is to allow males the certainty of their paternity, have the same importance as the goods produced by justice. The utility of chastity and modesty is contingent on a historically determined organisation of society in which only males provide for the material well-being of the family.[40] The utility of justice, on the other hand, is that it allows a more general, typically human problem to be solved, which concerns how to coordinate with regard to scarce fundamental resources. Secondly, Hume concludes his discussion of the rules concerning property (T 3.2.2-4) and those concerning the obligation of promises (*T* 3.2.2.5) with the section 'Some further reflections concerning justice and injustice' (*T* 3.2.2.6), which has all the air of being a concluding discussion of justice in which he takes up, and further clarifies, the main themes examined in the first section (*T* 3.2.1). Drawing on these considerations, I will consider justice in a less broad sense than that accepted by Sayre-McCord, treating it as the notion that incorporates the rules and motivational dispositions that relate to the respect of property and promises.

Before concluding this section, it is worth pointing out a theme that we shall see taken up in Adam Smith's and Mill's reflections on justice, which is specified precisely in the section 'Some further reflections concerning justice and injustice' mentioned

[39] See Sayre-McCord 2015: 443. See also Harrison 1981: 203 ff.
[40] See Baier 1979: 1ff. On whether the patriarchal society was not the only possible family model and whether a monogamous family consisting of only a mother and child was morally appreciable, see the essay *Of Moral Prejudices*. See D. Hume 1987 [1742/1752]: 538-44.

earlier. Hume clarifies that just actions differ from those expressing the natural virtues because unlike the latter they do not admit of degrees. While a charitable action may admit of degrees of excellence, an action that expresses justice is or is not just. Trivially: you cannot more or less respect a covenant or someone else's property; you either do it or you don't. Hume takes up here the theme, proper to the natural law tradition, that actions inspired by justice are different from those inspired by beneficence because, unlike those, they do not leave it to the agent to decide what circumstances or persons to be just with. The virtue of justice requires the performance of precise particular actions. Unlike Smith and Mill, Hume does not explain this property of justice through the distinction between imperfect and perfect obligations, nor even through the notion of right.[41] Yet the absence of this terminology cannot be taken as proof that his theory of justice did not incorporate this distinction.

1.4 The Emergence of Virtuous Motive to Justice. The Motive of Duty

Let us examine the nature of the virtuous motive of justice, leaving the assessment of how this explanation is compatible with the virtuous motive requirement set out in the Circle Argument to chapter 2. Since it will be dealt with in detail there, I will now examine only some less fundamental aspects of Hume's genealogy of justice set out in *T* 3.2.2.

Hume describes the natural condition of human beings prior to the invention of the rules of justice as characterised by a moderate scarcity which, combined with the limited generosity of human beings, leads us to squabble over material goods. Our natural

[41] Hume explicitly argues that the notion of right connected to the ownership of external goods is temporally, and thus logically, subsequent to the emergence and stabilisation of the convention concerning 'abstinence from the possessions of others' (*T* 3.2.2.11/SBN 490-91). This means that the notion of right is not an explanatory factor, but is instead an effect of convention.

ideas of vice and virtue offer no remedy because they tend to conform to the natural partiality of human affections that drive us to hoard resources for ourselves and our friends and relatives. The only remedy is in the passion of self-interest that, through reflection on the counterproductive outcomes of its free exercise and by repeated experience of possible forms of cooperation, drives us towards collective patterns of action. Gradually and unintentionally, human beings converge on the rule of letting each one enjoy the goods in his possession on condition that the others do the same for him (*T* 3.2.2.10/SBN 490). This sanctions the invention of property, which is progressively articulated through the rules governing its acquisition and transfer by consent (*T* 3.2.3.4/SBN 503-4). Hume explains how this should be regarded as the first stage of justice. This is characterised by forms of cooperation within small communities where any defection is discovered and results in either the exclusion of the violator from the practice or even the dissolution of the practice itself. This explains why redirected self-interest or 'the natural obligation to justice' is a suitable motivation to ensure stable compliance with the rules of property (*T* 3.2.2.23/SBN 498).

As society grows larger and trade becomes more extensive, the balance achieved becomes more precarious than ever. On the one hand, as Rachel Cohon has rightly pointed out, the increase in productive activity generates more goods, which represent continuous temptations to human greed, which becomes difficult to counteract by redirected self-interest.[42] On the other hand, redirected self-interest loses its effectiveness: in a large society, it is no longer immediately clear whether individual violations of justice will have negative effects on individuals, either because it is more difficult to uncover and punish the guilty and/or because unjust actions are not directly connected to the dissolution of the convenient practice of justice.

[42] See Cohon 2008: 173-4.

Hume thus introduces a second stage in his explanation of justice, in which it is now the motive of duty that guarantees conformity to the rules of property. As we shall see in more detail in chapters 2 and 3, moral sentiments constitute the crucial element of our psychology, which underlies the formation of this new motive. Through impartial sympathy, we at first approve and disapprove of behaviour that respectively complies with and violates the rules of property rights (*T* 3.2.2.24/SBN 499).[43] Initially, moral sentiments are directed exclusively at the actions of others, however, through a 'progression of sentiments' (*T* 3.2.2.25/SBN 500) that operates through the sympathy we feel for the moral sentiments that others have for our actions, we become moral spectators of our actions by evaluating them from the common point of view (*T* 3.2.2.24/SBN 499). As Rachel Cohon has argued, these sentiments have the potential to motivate conduct: as both agreeable and painful feelings, they are able to arouse desire for the pleasures of approval or aversion to the pain of disapproval that drive us towards behaviour that complies with the rules of justice. Although these sentiments are originally too weak to counteract the greed that drives us to satisfy more immediate interests by violating the rules of justice, they can nevertheless acquire considerable motivational force thanks to public and private education that encourages the young to associate their reputation and social pride with respect for justice (*T* 3.2.2.25-27/SBN 500-501). In this way, the initial practical weakness of moral sentiments is reinforced by the desire for one's reputation and the resulting pleasures of social pride. Moral sentiments thus constitute a new

[43] As Rachel Cohon (2008: ibid.) has rightly observed, this point marks a departure from Hume's virtue-based morality according to which the moral value of actions derives from that of motives, i.e. from their being vicious or virtuous. Following Cohon's approach, I will argue in chapters 2 and 3 that the moral evaluation of just actions represents a double violation of the virtues-based conception of morality. First, just actions are identifiable independently of the virtuous motive for justice (which in fact will emerge later in the evolutionary process) and only through the rules of property and promise that are constitutive of our idea of justice. Secondly, and the point is related, these actions are not only identifiable, but are also morally assessable from those rules. In particular, they are evaluated from the general effects that would be produced by imagining that the action under evaluation becomes a shared practice.

character trait or stable motive of duty that drives us to uniform compliance with the rules of justice.

Cohon's proposed interpretation of the virtuous motive of justice has the double advantage of being extremely faithful to the text of *Treatise* Book3, Part 2, and of offering an interpretation of the motivation of justice that is consistent with Hume's hedonistic conception of motivation[44]. Regarding the first point, the strength of Cohon's position depends on the fact that it takes into account two crucial expositions in Book 3, Part 2, namely that justice is a virtue and that the virtue of an action depends on the motive with which it is performed. However, it comes at the seemingly high cost of assuming that the undoubted maxim applies only to natural virtues - a position that Hume never explicitly states.

Since Hume's silence, some have concluded against Cohon that justice is not a virtue. James Harris argues that this assumption should not be particularly shocking if we bear in mind that just actions are identifiable without detecting the motives that produced them. Unlike natural virtues, they are approvable simply on the basis that they conform to the rules of justice. What matters morally in this case are the actions and not the motives. Harris does not, of course, dispute that Hume talks about justice as a virtue, but argues that this is merely a *façon de parler* where the alleged virtue is nothing more than a habit of action.

Harris's position accounts for the undoubted maxim, but at the cost of denying that justice is a virtue. This is too high a price that fails to take into account not only the way Hume explicitly characterises justice in the *Treatise*, but also what he says in his

[44]Some have argued that if the motive that gives merit to justice were a sense of duty, this would be incapable of motivating just actions. As Stephen Darwall (1993: 417) has argued, duty compels us to adhere to the rules of justice simply because they are 'authoritative', regardless of the pleasure and pain we derive from obedience. As I have shown, if we follow the way Cohon interprets *T* 3.2.2.25-27, this criticism is misplaced. It does not take into account the fact that duty is explained from the sentimental mixture of moral feelings, pride and the desire for moral reputation. In this way, the motivation for justice is explained by the desire to obtain the pleasures of approval and to avoid the pains of blame, and this is compatible with the hedonistic psychology of motivation set forth in *Treatise*, Book 2.

famous discussion of the sensible knave in the second *Enquiry*. This discussion, which we will examine more explicitly in the next section, introduces a type of free rider who only complies with justice when it is advantageous to him. The whole discussion can only be intelligible against the background of the thesis that one who performs just actions simply out of self-interest is not a person who is morally valuable as just. His actions, even if they have the external form of justice, are caused by psychological factors that have 'lost a considerable motive to virtue' (*EMP* 9.23/SBN 283). They are an expression either of the desire 'of profit or pecuniary advantage' or of the fear of punishment, and not of the virtuous disposition that leads us to rebel at 'the thoughts of villany or baseness and to seek the satisfaction that comes from consciousness of [our] integrity' (ibid.).

To reject Harris's solution and accept Cohon's proposal is, of course, to accept that the undoubted maxim only concerns natural virtues and that the motive of justice can confer merit on just actions even if it is a moral motive. Is this a fatal problem for Hume? There are several problems to consider. The first is that, as I have already stated, Hume never explicitly argues for this thesis. However, this does not mean that he believed the thesis to be false. The second is that in this way we need a broader theory of virtue in which the undoubted maxim does not appear as a necessary requirement.[45] Hume does indeed provide different descriptions of the nature of virtue, and I believe the description given in *T* 3.3.1.17-31 as a character trait assessed from the common point of view is broad enough to include justice.[46]

The interpretation of the first justice motive advocated here has not satisfied some interpreters who are unwilling to sacrifice the role of the undoubted maxim within Hume virtue theory. In the next section I will examine how this assumption has led some interpreters to attribute a crucial role to enlightened or redirected self-interest.

[45] See Sayre-McCord: 450-52.
[46] For a similar view, see Cohon 2008: chap. 6.

1.5 The First Virtuous Motive to Justice. Some Difficulties for Redirected Self-interest

According to some authors, the virtuous disposition of justice is enlightened or redirected interest. This position would solve the problem that the Circle Argument conclusion poses to the thesis that justice is a virtue. Indeed, the conclusion claims the first virtuous motive which bestows a merit on any action must be a non-moral motive. As we have seen, this poses a problem for the thesis that justice is a virtue because Hume had ruled out that the first virtuous motive of a just action can be identified with a non-moral, non-artificial motive (self-love, regard to public interest and private benevolence). According to these interpreters, however, this problem is not fatal to Hume's theory of justice as a virtue. There is in fact another possible candidate capable of fulfilling the constraint posed by the Circle Argument, which is precisely redirected or enlightened self-interest. As non-moral, it can perform the function of the first virtuous motive. Moreover, insofar as it is artificial, it does not belong to the class of motives that Hume considers unable to cause the 'full behavioural profile' of justice.[47]

According to this interpretation, therefore, the first non-moral virtuous motive is the one that prompts us to perform the right actions in view of the advantages that are made possible by a pattern of collective action.[48] There is no lack of textual evidence in support of this interpretation. As we will see in detail in Chapter 2, Hume argues that the conventions of justice are not only in each individual's interest, but are also perceived by each individual as beneficial (*T* 3.2.2.4/SBN 486; also *T* 3.2.2.22/SBN 497-98; *T* 3.2.2.24/SBN 499). Having said that, the question is whether these considerations are sufficient to justify the thesis that this motive is what gives merit to right actions, making

[47] An extensive discussion of this can be found in R. Cohon 2008: p. 184.
[48] It means that redirected self-interest is not only what we have when we consider our long-term benefits. The reflexive aspect of this passion depends on whether it considers the advantages of a pattern of collective action. On this point see Sayre-McCord 2015: p. 453.

them virtuous. To be that kind of motive, artificial self-interest must not only cause us to observe the rules of justice, but must do so inflexibly and universally. Otherwise, it would be just one of the possible motives for justice, such as natural self-interest, private benevolence and public benevolence. Unlike these motives, however, the first virtuous motive of justice must be able to cause the 'full behavioural profile' of this virtue - that is, to motivate compliance with rules of justice in all circumstances, even when the individual action is contrary to the interest of the agent, or of another person, or, at the limit, of society as a whole.

In T 3.2.22, Hume points out how artificial self-interest causes precisely this kind of inflexible adherence to the rules of justice. The different kinds of particular negative consequences associated with the performance of each particular just action are always compensated for by the fact that each action guarantees the continuation of the general adherence to the rules of justice that is a condition for the survival of society (*T* 3.2.2.22/SBN 497-98). Hume's assumption is that there is a link between the particular observance and the general observance such that the advantage deriving from the general observance motivates everyone to play their part in the practice. The link depends on the fact that each participant in the practice thinks that every single one of his defections could be imitated by the other participants with the risk of generating a domino effect[49] that produces the dissolution of the practice itself (*T* 3.2.2.22/SBN 497-498). Whenever each participant in the practice of justice, governed by redirected self-interest, is faced for example with the choice of whether or not to comply with the rules of property, he will have to take into account not only the negative effects of compliance but also the negative effects of non-compliance, in particular the impact that his choice has on the inclination of others to imitate his non-compliance.

[49] For a discussion of this point, see Gauthier 1992: 409 ff.

In the *Treatise*, Hume takes the view that the balance between these considerations of different kinds of harmful consequences will always be in favour of the choice to comply with the rules of justice, whatever the immediate effects this action has on one's own interest, on the public interest, and on the interest of those in one's narrow circle. That means that redirected self-interest, unlike the other natural motives that Hume acknowledges may from time to time motivate just actions, can instead motivate full motivational profile of justice. In the second *Enquiry*, however, the argument of the *Treatise* breaks down because Hume questions its second point, i.e. that any particular non-observance affects the general observance of the participants in the practice of justice. Once this link is broken, it is possible to distinguish between two interests that coincide in the *Treatise*, i.e. the interest that each agent has in the general observance of the practice and the interest that each agent has in his own observance of the practice. In this new scenario, an agent who is driven by the motive of artificial self-interest may consistently want all participants in the practice of justice to respect its rules and be driven, under certain circumstances, to transgress them. As is well known, Hume examines this circumstance through the creation of the figure of the sensible knave.

> Treating vice with the greatest candour, and making it all possible concessions, we must acknowledge, that there is not, in any instance, the smallest pretext for giving it the preference above virtue, with a view to self-interest; except, perhaps, in the case of justice, where a man, taking things in a certain light, may often seem to be a loser by his integrity. And though it is allowed, that, without a regard to property, no society could subsist; yet, according to the imperfect way in which human affairs are conducted, a sensible knave, in particular incidents, may think, that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy. That *honesty is the best policy*, may be a good general rule; but is liable to many exceptions: And he, it may, perhaps, be thought, conducts himself with

most wisdom, who observes the general rule, and takes advantage of all the exceptions. (*EPM* 9.22/SBN 282-283)

The sensible knave admits that the existence of justice, as a condition of society's existence, is in his best interests. He also accepts that the maintenance of justice is conditional on the observance of these rules by his companions and fellow-citizens, and that he must therefore endeavour, so far as he is concerned, to ensure that this observance does not fall short. This is, however, compatible with the possibility that, when his particular interests are thwarted by respect for justice and when he is confident that by acting unjustly he will not be discovered, he is driven by the same artificial self-interest to act unjustly. This motivation, in fact, tells him to act justly only because this is a condition of society's existence. But if his acting justly, in some cases, is not a condition of that, he may violate its rules and act in ways that in that circumstance advance his own interests.

The objection posed by the sensible knave seems to be a serious problem for the possibility of considering redirected self-interest as the virtue-conferring motive of just actions. Indeed, David Gauthier has argued that to be such, artificial self-interest should be able to motivate just conduct in an inflexible and universal way, because this is exactly what the virtue of justice requires us to do. However, this is not the case since the knave, although motivated by artificial self-interest, acts justly only occasionally and not whenever circumstances require it.

There have been several reactions to this problem. Gauthier has, for example, argued that if we take into account both the *Treatise* and the second *Enquiry* we must conclude that the only motive for complying inflexibly with the rules of stability of possession, its transfer and respect for promises is the motive of duty. However, since this motive presupposes a non-moral motive for performing these actions other than a

sense of duty, and since this non-moral motive (artificial self-interest) does not exist, then there is no motive to perform just actions other than the mistaken apprehension that this is a duty. Hume thus has an error theory regarding the motive of justice. Gauthier concludes that this is a salutary error: it is in the interest of society that we believe we have an obligation to perform these kinds of actions and the error is one that Hume does not explicitly reveal in his work because he does not want to threaten its hold on us.[50] Marcia Baron, on the other hand, put forward a slightly different interpretation according to which redirected self-interest is the non-moral motive of just action, but we simply do not know that our belief about the devastating effects our non-compliance has on others is founded on a false judgment. However, just like Gauthier, she believes that this false judgement has positive effects and is therefore reinforced by public and private education.[51]

A similar position, advocated by Knud Haakonssen, argues that we know that we have no non-moral reason to inflexibly comply with the rules of justice. At the same time, we morally approve of those who behave justly and we assume that they have the very non-moral motive that is absent in our mind. We despise ourselves for this and act out of duty both in the hope of acquiring the phantom first motive of justice and to conceal this absence from ourselves. However, Haakonssen argues, we are mistaken in thinking that the motive indicated by the indubitable maxim exists and thus we are mistaken about justice being a virtue.[52]

A final position, argued by Christine Korsgaard, challenges the very premise of these readings: Hume provided a normative reason for the knave to comply with the rules of justice. Regardless of his particular beliefs about the positive consequences that sporadic violations of justice have on one's happiness, the knave has normative reasons

[50] See David Gauthier 1992: 402.
[51] See Baron 1982: 539 ff.
[52] See Haakonssen 1981: chap 2.

for respecting the rules of justice in a non-instrumental way. This is because Hume would argue that if we abide by the rules of justice we would be loved by others and as a consequence of this we would be able to experience that stable sympathetic pride which is the condition for individual happiness.[53]

Korsgaard touches here on a very complex issue concerning the general theme of the connection between virtue and happiness in Hume.[54] Regardless of this general theme, Korsgaard's argument seems to be based on a confusion between a moralized version of egoism, which is that of the just virtuous person, and a conception of happiness that possesses the sensible knave who does not possess the virtue of justice, in which non-instrumental compliance with its rules is not a source of happiness per se.[55] The mere appeal to individual happiness seems incapable of bridging this gap and therefore Korsgaard's argument is not compelling.

In the end, I believe that the thesis that justice coincides with redirected self-interest is a mistake. Not only did Hume never support it, limiting himself to considering this passion only as the basis for the emergence of the first convention of justice, but the very figure of the sensible knave shows that this motive cannot guarantee the full behavioural profile that this virtue must ensure. This theoretical option is motivated by a very specific fear: by rejecting the undoubted maxim there is no basis for arguing that justice is a virtue. Based on Cohon's line of argument, I have claimed that this fear is unjustified. While rejecting that maxim, the thesis that justice is a virtue remains in good shape.

[53] See Korsgaard 1996: 55-60.
[54] For a convincing analysis of the reasons to be sceptical about this connection, see R. Cohon 2020: 156 ff.
[55] For an examination of this difference, see the illuminating discussion by Hills 2010: chap. 4.

Chapter 2

Hume's Genealogy of Justice

Following a common approach to the moral and political philosophy of the 17th and 18th centuries, Hume considers political society as the outcome of a gradual process of civilization from a pre-political natural condition of human beings. Hume takes up the line developed by Locke in the *Second Treatise on Government*, later also taken up by Rousseau in the *Second Discourse*, according to which pre-political society is not a condition of permanent war but a state in which individuals participate in different social practices. Hume further elaborates this approach, distinguishing pre-political society into two distinct phases. The first identifies a living condition where the natural virtues regulate each person's behaviour within extended family communities. The second describes an intermediate condition of social life that is subsequent to the one just described, but precedes the political one governed by law. This new condition is characterised by respect for justice, which allows morality to extend its scope beyond that of family relations. The rules of justice, which pertain to property and the observance of covenants, in fact regulate social and economic interactions between strangers allowing each to coordinate with the others to avoid conflicts over scarce resources.

In *T* 3.2.2, Hume gives an account of the transition from the first to the second condition, which constitutes a theory of the emergence of the artificial virtues of justice. In this chapter I intend to examine three methodological aspects of this account. First, I consider whether it has a genuine explanatory dimension and, if so, what role the empirical data play in it. Secondly, I examine whether it has normative implications.

Thirdly, should it prove to have both dimensions, I will consider whether there is any connection between them, that is, whether the very content of the explanation provides those who are already abiding by the rules of justice with an additional motivation to respect them, or whether, conversely, this very content threatens to weaken the participation of community members in this practice. Once this examination has been carried out, I intend to consider whether Hume's theory of justice can be categorised as a form of genealogy.

2.1 The Two Conceptions of the State of Nature

Let us start with Hume's description of the natural state of human beings before rules of justice were invented. Hume actually formulates two different descriptions of his narrative starting point. My hypothesis, which I will try to prove in this section, is that this is part of a strategy aimed at highlighting what he considers to be a necessary feature of any genuine narrative starting point for the emergence of cooperative practices. I will argue that a plausible reason for this strategy is that Hume's narrative has an essential explanatory dimension that aims to account for the emergence of the notion of justice from a state of human life in which justice does not exist.

In *T*. 3.2.1, the examination of justice begins with a description of the extreme difficulties that characterize human life in its 'wild uncultivated state' (*T* 3.2.2.4/SBN 486). Unlike other animals, who have desires proportionate to their ability to satisfy them, human beings have an infinite number of desires and needs and a small number of means to fulfil them. Hume claims:

> In man alone, this unnatural conjunction of infirmity, and of necessity, may be observ'd in its greatest perfection. Not only the food, which is requir'd for his sustenance, flies his search and approach, or at least requires his labour to be produc'd, but he must be possess'd of cloaths and lodging, to defend him against the injuries of the weather; tho' to consider him only in himself, he is provided neither with arms, nor force, nor other natural abilities, which are in any degree answerable to so many necessities. (*T* 3.2.2.2/SBN 485)

Interestingly, Hume illustrates the difficulties that characterise the primitive state by referring to the way in which people organise their work and the negative impact this has on their lives. First, having to deal with countless activities, they cannot achieve a high level of competence in any of them. Second, even if they were competent, each human being does not have the strength to perform constantly all these activities. After all, relying solely on themselves, they cannot afford to make mistakes, and any single failure could lead to their ruin.

The only remedy for these difficulties is social grouping, in which labour acquires two new characteristics. In the first place, the labour is mainly divided among the members of the community, so that each one can carry out only certain activities, specialise in them, and develop skills and acquire experience that make him excellent in the exercise of these activities. Secondly, each kind of work is carried out by several people, which makes it possible to overcome the weakness of individuals and to distribute common resources for the benefit of those in difficulty. On this point, Hume claims:

> When every individual person labours apart, and only for himself, his force is too small to execute any considerable work; his labour being employ'd in supplying all his different necessities, he never attains a perfection in any particular art; and as his force and success are not at all times equal, the least failure in either of these particulars must be attended

> with inevitable ruin and misery. Society provides a remedy for these *three* inconveniences. By the conjunction of forces, our power is augmented: By the partition of employments, our ability encreases: And by mutual succour we are less expos'd to fortune and accidents. 'Tis by this additional *force*, *ability*, and *security*, that society becomes advantageous. (*T* 3.2.2.3/SBN 485)

From what Hume says, then, it would seem that the transition from the original solitary state to that of life in society is motivated by the fact that human beings, reflecting on possible future goods of '*strength*, *ability* and *security*', are motivated to modify their wild state and associate. But Hume points out that this hypothesis cannot work. And this leads him to formulate an entirely new starting point.

Understanding why this description of the State of Nature fails is essential to understanding what the function (or at least one of the main functions) of the virtue account of justice in T 3.2.2 is. Hume argues that although social life is causally related to living conditions better than those associated with the isolated state, solitary human beings, having experienced neither the benefits of cooperation nor anything remotely like them, cannot come to understand them even by reflection, and therefore cannot be induced to change their primitive state and adopt a cooperative life. Hume claims:

> But in order to form society, 'tis requisite not only that it be advantageous, but also that men be sensible of its advantages; and 'tis impossible, in their wild uncultivated state, that by study and reflection alone, they should ever be able to attain this knowledge (*T* 3.2.2.4/SBN 486)

The reason why the first Hobbesian hypothesis does not work, then, is that it is incapable of fulfilling the explanatory task that Hume clearly considered central to his investigation of justice. A new starting point is therefore needed to explain the transition

between the two stages of associative life. But is this really necessary? Could we not argue that the original stage is already governed by justice? As chapter one showed, this solution is not viable for Hume. Indeed, the point of the Circle Argument is to show that we run into insoluble difficulties if we think of the virtue of justice and property as something primitive in human animals. Hume then introduces a new starting point for his account of the origin of justice, where humans are not solitary, but already socialised within small family clans. Note that he does not simply assume this second condition, but explains it on the basis of two basic elements of human psychology, namely the sexual instinct (i.e. 'the natural appetite between the sexes') and 'concern for the common offspring' (*T* 3.2.2.4/SBN 486). The first instinct leads people to encounter each other, while the second leads them not to abandon each other when sexual interest fades. The love of offspring is therefore the reason for both the stable bond with the partner and the bond with the children, and thus for the condition of family life.[56]

Here is the reason for the second starting point: Hume observes that in family interactions people, especially children, who are the first and most important beneficiaries of parental care and protection, cannot help but experience the benefits of cooperation. They experience the importance of restraining their more violent and selfish tendencies

[56] On this point, Hume has a different position from what Rousseau put forward in his *Second Discourse*. Although they are similar in rejecting Hobbes's thesis that views the original condition of human beings as characterized by conflict, Rousseau and Hume differ in that, while the former describes the original condition of life as solitary, the latter characterizes it as social. The difference can be explained precisely by the fact that for Hume the starting point is made by human beings who have both sexual drive and love for their offspring, whereas for Rousseau they have only sexual drive. According to Rousseau, this passion explains the continuation of the species but cannot explain, on its own, the formation of relationships between human beings. Rousseau believes that the sexual drive is a psychological episode that has a short duration, and, once satisfied, makes us lose interest in the person with whom we are united. The sexual drive is therefore not associated with the desire to experience that pleasurable experience again with the same person. Unlike love, which is a product of social life, the sexual drive has no specific individual object and therefore cannot generate any preference for one person over another (Rousseau 2003 [1755]: 21-22). Hume's position on sexual desire is less cogent than Rousseau's. From what little he says, however, one thing is clear. That instinct is not even enough to guarantee a stable relationship with one's partner, so much so that Hume argues that love for one's offspring not only makes stable family union possible but also strengthens the bond with one's partner, which, if it were entrusted solely to sexual desire, would inevitably die out in a short time.

in order to avoid the evils of abandonment or distance and to receive protection and affection in return. Hume claims:

> In a little time, custom and habit operating on the tender minds of the children, makes them sensible of the advantages, which they may reap from society, as well as fashions them by degrees for it, by rubbing off those rough corners and untoward affections, which prevent their coalition (*T* 3.2.2.4/SBN 486)

This clearly solves the problem that Hume has just raised. Human beings are able to appreciate the benefits of cooperation because they have already learned within family unions the importance of respecting basic forms of cooperation.[57]

At this point, Hume's strategy raises two questions. First, what is the status of his claim on the proposed starting point? Second, why doesn't Hume begin his explanation directly from this second description of the State of Nature instead of from the first and then discredit it? Would that not be a way of making his argument sharper and clearer?

With regard to the first question, it could be argued, using contemporary terminology, that Hume is basing his description on an inference to the best explanation. In other words, Hume wants to explain the origin of social life by arguing that, of the possible starting points, the best one is the one in which there is a minimum of sociability that allows individuals to appreciate it and to anticipate how more sophisticated forms of cooperation can further improve their living conditions. If we had started from a state of solitude, the emergence of society would be inexplicable.

On the second question, Hume's indirect approach, far from being bad, is not only effective, but, as I have just mentioned, reveals a crucial aspect of his narrative. First, it

[57] Annette Baier has taken this fact as the basis for her interpretation that artificial virtues depend in an important sense on the natural virtue of 'equity' between human beings. See for example A. Baier (2010) *The Cautious Jealous Virtue: Hume on Justice*, pp. 56-82.

is effective because by showing what a starting point should not be, it gives us a clearer picture of what it should be. Secondly, and more importantly, it highlights the fundamental explanatory dimension of his narrative, by which he seeks to distinguish himself from other narratives of the state of nature, such as the one described by Hobbes in Leviathan, whose only function is to justify a particular practice or contract capable of securing social life. In order to understand this point, let us first ask the following question: under what circumstances can the characteristics of the new point of departure be considered a necessary feature of a narrative point of departure?

A plausible answer seems to be that this is necessary if the narrative is to explain how people came to adopt a particular practice that ensures a stable social life. Why? Because if that is the aim of the narrative, then the mere existence of a causal link between the practice and the best living conditions for all is not enough to explain the transition. What is needed is for the causal link to be perceived by those who eventually join the practice. And that is exactly what happens in the second case, but not in the first. Hume's choice of the second state of affairs as a starting point thus reveals the essential explanatory ambition of Hume's narrative.

This brings us back to Hume's second point. The reason why Hume begins his narrative with a point of departure which he immediately abandons is to emphasise that those who, like Hobbes, start from a wild and solitary state of life are not in a position to pursue Hume's project because their description of the original state condemns them to failure in an explanatory framework.

2.2 The Stage of Family Society and the Conflict over External Goods. Is Hume's State of Nature a Mere Fiction?

Family society is the first stage in explaining the emergence of justice. Hume describes it as characterised by the problem of conflict between family clans over scarce material

resources, to which justice is a solution. As we shall see, the explanation of the transition from family society to one in which justice has taken hold is only the first part of the explanation of the emergence of justice, i.e. the one in which justice is respected in an instrumental way because it furthers our interests in living in a peaceful society. Alongside this is a second part in which justice is respected non-instrumentally as a virtue. In this section I will consider only the first of these, leaving the second for the next section.

The stage of family society brings out some methodological features of Hume's overall account of the virtue of justice. First, it highlights the fundamental role of human psychology, and in particular his cogent conception of the passions as an explanatory factor in the emergence of justice. Indeed, Hume shows how the passion of self-interest can be aroused not only by what is pleasurable insofar as it is the object of our instincts, but also by thinking about the pleasure that comes from participating in complex patterns of collective action. Secondly, the first stage emphasises that his explanation is in the service of his naturalism.

As is well known, this label refers to many aspects of Hume's philosophy. With regard to what we are interested in here, two points need to be emphasised. First, the account of justice does not use elements or powers that presuppose or imply the divine providential will or principles peculiar to justice, i.e. invented ad hoc. Secondly, and in line with the last point, the *explanans* is intelligible independently of the *explanandum*, i.e. the notion of justice to be explained. Of course, Hume is particularly attentive to the dangers of circular explanations, but, as we shall see later in this chapter, this is not the only concern at work here: Hume seeks, as far as possible, to explain the high and complex in terms of the low and simple. This means that his account of justice makes use of elements of our psychology that we share with the rest of the animal kingdom.

Hume characterises the early stage of society in contrast to two fictional images of the state of nature, the golden age of the poets (*T* 3.2.2.15/SBN 493-4) and the state of

nature of Hobbes (*T* 3.2.2.14/SBN 492-3). The former presents human beings as characterised by a universal benevolence that always inclines them to sacrifice their interests when they conflict with those of others, even strangers. The second, by contrast, describes them as solitary and selfish. Hume argues that these conceptions of human nature, however contradictory, are similar because they are both mere fictions, lacking any support from an empirical study of human nature. Hume claims:

> This *state of nature*, therefore, is to be regarded as a mere fiction, not unlike that of the *golden age*, which poets have invented; only with this difference, that the former is describ'd as full of war, violence and injustice; whereas the latter is painted out to us, as the most charming and most peaceable condition, that can possibly be imagin'd. […] The storms and tempests were not alone remov'd from nature; but those more furious tempests were unknown to human breasts, which now cause such uproar, and engender such confusion. Avarice, ambition, cruelty, selfishness, were never heard of: Cordial affection, compassion, sympathy, were the only movements, with which the human mind was yet acquainted. (*T* 3.2.2.15/SBN 493-4)

In contrast to these views, Hume wants to offer an explanation based on a conception of human beings that is not rooted in fantasy. Note that this point differs slightly from that made about Hobbes earlier in Part 2. Whereas the former criticised the solitary state of life in the state of nature, so solitary that it could not explain the transition to society, the criticism now concerns Hobbes's fantastic elements per se. On the contrary, Hume does not consider his first stage to be purely imaginary, because it is composed of individuals whose psychology is based on the very passions that his empirical study of human nature had validated in *Treatise* Book 2.

Life in the first stage is characterised by the problem of the instability of material goods, for which justice is the solution. This instability depends on three 'inconveniences'

(T 3.2.2.16; SBN 494), which Hume describes as 1) 'selfishness' and 'limited generosity'; 2) the 'easy change' of 'external objects'; 3) their 'scarcity in comparison of the wants and desires of men' (*T* 3.2.2.16/SBN 494; *T* 3.2.2.18/SBN 495). Let us examine them, starting with the last. Hume claims that there are three kinds of goods available in natural society: the affective experience of pleasure and fulfilment associated with mental activities; the advantages of possessing a strong, agile, and resilient body; and finally, the possession of objects that produce pleasure or are the natural object of our instincts. Of these, only the last is at risk of being taken away from us, because it can pass from one hand to another without losing its value - a risk that becomes more than a possibility when you consider that these goods are not enough to satisfy everyone's desires. Hume claims:

> There are three different species of goods, which we are possess'd of; the internal satisfaction of our mind, the external advantages of our body, and the *enjoyment of such possessions as we have acquir'd by our industry and good fortune*. We are perfectly secure in the enjoyment of the first. The second may be ravish'd from us, but can be of no advantage to him who deprives us of them. The last only are both expos'd to the violence of others, and may be transferr'd without suffering any loss or alteration; while at the same time, there is not a sufficient quantity of them to supply every one's desires and necessities. As the improvement, therefore, of these goods is the chief advantage of society, so the *instability* of their possession … is the chief impediment. (*T* 3.2.2.7/SBN 487-8)

In addition to the easy transition of external goods and their relative scarcity, their instability is also determined by the so-called third inconvenience, i.e. the natural 'selfishness' and 'limited generosity' of the actors in the family society. Why does limited generosity aggravate the problem of the instability of possessions instead of alleviating it?

To answer this question, a brief digression into the nature of this motive and how it operates in family society is necessary. Hume famously disagrees with Hobbes's thesis that the natural human condition is that of individuals who focus exclusively on satisfying their own personal interests to the detriment of those of others. While human beings typically have a greater interest in themselves than in any other person, their general passion and benevolent instincts are stronger than those directed solely at their own welfare. Nevertheless, Hume shows that benevolent inclinations (benevolence, gratitude, pity) have a limited scope, confined to our nearest and dearest and not extending to strangers. While this trait leads us to restrain our greed when it comes to the material goods of our loved ones, it is completely powerless to restrain our acquisitive desires when these are directed towards goods used by strangers. On closer inspection, in these circumstances, limited generosity might even increase greed, since it might lead us to satisfy not only our own acquisitive desires, but also those of our loved ones. Thus, far from being a stabilising factor for material goods, limited generosity is, on the contrary, something that favours their mobility between different family clans.

So the fundamental problem of the primitive stage is that every family clan is under threat of having its possessions taken by members of other clans. Why is this a problem? The easy transfer of external goods does not involve the risk of one's life. Moreover, not only does Hume never describe a situation of mutual warfare of the Hobbesian type, but the possibility seems unlikely. As we will see in Chapter 3, the only passion that could ignite the conflict is resentment. But this passion could only arise if the theft were seen as a violation of a rule. In family society, however, there are no rules about property because property and the law that protects it do not yet exist. One might think, therefore, that even if it is a problem, it is not so serious as to make this stage unstable.

These considerations do not take into account the organisation of life in family society. Hume does not describe it in the *Treatise*, but he does mention it in the essay *Of Commerce*, where he describes a three-stage model of civilisation. Rejecting in advance Rousseau's claim in the *Second Discourse* that primitive humans subsist on what nature offers them and live in an eternal present, Hume argues that they are typically devoted to activities such as hunting and fishing, which they gradually abandon to devote themselves first to agriculture and then, without abandoning it altogether, to trade. Hume claims:

> The bulk of every state may be divided into *husbandmen* and *manufacturers*. The former are employed in the culture of the land; the latter work up the materials furnished by the former, into all the commodities which are necessary or ornamental to human life. As soon as men quit their savage state, where they live chiefly by hunting and fishing, they must fall into these two classes; though the arts of agriculture employ *at first* the most numerous part of the society. Time and experience improve so much these arts, that the land may easily maintain a much greater number of men, than those who are immediately employed in its culture, or who furnish the more necessary manufactures to such as are so employed (*E* 256).

The effectiveness of hunting and fishing, especially of animals that outstrip humans in strength and aggressiveness (*T* 3.2.2), presupposes the availability of offensive tools, i.e. external goods that enhance human capabilities to give them a better chance of winning an unequal fight. Hunting and fishing also require the planning of attacks at the right time. Offensive tools for hunting must be available not only in the present, but also in the future. And it is on this expectation that hunting strategies can work. Since hunting and fishing are not always possible, the survival of clans also depends on the accumulation of supplies, i.e. external goods, which can be used in lean times. Even if the mobility of external goods does not lead to a permanent war of all against all, it is

clear that it represents a serious threat to the survival of this primitive community, which is struggling to find a way out. The first stage of Hume's theory of justice is therefore characterised by a recurrent problem which makes it unstable and forces its members to find a new way of living together.

Hume's solution is that the members of these family communities will be induced, through slow mechanisms of trial and error, to converge on rules of conduct capable of solving the problem of the instability of external goods. In contrast to the Hobbesian contractualist position, Hume argues that coordination does not depend on the formulation of explicit promises, but is driven only by the awareness of having an interest in solving a common problem, and the reflection that the actions one takes to solve the problem may affect the actions of others. Hume describes the development of this tacit mutual accommodation with the famous image of two oarsmen who, without verbally declaring the direction to be taken, both having the desire not to stand still and each having the ability to observe and reflect on the other's movements, end up rowing in unison. Of this image, Hume claims:

> Two men, who pull the oars of a boat, do it by an agreement or convention, tho' they have never given promises to each other. […] it arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it. […] In like manner are languages gradually establish'd by human conventions without any promise. In like manner do gold and silver become the common measures of exchange, and are esteem'd sufficient payment for what is of a hundred times their value. (*T* 3.2.2.10/SBN 490)

Hume explains coordination in terms of the rules that establish and regulate the transfer of property through two main explanatory factors. The first is redirected or

enlightened self-interest. The second relates to the epistemic conditions that individual agents must fulfil in order to be able to engage in practices that respect property. As we saw in Chapter 1, Hume believes that mere self-interest can counteract itself: that is, its destructive tendencies may lead us to want to acquire as much property as possible for ourselves and our family clan.[58] We gradually learn that our interest 'is much better satisfy'd by its restraint, than by its liberty' (*T* 3.2.2.13/SBN 492). The preservation of social life is better guaranteed if we agree to 'bestow stability on the possession of those external goods, and leave every one in the peaceable enjoyment of what he may acquire by his fortune and industry' (*T* 3.2.2.9/SBN 489). Redirected self-interest thus drives us to obey this rule, and this gives rise to the first notion of justice, and with it the notions of 'right' and 'property', which Hume defines as 'those goods, whose constant possession is establish'd by the laws of society; that is, by the laws of justice' (*T* 3.2.2.11/SBN 490-1).[59] Thus, Hume's explanation shows us how self-interest is capable of generating a new reason to act, which concerns respect for the rules of property. This is a conditional reason: each agent believes that it is in his own interest to restrain his acquisitive desires by respecting property on condition that others do so. It is an instrumental reason, since each agent believes that it is in his or her own interest to respect justice, because this makes it possible to solve the problem of the instability of external goods, which worsens the living conditions of all.

The conditional nature of this instrumental reason presupposes the fulfilment of certain epistemic conditions. Following Matthieu Queloz's investigation, we can identify

[58] For an examination of how the account of justice highlights Hume's non-static but dynamic and progressive conception of human nature, see Gill 2000: 87 ff; Cohon 2008: 172 ff.

[59] In T 3.2.3, Hume breaks down the rule of respect for property into four sub-rules, according to which an object is my property if 1) I am the first to occupy and control it; 2) I have possessed it for a long time; 3) it is the product of goods or resources that are my property; 4) it has been consensually given to me by its owner. Hume seems to think that these rules have the same general validity as the one that establishes ownership. He does not seem to think that they have value or are useful only in some societies and not in others. To confirm this, Hume seems to derive our inclination to respect them directly from the imagination.

three.[60] First, each hypothetical participant in the practice must be sensitive to the fact that he or she has a conditional reason for participating in the practice, i.e. to refrain from trying to take someone else's property on the condition that others do the same with him or her. Secondly, each agent must be aware that others are also sensitive to this reason. Thirdly and finally, each agent must be aware that others are aware of (1) and (2). Queloz thus rightly observes that Hume was aware that in order for redirected self-interest to provide each agent with a reason to obey the rules of property, it is not enough that all have a conditional interest in cooperating; it is also necessary that all know that all have an interest in cooperating. Only then will each individual's conditional reason, combined with an expectation of the behaviour of others, motivate each to cooperate by respecting the rules of property.[61]

Now that the explanation for the emergence of justice is before us, it may be useful to return to the theme of naturalism mentioned at the beginning. As we have seen, Hume's proposed explanation does not rely on elements of psychology that are introduced ad hoc, but uses a passion, i.e. self-interest, that exists anyway and operates in human nature to explain a wide range of phenomena other than justice. Moreover, the explanation is not circular because it never presupposes the idea it seeks to explain. If it did, the possession of the reasons for participating in the practice of justice would have to depend on the potential participant's recognition of the instrumental link between the institution of property and the general welfare associated with the creation of a stable and peaceful society. But there is no need for such complex knowledge. All that is needed for cooperation is for people to be aware that they have a conditional reason for abstaining

[60] See Queloz 2021: 80.
[61] Queloz argues that Hume is fully aware that cooperation on the rules governing property only arises when there is not simply a 'shared knowledge', but when there is a 'common knowledge' that it is in one's interest to respect the property of others on condition that others do the same, see Queloz 2021: 81.

from the property of others, and to know that others not only have the same reason, but also know that others have it.

At this point we might want to go a little further and ask whether the first stage is fictional or not. Some might object that all I have done so far is to show that Hume offers a description of human psychology that is not fictional. Indeed, Hume contrasts his description of the state of nature with that of Hobbes or the poets by arguing that, unlike them, his is based on a realistic rather than a fictional conception of human psychology. But this is compatible, the objection continues, with the fact that this conception of the state of nature is fictitious in another sense, namely that this stage before the invention of justice can never have existed. According to this position, not only does Hume offer no empirical data for its existence, but such data could never have been available, since there is no state of social life prior to the invention of justice. This position has recently been taken up by Matthieu Queloz, who argues that since justice is necessary for the existence of society, it is a futile exercise to look for empirical evidence of a society in which justice has not yet been invented.[62] Queloz supports his argument with several textual data, the most important of which is that Hume states that since it is 'utterly impossible for men to remain any considerable time in this savage state', he conceives of the state of nature as 'a mere philosophical fiction, which never had, and never could have, any reality' (*T* 3.2.2.14/SBN 492-3). Is this reconstruction of the fictional character of Hume's starting point correct?

First of all, Queloz's arguments are unconvincing. When Hume claims that the state of nature is a 'philosophical fiction', he is not referring to the stage of family clans, but to the Hobbesian stage of war of all against all. A judgement about the non-existence of a particular conception of the state of nature is not the same as a judgement about the

[62] See Queloz 2021: 77.

non-existence of any form of the state of nature, i.e. before the advent of justice. Secondly, the fact that a state before justice could not have lasted long does not mean that it never existed; it may more plausibly mean that, given that human beings invented justice as a remedy for a situation of conflict, and that this invention was necessary, as we shall see, it could not have taken long to be discovered. However, if we move away from Queloz's hypothesis and consider the question of the status of the state of nature in a different explanatory context, the idea that it is fictitious is not so incontrovertible. Take, for example, the case of the explanation of the emergence of the concept of polytheism in the *Natural History of Religion*. Again, as with justice, Hume explains the emergence of a concept from a stage of life in which the concept does not yet exist.

Hume states that the causes of polytheistic beliefs do not arise from the 'contemplation of the works of nature, but from a concern with regard to the events of life', especially 'from the incessant hopes and fears'(*NHR*: sec.II) concerning the course of future causes that cannot be controlled. In the first three chapters, Hume describes the causes of these passions, referring not only to natural events ('Storms and tempests [that] ruin what is nourished by the sun'[ibid.]) but also to the uncertainty that characterises the outcome of a 'war', the success of a nation, or the consequences of 'sickness and pestilence'(ibid.) - facts for which historical evidence can be gathered. Although in the case of justice the quasi-historical references to the state of nature are less explicit than in this text, I believe that since Hume follows the same method of explanation in both cases, there is no reason to conclude that the description of the social state of the family never existed. On the contrary, I think it plausible that in the case of justice, too, there is a sense in which this explanation can be considered at least potentially factual rather than

fictional, since it refers to events set against the background of practices or institutions for which we can, in principle, gather historical evidence.[63]

Summing up what's been said so far. The description of family society in T 3.2.2 highlights four important features of Hume's explanation of justice. First, it starts from a human condition that is not outside history. Although it is not located by Hume in a precise time and place, it is underpinned by references on which we have or can gather historical evidence. Second, its members are characterized by a psychology that is sufficiently basic to be attributed to any human being, even those living in ancient times. Third, this condition is characterized by a problem that imposes the need to escape that condition. Fourth, the problem described, precisely because it depends on general psychological characteristics and circumstances that we share with the inhabitants of that condition, is fully intelligible to present human beings.

2.3 The Non-instrumental Virtuous Motive of Justice. Is Hume's Narrative Vindicatory?

As I suggested in Chapter 1, it is implausible that prudence is the first virtuous motive for justice. The reason for Hume's rejection of this hypothesis is that, in the commercial societies of his time, there are at least two circumstances that make the violation of justice convenient in some cases. The first is that the growth of wealth encourages the production of more and more diverse goods, which multiplies the incentives to violate the rules of property. The second is that, in a large society, social interactions with strangers become more frequent, and with them the likelihood that the people with whom we deal have no

[63] For an interpretation that departs from that of Queloz and interprets Hume's explanatory method as a form of conjectural history, see Evnine 1993: 589 ff. See also Kail 2009: 113 ff.

way of knowing whether we have been unfair or not. The effect of individual wrongs on individual reputations is therefore uncertain. Moreover, if we are not discovered, individual wrongs will not cause others to imitate our behaviour, and therefore there is no risk of the practice collapsing as there would be if the community were small.

If there are no strong self-interested reasons for following the rules of justice, and if in some cases there are strong self-interested reasons for doing injustice, it is likely that injustice will be done. As we saw in Chapter 1, Hume illustrates this possibility with the figure of the sensible knave, who, although he has reasons of self-interest in favour of maintaining the practice of justice, does not always have reasons of self-interest in favour of being the one to uphold the practice; on the contrary, he is sometimes driven to violate it.

In Chapter 1, I argued that this circumstance underlies Hume's refusal failure to believe that the motive behind the constitution of justice can be the reason why he and his contemporaries considered just actions to be virtuous. Since the motive that compels us to respect justice as an instrument of our interests is incapable of guaranteeing the 'full behavioural profile' of justice, Hume looks for a non-instrumental motive for respecting justice, which he identifies with the motive of duty. I have already examined how this new element emerges in the psychology of the individual, and how Hume's account of its motivational efficacy is consistent with the theory of motivation set out in T 2.3.3. Let us now consider what this new stage of development tells us about Hume's overall account of the virtue of justice.

Hume's account is distinguished by its ability to enrich both the description of the problem of resource instability and the description of justice with historically verifiable data. Hume moves from a very general description of the value of justice as an instrument capable of resolving conflicts that human beings cannot help but encounter, given the contingent features of their environment and psychology, to a more particular and

historically determined description that frames the value of justice in the commercial society in which Hume's contemporaries lived. The general description of the value of justice is important because it shows that the model has great explanatory fertility that can be applied to any community in which, despite socio-cultural variations, the parameters of limited human benevolence and resource scarcity remain constant. However, the explanation of the value of justice does not stop at this general level. By enriching the description of the problem of resource scarcity with empirical data, Hume's explanatory model also manages to account for the particular way in which justice is conceived in commercial societies. In particular, the fact that his fellow citizens regarded the repayment of a loan as a virtuous act of justice that must be performed regardless of the effects of that particular act on the interests of the individuals involved or on the interests of society.

Of course, the general and the specific explanations are not unrelated. But in what sense are they related? Both form an overall account that shows that, if we examine the various causes that have led us to ascribe value to this concept since its emergence, there is nothing in them that leads us to doubt the value that we ascribe to it today.[64] On the contrary, the content of the explanation of this stage strengthens our commitment to justice. This does not mean, however, that this content can form the basis of an argument to persuade the sensible knave to respect justice as a virtue rather than instrumentally. Rather, Hume's explanation provides additional reasons for those who already respect justice as a virtue to continue to do so. As I have said, the account shows that there is nothing in justice, whether we regard it as a means of resolving inevitable human conflicts, or as something which we approve by our impartial sympathy with its effects on society, which threatens our confidence in being guided by it in the way it does.

[64] Queloz described this using the expression '*negative* vindication' and argued that this makes Hume an epigone of the method B. Williams (2002: chap 2) see Queloz 2021: 98.

2.4. Justice as Respect for Promises

In addition to the rules of property and the virtue that motivates us to respect them, justice also includes the institution of promising and the virtue that identifies 'trustworthiness with respect to one's word'.[65] Promising is a central aspect of our social interactions with others, without which the value of the institution of property itself would be severely limited. The transfer of property and trade often depend on people committing themselves now to perform certain actions at a later date. These forms of social interaction, typically between strangers, are based on trust that the promisor will keep his word. For Hume, trust thus rests on the common-sense assumption that there is such a thing as a stable disposition to do what we have promised to do, which operates independently of our interests at the time when we have to perform what we have agreed to do. In treatise 3.2.5, Hume sets out to explain the emergence of this virtuous trait, which, like respect for property, cannot be explained by any natural motive.

As in the case of respect for property, Hume believes that trustworthiness in one's word is generally recognised as a virtue. In other words, it is not just a behavioural disposition but, as we saw when analysing the Circle Argument, a disposition to have certain passions and certain sentiments that lead us to act in certain ways in certain circumstances. Thus, fulfilling promises is only virtuous if it is done for certain reasons and not for others. For example, Hume rules out the possibility that the 'full behavioural profile' of the person expressing this virtue can be caused by any natural non-moral motive. The proper motive, capable of ensuring respect for the given word in all circumstances, is that we must keep our promises precisely because they are promises and it is our duty to do so.

[65]The expression is used by Cohon (2008: 193) to describe the virtue of keeping promises.

In *Treatise* 3.2.5, Hume addresses in particular two difficulties concerning the keeping of promises. The first is to explain the emergence of this non-natural moral motive which makes the act of keeping one's word virtuous. The second, a problem peculiar to this virtue, is how it is possible for us to oblige ourselves to do something in certain circumstances simply by saying certain words. I will leave the second question aside,[66] as it is not directly relevant to the discussion in this chapter, and concentrate on the first.

The relevant aspect of Hume's investigation is that he uses exactly the same explanatory model as for the virtue of respect for property; in particular he resorts to a staged explanation that revolves first around self-interest and then around the explanatory contribution of moral sentiments. The first stage is thus based on a general problem that characterises the socialised state before the institution of the promise. Because of their selfishness and limited generosity, people do not trust each other to exchange goods and services that require actions to be performed at different times. They do, however, recognise that it is in their interest to exchange a greater quantity of goods than would be possible through simultaneous exchange without the use of promises. As in the case of property, each person expresses this interest to the others and gradually becomes aware that this interest has become common knowledge. In this way, people begin to agree on the need to use a linguistic formula to distinguish this useful type of exchange of goods from those that take place between blood relatives or friends, which are simply based on the trust that arises from the natural virtues of mutual benevolence and gratitude. Thus, it is stated that whenever 'a man says *he promises any thing*', he is expressing a determination to act in accordance with that promise and 'subjects himself to the penalty of never being trusted again in case of failure' (*T* 3.2.5.10/SBN 522). As with the

[66] For a discussion of this aspect and the various uses of the term 'promise' in Hume, see Baier 1985: 174 ff. See also Cohon 2008: chap. 8.

institution of property, self-interest not only drives us to seek a solution to a problem determined by the interplay of human psychology and circumstances, but also gives us a motive to conform to the practice that allows the problem to be solved. In this case, the motive depends on our interest in not being excluded from mutually beneficial forms of cooperation. As society and consumer goods grow, this motive proves insufficient to ensure stable compliance with the practice. Hume thus introduces a second stage in the explanation of the obligation of promises. As in the case of respect for property, the instrumental respect for the covenant of promising is gradually replaced by a non-instrumental motive which, created by moral sentiments and reinforced by public and private education, is transformed into a stable trait or virtue which motivates us to respect promises simply as promises.

The discussion of the traits of the trustworthiness of one's word and respect for property shows that Hume has a unified and internally consistent account of the virtue of justice. In the next section I will consider whether and to what extent this can be seen as a form of genealogy.

2.5 Is Hume's Theory of Justice a Genealogy of Justice?

In this concluding section I will argue that Hume's specific explanation of justice can be considered a form of genealogy. Furthermore, I will argue that Hume brings an element of originality to this method. While *vindicatory* genealogies typically provide only one kind of justification of the practice they explain, Hume's method allows for two different levels of justification. Before discussing this thesis, I will briefly describe the main elements of the genealogical method. Much of what I shall say depends on the way in which the analysis of this concept has been conducted in the contemporary debate

involving Bernard Williams and Edward Craig. This means that my discussion of that concept will contain nothing more than a set of answers to purely methodological issues.

Genealogy is a method of philosophical investigation that aims to explain how a concept, a practice, a belief system or a virtue emerged in human culture. Genealogy seeks to provide the history of the process that has led to the phenomenon under investigation, though it is a distinctive type of history. It is not just a description of how the phenomenon came about and was used in the past, but - as Edward Craig claimed – it is an 'historical narrative' that 'should throw some light' on the concept as we currently understand it.[67] This means that genealogy should give some sense of the fact that the concept in its present form is an evolution from previous stages.

This first characterization gives us a first element for distinguishing between genealogy and historical chronology. It is too generic, however, since, as it is, it still does not allow us to discriminate between different forms of narratives. We might ask: what exactly is the status of genealogy? Is it a factual history or can it be merely imaginary?

In *Truth and Truthfulness*, Bernard Williams argued that it can be either, as this passage suggests:

> A genealogy is a narrative that tries to explain a cultural phenomenon by describing a way in which it came about, or could have come about, or might be imagined to have come about.[68]

This interpretation of the genealogical method might be criticized as too liberal. One could object: in what way can a purely imaginary or fictitious account actually explain the emergence of a phenomenon? An account that aspires to shed light on a

[67] See Craig 2007: 184.
[68] See Williams 2002: 20.

phenomenon should be, at least in the author's intentions, a factual and not fictitious account. This objection certainly makes an interesting point, one that helps us to introduce a further feature of the genealogical method. Let us first see what the feature is and then go back to the point.

Genealogies are intended not only to shed light on a phenomenon but also to affect, through that narrative, our attitudes towards that phenomenon.[69] In relation to this point, genealogies are roughly divided into two types: *subversive* and *vindicatory* of the phenomenon they aim to describe.[70] In the first case, the account shows how the target phenomenon cannot be accepted without undermining our esteem or our epistemic certainty regarding it. In the second case, the story serves as a recommendation: showing how the phenomenon in question stems from the need to find a solution to a human problem that any human society is facing.[71]

In the face of this additional characteristic, the problem raised above is clear. If genealogy is not simply an account of a phenomenon, but something that aims to affect our attitudes towards that phenomenon, how could an imaginary story perform this function? How, for example, could a story about the emergence of the virtue of honour lead us to undermine our confidence in this virtue or, conversely, to reinforce it, if this story is openly fictitious?

Certainly, there is a sense in which genealogy must be factual rather than fictional, that is to say, it must refer to events set against the background of practices or institutions for which we can in principle gather evidence, as I suggested in relation to the origin of religious belief in Hume's *Natural History of Religion*. Nevertheless, there is a sense in which Williams's statement that genealogy is a narrative that describes how a given

[69] See Craig 2007: 186.
[70] See Craig 2007: 182-183. By that I don't mean that there cannot also be neutral genealogies, that is, that they are neither vindictive nor subversive, but I shall not deal with them in this chapter.
[71] See Queloz 2018: 1 ff.

phenomenon 'might have come about, or might be imagined to have come about' sheds light on a crucial fact. Genealogy, typically, gives an account of the emergence of a phenomenon out of a thesis on how human beings, in an unspecified time and place, have reacted psychologically to a certain practical problem. Due to its generic nature, this thesis can never be *completely* based on empirical evidence. This poses a very precise problem: unless the characteristics of human beings described in the initial stage are very general, i.e. potentially shared by any human being regardless of his or her culture, genealogy finds it difficult to justify its statements. This point has been well made by Edward Craig. Commenting on Hume's *Natural History of Religion*, which Craig rightly considers to be a non-fictitious story, he notes the following potential problem:

> Initially, we were worried by the question 'If the state of nature is something imaginary, how can it explain anything?' But it seems – for the moment at least – that that may not be the problem. Where, as in this example from Hume, the posited state of nature isn't imaginary, it can't be the problem. But there surely is one. […] Sticking with Hume's Natural History of Religion, suppose we are asked what reason we have to think that human beings reacted to the experience of those facts by imagining, and coming to believe in, a number of invisible person-like powers manipulating nature. *We aren't talking about any particular people, so our answer must take the form 'Human beings are like that'* or, rather, as it can hardly be maintained that all humans would react in that way (most of us wouldn't for a start) 'Human beings with property X (e.g., untouched by the cultural developments of the last three thousand years) are like that.' ... [this] theorist has an epistemic hill to climb, if not a mountain. Unless we are dealing with the most basic, almost animal, reactions, or those without which their very survival would have been threatened, how sure can we be that they were indeed like that?[72]

[72] See Craig 2007: 187-188.

If the origin of the phenomenon lies in a distant past, the narrative cannot but rely on conjecture about how human beings came to have certain beliefs and develop certain attitudes that resulted in the phenomenon in question or a basic version of it. Is this a problem for the genealogy that aims to shed light on or explain a phenomenon?

It is not, and so we come to the third features of genealogy, if conjectures cling to historically documented fixed points (e.g. Hume's military reversals, political institutions, etc.) that bring this narrative closer to a true story, and conjectures are made starting from such a minimal conception of human nature that it can be assumed that any human being, however distant in time from us, must have it.

Let us now come to a question left open until now: what is the function that the past, or human prehistory, plays in this method? There seem to be at least two possibilities: the first is that the past serves to make the point of the practice more perspicuous; the second is that the past serves to illuminate the origins of a simplified version of the practice and this sheds light on how humans participate in the present practice.

In the first case, we assume a general hypothesis of what is the role of practice in human life starting from a description of certain general facts about human nature. This hypothesis is then illustrated through a description of a distant past in which the conditions of human life are artificially simplified to bring out the contribution of practice to human need. In this version, the past does not serve to describe how the idea of the practice took hold or how this idea could produce a certain behaviour: the function of the past is simply to make the relationship between the core of the target phenomenon and human needs more perspicuous. In the second and third case, on the other hand, via the past, we can see how the idea emerged and how it became a practice.

It is clear that these two ways are linked to two distinct methodological demands. The function of the first mode is simply to recommend something: for example, that human nature is such that if we do not embrace that practice there will be negative effects

for all of us. The second mode has explanatory ambitions. The past indicates a stage of human life prior to the emergence of the target practice that serves to explain how it came about. For these purposes, it is not necessary, as in the first case, for the original practice to be identical with the present one. What is needed is that the original practice contains the basic elements of the current one, and that there is a narrative which, through different stages, illustrates the passage between the two.

Are both uses congenial to genealogy? On the basis of what we stated at the outset, it is now clear that genealogy cannot make an exclusively exhortative and recommendatory use of the state of nature, but it must also play an explanatory function. This does not mean that the reference to the state of nature should be an essential aspect of this method; indeed, this may not be the case. But it does mean that if there is such a reference, this is not merely a literary instrument, a way of presenting a generalization about human nature that serves to justify something, but is instead a component of its explanatory strategy.[73] It can therefore be argued that a third characteristic of the genealogical method is a particular use of the state of nature through which its constitutive explanatory dimension is implemented.

A fourth characteristic that defines the genealogical method and that differentiates it from other forms of functional explanation of a concept, practice, etc. has to do with the content of this explanation. As I mentioned at the beginning, genealogy explains a practice, a concept, an institution on the basis of what it does for us, i. e. the way it allows us to solve a typically human problem that depends on the interaction between very general characteristics of human nature and the environment. Genealogy explains the emergence of a concept out of the function of that concept. But that is not all. Indeed, genealogy is considered a method that brings to light a function that is typically hidden.

[73] See Craig 2007: 193-196.

As Williams argued, the practices that are the object of genealogical investigation, typically, are not respected because their existence satisfies human needs, but because doing so has a value in itself. In other words, these practices are not perceived to have an instrumental value, but an intrinsic one. As Craig has repeatedly argued, the specificity of the genealogical explanation is precisely its ability to explain how the attribution of the intrinsic value is justified out of the hidden instrumental function of the practice, which the genealogical method has unveiled. The outcome of the genealogical enterprise is in fact to show that the practice performs its function more effectively if it is respected, independently of that function. In this way, unmasking of the hidden function, far from weakening our respect for the practice, makes it intelligible and strengthens it. Following Craig's line one can then say that the genealogical enterprise is able to explain both the instrumental character of a practice and its intrinsic value, and is also able to explain the connection between these two.

In the light of the above, the account of justice that Hume develops in Part II of Book III of the *Treatise* can be considered a form of genealogical explanation. First, Hume's theory has a clear explanatory dimension. Through his discussion of the two possible starting points, Hume wants to emphasize not only that he has the resources to explain the benefits of the practice of justice, but also how human beings come to discover these benefits. Moreover, the scope of Hume's explanation is twofold. On the one hand, it gives an account of our natural obligation to justice, which arises from self-interest and leads us to limit our natural greed in order to avoid conflict between external goods. On the other hand, he explains how, when society becomes larger, it is vital that respect for justice is entrusted to a virtuous motive which approves and respects justice in a non-instrumental way. Hume's theory is therefore able to account for both the instrumental and the non-instrumental value that we attribute to justice.

Finally, the vindicatory dimension. The way in which Hume takes into account the link between instrumental and virtuous motivation certainly leaves room for a justification of the virtue of justice that rests on the fact that it allows larger societies not to dissolve, and this satisfies the interests of all. This is clearly in line with the model of genealogical justification indicated so far. Hume, however, offers a further dimension for the justification of the virtue of justice which provides a new line to the genealogical model. Justice as a virtue can be further endorsed not only because we have an interest in doing so, but also because it is based on our ability to impartially sympathize with the suffering of others that we are able to approve reflectively (*T* 3.3.6.6/SBN 620-1).

Chapter 3

Resentment and Injustice. Is Resentment just a 'Third Condition of Justice'?

In the previous chapters, I have pointed out how human passions are crucial explanatory factors in Hume's genealogy of justice. This is true for at least three different passions, namely benevolence, enlightened self-interest, and pride. With regard to the former, Hume argues that the idea of justice is intelligible only in the condition, which in fact characterizes human nature, in which each person has a moderate tendency to desire the good of those with whom he/she has relations. Without this psychological condition of a medium level of altruism, justice could not have been invented. If human beings were motivated exclusively by universal benevolence, moderate scarcity of goods would not have been a problem for human beings since everyone would always be instinctively driven to sacrifice his own well-being for that of his fellows. Similarly, and contrariwise, if no one were capable of feeling any altruistic inclination, no one would have been able to form the family units that are crucial to gaining the experience of the benefits of cooperation, without which no one could have imagined the benefits of the conventions of justice. Private benevolence, however, is not sufficient to explain the origin of justice. Hume argues that the only passion capable of counteracting the destructive tendency of greed is greed itself or enlightened self-interest, which, through experience and reflection, is capable of changing its natural direction and inclining us towards cooperative behaviour. Finally, the idea of justice as a virtue, the possession of which ensures respect for property and promises in advanced commercial societies, can only be acquired

because private and public education latches onto the indirect passions of pride and humility.

The passions are thus an explanatory factor that is crucial in both phases of the genealogy of justice, not only in the account of the emergence of natural obligation, but also the later phase that focuses on the moralisation of this idea and respect for the rules of justice as ends in themselves. In this chapter I intend to add to this line by arguing that Hume's description of the process of the moralisation of justice would be incomplete if it did not also include the dark passion of resentment.

The connection between resentment and justice has only recently become the subject of some interest in Hume scholarship. It is by no means an easy topic, not least because Hume does not offer any systematic examination of resentment in Book II of the *Treatise*, where it seems to be merely mentioned as an instinctive antisocial tendency. Nor does Section II of Book III ever explicitly mention this passion either, suggesting that it has no explanatory role to play in the account of society's evolution. It is, however, present in Section III of the *Enquiry Concerning the Principles of Morals*, where it is even claimed by some to be the third of the so-called 'conditions of justice'[74], that is, one of the factors without which justice could not have arisen among human beings. Martha Nussbaum has argued, for example, that alongside limited benevolence and scarcity of resources, the ability to make the effects of one's resentment felt would be a new necessary condition for joining the community of people whose relationships are governed by the rules of justice.[75]

[74] The expression is by John Rawls (Rawls 2020: 126) and indicates the 'normal conditions under which human cooperation is both possible and necessary'. Rawls argues that his account 'largely follows that of Hume in A Treatise of Human Nature, bk. Ill, pt. II, sec. ii, and An Enquiry Concerning the Principles of Morals, sec. III'.

[75] Nussbaum argues that this condition of justice is derived from what John Rawls, commenting on Hume, calls the condition of equality in the physical and mental capacities of those who enter the convention of justice. As Rawls claims: 'Individuals are roughly similar in physical and mental powers; or at any rate, their capacities are comparable in that no one among them can dominate the rest' (Rawls 2020: 126-7.)

This integration into Hume's so-called conditions of justice raises several questions. Why, for instance, is this condition not present in *Treatise* Book 3, Part 2, where Hume elaborates his most detailed reconstruction of justice? Moreover - leaving this question aside - if this capacity plays such an important role, why is resentment not given any systematic treatment in Book II as is the case with the other elements of his moral psychology?

Beyond these issues, it is the very idea that resentment is a condition of justice that underlies a powerful critique of the Humean model, which questions its inclusive capacity. Martha Nussbaum has argued that Hume's integration makes his conventionalism vulnerable to the same difficulties as Rawlsian contractual liberalism. In *New Frontiers of Justice*, she argued that Rawls' theory of justice is guilty of formulating the principles of justice on the basis of the interests only of those who are rational, equal, free and independent, to the exclusion of those who, due to cognitive or physical disabilities, are not equal to or are dependent on others.[76] In this way, not only are those with disabilities excluded from the circle of contractors who choose the fundamental principles that are to govern just institutions, but, furthermore, these principles are not even formulated with the interests of those who do not participate in the deliberative process in mind.[77] The interests of these subjects may only be taken into account later, i.e. by secondary principles, and not by the primary principles that for Rawls have the task of designing the institutions of the just society. This procedure renders just institutions inevitably unjust, insofar as they are not sufficiently inclusive with respect to the interests of some of the subjects living under those institutions.

[76] See Nussbaum 2006: chap 1.

[77] Annette Baier (Baier 1980: 133ff.) noted that if this 'almost Hobbesian condition' were true, it would risk depriving not only animals but also women, the elderly and children of their rights. Unlike Nussbaum, however, Baier argued that Hume is quite capable of explaining how 'the weak' by uniting in confederations are able to make their voices heard against their 'masters'.

Nussbaum believes that the constraint with respect to equality in the power to retaliate, i.e. the 'ability to make the effects of one's resentment felt', introduced in Hume's second *Enquiry* makes the Humean sentimentalist model of justice vulnerable to the same criticism as that of Rawlsian contractualism. Again, those who do not have the same strength (typically that of males!) in creating problems for those who attack them by making the effects of their resentment felt will be excluded from relationships and pacts governed by justice. The use of violence against them and their property by the strongest 'equals' will find no restraint in the rights and rules of justice, but can only be tempered by their compassion and benevolence.

Given these difficulties, it is therefore not surprising that the topic of the relationship between justice and resentment has been much neglected by Hume's scholarship. Not only does it raise delicate interpretative problems that also affect the relationship between the moral and political philosophy of the *Treatise* and that of the second *Enquiry*, but it seems to condemn the Humean conception of justice to a particularly unpleasant conservatism. In this chapter I intend to address this issue and I will do so by attempting to answer three questions. The first is what exactly Hume's conception of resentment is. The second is whether or not it is a plausible conception. The third, finally, is what role it plays in his theory of justice, not only in the second *Enquiry*, but also in the *Treatise* where it seems *prima facie* to have no function to play. In answering these questions, I will argue the controversial thesis that not only does Hume present a coherent and defensible conception of this passion, but that it plays a non-secondary role, both because it underpins the moralisation of justice and because, once the moralisation process is established, it reinforces the feeling of moral disapproval of injustice that would risk being an ineffective sanction if it were based solely on sympathy with the public good.

Let us start with the first issue, i.e. the investigation into the nature of resentment in Hume. In order to perform this task, however, it is useful to start with a general notion of resentment and from this specify the peculiar elements of Hume's description.

3.1 Resentment: an Outline

The passion of resentment has only recently begun to receive the attention it deserves from contemporary philosophical psychology. In fact, if we look at two important essays of the last century that dealt with negative emotions, although they identify some elements in the passion of resentment, they do not offer a systematic analysis of its constituent components. Peter Strawson's influential analysis, for example, merely lists it, along with indignation and guilt, in the class of 'reactive attitude and feelings' all of which track 'the great importance that we attach to the attitudes and intentions towards us of other human being'.[78] In a similar way, R. J. Wallace characterized resentment as a member of 'reactive emotions', which are connected to evaluation.[79] More recently, resentment has begun to be explored both in relation to other negative emotions and with respect to the function this passion has played and can play in the liberal political tradition. Charles Griswold, for example, has developed an analysis of resentment that, in the wake of some of Joseph Butler's reflections, shows how it is not incompatible with the positive emotion of forgiveness.[80] Subsequently, Martha Nussbaum presented a cogent description of anger, in which resentment appears as a specific case. Following a different line of analysis, Michelle Schwarze illuminated the role that 'sympathetic resentment' plays in the political philosophy of the Scottish Enlightenment and O. Flanagan examined the

[78] See Strawson 2008: 24.
[79] See Wallace 1994: 15ff.
[80] See Griswold 2007:19-21.

positive role that anger and resentment, purified of its element of revenge, can play in complex liberal democracies.[81]

Although these considerations have captured some aspects of resentment, they they merely consider it a species of the genus anger, but do not attempt to indicate its specific elements. Therefore, based on the analysis of these recent studies, I will try to set out a more precise description of this passion.

Resentment can be characterized as a painful reaction of indignation towards the author of a wrongful act against us that makes us wish for his/her suffering.[82] From this first rough description, resentment is a passion that has an intentional object and is aroused against the background of beliefs and evaluations. Let us examine these two aspects separately.

Intentionality is the property of being directed toward something or someone.[83] Resentment, as well as other passions such as anger or fear, possesses this property. When I feel resentful, I am always resentful towards someone, i.e. the one I believe has acted wrongly towards me. This someone is the person towards whom resentment is directed, i.e. its intentional object. That feature distinguishes this painful passion from bodily sensations like pain. When I have a headache or experience acute pain from breaking a hand, those feelings are not directed at anything or anyone. They are located in a specific region of our body, but they do not have an intentional object.

Examining the intentionality of resentment allows us to specify a first element of distinction with respect to anger. The presence of an agent is crucial to the intentionality of resentment: this emotion is always directed to someone and not to something. We do not feel resentful against a thunderstorm that prevents us from going to our appointment or against our car that seems unwilling to start when we are in a hurry. However, I can

[81] See Schwarze 2020: Intro. See Flanagan 2021: chap. 3.
[82] See Nussbaum (2016: 261) for this definition.
[83] For an analysis of this notion of 'intentionality', see Deigh 1994: 39-71.

feel resentment if the defective car was sold to us at a high price by one of our family members who betrayed my trust in him. By contrast, anger, besides being directed against someone, can also be directed against things, as when we have a burst of anger when the computer suddenly freezes.[84]

In addition to having an intentional object, resentment is typically dependent on a set of related beliefs and evaluations. Let us examine each of these two requirements.

Resentment is a painful emotion which is aroused by the perception or belief that a wrongful act has been committed against us. I do not, therefore, need to have actually been injured to be resentful; what matters is that injury is part of the content of my perception. The subsequent discovery that I misinterpreted the action generally has an effect on the passion. It may be a reason to stop wishing to take revenge or, having already avenged myself, to apologize to the alleged attacker. But that does not mean that the perception of the harm is not in itself a sufficient basis for making me feel this emotion.

In addition to the perception of injury, the presence of a set of causal beliefs is also necessary.[85] As Martha Nussbaum has acutely argued, although the *focus* of this passion is on actions, its *target* is not actions but persons.[86] Therefore, to feel resentment we must be able to master some form of even elementary causal reasoning, by which to attribute a certain injurious action to its author. This does not mean that we must be able identify the injurer, it is indeed quite sufficient to assume that there is someone responsible for the wrong. If I discover that my home has been robbed and even damaged

[84] Similarly, Griswold, for example, argues that 'the objects of hatred are of wider scope than those of resentment, including inanimate things, conditions such as illness, theories or principles, groups of people (all rapists, for example), states of affairs such as poverty'. See Griswold 2007: 25.
[85] Discussing anger, Nussbaum has argued that to feel this passion we need to have a 'causal thinking' through which we attribute the wrongdoing to its author. Although she did not explicitly formulate this thesis for resentment, the brief observations she made suggest that she considered them valid for this emotion as well, see Nussbaum 2016: 17.
[86] See Nussbaum 2016: 17. On this point I follow Nussbaum and depart from Jean Hampton (1988: 60) who instead argues that the object of resentment is always an action. Hampton uses this element precisely to distinguish hatred from resentment, arguing that the former passion, unlike the latter, can also have people among its objects.

for the simple malicious pleasure of doing so, I will tend to resent the offenders even if I do not know who they are.

Finally, the perception that one has been injured presupposes a negative evaluation of the action, assessed against the background of norms that the agent endorses. This feature highlights an important property of this passion. Resentment does not seem to be identifiable as an arational and merely idiosyncratic feeling of aversion or dislike, but rather as an emotional reaction that has a rationale.[87] The person who experiences it tends in fact to believe that she has some justification for feeling as she does. This line of thinking can be and has been expressed in different ways. Thomas Nagel, for example, has argued that when we feel resentment we believe that our suffering gives the person who causes it a reason to quit and that if he does not, he acts against a shared reason that is fully available to him. Resentment is therefore a passion that is based on a judgment that recognizes the violation of a general principle, the validity of which is shared between the parties.[88] Again, Martha Nussbaum, who considers resentment a species belonging to the anger genus, has argued that it involves a judgement of wrongfulness, not only moral but also about perceived slights to social status. She argues:

> So the question for me, then, is whether 'resentment' contains a particular type of judgment of wrongfulness, namely, a moral type? I think our linguistic intuitions just do not support that claim. When a person describes her emotion as resentment, that typically would suggest that she believes it has some grounds. But must those always be moral grounds? If a person is insulted in a typical down-ranking way, she might well say, 'I resent that'. If a school rejects one's child as an applicant, the aggrieved parent, believing that the school was both careless and mistaken, might say that she resents the way the

[87] For a similar point, see Griswold 2007: 26. He argues that 'resentment is aroused by the perception of what we (the spectator of the scene or the victim) - regard as "unwarranted injury" '. From this analysis, Schwarze draws the conclusion that resentment is a 'moral sentiment'.
[88] See Nagel 1970: 83.

> school acted—without even raising the question whether a moral principle was involved. 'Indignation' is similarly slippery. I can be 'indignant' about insults to status and rank, about nonmoral affronts of many kinds. So, although many cases of resentment and indignation are surely moral, I don't think they all are. We can do better, I think, by focusing on the generic term with its implied judgment of wrongfulness, and then getting clear, in each case, about what type of judgment it is.[89]

Accepting both Nagel's and Nussbaum's suggestion, then: when a person feels resentful, he is able to justify his/her passion to himself and others. The reason may be moral, but it does not have to be so. As Nussbaum points out, resentment can and is often caused by actions that we perceive as wrong because they do not respect or they degrade our social status. In the first case, the justification may concern various types of immorality. For example, it could be about the offender violating a moral principle such as not respecting a person's dignity by manipulating him/her, or even committing physical violence on him/her. But it can also be about the action showing the lack of an important virtue in the character of the offender, as when we are in the presence of openly disloyal behaviour or lack of gratitude. In the second case, however, the justification will concern the fact that the reputation or social status of the victim has not been taken into account by the wrongdoer. This can take place in different kind of social interaction. They may be our peers who do not recognize us as their peers and exclude us from their social circle. Or they may be people we consider inferior to us who treat us as if we were at their level.

Taking Nussbaum's point for granted, let us distinguish between two grounds for resentment: those actions which bring about a harm that is constitutively related to the person taking offence at a straightforward insult and those that produce a harm that is

[89] See Nussbaum 2016: 262.

constitutively related to not recognising the offended person's rights.[90] As it stands, this description is incomplete as it makes no reference to the willingness to harm by the author of the wrong act. But this seems to be a crucial element to consider: an essential part of the pain caused by the injury, that make us wish for its author to suffer, is precisely the fact that that person wanted me to suffer.[91] Taking this aspect into account, Peter Strawson suggests a slightly different description of the grounds of resentment. Indeed, he distinguishes between two types of actions that cause pain: those whose pain is describable independently of the intention to offend and those in which it is not describable independently of this intention. Resentment only arises in the latter case, but not in the former. Strawson claims:

> If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard of my existence or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment that I shall not feel in the first […] These examples are of actions which [...] inflict injuries over and above any […] inflicted by the mere manifestation of attitude and intention themselves. We should consider also in how much of our behaviour […] injury resides mainly or entirely in the manifestation of attitude itself. So it is with good manners […] with deliberate rudeness, studied indifference, or insult on the other.[92]

Strawson's description maintains Nussbaum's distinction between moral and non-moral grounds, but formulates it by giving a central role to the wrongdoer's intention,

[90] Resentment is not caused by dispositions to act or character traits. This distinguishes resentment from other painful passions that are directed against someone and not against something. Disgust, although directed towards individuals (as well as ethnic groups), is typically caused by bodily secretions considered impure. Similarly, hatred, which targets a person, is characteristically caused by personality traits.
[91] It could be argued that this is not a constitutive characteristic of resentment. I can resent someone's being more successful than me, independently of whether I think they have deliberately sought to harm me. Whether or not this objection is justified depends on whether or not one thinks that resentment is similar to envy. In the remainder of this chapter, I will assume that the two passions are distinct.
[92] See Strawson 2008: 23.

which makes it a better description than the one given by Nussbaum. To explain this point, we must examine a further element of the content of resentment, namely its painful feeling. Unlike anger, which is accompanied by a wide range of subjective feelings (and bodily reactions), resentment seems to be identified by a specific painful one, a kind of mortification that generates a desire for revenge. This is not a feeling related only contingently to the passion, but constitutes it and therefore can enter into its definition. Now this particular mortification cannot be explained otherwise than by assuming the injurer's will to deride our reputation or - when a moral injury is involved - to trample on the subject's humanity. But this is precisely Strawson's point about grounds, when he says that whether or not it is the case that the suffering depends entirely on malevolent attitudes or deliberate lack of respect, their perceived presence seems to be a necessary condition for this passion to arise.[93]

This thesis contrasts with the famous and detailed description of the resentment that Joseph Butler provided in chapter seven of the *Fifteen Sermons Preached at Chapel Hill*. Butler argues that human beings can experience two types of resentment. The first is 'hasty and sudden', while the second is 'settled and deliberate'.[94] The first type is described as a 'sudden anger'[95] which is typically caused by violent and 'sudden hurt' [96], that is by intense pain. This kind of resentment arises in the mind without the need for the intervention of reason[97]: its presence does not require the offended person to form any representation of the malicious intention[98] of the aggressor or, possibly, of the contempt he held us in[99]. Butler claims:

[93] Strawson believes in fact that the discovery that the injurious action was accidental, or that the author acted because forced, cancels resentment.
[94] See Butler [1736]: 196.
[95] Ibid.
[96] See Butler [1736]: 198.
[97] See Butler [1736]: 197.
[98] See Butler [1736]: 198; 204.
[99] See Butler [1736]: 197.

> Sudden anger, upon certain occasions, is mere instinct; as merely so as the disposition to close our eyes upon the apprehension of somewhat falling into them; and no more necessarily implies any degree of reason. […] Now, momentary anger is frequently raised, not only without any real, but without any apparent reason; that is, without any appearance of injury, as distinct from hurt or pain. It cannot, I suppose, be thought that this passion, in infants, in the lower species of animals, and, which is often seen, in men towards them; it cannot, I say, be imagined, that these instances of this passion are the effect of reason: no, they are occasioned by mere sensation and feeling.[100]

Anticipating a theme that will be taken up by Adam Smith[101], Butler emphasises the usefulness of this instinctive reaction that enables human beings to defend themselves against the sudden violence of others immediately, even before being able to ask whether or not it depends on an evil intent on the part of its perpetrator.[102] Contrary to Strawson, Butler believes that there is a type of resentment that is aroused simply by the perception of physical harm, without it being accompanied by the representation 'of the evil designed or premeditated' [103] against ourselves.[104]

Does Butler's analysis, particularly the existence of so-called hasty resentment, constitute an objection to the Strawsonian thesis I am arguing? If there is an immediate form of resentment, similar to anger, is it still possible to argue that establishing the malevolent will of the injurer is a necessary condition for feeling resentment?

---

[100] Ibid.

[101] As we shall see in chapter 4 of this thesis, Smith claims that resentment is a reaction supplied by a provident Nature for our personal defence. Reference References to Smith's *Theory of Moral Sentiments* is cited with notations of the form *TMS* j .k.m.n, the lower-case letters here standing for arabic numerals. Numerals immediately following *TMS* indicate part, section, chapter, and paragraph in Smith 2002 (1790; 1st edn. 1759). On this point see *TMS* II.ii.I.4.

[102] On this aspect, see Griswold 2007: 22.

[103] See Butler [1736]: 202.

[104] I am not going to dwell here, because the subject is not relevant for the purposes of this investigation, on the elements of difference and continuity between sudden resentment and stable resentment. On this issue see Griswold 2007: 22-31. See also, Schwarze 2020: chap. 2.

Butler appears to have a broader notion of resentment than our ordinary conception of this passion. I believe his notion is too broad. As I will examine shortly, our ordinary notion of resentment is constructively related to the desire for revenge, but that is precisely what hasty resentment lacks. As Butler stated, it defends us from other's attacks not through the threat of revenge, but by putting an end to our suffering in any way possible, not least by moving away from the source of the pain.[105] In the light of our ordinary conception, the Strawsonian thesis seems plausible: the idea that the cause of resentment must be related to some desire to humiliate us makes it intelligible that resentment is not simply connected to the impulse to put an end to our pain, but is typically connected to the desire to get even with the injurer.

That said, the broad notion of resentment proposed by Butler is instructive for our analysis. The terminological choice of treating the two passions as species of a single genus highlights a connection between the two emotions. Butler seems to relate it merely to the fact that both inclinations are psychological instruments through which nature enables human beings to defend themselves from harm by others. The connection, however, is important for another reason that Butler does not consider and that would become after him a shared view: resentment can be explained as an evolutionary product of a more basic and general instinct of self-defence that human beings share with the rest of non-human animals. This thesis, not explicitly formulated by Hume but nevertheless compatible with his discussion, is however present in both Adam Smith and later John Stuart Mill. In this picture, resentment indicates a passion of moral indignation, which

[105] Griswold emphasises this distinction. He also implicitly defends what I call the ordinary conception of resentment, which thus distinguishes this passion from anger, through the etymological analysis of the term resentment. Griswold argues that 'to resent' means 'to feel a sentiment again' a feeling that lingers beyond the event that triggered it. Reproducing this feeling does not simply require remembering the event, but a memory that continues to provoke a feeling of anger that sustains the desire to want to get even with our aggressor. This 'temporal projection of self into the future', who continues to suffer until he has satisfied his desire for revenge, is not typically present in the case of anger, which is generally confined to the time of the harm. See Griswold 2007: 23.

arises in response to the violation of rules concerning fair behaviour among people that are necessary conditions for the stability of society. Smith and Mill believe that this highly useful passion for society can be explained out of a simpler instinct that human beings share with the animal world. Butler's option, therefore, is important because, by highlighting the similarity between these two emotions, it indirectly paves the way for a naturalist explanation of resentment that is a key element of the genealogies of justice that are examined in these chapters.

Yet, at this point an objector might still have some misgivings about Strawson's thesis. Granted that our ordinary notion of resentment is the one just described, the objector might argue that there are circumstances in which, even though we do not think in any way about the intentions of the agent, we are outraged and resentful towards him/her. Someone might therefore argue that the idea that resentment is constitutively linked to the desire for revenge is not a sufficient condition for arguing that resentment arises only after ascertaining the malevolent intentions of the one who harmed us. Think of Nussbaum's case of parents who are resentful towards the Director who refuses to admit their son in the prestigious school. To his parents, he is a talented teenager who met the entry threshold. They probably think the school is denying an important opportunity to him without even explaining the rejection (maybe the refusal is perceived as an offense to the whole family!). They feel resentment towards the examiners and want to make it a bad publicity to the school, imagining that this will lessen their anger. This is an intelligible reaction on the part of the parents. Moreover, they will probably have no belief at all as to the examiners' intentions when they excluded the boy.[106]

Strawson's description of the grounds of the resentment must be adjusted to take account of these circumstances. Although it is typically caused by the belief that the

[106] It could be argued that the parents' resentment is due to the school's lack of care for their child. So, at the very least, they are guilty of negligence. In the rest of this chapter, I assume that this is not a necessary condition for explaining the parents' resentment.

offender has a malevolent intention to offend us, this emotion can arise when either this judgment is not explicitly present or even when this intention is clearly absent. In these circumstances, it is sufficient for the action to be described as a breach of a rule governing mutual respect. To return to the example of the school, give a reasonable explanation of a rejection that goes beyond the fact that the school has the final say on admissions.

An interesting further aspect of resentment is that this passion is not restricted to the immediate victims since we feel resentment when other people are victims of brutality or humiliation. We are often outraged and ask for revenge because a member of our family has been wronged or an innocent and defenceless stranger has been physically attacked. However, we do not feel resentment for any injury we perceive. How to explain this fact? Discussing this aspect in relation to anger, Nussbaum argued that we 'get angry only at those [instances of wrongdoing] that touch on core values of the self'.[107] This certainly captures one aspect of the problem. If, for example, I believe that Alice's resentment against her parents over inheritance issues depends on a misconception of what parents have to give their children, I am unlikely to resent them. Although correct, Nussbaum's response needs to be further articulated. In order to feel resentment, the spectator must feel an intense non-trivial pain (it is to be connected to the desire for revenge) which must be similar in kind, i.e. a form of indignation, to that of the injured person. A natural way to explain this similarity is that the spectator feel resentment because he/she is able to sympathize with the original feeling of pain of the injured person. Sympathetic communication is subject to several constraints. As Michael Slote has argued, these are typically of two types. Spatial and/or temporal proximity, and cultural similarity.[108] The two factors can be mutually supportive or weakened. The use of this explanatory principle makes it possible to explain the variations in the ability to feel

[107] See Nussbaum 2016: 19.
[108] See Slote 2006: chapters 1-2.

resentment: for example, the same injury suffered by a loved one arouses outrage and a desire for revenge more violent than what we feel for a stranger.

Having said that, sympathetic resentment is not necessarily subject to the effects of the contingent relationships a spectator has with the person being injured. Imagine, for example, a sick person asks to be accepted at an emergency room that rejects him because his documents are not in order (perhaps because he is an illegal immigrant!). If we were to observe this scene, we may resent the doctor who has rejected him. This sympathetic reaction does not necessarily depend on whether the sick person is outraged (he may not be) nor on our relationship with him (he may be a stranger). It may depend on whether we think he has been denied the right to be healed, a right we believe all human beings are entitled to, regardless of their nationality. This type of sympathetic impartial resentment - as Strawson has famously observed [109] - is typically associated with indignation, although indignation, as Adam Smith acutely noted, can also be experienced by the offended person himself provided he takes an impartial perspective on the offence received.

The last crucial element of resentment to consider is the desire for revenge. As I said at the beginning of the section, this passion is characterized not only by a peculiarly painful sensation of pain, but is also associated with the desire that the author of the injury experience suffering. The connection between this passion and this desire does not seem simply contingent but of a conceptual nature, something without which one cannot say one is experiencing resentment.[110] Adam Smith clearly states this point when he contrasts resentment with gratitude:

[109] See Strawson 2008: 14-15. See also Griswold 2007: 26.
[110] This thesis is discussed for example in Nussbaum 2016: 15.

> The sentiment which most immediately and directly prompts us to reward, is gratitude; that which most immediately and directly prompts us to punish, is resentment.
>
> […] To reward, is to recompense, to remunerate, to return good for good received. To punish, too, is to recompense, to remunerate, though in a different manner; it is to return evil for evil that has been done.
>
> There are some other passions, besides gratitude and resentment, which interest us in the happiness or misery of others; but there are none which so directly excite us to be the instruments of either. (*TMS* II, I, I, 2-5)

With regard to the link between resentment and the desire for revenge, there are at least three issues to be examined that can be summarised in three points: 1) whether or not it is a necessary element of the content of the desire for revenge that it is the victim herself who takes revenge; 2) whether the connection between this desire and resentment is hardwired or not; 3) whether it is an element of the content of the desire that there is a proportionality between suffering received and suffering to be inflicted.

On the first point, there are conflicting views. Adam Smith, for example, famously argued that what distinguishes hatred from resentment is the fact that while a hater only hopes that the hated person will suffer, no matter how, in resentment the victim wants to inflict the suffering himself and wants to do so with respect to the injury he has received.

> The hatred and dislike […] would often lead us to take a malicious pleasure in the misfortune of the man whose conduct and character excite so painful a passion. But though dislike and hatred harden us against all sympathy, and sometimes dispose us even to rejoice at the distress of another, yet, if there is no resentment in the case, if neither we nor our friends have received any great personal provocation, these passions would not naturally lead us to wish to be instrumental in bringing it about. […] But it is quite otherwise with resentment […]. Resentment would prompt us to desire, not only that he

> [injurer] should be punished, but that he should be punished by our means, and upon account of that particular injury which he had done to us. (*TMS* II, I, I, 6)

However, as Martha Nussbaum argued, this seems to be too strict a requirement on the content of the wish.[111] The victim's desire for revenge may find satisfaction even if it is not she/he who materially punishes the wrongdoer. Think of when we decide to take legal action against someone who is then punished by law or of women who are victims of an honour offence, especially those who belong to a traditional society, and wants their families to bear the burden of revenge. In such cases, it is not the victim who takes vengeance and yet it is plausible to think that his/her desire for compensation is still satisfied.

It could be argued that in such cases the avenger remains an instrument of the victim's will and therefore this does not change Smith's thesis much. To these cases, however, we can add others in which this instrumental link is absent. It is not uncommon, for example, that as a result of emotional damage such as betrayal and abandonment from our partner, we hope that he/she will be treated in the future the same way that he has dealt with us, i.e. that he will be abandoned by other partners. Here we are not the ultimate authors of the revenge, which is instead entrusted to a reparatory fate.[112] It does not therefore seem necessary that, in order to satisfy the victim's desire for revenge, it should be the victim himself who inflicts suffering on the offender. Admittedly, the fact that in the future it is I who will reject the partner who betrayed me could give me more pleasure

[111] Although Nussbaum mainly discusses anger, she seems to think that anger shares this characteristic with resentment. She argues: 'The claim is not that anger conceptually involves a wish for violent revenge; nor is it that anger involves the wish to inflict suffering oneself upon the offender. For I may not want to get involved in revenge myself: I may want someone else, or the law, or life itself, to do it for me. I just want the doer to suffer. And the suffering can be quite subtle. One might wish for a physical injury; one might wish for psychological unhappiness; one might wish for unpopularity. One might simply wish for the perpetrator's future (your unfaithful ex's new marriage, for example) to turn out really badly. And one can even imagine as a type of punishment the sheer continued existence of the person as the bad and benighted person he or she is: that is how Dante imagines hell'. See Nussbaum 2016: 22-23.

[112] See D'Urso 2013: 115 ff.

than if he/she were betrayed by someone else. But the fact that this eventuality gives me more pleasure does not mean that this is the only way to satisfy my desire. All in all, even if these considerations do not prove that resentment is never connected with a desire for the victim to punish himself, they show, against what Smith seems to claim, that there are cases where this does not happen. Therefore, it seems plausible to argue that what is necessary to identify resentment is the presence of the offender's desire for suffering, regardless of the person or event inflicting it.

Moving on to the second question, concerning the type of link between resentment and a desire for revenge, there seem to be at least three possible explanations for that connection. The first is that it is a hardwired bond: human beings are made in such a way that when they experience that pain in those particular circumstances, they also have that desire. It is simply a fundamental psychological fact, like Hume's thesis that pride has the self as its object, that is, a fact that we cannot explain further and that we must merely observe, that the passion of resentment is connected to a desire for revenge. The second explanation rejects this idea and holds instead that the connection should be explained by taking into account the fact that the suffering of the aggressor in some way compensates for the victim's pain. The question then is: why should inflicting pain have such a healthy effect on the victim? After all, it seems that the vengeance does not cancel out the suffering the victim is experiencing. According to its proponents, however, this observation does not get to the point because it does not take into account what has so far emerged as a crucial element of the feeling of resentment, that is, the peculiar sense of diminishment (real or feared) of his own value that the victim feels as a result of the injury. If this element is taken into account, then a new perspective opens up: retaliation compensates for the victim's suffering (and with it the injury) because it puts the injurer below him/her, restoring his/her confidence that his/her social standing has not been affected. Nussbaum, who rejected the hardwired explanation, has been a proponent of the

latter explanation of the link between resentment and desire for revenge. Discussing the imaginary case of Angela who was the victim of rape, Nussbaum says:

> All of a sudden, the retaliatory tendency makes sense and is no longer merely magical. To someone who thinks this way, in terms of diminution and status-ranking, it is not only plausible to think that retaliation atones for or annuls the damage, it is actually true. If Angela retaliates successfully (whether through law or in some other way, but always focusing on status-injury), the retaliation really does effect a reversal that annuls the injury, seen as an injury of down-ranking. Angela is victorious, and the previously powerful offender is suffering in prison. Insofar as the salient feature of O's act is its low ranking of Angela, the turnabout effected by the retaliation really does put him down and her (relatively) up.[113]

A third explanation, finally, is that the desire for retaliation is an expression of a moral principle that people cannot be treated in certain ways and that those who dare to do so must be brought to recognise through suffering similar to that suffered by the wrongdoer that they have committed a wrongful act. The idea is clearly formulated by Adam Smith, who points out that the desire for retaliation is satisfied if the wrongdoer is aware that he suffers from the particular harm he has inflicted on the victim.[114] It is not clear, however, how this third explanation is neatly distinct from the second. Indeed, the fact that Smith specifies that it must be the victim himself who inflicts the punishment seems to be related to the fact that the wronged person must be able to take pleasure in the sight of the aggressor's pain. However, some have insisted that the Smithian line nevertheless expresses a moral point of view that should not be confused with the mere

[113] See Nussbaum 2016: 26. In the course of his book, however, Nussbaum shows that resentment is a primitive and useless reaction to cases like these.
[114] On this aspect see TMS II.I.I.6. I examine this issue in more detail in Chapter 4.

pleasure of revenge. According to Griswold, for example, resentment is a kind of 'moral protest' against those who think they can 'treat others as if they count for nothing'.[115] Taking revenge is therefore an effective way of forcing the aggressor to acknowledge what he has done and to take responsibility for it.

Each of the three different explanations of the connection between resentment and revenge seems to have elements of plausibility. Without going into the matter at this stage, it is noteworthy that the three positions are not mutually incompatible. Indeed, it is possible to argue that there is probably a very basic level of the passion of resentment, coinciding with the innate instinct of self-defence, where the connection is hardwired. As a result of social pressures, however, resentment can change its structure, broadening the scope of its causes as well as the types of defensive response appropriate to them. At these stages, the connection between resentment and the desire to punish can be explained by further elements, such as the pursuit of the pleasure of revenge or, in more complex cases, the desire for the offender to take responsibility before the community for what he has done.

Now to the last question. Is there a proportionality between the offence and the punitive response? Henry Home was notoriously sceptical on this point arguing that resentment is a passion subject to various kinds of 'irregularity'.[116] Home dwells on three aspects in particular. First, resentment is often caused by harmful actions that are not intentional. He observes:

> Injury, or voluntary wrong, is commonly the cause of resentment; we are taught, however, by experience, that sudden pain is sufficient sometimes to raise this passion, even where injury is not intended.[117]

[115] See Griswold 2007: 28.
[116] See Home 1792: 8.
[117] Ibid.

Secondly, as a further irregularity, resentment can also channel our thirst for vengeance towards those who have no responsibility for the voluntary injuries, but are related in some way to the guilty party. On this point, Home claims:

> But all the irregularities of this passion are not yet exhausted. It is still more savage and irrational, when, without distinguishing the innocent from the guilty, it is exerted against the relations of the criminal, and even against the brute creatures that belong to him.[118]

To which is added the fact that natural human partiality can only increase the probability that these two circumstances frequently occur, transforming resentment into a passion that endangers social peace.

> A partiality rooted in the nature of man, makes private revenge a most dangerous privilege. The man who is injured, having a strong sense of the wrong done him, never dreams of putting bounds to his resentment. The offender, on the other hand, under-rating the injury, judges a slight atonement sufficient. Further, the man who suffers is apt to judge rashly, and to blame persons without cause.[119]

Home believes that, although resentment is the only passion able to perform a useful deterrent function in the early stages of human social life, it is marked by a 'destructive tendency to excess' that leads it to expand 'beyond its proper objects' and strengthen the intensity of the 'malevolent', 'anti-social passions'[120] (such as hatred and envy, for example). That is why society cannot be created until individuals yield their

[118] Home 1792: 10.
[119] Home: 23.
[120] Home: 14.

natural right to avenge their own wrongs to impartial judges that, unlike any arbiters who may be nominated by the parties in conflict, cannot be discharged by the parties.

Home grasps an important point on which the analyses carried out so far cannot help but agree. Resentment is not an arational passion since the agent who feels it has always a certain ground that justifies it. But this does not mean that the reaction is effectively proportional to the offence. On the one hand, it is still a perceived offence that may not be real. On the other hand, there is no constraint on the measures that the offended party believes he should take to free himself from pain and restore his social position by lowering that of the offender.

From these preliminary considerations, we can now propose a first rough definition of the object of our investigation. We will define resentment as a *prima facie* justified feeling of anger, indignation or grievance towards the author of a harm or injury against us that provokes a desire for revenge against him. This general conception can be used as a useful working notion to understand what resentment is in Hume.

3.2 The 'Dark Passion' of Resentment in Hume

In Book 2 of the *Treatise*, Hume mentions the passion of resentment in two different places. In the first, resentment is identified with the mental cause of the desire to punish the person who has caused us an *injury*. This appears to have a great deal of similarity to the contemporary characterization of resentment put forward by Nagel[121], Nussbaum[122] and Strawson[123]. I suggest that this similarity supports the plausibility of Hume's view. Hume claims:

[121] See Nagel 1970: 83.
[122] See Nussbaum 2016: chap. 2; see also *Appendix* C.
[123] See Strawson 2008: 23.

> Beside these calm passions, which often determine the will, there are certain *violent* emotions of the same kind, which have likewise a great influence on that faculty. When I *receive any injury* from another, I often feel a violent passion of resentment, which *makes me desire* his evil and punishment, independent of all considerations of pleasure and advantage to myself. (*T* 2.3.3.9/SBN 417-418, my italics)

Secondly, Hume includes resentment in the list of those passions that can be calm:

> Now 'tis certain, there are certain *calm* desires and tendencies, which, tho' they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation. These desires are of two kinds, either certain *instincts originally implanted* in our natures, such as *benevolence* and resentment, the love of life, and kindness to children; or the general appetite to good, and aversion to evil, consider'd merely as such. (*T* 2.3.3.8/SBN 417, my italics)

In addition to these quotations, there is also a third one that is relevant here. Hume mentions the desire to punish his enemies as part of the impressions of reflection he calls instincts. Although he does not use the term 'resentment' here, the context suggests that it is to this passion that he is referring. The desire to punish our enemies is clearly a synonym of the desire to punish the person who has injured us, which is a constitutive part of resentment.

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

Together these passages describe resentment as: 1) a violent passion against an injurer causing a desire for vengeance; 2) a passion that can be in some cases calm and in others violent; 3) an original instinct, that is, a passion which is not caused by pleasure and pains but which produces them. I will leave to one side the distinction between calm and violent because it doesn't tell us anything about this passion itself and I am going to focus instead on resentment's other aspects.

In the first passage, resentment is tied to the notion of injury. Let us try to understand this notion and find if it helps to understand that of resentment. Hume claims:

> Sometimes scurrility is less displeasing than delicate satire, because it revenges us in a manner for the injury at the very time it is committed, by affording us a just reason to *blame* and *contemn* the person, who injures us. But this phænomenon likewise depends upon the same principle. For why do we blame all gross and injurious language, unless it be, because we esteem it contrary to *good breeding* and *humanity*? And why is it contrary, unless it be more shocking than any delicate satire? The rules of good-breeding condemn whatever is openly disobliging, and gives a sensible pain and confusion to those, with whom we converse. (*T* 1.3.13.15/SBN 151-152, my italics)

The first use of injury in the *Treatise* refers to speech acts which are perceived as violations of rules of good breeding and humanity. Such violations are painful and it is in this that injury is constituted. Note that Hume makes no reference to the fact that the harmful action, in order to be qualified as injurious, must be intentionally directed against someone in particular or indeed an intentional violation of some social rule. Hume therefore seems to depart from Strawson's condition on the intentional nature of injury.[124]

[124] See Strawson 2008: 23.

On the contrary, it seems that in order to explain the injured person's reaction it is enough that the action causes his pain simply by being perceived as a violation of some rules, whether intentionally or not.

That would be too hasty a conclusion to make. *Treatise* Book 2 adds an important qualification to this first description. Hume says the 'the principal part' of an injury is pain felt by the harmed person in response to the contempt expressed by the injurer.

> But we must farther consider, that an intention, besides its strengthening the relation of ideas, is often necessary to produce a relation of impressions, and give rise to pleasure and uneasiness. For 'tis observable, that the *principal part of an injury is the contempt and hatred, which it shews in the person, that injures us; and without that, the mere harm gives us a less sensible uneasiness*. (*T*. 2.2.3.5/SBN 349-350, my italics)

Hume connects the perceived violation of a rule with the expression of contempt through such violation. Contempt is an emotion of disapprobation (*T* 1.3.14.24/SBN 166-7) that is partially constituted by a negative evaluation of the status of its target, i.e. a human being.[125] Contempt presents its object failing to meet a standard that the contemnor endorses and so it is an expression of conceiving the object as inferior. Further, contempt has an important comparative element: the person being despised is perceived as inferior by the person experiencing the passion.[126] On this aspect Hume writes:

> In considering the qualities and circumstances of others, we may either regard them as they really are in themselves; or may make a comparison betwixt them and our own

[125] Hume believes that contempt is a very similar, if not almost identical, passion to hatred, from which it differs only by some differences in the causes that arouse it (*T* 2.2.2.10/SBN 337). The difference probably lies in the fact that contempt is typically produced by our poverty and the physical and mental qualities that have to do with our shabbiness, lack of status and taste (*T* 2.2.5.1/SBN 357; *T* 2.2.5.14/SBN 362).
[126] I take this description of contempt from Bell 2013: 8-11.

> qualities and circumstances; or may join these two methods of consideration. The good qualities of others, from the first point of view, produce love; from the second, humility; and from the third, respect; which is a mixture of these two passions. Their bad qualities, after the same manner, cause either hatred, or pride, or contempt, according to the light in which we survey them. (*T*. 2.2.10.2/SBN 389-90; see also *T* 2.2.10.3/SBN 390)

Moreover, Hume associates contempt with dislike (*EPM* 6.33n.34/SBN 248), suggesting the intention to maintain the distance from the object of the contempt. Given all these properties, contempt is clearly intentional.[127] This means that Hume takes the relevant injury as a typically intentional behaviour: a perceived violation of something normative (for example, contrary to good breading and humanity) made with the intention of expressing contempt towards a person that the despiser regards as inadequate in relation to a standard which he endorses and inferior to him.

Although intention is conceptually central for injury, Hume still allows for an unintentional causation of resentment through a violation of a rule. This emerges in the complex section *T* 2.2.3, where Hume examines whether actions causing anger or hatred should be intentional, i.e. done with the aim of displeasing us (*T* 2.2.3.4/SBN 348-9). Hume clearly rejects this thesis, arguing that the absence of intention can weaken the passion or make it transient, but does not eliminate it at all. Let us consider Hume's argument in detail.

The first part of Hume's answer touches on the broad issue of how to assign an action to an agent. In order to be able to blame someone for the action he has carried out, we must be able to consider him as its immediate cause. In the specific context of the indirect passion of hatred, the problem concerns the difficulty for the imagination to pass

[127] I use 'intentional' not in the sense that passion has an object but in the sense that it involves an intention to distance oneself.

from the idea of an action, by its nature perishable, to that of the character of the agent, by its nature stable. That is where the intention comes in. The presence of an intention to act in the mind of the offender facilitates the imaginative passage from the idea of attacking action to that of the offender in the mind of the offended person. The intention therefore carries out the important task of strengthening the relation of ideas underlying the hatred of the offences of others. Having said that, Hume points out, that a strong relation of ideas is not a necessary condition to feel hatred for the person who damages us. Indeed, this passion arises even in the presence of a weak relation of ideas as long as the painful impression is sufficiently intense. Hume admits that hatred is a fleeting emotion here and does not give rise to any hostile relation with the author of the involuntary aggression.

The second part of the answer is where the notion of injury came in. After claiming, as pointed out a few lines above, that the major part of the pain caused by an injury is not the harm itself, but the mortification produced by the contempt with which the injurer acted, Hume adds a crucial clarification. He points out that it is a fact of our shared experience of human beings feeling anger at injury that they know to be involuntary. Hume initially explains this fact by simply saying that there is a natural connection between uneasiness and anger.

> But then I ask, if the removal of design be able entirely to remove the passions of […] hatred? Experience, I am sure, informs us of the contrary, nor is there any thing more certain, than that men often fall into a violent *anger* for *injuries*, which they themselves must own to be entirely involuntary and accidental. This emotion, indeed, cannot be of long continuance; but still is sufficient to shew, that there is a natural connexion betwixt uneasiness and *anger*, and that the relation of impressions will operate upon a very small relation of ideas. But when the violence of the impression is once a little abated, the defect of the relation begins to be better felt; and as the character of a person is no wise interested

> in such injuries as are casual and involuntary, it seldom happens that on their account, we entertain a lasting enmity. (*T* 2.2.3.6/SBN 350 my italics).

This explanation is not fully satisfactory as it fails to explain what the notion of involuntary injury connected with anger consists of. However, this point is clarified by Hume a few lines later. He points out that, in the given circumstances, it is the passion for anger that produces the opinion of injury and not the other way around. More precisely, after uneasiness has produced anger, we naturally try to stabilize this passion and we do so by trying to justify it through the belief that the offender intentionally wanted to cause us pain.

> […] 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. Here the idea of injury produces not the passion, but arises from it. (*T* 2.2.3.9/SBN 351)

At this point one could object: why should all this be relevant to resentment? Anger and resentment are different after all, and Hume's clarification concerns the connection between injury and anger and not that between injury and resentment. From what has been said so far, it might be the case that while unintentional injury is sufficient to generate anger, resentment arises only after the victim has perceived the injurer's desire to humiliate him.

This objection fails to grasp the general point Hume is trying to make. The idea is that when harm is associated with a certain malevolent passion - as in the case of injury with contempt - this harm will continue to be associated with that passion even in circumstances where it is not actually present. This thesis applies both to anger, which

Hume explicitly discusses, and to resentment, which he does not discuss. Specifically, as far as resentment is concerned, this argument clarifies an important aspect regarding the causes of this passion. Although the intention to offend - i.e. express contempt - is conceptually central to the injury that causes resentment, Hume still allows for an unintentional injury which causes resentment through violation of a rule. Once a violation of a certain rule is identified as an action that expresses contempt for someone, any action that is perceived as a violation of that rule will be considered in the same way and will arouse resentment, regardless of whether the injurer is aware that he is violating a rule and/or has wanted to show contempt by acting in this way. In this way, Hume's account is perfectly capable of explaining a central feature of resentment that has emerged in the contemporary literature. By reference to contempt, Hume succeeds in explaining the typically intentional character of the causes of resentment. On the other hand, by the notion of violation of a rule, Hume is able to account even for those cases in which, although there is no malicious intention on the part of the aggressor, resentment is a perfectly intelligible emotional reaction on the part of the victim of an injury.

The second and final characteristic of Hume's resentment is its being an instinct. Instincts are described as those impressions or those characteristics of our impressions that (1) cannot be explained by further principles or that (2) are describable as facts that constitute our nature and that it is impossible for us to modify unless we change our own constitution. An example of an instinctive opinion would be the one that sensation impressions are identical to external objects, which we embrace only "on account of their suitableness and conformity to the mind", and that they always prevail over any philosophical principle that tries to question them (*T* 1.4.2.51/SBN 214). Similarly, 'instinct' describes the connection that pride has with its object - the "self" – which is "absolutely impossible" to change and of which no explanation can be given (*T* 2.1.5.3/SBN 285-86). In the context of direct passions, the term 'instinct' has a third

meaning. This describes a *sui generis* class of direct passions which, unlike direct passions, do not arise from pleasure or pain but instead produce these sensations.

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

As I said in the opening, this third meaning is particularly relevant to my analysis since this class of passions includes the "desire to punish our enemies", i.e. the desire that is connected to and qualifies the passion of resentment. The connection between resentment and instinct is stressed in a passage in the second *Enquiry*, where Hume explicitly describes resentment as an instinct.

> The dilemma seems obvious: As justice evidently tends to promote public utility and to support civil society, the sentiment of justice is either derived from our reflecting on that tendency, or like hunger, thirst, and other appetites, *resentment*, love of life, attachment to offspring, and other passions, arises from a simple original *instinct* in the human breast, which nature has implanted for like salutary purposes. (*EPM* 3.40/SBN 201, my italics)

The inclusion of resentment in the class of instincts adds a further significant qualification to the analysis conducted so far that gives us a key to explaining the link between resentment and the desire to punish the offender. We have seen resentment as the pain caused by the perceived violation of a norm that is typically linked to the manifestation of contempt, but we have not yet examined the second part of Hume's characterisation of this passion, i.e. that 'makes me desire' the offender's 'evil and

punishment'. The two passages just considered give us two important indications. First, characterizing not only the '*desire of punishment to our enemies*' as an instinct, but also the resentment itself, indicates that this desire for revenge is not simply an impression associated with resentment, but something that is robustly connected to this passion and somehow identifiable with it. Second, arguing that the rise of this desire does not proceed from pleasure and pain indicates the direction to be taken to explain the link between the painful sensation caused by injury and the desire at issue. Both indications are relevant to the plausibility of Hume's position against the background of the general analysis conducted in the first section. Let us examine them individually.

Hume agrees with the contemporary analysis that the desire for the aggressor's suffering is a constituent element of this emotion, without which we cannot experience resentment. Moreover, as an implicit consequence of this point, he agrees that the aggressor is the object of this passion: if resentment is identified by the desire for the aggressor to suffer (or at least strongly connected with it), resentment is an emotion directed at the aggressor.[128]

Let us now examine the second problem identified earlier, namely how resentment, as a painful feeling, is connected with the desire for punishment of the injurer. As we shall see, Hume gives a more cogent account than that provided by contemporary theories, especially Nussbaum. Let us return to the characterization of desire for revenge as an instinct. That means that this psychological state is motivated neither by the search for pleasure nor by the avoidance of pain. This observation leaves one puzzled, to say the least, as in this way Hume seems not to take into account the fact that the desire arises

[128] That passions have an object is not an unusual fact in Hume's descriptive psychology. He famously characterizes indirect passions as having causes and objects and describes some desires as having specific objects, i.e. 'desire of fame' (*T* 2.1.11.11/SBN 321) or 'of society' (*T* 2.2.5.15/SBN 362-3). What however has puzzled more than one commentator is that Hume does not seem to have the resources to account for the intentionality of emotions, on the contrary he seems to exclude it. A fact that would seem to undermine his ability to account for resentment in a plausible way. For an examination of this aspect of Hume's passion theory, see Radcliffe 2018: chap. 5. See also Bricke 1996: 5; 26-27.

from the painful sensation caused by the injury. Annette Baier, who offered a more charitable interpretation, argued instead that Hume's claim is unproblematic and accounts for an important fact: the expression of resentment in revenge is "something that is wanted in itself and not because of its hedonic promise" since the pleasure we derive from the satisfaction of revenge is not necessarily something we remember from previous successful revenges.[129] Unlike Baier, I think Hume's clarification expresses a stronger thesis than Baier's: Hume does not merely deny that revenge is motivated by the prospect of pleasure, but adds that it is not motivated by the avoidance of pain either. What does all this mean?

To clarify this point we should examine not only Book II of the Treatise, but also Appendix II of *An Enquiry concerning the Principles of Morals*, illustrating a further aspect of the notion of instinct. This identifies those psychological states that direct the human mind towards certain objects, where these are not instrumental to the search for pleasure or the avoidance of pain, but are the immediate ends of those inclinations or passions. To this first characterization Hume makes an important additional point about the subject of these states. Once satisfied, these instincts produce pleasure, making possible the formation of further psychological states that seek those objects for the pleasure attached to them. Hume first illustrates this aspect of his theory of instincts by bodily appetites:

> There are bodily wants or appetites, acknowledged be every one, which necessarily precede all sensual enjoyment, and carry us directly to seek possession of the object. Thus, hunger and thirst have eating and drinking for their end; and from the gratification of these primary appetites arises a pleasure, which may become the object of another species of desire or inclination that is secondary and interested. (*EPM* App.2.12/SBN 301)

[129] See Baier 1980: 137.

Hume argues that this same explanation can also be extended to some passions, among which there is precisely the particular anger of resentment, which has revenge as its object.

> In the same manner, there are mental passions, by which we are impelled immediately to seek particular objects, such as fame, or power, or *vengeance*, without any regard to interest; and when these objects are attained, a pleasing enjoyment ensues, as the consequence of our indulged affections. Nature must, by internal frame and constitution of the mind, give an original propensity to fame, where we can reap any pleasure from that acquisition, or pursue it from motives of self-love, and a desire for happiness. If I have no vanity, I take no delight in praise: If I be void of ambition, power gives me no enjoyment: *If I be not angry, the punishment of an adversary is totally indifferent to me*. In all these cases, there is a passion, which points immediately to the object, and constitutes it our good or happiness; as there are other secondary passions, which afterwards arise, and pursue it as a part of our happiness, when once it is constituted such by our original affections. (*EPM* App.2.12/SBN 301, my italics)

The origin of the desire to punish is not the expectation of pleasure that comes from a successful revenge, but is the peculiar painful passion of the resentment itself. This passion determines the human mind towards an end that is revenge, no less, an end that is sought regardless of considerations concerning pleasure and pain. We do not take revenge because we desire a particular pleasure nor because we want to avoid some kind of pain, but because we are made in such a way that when we feel that particular angry pain of resentment we cannot help but want to take revenge, even at the cost of adding further suffering to the suffering received.

That said, Hume adds that revenge can also be sought for the expected pleasure. This does not contradict his naturalist explanation, but rather shows that this further desire is only possible because we are made in such a way that when we receive an injury and feel resentment we are inevitably driven to revenge.

This concerns pleasure, but what about the avoidance of pain? Does not this desire arise from the mortification of insult? Desire is certainly connected to that pain, but that does not mean that its cause is the prospect of avoiding pain. I believe, however, that similarly to what has been said about the anticipation of pleasure, the prospect of pain avoidance can also become a secondary cause of revenge. Although we are initially driven to revenge instinctively, through experience we understand how the fear of our revenge is a deterrent to others who will be discouraged from insulting us. Having acquired this awareness, we may then be driven to revenge not only by the prospect of anticipated pleasure from successful revenge, but also by the prospect of avoiding the pain we would receive if we did not let others know that we are not to be messed with.

As such, Hume's position on the connection between resentment and vengeance does not conflict with Nussbaum's thesis. But it is more cogent than hers. It shows that anticipated pleasure is not the only reason for revenge because we can do so even in circumstances in which we do not consider pleasure at all or know that we would incur terrible suffering.

In conclusion, Hume is also able to account for our ability to sympathize with the resentment of others that while being a central aspect of this passion is completely absent from Nussbaum's account. In Book 2, when Hume claims that resentment is felt by sympathy:

> A cheerful countenance infuses a sensible complacency and serenity into my mind; as an angry or sorrowful one throws a sudden damp upon me. Hatred, *resentment*, esteem, love,

> courage, mirth and melancholy; all these passions I feel more from communication than from my own natural temper and disposition. So remarkable a phenomenon merits our attention, and must be trac'd up to its first principles (*T* 2.1.11.2/SBN 316-317, my italics).

The examination conducted so far, therefore, shows that Hume presents a coherent and plausible account of resentment. After examining the different elements of this passion, we have shown the adequacy of his theory by comparing it with contemporary discussion gauging its plausibility in the light of that.

3.3 Resentment and Justice in the *Treatise* and the *Enquiry Concerning the Principles of Morals*

Having addressed the two questions posed at the beginning of the chapter, namely what Hume's conception of resentment is and whether or not it is a plausible conception, we now examine the third and final question of whether resentment plays a role in his theory of justice. In Book III of the *Treatise*, as we have seen in the previous two chapters, resentment does not appear among the social passions nor does it appear to be among the psychological factors that explain convention. In *Enquiry Concerning the Principles of Morals,* however, the picture seems to have changed. Indeed, in *EMP* 3, resentment seems to have a role to play as a condition for entering the convention that generates justice and becoming part of a society governed by rights.

This function has been recognised by both philosophers and interpreters of Hume, who have offered different interpretations.[130] Unlike those authors, I will argue that the

[130] See, for example, Baier 1980: 133 ff.; Nussbaum 2006: 45-9; Pritchard 2008: 59 ff.; Zagorac 2015: 189 ff.; Pollock 2016: 107 ff.

importance of this passage lies in that it specifies certain characteristics of resentment, not discussed in Book II of the *Treatise*, that illuminate two different functions that this passion plays in the genealogy of justice. First, resentment originates the process of moralising justice, in the course of which it is no longer evaluated solely through the advantage of participating in its practices and the disadvantage of being excluded from them, but on the basis that justice is morally appreciable and injustice blameworthy in itself, i.e. not instrumentally. Second, resentment plays a role in reinforcing the feeling of moral disapproval of injustice, which would risk being an ineffective sanction if it were based solely on sympathy with the general welfare.

Before that, let us briefly examine the meaning of the passage in EMP 3.18 and the characteristics of resentment it highlights.

3.3.1 Resentment as a 'Third Condition' of Justice

According to a widely accepted reconstruction, Hume supplemented his conception of the two conditions of justice set out in T 3.2.2.5-7 with a third one stated in *Enquiry Concerning the Principles of Morals*. Using Rawls's terminology, which attributes this discovery to Hume, the conditions of justice are those factors without which justice could not have been invented by human beings.[131] In the *Treatise*, as argued in Chapter 2, Hume points to two, one internal and one external. The internal one states that justice is only conceivable if human beings are endowed with limited benevolence since it would not have been invented if they were either completely selfish or had universal benevolence. Similarly, the external condition holds that justice is only conceivable in a condition of mediation between two extremes. The external condition holds that justice is intelligible

[131] In his exposition of the conditions of justice, Rawls claims that his "account largely follows" the one Hume laid out in the *Treatise* and in the second *Enquiry*. See Rawls 1971: 126, n.3.

only in a context of limited scarcity of resources. For if resources were unlimited, it would be superfluous; while if they were excessively scarce, not being sufficient for the basic needs of all, no cooperation would be possible. According to Rawls and Nussbaum, in his *Enquiry Concerning the Principles of Morals* Hume adds a third condition, specifically relating to the equality in strength among those who can join the convention of justice. This form of equality, as Annette Baier has nicely pointed out, concerns the ability to make 'one's resentment felt' against those who have injured us. Those lacking this capacity could only have with 'equals' strongly asymmetrical relations, made up of absolute command on the one hand and obedience on the other where the only restraint on the power of the former would only be their compassion and never the rights of justice.[132] Hume claims:

> Were there a species of creatures, intermingled with men, which, *though rational, were possessed of such inferior strength, both of body and mind*, that they were incapable of all resistance, and could never, upon the highest provocation, make us feel the effects of their resentment; the necessary consequence, I think, is, that we should be bound, by the laws of humanity, to give gentle usage to these creatures, but should not, properly speaking, lie under any restraint of justice with regard to them, nor could they possess any right or property, exclusive of such arbitrary lords. *Our intercourse with them could not be called society*, which supposes a degree of equality; *but absolute command on the one side, and servile obedience on the other*. Whatever we covet, they must instantly resign: *Our permission is the only tenure, by which they hold their possessions*: *Our compassion and kindness the only check, by which they curb our lawless will*: And as no inconvenience ever results from the exercise of a power, so firmly established in nature,

[132] Baier 1980: 135.

> the restraints of justice and property, being totally *useless*, would never have place in so unequal a confederacy. (*EPM* 3.18/SBN 190-1, my italics)

The passage raises several questions. Who does Hume have in mind when he refers to those who, although rational, have inferior strength in body and mind? Is he alluding to the inequality, we would say today, of gender between men and women, or rather the inequality of species between men and animals? If he is thinking of women, does this mean that he thought the conventions on which he based the virtues of chastity and modesty were unfair and therefore outside society? And again: how can those who cannot make the effects of their resentment felt do so? By becoming aware that their masters' power depends on them, or by simply joining forces?

These issues have given rise to conflicting assessments of the inclusive capacity of the Humean theory of justice, making the passage on resentment a kind of trial by fire to assess the overall viability of his theory. This debate can be schematically illustrated by following the contrasting position of A. Baier and Marta Nussbaum. The first line, set out by Baier, has famously argued that Hume's resentment can be considered a kind of watchdog of our moral subjectivity: we make others feel the effects of our resentment whenever they exclude us from the practices of moral recognition. Resentment, according to Baier, has thus an important function in Hume's moral philosophy since it is the most important weapon one can use against those who do not recognise our rights to justice. Baier argued that Hume's third condition does not prove that he had a narrow conception of those who are protected by justice because it should not be underestimated that the isolated inferiors, when indispensable to the recognition of the moral status of the superior, can join together and together make felt the effects of their resentment against their rulers.[133] The second line, however, clearly put forward in Nussbaum's *New Frontiers of*

[133] See Baier 1980: p. 147.

Justice, argues instead a different view according to which the constraint on equality in the capacity for revenge shows unequivocally that Hume's theory of justice has a poor capacity for inclusiveness. The rules of justice are in fact not only invented by but also designed to benefit those who are equal, in strength and interest, thus leaving the unequal at the mercy of the equal. Nussbaum thus interprets Hume's constraint in a similar way to that put forward by Rawls in his theory of justice, in which the subjects of justice are only free, equal and independent individuals. Both models must be rejected, according to Nussbaum, because they are not equipped to include in the community governed by rights those, such as animals or the handicapped, who are not equal to others in rationality and agency.[134]

As interesting as these issues are, they are beyond the scope of the problems addressed in this chapter, the objective of which is more modest than that indicated in these lines of research. My aim is not to examine the role that resentment might play in the claims of those who are not accorded full moral status, nor even to explore the question of how far a sentimentalistic Humean model of justice is vulnerable to Nussbaum's criticism of Rawls. My aim is to ascertain what role resentment plays in Hume's notion of justice understood as a set of rules that protect property and the observance of covenants, in particular, whether it plays any role in the process of moralising the rules of justice and strengthening their observance by all. In the light of this aim, although important, it is not as central as in these studies. Its relevance to me depends on the fact that it clarifies two further aspects of the passion of resentment not hitherto examined. First, the injury that causes resentment consists not only in the violation of a rule of good breeding that generates generic social frustration, but also in the violation of a right secured by justice, identifiable with an attack on property or with

[134] See Nussbaum 2006: chap. 1.

the failure to keep one's word. Second, this type of resentment can only be relevant as a deterrent to injustice if the offended party has the power to make the offender feel the effects of his resentment. In itself, therefore, resentment may not only be of no use to those participating in the conventions of justice, but it may also be a merely toxic passion that reinforces the pain of the injury with the peculiar pain generated by seeing our desire for reprisal unfulfilled.

As I said, I will not deal with the subject of the social and individual conditions in which a so-called 'passive' resentment (i.e., unable to express itself in revenge) can become 'active', but I do want to focus on the role that active resentment plays in moralising and strengthening respect for the rules of justice. Before examining this issue, however, there is a preliminary problem to be solved. In the previous section, I have argued that the passion of resentment *qua* feeling is a peculiar pain, whose phenomenological quality is explained by Hume with the circumstance that its cause is perceived to express contempt towards the injured person. If resentment can be produced by injustice, then injustice - i.e. the violation of property or covenants - must somehow be perceived by the injured person as a wrong that expresses contempt towards him. How can we explain this connection?

To address this point, it is appropriate to examine Hume's idea of property in the context of causes of pride.

3.3.2 Property as a Source of Pride

Unlike Locke, Hume explains property not through natural relations[135], but artificial ones that are dependent on psychological attitudes emerging from the convention of justice.

[135] See Locke 1997 (1662): chap 5.

Hume's artificialist conception of property is, however, not immediately clear to the reader of the *Treatise* as he first examines this notion well before his treatment of justice as an artificial virtue (the logical presupposition of property). Prior to the section *Of the rules, which determine property* (T 3.2.3), Hume in fact examines property in the section *Of property and riches*, contained in Book 2, Part 1, of dedicated to the indirect passions of pride and humility.

Hume was certainly aware of this unorthodox way of proceeding. Indeed, he describes property as a '*relation betwixt a person and an object as permits him, but forbids any other, the free use and possession of it, without violating the laws of justice and moral equity*' (*T* 2.1.10.1/SBN 310). In this way, he presented a formula that was deliberately neutral on the question of whether justice was dependent on 'natural conscience' or on 'honour, and custom, and civil law' (*T* 2.1.10.1/SBN 309-10).

The reason for this order of presentation depends on the crucial importance of property as a cause of pride. I intend to illustrate this by showing how injustice, whether in the form of the taking away of our property or as the non-return of what is owed to us as a result of a promise made, constitutes a subtraction of a crucial source of pride.[136] Indeed, as we shall see later, Hume correctly describes this kind of injustice as a form of 'injury' (*T* 3.2.2.8/SBN 488-9), thereby emphasising an important fact. Not only is injustice a violation of a rule that is highly useful because it is necessary for the maintenance of peace in society. It is also that particular kind of harm that arouses the resentment of the victim with whom, as observers, we are able to sympathise impartially. I shall argue that it is precisely this fact, namely injustice being an injury, that is the origin of the process that leads us to consider the violation of justice as the violation of a moral obligation.

[136] In some extreme cases, injustice can also be the cause of a new social condition, which arouses humility that is heightened by comparing the present and past condition of prosperity. However, I will not deal with this further specific case in the remainder of this thesis.

Let us therefore examine property as one of the causes of pride. A few brief remarks on pride in general are in order first. Together with love, hate and humility, it forms the core of the indirect passions, to whose examination Parts 1 and 2 of Book 2 of the *Treatise* are devoted. They are impressions of reflections that arise in the human mind from other perceptions, internal or of sensations, that precede them. Indirect passions, in particular, are preceded by pleasurable or painful impressions but, unlike the reflective impressions that Hume calls direct passions, which arise immediately from pleasure and pain, indirect passions arise through a more complex causal process involving several associative relations. Although Hume considers the indirect passions as not susceptible of exact definition, viewing them as consisting of peculiar sensations without parts, he believes, in accordance with the general explanatory approach of Book 2 of the *Treatise*, that this does not prevent him from putting forward a coherent explanation of the origin and effects of these passions.

Hume examines pride through both psychological introspection and the philosophical notions of 'cause', 'object', and associative relations.[137] Introspection reveals that pride is constituted by a peculiar pleasurable feeling that we experience when considering our self in an advantageous light (*T* 2.1.2.2/SBN 277). The object of pride is thus the self, i.e. 'succession of related ideas and impressions, of which we have an intimate memory and consciousness' (*T* 2.1.2.2; SBN 277). Whenever we experience this passion, the mind automatically fixes our attention on no other object but the self and feels a pleasant sensation of elevation in considering it positively.[138]

[137] On the use of complex philosophical ideas to describe human passions, Taylor 2015: Chap. 1.

[138] A much-discussed issue that I will not address here is whether the object is part of the passion, i.e. a constituent element of it, or whether it is instead something that the passion produces, i.e. an effect of it. In favour of the first hypothesis is the consideration that Hume claims that pride is a peculiar pleasurable sensation (*T* 2.1.2.1/SBN 277), where its peculiarity consists in a sense of being "elated" in considering the idea of the self in an advantageous light. In this way, considering the self in a certain way seems to be part of the content of that feeling. In support of the second hypothesis, however, is the fact that in more than one passage Hume argues that pride, understood as the pleasurable sensation that constitutes its essence, 'produces' a certain positive idea of the 'I' (*T* 2.1.2.4/SBN 278). Following this consideration, the positive

The direction of pride towards the self depends on an original principle, i.e. a brute fact that characterises human pride and cannot be explained further. This marks an important point of difference with the causes of pride, whose connection is instead explained through the principles governing the association of impressions and ideas in the human mind. Hume distinguishes between two 'properties' of the cause: a 'quality', which typically arouses a pleasurable sensation, and a 'subject' on which the property is inherent (*T* 2.1.5/SBN 285-6). A cause of pride is such only when it satisfies the following twofold condition: its ownership arouses a pleasurable sensation that bears a relation of resemblance to the pleasurable sensation of elevation that constitutes the essence of pride, and the idea of the object, on which that quality inheres, is related to the idea of the self that constitutes the object of pride.

This explanation of the origin of pride establishes a pivotal characteristic of its causes without which Hume could not have explained pride in property. If the link between causes and pride depended on an original principle, the causes of pride would be part of a closed set and it would not be possible for something that was not originally part of the set to subsequently become part of it at a certain stage of human development. The explanatory principle of a double relation of impressions and ideas, on the other hand, makes it possible to account not only for the fact that the causes of pride, as opposed to its object, are manifold, but also that they form part of an open set that is contingent on what, from the changing conditions of life and human practices, arouses pleasure and it is able to acquire a stable relationship with the self.[139]

Hume discusses property as a cause of pride in *Of property and riches* (*T* 2.1.10). This concludes an examination of other types of causes that concern in order *Of vice and virtue* (*T* 2.1.7), *Of beauty and deformity* (*T* 2.1.8) and *Of external advantages and*

idea of the self is not part of pride any more than an effect is part of a cause. For a very recent discussion of this complex topic see Radcliffe 2018: chap 7.

[139] See, for example, *T* 2.1.3.4-5/SBN 280-2, see also *T* 2.1.6.9/SBN 293-4.

disadvantages (*T* 2.1.9). The order of position should not be misleading. It indicates neither some evaluative perspective on the causes of pride [140] or any statistical order that records the frequency with which a type of cause produces pride. Rather, the order seems to be dictated by the fact that this explanation in some sense presupposes that of the causes, e.g. beauty or vices and virtues, examined in the previous sections.

Hume considers property to be one of the most common and most stable causes of pride. This emerges from his detailed description of the relationship of impressions and ideas that connects goods of property to the pride of its possessor. Let us start examining the relationship of ideas, which is illustrated through the following three considerations. First, it is described in language very similar to that used to describe the connection between the idea of the self and that of its vicious/virtuous character traits, suggesting that the property we own has as strong a relation to ourselves as we have to what is part of our mind (*T* 2.1.10.1/SBN 309). In this respect, the relation of ideas between property and its proud owner is considered to be very strong and stable as analogous to that of contiguity between the idea of an object and the idea of what is part of it.[141] Second, Hume adds shortly afterwards that it can also be compared to a 'particular species of causation' considering 'the liberty it gives the proprietor to operate as he pleases upon the object, or the advantages, which he reaps from it'. Hume therefore explains this connection through what in Book 1 is considered the strongest relation of ideas that can be entertained by the human mind. Third, concluding the paragraph, Hume argues that what is said so far proves that it is a 'perfect relation of ideas', meaning a

[140] J. Taylor, for example, observes that *Treatise* Book 2 examines passions from an entirely descriptive rather than normative perspective. See Taylor 2015: chap. 2 and 3.

[141] In T 2.1.5, Hume describes vice and virtue by arguing that 'Thus the good and bad qualities of our actions and manners constitute virtue and vice, and determine our personal character, than which nothing operates more strongly on these passions.' (*T* 2.1.5.2/SBN 285). A few sections later, describing property, he states that 'but the relation, which is esteem'd the closest, and which of all others most commonly produces the passion of pride, is that of property' (*T* 2.1.10.1/SBN 309). Beyond the possible conflict between these two passages, both of which seem to establish a respective primacy for their causes, a charitable reading of both indicates that character traits and property have the same strong propensity to cause these passions.

symmetrical relation whereby the idea of property 'naturally carries our thought to the proprietor, and of the proprietor to the property'.

A relation of ideas, however strong and stable, is never by itself sufficient to cause any single episode of pride, which also requires a relation of pleasant impressions. Hume argues that everything that is 'either useful, beautiful, or surprising' arouses pleasure and is therefore capable of entering into a relation with the pleasurable passion of pride (*T* 2.1.8.5/SBN 300-1). It seems then that if something possesses one of these properties and is also in our possession then it will automatically tend to arouse our pride. But is this really the case?

In the much-discussed section *Limitations of this system* (*T* 2.1.6), Hume introduces a requirement that seems to point to an obstacle to the thesis just formulated. To arouse pride its cause must be 'common to us with a few persons' (*T* 2.1.6.4/SBN 291). Following this constraint, a commonly used piece of furniture such as a bookcase will not arouse pride in its owner simply because it is an undoubtedly useful object. The bookcase may be a source of pride provided it has characteristics that make it excellent in its kind, such as its functionality, i.e. its ability to hold many books in relation to the space it occupies, or its environmental sustainability, which is expressed in the use of green materials within its production process. Possessing these characteristics, Hume seems to imply, enables the bookcase to acquire a closer relationship with its owner since very few people own an object with these characteristics. It goes without saying that such a constraint will inevitably tend to restrict the number of useful or beautiful objects we own that are capable of arousing pride, as only a small number or even none can have this capacity. Following this limitation, the initial thesis that possessions are the most common cause of pride seems to be clearly false.

However, this is too hasty a conclusion. Indeed, there are three further claims to bear in mind. Firstly, what makes the relation of ideas sufficiently close is not so much

the excellence of the object but rather, since we tend to 'judge of objects more from comparison than from their real and intrinsic merit' (*T* 2.1.6. 4/SBN 291), that the object is judged as better than those possessed by others. Secondly, Hume claims that any object that is beautiful or useful is, by the very fact of being our property, judged to be better than other similar objects that we do not possess. In other words, in order for evaluation by comparison to reward our objects, it is sufficient that the objects are beautiful and useful in their kind because they, by the very fact of being ours, will automatically be perceived by us as more beautiful or more useful than others. Hume claims:

> Every thing belonging to a vain man is the best that is any where to be found. His houses, equipage, furniture, cloaths, horses, hounds, excel all others in his conceit; and 'tis easy to observe, that from the least advantage in any of these, he draws a new subject of pride and vanity. His wine, if you'll believe him, has a finer flavour than any other; his cookery is more exquisite; his table more orderly; his servants more expert; the air, in which he lives, more healthful; the soil he cultivates more fertile; his fruits ripen earlier and to greater perfection (*T* 2.1.10.2/SBN 310).

It is clear from these first two claims that the constraint Hume specifies in the section on *Limitations of this system* does not represent a threat to the thesis that property goods are a widespread cause of pride. On the contrary, that constraint is compatible with the idea we argued earlier that property is a numerically important part of the sources of pride, since each of them will constitute an autonomous cause of pride.

There is finally a third consideration to bear in mind, which concerns the positive effects of sympathy on pride in property that contribute to making this type of cause particularly stable and phenomenologically intense. In the well-known section, *Of the love of fame*, Hume returns to the subject of the causes of indirect passions, claiming that,

apart from the *original causes* of pride and humility, there is a secondary cause in the 'opinions of others' (*T* 2.1.11.1/SBN 316). He adds that the 'vast weight and importance' of this cause depends on a characteristic that distinguishes it from the other causes. The opinions that others have of us, which form our 'reputation' and our 'name,' are not only an independent cause of pride and humility, but also the condition that makes the other causes effective. Without the support of the opinions of others, such causes would have 'little influence' on those passions. In the *Treatise*, and later in the *Dissertation on the Passions*, Hume explains this fact, claiming that our judgments 'of our own worth and character' (*T* 2.1.11.9/SBN 320–1; but see also *T* 2.1.8.9/SBN 303) are particularly shaky and in need of confirmation. On the other, as human beings are also partial to themselves, they always seek views that confirm 'the good opinion' they have of themselves (ibid.). To give stability to our ideas on what we regard as the causes of our pride, human beings therefore need to receive confirmation from others.

These remarks in his discussion on the causes of pride do not concern property in particular, but every type of cause in general. Yet, it is pride in property that Hume uses to illustrate how sympathy operates in this passion. In his famous statement that the minds of human beings are 'mirrors to one another' (*T* 2.2.5.21/SBN 365), Hume shows that a standard episode of pride in property is interwoven with the esteem of others in a dynamic process that entails at least two waves of mutual support. So, the ease and joy that characterize, for example, the rich man's life arouse esteem in those who come into contact with him. This is transmitted by sympathy to the 'possessor,' who receives a 'second satisfaction' of his wealth. In turn, this second pleasing sensation, 'once more reflected,' becomes a new form of esteem in the observer (*T* 2.2.5.21/SBN 365).

For Hume to choose this example to illustrate the role of sympathy in pride might seem a little strange. As we have seen, we naturally tend to believe that objects we own are better than others. Our opinions on this matter therefore do not seem to be as shaky

and in need of confirmation as are those concerning the value of our other qualities, where instead the role of sympathy seems more crucial in stabilising our evaluative opinions underlying our pride.

That said, I believe Hume's choice is justified in another respect, which reveals an important aspect of his treatment of pride examined in *Treatise* Book 2. It has been written that the description of the passions that Hume presents in Book 2 is strongly influenced by the fact that the common life that Hume observes is that of the nascent Scottish commercial society of the eighteenth century.[142] In this scenario, property is regarded not only as a sure viaticum for the improvement of human living conditions but also as an important sign of social distinction and prestige, which is not detached from the possession of skills, which Hume repeatedly describes as moral virtues, such as 'enterprise' (*EPM* 6.21/SBN 242-3), 'prudence' (ibid.), 'discretion' (*EPM* 6.8/SBN 236) and 'frugality' (*EPM* 6.11/SBN 237). Property goods are not only immediately visible to any observer, but are also immediately associated with happiness and the positive qualities of its possessor who experiences satisfaction by reflecting on them. Ownership, more easily than other causes of pride, is thus able to set in motion that dynamic sympathetic process that strengthens pride, transforming this passion into a disposition to positively self-evaluate.

Based on what has been said, one might conclude that once property has been established through the rules of justice, the infringement of individual property is likely to adversely affect the sources of our pride and our public image. Violation of property can thus rightly be considered an injury, i.e. an act that violates the rules of 'good breeding and humanity' (*T* 1.3.13.15/SBN 151-152), a violation that is deemed to express contempt towards the offended party and which arouses his or her resentment towards the aggressor.

[142] See Taylor 2015: chap 2.

As we shall see, not only does this thesis appear plausible in the light of the examination just conducted on pride, but it is confirmed by the fact that 'injury' is the term Hume uses in his discussion of injustice as the taking of another's property.

3.3.3 Resentment and Moral Sentiments

Resentment was deemed to play a limited role in Hume's theory of justice, which is exclusively concerned with the discussion of the conditions of justice examined in *EPM* 3.18. I intend to show that resentment plays a broader role than this: not only because this passion is a crucial factor in the explanation of the origin of the moral evaluation of justice, but also because, once moral sentiments towards justice have been consolidated, resentment offers a further motivation that strengthens their effectiveness.

Let us start with the moralisation of justice. Hume addresses this issue in the final six paragraphs of *T* 3.2.2, when, having concluded his explanation of the natural obligation to justice, he turns to the question of its 'moral obligation', i.e. the 'sentiments of right and wrong' (T 3.2.23/SBN 498) that attend just and unjust actions.

In a curiously overlooked passage, Hume explains the origin of this phenomenon through the sentiment of disapproval of the unjust action. Importantly, disapproval does not depend on the fact that we sympathise with the general pain produced by the class of actions to which the individual injustice belongs, but on the unbiased sympathy with the particular suffering experienced by the person who is harmed by the injustice. This is noteworthy. Although Hume considers that the moral value of justice depends on the fact that it promotes general utility, this consideration is not at the origin of the process that leads us to attribute that value to justice. Hume claims:

> But tho' in our own actions we may frequently lose sight of that interest, which we have in maintaining order, and may follow a lesser and more present interest, we never fail to observe *the prejudice we receive, either mediately or immediately, from the injustice of others*; [...]. Nay when the injustice is *so distant from us, as no way to affect our interest*, it still displeases us; because we consider it as prejudicial to human society, and pernicious to every one that approaches the person guilty of it. We partake of their uneasiness by *sympathy*; and as every thing, which gives uneasiness in human actions, upon the general survey, is call'd Vice, and whatever produces satisfaction, in the same manner, is denominated Virtue; this is the reason why the sense of moral good and evil follows upon justice and injustice. (*T* 3.2.2.24/SBN 499, my italics)

Hume specifies some significant elements of the object of this type of sympathetic episode. The victim of the injustice is not described as someone who has family or friendship relations with the spectator. Indeed, it is an injustice that does not affect his interests and is not even capable of arousing his limited benevolence. The spectator can, however, sympathise impartially with that peculiar suffering that the victim feels when someone has violated the rules of justice against him. These rules have a strong social significance: they protect property, which, as I explained in the previous section, is one of the most important sources of pride and esteem from our fellow human beings. Violation of these norms is perceived by those who suffer it as an attack that expresses contempt. Part of the moral disapproval of injustice thus crucially seems to depend on our ability to participate emotionally, as spectators, in the resentment that victims feel towards those who violate their property and threaten their social position. Therefore, although resentment is different from moral blame, Hume believed that the latter feeling emerged from the former and sympathy combined. Sympathy enables one to feel *as one would* were it oneself whose rights were being violated.

Some might rightly object: the analysis seems plausible, but is it really what the treatise is telling us? Why does Hume not explicitly use the term resentment in this context? Why does he not mention the resentment of the victim or even the sympathetic indignation felt by the observer towards the perpetrator of the injustice? Hume certainly does not mention resentment in this section, but this fact is not decisive for my interpretation since he states shortly before in the same section that 'injury' and 'injustice' are equivalent terms (*T* 3.2.2.8/SBN 488). This is tantamount to implicitly stating what we are arguing: namely, both that injustice, *qua* injury, arouses the resentment of the victim and that this sentiment can be the object of impartial sympathy on the part of a spectator.

In addition to the origin of the moral obligation of justice, the passion of impartial resentment is also taken up in connection with the particular intensity of moral disapproval of injustice. Indeed in the second *Enquiry* Hume argues that if that disapproval would depend solely on an abstract consideration of its negative effects on the well-being of society, it would not have the force that characterizes it in our everyday social interactions. Hume argues that, once justice is respected and approved of in modern commercial societies because it is useful, this feeling is reinforced by our sympathy for the particular injuries caused by injustice. Hume claims:

> We may just observe, before we conclude this subject, that, after the laws of justice are fixed by views of general utility, the injury, the hardship, the harm, which result to any individual from a violation of them, enter very much into consideration, and are a great source of that universal blame, which attends every wrong or iniquity. By the laws of society, this coat, this horse is mine, and *ought to* remain perpetually in my possession: I reckon on the secure enjoyment of it: By depriving me of it, you disappoint my expectations, and doubly displease me, and offend every bystander. It is a public wrong, so far as the rules of equity are violated: It is a private harm, so far as an individual is

> injured. And though the second consideration could have no place, were not the former previously established: For otherwise the distinction of *mine* and *thine* would be unknown in society: Yet there is no question, but the regard to general good is much enforced by the respect to particular. What injures the community, without hurting any individual, is often more lightly thought of. But where the greatest public wrong is also conjoined with a considerable private one, no wonder the highest disapprobation attends so iniquitous a behaviour. (*EPM* App3.11/SBN 310-1)

Again, resentment plays an important role in Hume's theory of justice: once respect for justice has spread through society, sympathy with the victim's resentment strengthens our attachment to this virtue.

In this chapter, I have argued that Hume has a plausible conception of resentment, the strengths of which emerge from the contemporary debate. I have also argued that resentment plays a non-secondary role in the theory of justice. Not only does it underpin the genealogy from which the moralisation of this virtue emerges, but it also helps to explain the particular strength of contempt for injustice. The observations made thus make it possible to outline a new perspective on the emotions that are involved in the genealogy of justice in which, in addition to the positive emotions associated with sympathy for general happiness, the negative emotion of resentment also finds a place. This is an issue that concerns not only the understanding of Hume's moral theory, but also the very idea of justice within the broad empiricist line that, starting with Hume, and through Adam Smith, reaches J.S. Mill's reformed utilitarianism.

Chapter 4

Justice, Sympathy and Resentment in Adam Smith

In the previous chapters I have explained how Hume's genealogical account of the idea of justice shows how it depends in different ways on human passions. Following a fairly consensus approach, I have argued that this is true for at least three different passions. First, the idea of justice would be incomprehensible if human beings were not endowed with a limited amount of benevolence. Second, it is the passion of enlightened self-interest that leads people to enter into the convention that binds them to obey the rules of justice, and that leads them to engage in mutually binding patterns of behaviour that promote the long-term happiness of each person. Third, the idea of justice as a virtue, the possession of which ensures respect for property and contracts in advanced commercial societies, can only be acquired because private and public education relies on the indirect passions of pride and humility. Finally, I have argued that this analysis would be incomplete if it did not include the passion of resentment. In particular, in chapter three, I argued the controversial thesis that Hume not only presents a coherent and defensible conception of this passion, but that it plays a not insignificant role both in explaining the emergence of moral blame for injustice and in reinforcing it once this sentiment is ingrained in our moral practices.

In this chapter, I intend to examine the fortunes of this last idea in Adam Smith's theory of justice, set out in Part II of *The Theory of Moral Sentiments*.[143] Unlike Hume,

[143] References to Smith's *The Theory of Moral Sentiments* is cited with notations of the form *TMS* j.k.m.n, the lower-case letters here standing for Arabic numerals. Numerals immediately

Smith expands the explanatory role that resentment plays in his theory of justice. In contrast to a recent interpretation, however, I will argue that this does not prevent Smith from following Hume in giving the notion of utility a significant role: not only in justifying the practices of justice, but also in resolving moral conflicts that arise from the application of its rules in some specific contexts. I will finally examine whether this last aspect creates tensions within his moral theory. Alex Voorhoeve and Michael Frazer, for example, have argued that Smithian sympathy operates in a non-aggregative way, allowing us to put ourselves in the position of one person rather than a multitude of individuals.[144] This would seem to place serious limits on the function of utility as a justificatory principle of justice.

Before addressing these issues, it is necessary to examine albeit briefly the psychological assumptions of Smith's theory of justice. I will therefore focus in the first sections on his conception of sympathy and his theory of moral evaluation. The study of sympathy is a prelude to the study of moral sentiments, the analysis of which is itself crucial for determining the moral status of duties and the virtue of justice, and for distinguishing them from the rest of morality, which Smith identifies with the principle of beneficence.

4.1 Smith's Account of Sympathy

Smith defines sympathy as 'our fellow-feeling ... with any passion whatsoever' (*TMS* I.I.I.5), whether pleasant or painful. Although this is reminiscent of the notion used by

following *TMS* indicate Part, Section, Chapter, and Paragraph in Adam Smith, *The Theory of Moral Sentiments*, ed. By K. Haakonssen, Cambridge University Press, Cambridge, 2002.

[144] See Frazer 2010: 94. See also Voorhoeve 2014: 69, 73.

Hume in Books 2 and 3 of the *Treatise*, it differs from it in some important respects, which are not always easy to identify because of the sometimes Smith's ambiguous use of the term 'sympathy'.

A first clear distinction with Hume concerns the content of sympathetic feeling. Whereas for Hume to sympathise is typically to feel what the person who is the object of the sympathetic episode actually feels or has felt, for Smith it is instead to feel what a spectator would feel if *he* imagined himself to be in the circumstances of the person with whom he sympathises. In the first case, I feel what someone else has felt or is feeling whose content is communicated to me during the sympathetic episode. In the second case, on the other hand, I feel something that depends on my putting myself in someone else's shoes. This distinction underlies a second element of difference, which concerns the epistemic conditions that must be met in order to sympathize. For Smith, it is necessary to have accurate knowledge of the circumstances in which the agent finds himself[145], that is, the factors that caused the passion with which the spectator sympathizes (*TMS* I.I.I.10).[146] For Hume, however, at least in the most basic forms of sympathy as emotional contagion, this is not necessary. What is necessary instead is the possession of beliefs about the agent's external behavior, without which the sympathetic spectator could not infer any idea of the other person's passion.[147]

[145] This account might be seen as inaccurate. It might be objected that Smith does indeed recognise that there are cases in which we seem to sympathise immediately through what Hume regarded as emotional contagion. For Smith, however, these cases correspond to a very partial sympathy, which can at best arouse in the spectator a curiosity about the causes that produced the agent's emotions, and it is only from this knowledge that a sympathetic episode can be truly complete. See *TMS* I.I.I.9-10.

[146] For a discussion of this difference and the fact that it was recognized as early as 1700, in authors such as Smith's student Dugald Stewart, see Frazer 2010: 113. On the point that Smithian sympathy requires a greater degree of cognitive and imaginative engagement, see again Frazer 2010: 97. See also Darwall 1998: 264.

[147] For a discussion arguing for the importance of inference in the different uses of sympathies in Hume, see Ainslie 2005: 143 ff.

Smith believed that his conception of sympathy was able to explain a greater number of phenomena than did Hume's conception of sympathy.[148] Although he does not explicitly discuss Hume, Smith illustrates his conception through examples that might be considered problematic from a Humean perspective, like the one in which a spectator sympathises with the plight of a corpse locked in a tomb (*TMS* I.I.I.13).[149] Whether Hume's sympathy is actually able to account for these threshold cases as well is irrelevant for the purposes of this chapter; what is important, however, is to point out that the Smith's approach allows for a clear distinction between the content of the sympathetic feeling of a spectator who imagines that he/she is in the agent's circumstances and the content of the feeling that in those circumstances the agent is actually experiencing or has experienced.[150] The possibility of this distinction opens the space in Smith's work for a second use of the term sympathy, this time indicating not so much the operation of imaginatively placing oneself in the circumstances of the agent, but rather, once this perspective is adopted, the perception of a concordance between one's own feelings and those of the agent. On the basis of this second usage, Smith argues that the spectator sympathises with the agent's feelings only when he actually experiences feelings similar to his own, and instead does not sympathise or refuses to sympathise when he either experiences no feelings or experiences feelings of a similar kind but of much less intensity than the agent. Of the passion of violent anger, for example, Smith claims:

[148] On this aspect see Frazer 2010: 98
[149] Ibid.
[150] This distinction creates an internal problem with Smith's conception of sympathy, which I will leave in the background as it is not directly relevant for the purposes of this chapter. The problem stems from the fact that Smith, unlike Hume, is sceptical about the possibility of knowing another person's feelings. Smith argues that it is precisely this epistemological difficulty that underlies his thesis that the only way to sympathise with others is to imagine what we would feel in their circumstances. In the light of what we have said so far, however, this observation shows that Smith is forced into this dilemma: either projective imagination lets me know what the agent feels, or only what I would feel if I were actually there. If Smith follows the first horn of the dilemma, it is not clear how he can distinguish what the agent feels from what the observer feels. If he follows the second horn, then it is unclear how the spectator could know the agent's feelings.

> There are some passions of which the expressions excite no sort of sympathy, but, before we are acquainted with what gave occasion to them, serve rather to disgust and provoke us against them. The furious behaviour of an angry man is more likely to exasperate us against himself than against his enemies. (*TMS* I.I.I.7)

As Michael Frazer has recently pointed out, echoing an analysis originally formulated by Stephen Darwall, Smith's observation adds a third element of distinction from Hume's theory. Sympathizing does not depend on the vividness of the spectator's conception of another's passion, but instead typically depends on the agent's decision as to whether or not the agent's passion is appropriate to the circumstance in which he finds himself.[151]

Smith argues importantly that just as the perception of concordance between the sympathetic feeling and that of the agent is always pleasant, so the perception of discordance is always painful. This is the fourth major difference with Hume. Although his conception of sympathy has little difficulty in explaining the pleasant reflexive reinforcement that each subject receives from the fact that his pleasant feelings are shared by others and bounce back to him, this is not so true in the case of the sharing of painful feelings.[152] More precisely, in this case, the effect of sympathetic sharing on the agent is opposite to the effect on the spectator. In the first case, the agent is relieved to share his pain with his loved ones. In the second case, however, the pain that a spectator may feel at the sight of someone suffering is added to and intensified by the suffering that is

[151] See Frazer 2010: 97-100. This meaning of sympathy is as we shall see identical to the meaning of moral approval. According to this usage, therefore, to sympathize means to approve and not to sympathize means to disapprove.
[152] Ibid.

conveyed to him/her when we sympathise with his/her feelings.[153] In contrast, Smith believed the perception of concordance of emotions was always pleasant for both agents and spectators and that this did not depend on the fact that they tend to reinforce each other, but simply on the perception of emotional concordance.

So far, I have shown how the sympathetic process described by Smith differs from Hume's in four ways. It suggests an imaginative process which, on the basis of knowledge of the circumstances in which an agent finds himself, allows an observer to feel an emotion related to that circumstance which may not be at all like the one the agent feels. Smith believes that only when the emotion felt by the agent is similar to that felt by the spectator as a result of this imaginative process does one have full sympathy with the agent. The perception of this concordance by both people involved is always pleasant for both, just as the perception of discordance is always painful. This happens regardless of the hedonic tone of the initial emotions.

Before concluding, let us address a final question that will be relevant to the discussion of justice. Is this form of sympathy capable of taking into account the distinction between different individuals? Two questions need to be distinguished. The first is whether it allows a distinction to be made between the feelings of the agent and those of the spectator. The second is whether it allows for a form of general welfare sympathy in which the feelings of a few individuals are likely to be overlooked in favour of the sum of all.

The answer to the first question is certainly in the affirmative. As we have seen, Smith is discussing a form of projective sympathy that concludes with a comparison between the feeling that the spectator experiences by imagining himself in the situation

[153] In a letter to Smith, Hume observed that this position is somewhat of a pillar of the Smithian conception of sympathy. Lamenting his disagreement, Hume ironically wrote that 'If all sympathy were agreeable' then 'a hospital would be a more entertaining place than a ball' (*L* 36: 43). On this point see also Frazer 2010: 99.

of the agent and the feeling that the agent experiences. The possibility of comparing these two feelings is obviously based on the spectator's awareness that what he feels and what the agent feels are two different feelings.

The answer to the second question is more complex. Given the way it is presented, it seems that the imaginative mechanism of sympathy can only be activated if the viewer imagines the circumstances in which an individual finds himself. This seems to rule out the possibility that one can sympathise directly with an idea such as general happiness. However, it does not seem to rule out the possibility that one can construct the idea of the happiness of an aggregate of people by sympathising with the reactions of each of them.

Let us now examine how Smith uses these elements to account for moral sentiments of approval and disapproval.

4.2 Moral Sentiments: on Propriety and Impropriety of Action

Smith describes his conception of moral evaluation as a twofold mode of virtue evaluation. In our common moral practice, virtues are evaluated in two different ways: in terms of their effects, or in terms of the causes that give rise to them. Smith claims:

> The sentiment or affection of the heart from which any action proceeds, and upon which its whole virtue or vice must ultimately depend, may be considered under two different aspects, or in two different relations; first, *in relation to the cause which excites it*, or the motive which gives occasion to it; and secondly, *in relation to the end which it proposes*, or the effects which it tends to produce. (*TSM* I.I.III.5, my italics)[154]

[154] Smith restates this same idea in *TMS* II.I.*Introduction*.2.

Before examining these two different modes of moral evaluation, it is important to introduce a clarification, without which the above passage is in danger of being misunderstood. As it stands, it suggests that Smith subscribes to a meta-ethics of moral evaluation which we have seen to be peculiar to virtue ethics. According to this conception, the moral value of actions depends entirely on the value of the virtuous motives that cause them. The only moral value that actions can have is that of being virtuous, and this quality in turn depends on that of the motivational dispositions that produce them. However, as we shall see in more detail below when we consider beneficence and justice, Smith's conception of the moral valuation of actions (and motives) is broader than this. While Smith argues that the moral value of actions depends on their motives, he adds that only some of them are virtuous (*TMS* II.II.I.6). In fact, there are motives whose moral value lies not in possessing the quality of virtue, but in being appropriate to the circumstances in which they arise. This means that although an appropriate motive can in principle also be virtuous, the two motives may diverge, because while the virtuous covers the realm of extraordinary behaviour that deserves admiration, the appropriate motive typically covers ordinary behaviour that is just above the threshold of what is morally blameworthy (ibid.)

Let us now examine the two different modes of moral evaluation mentioned in the passage. The former will be examined in this section, while the latter will be discussed in the next one.

The first mode of moral evaluation is to approve or disapprove a motive, and consequently the action it produces, as proper or improper on the basis of whether or not it is suitable to its object, i.e. the circumstance that caused it (TMS I.I.III.1). These relations are not objective qualities that exist in the world independently of human feelings, but depend on the kind of sympathetic reactions that the motive to be evaluated arouses in the spectator, given the circumstances that prompted that motive. Smith claims:

> When we judge in this manner of any affection, as proportioned or disproportioned to the cause which excites it, it is scarce possible that we should make use of any other rule or canon but the correspondent affection in ourselves. If, upon bringing the case home to our own breast, we find that the sentiments which it gives occasion to coincide and tally with our own, we necessarily approve of them, as proportioned and suitable to their objects; if otherwise, we necessarily disapprove of them, as extravagant and out of proportion. (*TMS* I.I.III.9)

'Propriety', then, consists in the fact that a sympathetic spectator, imagining that he or she is in the situation of the agent, is able to sympathise with the agent's emotion, i.e. to experience a feeling similar to the one he or she is experiencing. The 'impropriety', on the other hand, consists in the fact that this sympathy either does not take place at all or is very imperfect, in the sense that the emotion felt by the spectator has a significantly different intensity from that felt by the agent.

Although the sympathetic exchange of positions requires an imaginative effort capable of representing to us every detail of the other, the spectator cannot aspire to be a perfect sympathiser. Smith argues that it is a good thing that this is so, otherwise the distance of perspective necessary to judge the appropriateness of emotions to the circumstances that evoke them would be lacking. However, Smith adds, importantly, that this mode of moral evaluation requires the spectator to take an unbiased perspective on the circumstances in which the agent who is the object of the sympathetic process finds himself. How do we account for this additional constraint on the moral evaluation of the 'propriety' of emotions?

According to Frazer, Smith needs an unbiased perspective to avoid the possible psychological and social contradictions and conflicts that would arise if we let our

imagination be affected by the relationships we have with the people we sympathize with. According to Frazer, Smith would here follow Hume's account of the emergence of the common point of view.

> Yet Smith, like Hume, realized that our sympathy varies along with the closeness of our relationship to the objects of our feelings and hence that our judgments of propriety will be biased in favor of those closest to us. To avoid the social and psychological contradictions that result […], we correct our biased judgments through appeal to an imagined impartial spectator within, the functional equivalent of Hume's appeal to the general point of view.[155]

Frazer's hypothesis, however suggestive, fails to take account of an important aspect of the analysis of 'propriety' undertaken in Part I, Chapter IV, of of *TMS*. Unlike Hume, Smith argues that the conflicts caused by fluctuations in our evaluations of motives arise not so much from the fact that the operations of sympathy vary in intensity as the relations - of similarity, of space-time contiguity, or of causality - to the object of our sympathy vary, but rather from the fact that in some circumstances we do not sympathise with others at all. This inability to sympathize is typically manifested when people experience violent passions over offenses received. Spectators have an instinctive difficulty in sharing the violence of the reactions of the offended, and this generates mutual animosity and ill will. Smith notes:

> Though your judgments in matters of speculation, though your sentiments in matters of taste, are quite opposite to mine, I can easily overlook this opposition; and if I have any degree of temper, I may still find some entertainment in your conversation, even upon

[155] On this aspect see Frazer 2010: 100.

> those very subjects. But if you have either no fellow-feeling for the misfortunes I have met with, or none that bears any proportion to the grief which distracts me; or if you have either no indignation at the injuries I have suffered, or none that bears any proportion to the resentment which transports me, we can no longer converse upon these subjects. We become intolerable to one another. I can neither support your company, nor you mine. You are confounded at my violence and passion, and I am enraged at your cold insensibility and want of feeling. (TMS I.I.IV.5)

In order to avoid this type of conflict, the spectator must strive to acquire true beliefs about the psychological situation and circumstances of the agent with whom he or she sympathises. To accomplish this task, it is not necessary to adopt an impartial perspective; rather, it is a matter of striving to reconstruct all the 'minutest incidents' of the case of the person with whom we sympathise (*TMS* I.I.IV.6).

Contrary to Frazer's hypothesis, the need for an unbiased perspective seems to be explained not so much by the need to ensure the consistency of sympathetic judgments that would be sufficient to avoid social conflict, but rather by the more difficult task of achieving the requirement that all members of the community experience the same passions according to a shared measure of propriety. Smith's idea is that the greater the distance between the spectator and the agent, the more the agent must control his or her passions in order to hope to be approved by the spectator. Since everyone desires social approval, the only way to obtain it is to strive to tone down the intensity of our passions to the point where we can hope that everyone else, not just our close friends, will share them (*TMS* I.1.IV.9). In this way, the assumption of an impartial perspective is primarily one that the agent assumes in order to make himself presentable to an audience of outsiders who, by virtue of the distance that separates them from him or her, will in turn tend to be naturally impartial towards him or her. By incorporating this impartial

perspective, the evaluation of motives by their properties thus becomes a means of equalising human passions, depriving them of the roughness that makes them unsuitable for peace and social harmony.

Before turning to the second mode of valuing motives, a further observation about the comparison with Hume is in order. Smith claims the originality and importance of his proposal by arguing that it is contrary to that of philosophers who base moral evaluation on the tendencies of affections to the well-being of men, Smith claims:

> Philosophers have, of late years, considered chiefly the *tendency* of affections, and have given little attention to the relation which they stand in to the cause which excites them. In common life, however, when we judge of any person's conduct, and of the sentiments which directed it, we constantly consider them under both these aspects. When we blame in another man the excesses of life, of grief, of resentment, we not only consider the ruinous effects which they tend to produce, but the little occasion which was given for them. The merit of his favourite, we say, is not so great, his misfortune is not so dreadful, his provocation is not so extraordinary, as to justify so violent passion. We should have indulged, we say; perhaps have approved of the violence of his emotion, had the cause been in any respect proportioned to it. (*TMS* I.I.III.8, my italics)

Smith's form of moral evaluation evaluates emotions in terms of whether they are appropriate to the circumstances in which they are felt, and this is said to be assessable independently of their typical effects on human well-being. This might lead us to ask whether, using contemporary metaethics, we can characterise it as a non-consequentialist conception of moral evaluation. According to this conception, the moral value of something does not depend on the value of its consequences, but on some properties of actions/emotions.

Although this may seem a plausible reading, the hypothesis does not take into account that the property of a motive to be evaluated depends in turn on whether it produces pleasure in the spectators who are able to sympathise with it. Smith's idea, then, is not that property is independent of pleasurable consequences tout court, but only of the pleasurable consequences that the motive being evaluated has on those who are the objects or targets of that motive. In other words, the moral quality of a benevolent desire does not depend on the consequences - pleasant or painful - that motive has for the person benefited, but on the consequences that it has for bystanders who are able to sympathise with that motive. Viewed in this light, the dynamic of the judgement of property does not seem to differ much from that of Hume's moral judgement, which is based on the criterion of agreeableness, according to which a quality is valued simply because, regardless of its consequences for human welfare, it produces immediate pleasure in those who contemplate it from a general point of view. Unlike Hume, Smith argues that a proper motive is not typically a virtue. Beyond this difference, however, the importance that both authors attach to the pleasure that the object of approval inspires in its impartial spectators remains identical.

Although the judgment of 'propriety' is not directly concerned with justice, it is nonetheless in an important sense the basis of the evaluation that is concerned with this virtue. As we shall see in the next section, the judgment of merit and demerit that evaluates actions as just or unjust could not be made without a preceding judgment that asses the 'propriety' of their motives in relation to the circumstances of action. In this way, the judgment of 'propriety' ends up playing a constitutive role in the complex psychological mechanism that accounts for the value of justice.

4.3 Moral Sentiments: Merit and Demerit. The Evaluation of Justice and Beneficence

In Part II of the *Theory of Moral Sentiments* Smith discusses a second type of evaluation that assesses motives according to the beneficial or harmful effects they tend or are intended to produce, i. e., according to the properties of 'merit' and 'demerit' (*TMS* II.I.2).[156] From an examination of these sentiments, I will begin to examine the theme of justice. Looking at justice from this perspective will allow me to identify the specific character of this part of morality, i.e. the status of duties and the virtue of justice, and to contrast it with the remaining part, which Smith traces back to beneficence.

Unlike 'propriety', the judgment of 'merit' is based not on one, but on two sympathetic operations that are directed at two different objects. The first, which Smith calls 'direct sympathy', is directed at the motives of the agent being evaluated. The second, which Smith calls 'indirect sympathy', is instead directed at the persons who are the recipients of the agent's motives and actions. Why are both operations necessary to make a judgment of merit? Smith's explanation involves three steps.

First, he holds that 'merit' is the quality that a motive/action possesses when it arouses appropriate gratitude in the person to whom it is directed. Similarly, 'demerit' is the quality that a motive/action possesses if it produces an appropriate feeling of resentment in the recipient of the harmful action. Second, in order to assess whether an action/motive possesses the quality of 'merit'/'demerit' we must first assess whether the reactive passion of gratitude/resentment aroused in the recipient is indeed 'proper'. Smith thus does not exactly follow the Humean line of judging the consequences of a motive by sympathising with the pleasure and pain felt by those affected by it. Sympathy remains

[156] Smith surprisingly seems uninterested in distinguishing real effects from those merely intended perhaps suggesting that our knowledge of purposes of actions is obtainable only from knowledge of the tendencies of actions.

central but narrows its object to the two reactive passions felt by the recipients toward the agent. This is a sympathetic operation that Smith calls 'indirect' because it evaluates the agent's motives/actions not by sympathising directly with them (as in the case of 'propriety'/'impropriety'), but with the reactions they elicit in the recipients. Third, this indirect form of sympathy and evaluation is in turn based on direct sympathy, which verifies the 'propriety'/'impropriety' of the agent's original motive being evaluated.

Points two and three make it clear why the two sympathetic operations are both crucial. While it is true that the evaluation of 'merit'/'demerit' consists of the sympathetic feeling an observer experiences when attuned to the feelings of gratitude and revenge we feel towards the one who has injured or benefited us, this indirect sympathetic feeling would not be possible if the observer were not first able to experience direct sympathy, or lack thereof, for the motives of the benevolent or malevolent agent. This means, for example, that if I thought it was totally inappropriate to give a stranger a large sum of money simply because he or she shared the same quirky preferences about which flavours go well together in an ice-cream cone, I would not even be able to sympathise with the feeling of gratitude that the recipient of the fortune feels towards his or her benefactor. Although the bizarre benefactor's action is indeed beneficial, my inability to sympathise with it does not entitle me to regard it as meritorious (*TMS* II.I.V.2; see also *TMS* II.I.V.5). Similarly, and conversely, if I think that a teacher's punishment of a pupil who bullies his classmates is an appropriate action, I cannot sympathise with the pupil's pain or with his possible desire to retaliate against the teacher, and so I will not judge the retaliation as having merit. These examples show a very clear relationship between the two forms of sympathy: judging the appropriateness of a motive is somehow not only a necessary condition, but also logically and temporally prior to judging actions and motives in terms of their consequences.

What are the human motivations that are typically the objects of merit-based evaluation? It is in the context of answering this question, set out in *TMS* Part II, Section II, that Smith introduces his treatment of justice. Following a line typical of the natural law tradition and codified by Grotius in his *De iure ac belli et pacis* (1625), Smith characterises justice by contrasting it with beneficence. Although both are the subject of judgments of moral merit, justice and beneficence are distinguished not only by the kinds of duties that characterise these two moral principles, but also by the human passions that sustain them. Smith further complicates this distinction by also relating justice and beneficence to two different virtues, i.e. two kinds of motivational dispositions that produce just and beneficent actions respectively. Examining this distinction in its various dimensions is essential to understanding the nature of Smithian justice. We begin by examining beneficence and then move on to justice.

Beneficence denotes a broad class of duties concerning the promotion of the good of others in a variety of circumstances, ranging from the partial duties we owe to those with whom we have loving and friendly relations, to those of charity for human beings in general, to those that require us to repay those who have benefited us (*TMS* II.II.I.3). Compliance with these duties cannot be imposed by force, and their possible violation is never the subject of punishment. In other words, individuals are free to observe them or not, and if they do observe them, they are also free to choose both the greater or lesser extent of their benefit and the person who is the vessel of their good. This feature determines the kind of moral evaluation that is appropriate for beneficence. Let us begin by examining moral blame for ingratitude, which Smith considers the most serious case of violating beneficence.

In this case, moral disapproval is based on the impartial observer's failure to sympathise with the selfish motives expressed in the behaviour. We therefore blame the action that expresses these motives as 'improper' and tend to hate its perpetrator. To

generalise, actions that do not express the virtue of beneficence are amenable to the first type of moral evaluation, that which assesses the 'impropriety' of the agent's motives. But what is to be said about the second type of disapproval: are these actions also amenable to having a 'demerit'?

Smith's answer is negative. An action has the quality of demerit only if it arouses the passion of proper resentment in the one who suffers the effects of the action. But resentment is an appropriate response only if it is directed against the perpetrator of the actual injury. However, actions that express a lack of beneficence do not cause individual injury. Therefore, the impartial observer does not sympathise with the resentment felt by the betrayed benefactor towards the ungrateful beneficiary. If the resentment is 'improper', this means, importantly, that the breach of the duty of gratitude cannot be disapproved as an expression of 'demerit'. Note how this marks a clear distinction from Hume, who holds that a character lacking good and benevolent virtues, especially gratitude (T 3.1.1.24/SBN 466; see also T 3.2.1.8/479), is morally blameworthy precisely because of its painful effects on others. For Hume, by impartially sympathising with the sufferings of the ungrateful person's inner circle due to the tendencies of his character, each spectator feels a painful feeling and disapproves of the ingratitude.

Let us now consider the nature of the moral approval of beneficence. Unlike the previous case, benevolent motives elicit both forms of moral evaluation. On the one hand, the impartial bystander sympathises with the benevolent motivation that leads the agent to care for others when they are in need; on the other hand, he sympathises with the feeling of gratitude that the benefited person feels towards his benefactor, and thus approves of the benefactor's actions as meritorious. The approval of beneficent motives thus involves both types of evaluation, that of 'propriety' and that of 'merit'. To this, Smith adds an important clarification that, again, marks a distinction from Hume. A beneficent motive is accepted as meritorious only if the agent's motives and consequent actions transcend

the common measure and concern what is 'praiseworthy' rather than what is ordinary (*TMS* II.II.I.6). As I have already mentioned, Smith retains here a feature of the ethics of classical and Hellenistic Greece of identifying virtue with excellent behaviour. This contrasts, with the Humean approach that identifies natural virtues with the degree of solicitude for the welfare of others that enables the various spheres of social and affective relations that constitute our common life to flourish. For Smith, however, if going below this threshold identifies what is 'improper' and blameworthy, staying within it circumscribes a sphere of actions and motivations that is morally neutral from the point of view of 'merit'.

Let us now turn to the consideration of justice. Unlike beneficence, it is a set of duties that can be imposed by force. The reason for this distinction lies in the fact that unjust actions not only cause suffering to others and are motivated by what is considered an improper motive, but also tend to cause harm or injury to those to whom they are directed. Echoing the distinction from the natural law tradition later taken up by Mill, Smith argues that injury corresponds to an individual right in the offended person, 'antecedent to the institution of civil government', which entitles him to defend himself 'against wrongs' and 'to inflict some measure of punishment on those who have done him wrong' (*TMS* II.II.I.7). Natural resentment thus identifies and delimits the content of justice, which Smith identifies as that which violates a person's bodily integrity, his property and, ultimately, his word. Justice thus denotes the set of rights and duties that require us both to respect human life, property, and the given word, and to punish acts that violate them. Smith thus has a broader conception of justice than Hume, which includes not only respect for property and promises, but also respect for the physical integrity of human beings.

Like beneficence, however, justice is the name given to the motives behind the actions that fulfil these duties. Smith thus offers an explanation of moral evaluation that

takes account of the moral status of these motives. Again, it is useful to distinguish between approval of their presence and disapproval of their absence in an individual's character.

Unlike ingratitude, injustice gives rise to two different forms of moral blame. The first, like the disapproval of ingratitude, depends on the impartial observer's being unable to sympathise with the motive for violating the rules of justice, and judging that motive to be 'improper' to the circumstances of the action. The second, unlike in the case of ingratitude, depends instead on the impartial observer this time being able to sympathise with the injured agent's resentment and thus to disapprove of the injurer's actions as having 'demerit'.

When we act justly, however, although the motives are approved as 'proper' to the circumstances, they are not deserving of gratitude. There is nothing exceptional about just behaviour for Smith and therefore it is not approved as deserving 'merit'. This underlies the well-known Smithian thesis that justice can at best be considered a negative virtue, that is, a disposition to perform actions that prevent us from harming our neighbour. Smith claims:

> Though the breach of justice, on the contrary, exposes to punishment, the observance of the rules of that virtue seems scarce to deserve any reward. There is, no doubt, a propriety in the practice of justice, and it merits, upon that account, all the approbation which is due to propriety. But as it does no real positive good, it is entitled to very little gratitude. Mere justice is, upon most occasions, but a *negative virtue*, and only hinders us from hurting our neighbour. The man who barely abstains from violating either the person or the estate, or the reputation, of his neighbours, has surely very little positive merit. He fulfils, however, all the rules of what is peculiarly called justice, and does every thing which his equals can with propriety force him to do, or which they can punish him for not

doing. We may often fulfil all the rules of justice by sitting still and doing nothing. (*TMS* II.II.I.9, my italics)

Given the examination of moral evaluation in this and the previous section, we are in a position to identify some important aspects of the moral status of justice. As I have shown, justice refers both to a set of moral principles - such as respect for bodily integrity, property and promises - and to the motives that lead people to act in ways that respect these principles. In this first phase of the investigation of justice, which is carried out as part of the examination of the two modes of moral evaluation, Smith is not yet concerned with the value of the rules of justice. Instead, he is concerned with the moral status of the motives that cause just and unjust actions, and already at this level he provides some useful elements for comparison with Hume's theory, which we have examined in previous chapters.

First, unlike Hume, Smith does not question the nature of the motive in which the virtue of justice consists. This is probably due to two factors, which we have examined so far, which distinguish Smith's treatment from that of Hume. First, Smith does not argue that justice is a virtue sic et simpliciter, because there is no excellence in acting justly. Rather, justice is a virtue sui generis, which consists simply in not being moved by motives that lead us to violate the rules of justice. This makes positive discourse on this motive of little interest, since there is no single morally appropriate motive to identify, but appropriate motives to the circumstances of justice can be manifold. Second, unlike Hume, who has Book 3 revolve around the distinction between artificial and natural virtues, Smith does not pursue this distinction at all, and his two modes of approbation, if anything, echo the distinction between useful and immediately pleasing virtues. Here, too, the examination of the nature of the motive of justice loses its appeal because the question of whether or not it is an original motive of human nature is abandoned.

Second, beyond the difference in terminology, Smith agrees with Hume that unjust motives arouse more complete and stronger moral disapproval than motives expressing the absence of benevolent affections (*TMS* II.II.I.5). Indeed, Smith argues that while an unjust motive is both improper and deserving of resentment on the part of the injured party, the same is not true of the absence of a benevolent motive. Indeed, in this case, any resentment felt by the person who is not the object of generosity or gratitude could not be sanctioned by an impartial observer of the situation.

4.4 The Sanctions of Injustice. Resentment, Remorse, and Shame

So far, we have examined the role that the passion of resentment - as the object of the impartial spectator's indirect sympathy - plays in Smith's account of moral disapproval of unjust actions. What I want to emphasise now is that Smith believes that resentment underlies the invention and observance of the rules of justice. This argument is developed in several parts of *TMS*, Section II, where Smith distinguishes what he calls the 'efficient cause' of justice from its 'final cause' (*TMS* II.II.III.5). Let us examine this aspect in its various components.

Smith takes as his starting point the thesis, widely accepted in modern philosophy, that human beings can only survive in society because such 'weak and imperfect' creatures (*TMS* II.I.V.10), equal in strength and vulnerable to each other's offence, need mutual support. Although this support could be provided by the passions of love, gratitude and friendship, and although society will be prosperous and happy when this happens, society can exist without these social passions. In a way, this is fortunate, because while human beings are naturally inclined to be social, they also have a natural tendency to be selfish. Smith argues:

> Every man, therefore, is much more deeply interested in whatever immediately concerns himself, than in what concerns any other man: and to hear, perhaps, of the death of another person, with whom we have no particular connection, will give us less concern, will spoil our stomach, or break our rest, much less than a very insignificant disaster which has befallen ourselves. (*TMS* II.II.1)

A little further on, he adds:

> Men, though naturally sympathetic, feel so little for another, with whom they have no particular connection, in comparison of what they feel for themselves; the misery of one, who is merely their fellow-creature, is of so little importance to them in comparison even of a small conveniency of their own; they have it so much in their power to hurt him, and may have so many temptations to do so [...] they [...], like wild beasts, be at all times ready to fly upon him; and a man would enter an assembly of men as he enters a den of lions. (*TMS* II.II.III.4).

Respect for beneficence, with all the duties it entails, is therefore not a necessary condition for the existence and maintenance of society. Instead, what is necessary is that people do not harm each other by respecting each other's contracts and property. Smith claims:

> But though the necessary assistance should not be afforded from such generous and disinterested motives [...] the society, though less happy and agreeable, will not necessarily be dissolved. [...] Society, however, cannot subsist among those who are at all times ready to hurt and injure one another. (*TMS* II.II.III.2-3)

Society, Smith astutely observes, can exist even among thieves and murderers who are utterly devoid of benevolence, provided they respect the rules of justice among themselves.

> If there is any society among robbers and murderers, they must at least, according to the trite observation, abstain from robbing and murdering one another. Beneficence, therefore, is less essential to the existence of society than justice. Society may subsist, though not in the most comfortable state, without beneficence; but the prevalence of injustice must utterly destroy it. (*TMS* II.II.III.3)

Fortunately, human nature is constituted in such a way as to have endowed us with a strong natural instinct to protect justice, which, unlike beneficence, is the main pillar of society. Although human beings have weak instincts to defend the principle of beneficence, they possess the strong passion of resentment to safeguard justice and thus the human association necessary for mutual assistance. Beginning in *TMS*, Part II Chapter IV, Smith explains in detail how this instinct functions in relation to safeguarding justice, highlighting how the function of this passion is assisted by the passions of indignation, remorse and shame. Before examining this account, it is worth noting a methodological clarification. Smith emphasises that this inquiry 'is not concerning a matter of right, if I may say so, but concerning a matter of fact' (*TMS* II.II.V.10). He points out that he is not interested in discovering which principles 'a perfect being would approve of the punishment of bad actions', but rather intends to indicate 'upon what principles so weak and imperfect a creature as man actually and in fact approves of it'. In line with Hume's approach, Smith thus primarily wants to offer an account that explains the emergence of justice from a conception of human beings grounded in experience. Let us now look in detail at the main aspects of this explanation.

First, Smith emphasises the crucial importance of the existence of this instinct in its ability to infallibly guide us towards the discovery of the means necessary for the maintenance of society. According to this argument, although man is naturally endowed with the desire to preserve society, and although this desire, aided by reason, can in principle show us which courses of action would be conducive to a peaceful social form of life, the activity of reason would be an inadequate instrument for this purpose, since its judgments would still be slow and uncertain. Smith therefore rejects the Humean line, which explains the natural obligation to justice in terms of the passion of self-interest aided by reason, and considers it more plausible to explain the emergence of justice and society in terms of resentment. Indeed, this passion infallibly drives us to punish actions that harm others, regardless of whether these sanctions are perceived as useful for the maintenance of society. Through resentment, the institution of punishment for harmful actions is sought as an end in itself, rather than as a means to something else.[157]

How exactly can resentment, which Smith defines as an asocial passion (*TMS* I.II.III), perform this function?

Let us begin with some of the characteristics of the natural passion of resentment, to which we pointed in Chapter 3. Similarly to Hume, Smith argues that resentment is a

[157] Smith claims: 'Though man, therefore, be naturally endowed with a desire of the welfare and preservation of society, yet the Author of nature has not entrusted it to his reason to find out that a certain application of punishments is the proper means of attaining this end; but has endowed him with an immediate and instinctive approbation of that very application which is most proper to attain it. The economy of nature is in this respect exactly of a piece with what it is upon many other occasions. With regard to all those ends which, upon account of their peculiar importance, may be regarded, if such an expression is allowable, as the favourite ends of nature, she has constantly in this manner not only endowed mankind with an appetite for the end which she proposes, but likewise with an appetite for the means by which alone this end can be brought about, for their own sakes, and independent of their tendency to produce it. Thus self-preservation, and the propagation of the species, are the great ends which nature seems to have proposed in the formation of all animals. Mankind is endowed with a desire of those ends, and an aversion to the contrary; with a love of life, and a dread of dissolution; with a desire of the continuance and perpetuity of the species, and with an aversion to the thoughts of its entire extinction. But though we are in this manner endowed with a very strong desire of those ends, it has not been entrusted to the slow and uncertain determinations of our reason, to find out the proper means of bringing them about. Nature has directed us to the greater part of these by original and immediate instincts. Hunger, thirst, the passion which unites the two sexes, the love of pleasure, and the dread of pain, prompt us to apply those means for their own sakes, and without any consideration of their tendency to those beneficent ends which the great Director of nature intended to produce by them'. (*TMS* II.II.V.10)

painful passion caused by actions that are perceived as offensive to one's status and security, and is associated with a desire for revenge, which is seen as compensation for an injury suffered. Furthermore, Smith argues that resentment is an original and innate human instinct, rather than an inclination learned through education. Its connection with revenge marks a crucial distinction from hatred, which, together with resentment, forms the group of so-called antisocial passions. Unlike hatred, which is satisfied simply by the suffering of the hated person, the satisfaction of resentment requires that it be the wronged person who causes the wrongdoer to suffer, and that the wrongdoer suffer an injury of the same kind as that which he inflicted in the first place. Smith claims:

> Resentment would prompt us to desire, not only that he should be punished, but that he should be punished by our means, and upon account of that particular injury which he had done to us. Resentment cannot be fully gratified, unless the offender is not only made to grieve in his turn, but to grieve for that particular wrong which we have suffered from him. He must be made to repent and be sorry for this very action, that others, through fear of the like punishment, may be terrified from being guilty of the like offence. (*TMS* II.I.6)

Smith presents a conception of natural resentment not too dissimilar to that of Hume. In particular, the desire for revenge is satisfied when the following three conditions are met: 1) the offender suffers at the hands of the offended; 2) the suffering inflicted by the offender must be of the same kind as that originally inflicted by the offended; 3) point (2) must be clear to the offender. Smith argues that actions that cause harm to others, i.e. those that will constitute violations of the rules of justice once society is formed, typically arouse the aggrieved person's resentment towards the offender. Having said this, Smith adds, importantly, that not only can resentment arise from acts that are only perceived by the aggrieved as harm, but that even when they are genuine

harms, they can lead the aggrieved to inflict punishment that is disproportionate to the harm. Fortunately, the desire for sympathy and approval from others, which everyone craves, especially when their own suffering from offence is at stake (*TMS* I.I.II.2), causes those who wish to take revenge to moderate their passion to a level that can be sympathetically shared not only by our friends, but by anyone who looks at our condition with an unbiased mind. Since we wish to be esteemed by society, and since this presupposes that others disapprove of our offences as deserving of our resentment, men are apt, long before the creation of stable laws of justice, to inflict punishments which are justified not so much by the blind vengeance of the individual as by the noble indignation of all who consider the circumstances of our injury. Smith does not fail to emphasise that the 'indignation' of the impartial spectator, or what he will on other occasions call 'noble and generous resentment' (*TMS* I.II.III.7), nevertheless retains the constituent features of resentment, in the sense that we wish the person to be punished precisely for the offence done to us. Noble resentment cannot be fully satisfied unless the offender suffers that particular evil from us and is moved to repent for that very act (*TMS* II.I.6).

Smith notes that although the offender can escape from society, he will tend not to do so because he believes that social death is more terrible than punishment. Within society, regardless of the pain of punishment, the offender's psychological state is characterised by revolving around the passion of remorse, a passion that combines compassion for those who suffer because of him with the shame aroused by the disapproval of others (*TMS* II.II.3). The fear of being in this state, far from one's fellow human beings and yet in their midst, tormented by remorse and fear of punishment, is the strongest deterrent to injustice.

In conclusion, Smith's account argues that the explanation for the emergence of the rules of justice, or what he calls in *TMS* their 'efficient cause', depends on the instinct of resentment. This initially antisocial passion is transformed by sympathy into the

passion of ‘noble and generous resentment’ or impartial ‘indignation’, which advocates punishment only for actions that harm others. The fear of incurring the social punishments that resentment produces, coupled with the anticipated fear of remorse, is the surest guardian that drives men to respect justice and to preserve society, regardless of uncertain judgments as to its long-term utility. This psychological mixture ultimately tends progressively to produce the formation of laws of justice that will more precisely fulfil the political purposes of punishment, namely the correction of the criminal and the public example for others (*TMS* II.I.6).

4.5 Justice and Utility

Smith examines the relationship between justice and utility from *TMS* II.II.6, and takes up this issue in *TMS* IV as part of the more general question of the effect of utility on feelings of approval.

In TMS II.II.6, probably echoing Hume’s analysis in the third section of *An Inquiry Concerning the Principles of Morals*, Smith states that ‘every man is conscious that his interest is connected with the prosperity of society’ and thus with the observance of justice, since ‘injustice necessarily tends to destroy society’. As we have seen in the previous section, the status of these statements should not be misunderstood: self-interest and the knowledge that society cannot survive unless the laws of justice are obeyed are not the source of establishing and obeying justice, but can only be the source of a further motive to obey its rules in addition to the original one. In Smith’s terminology, the consideration of utility is never the efficient cause, but only the final cause of justice (*TMS* II.II.6).

But Smith’s view raises two important questions. First, is this all that Smith has to say about the function of utility in his theory of justice? Second, if the answer to this

question were a positive one, i.e. if the role of utility were merely to reinforce motives not to harm others, arising from fear of their resentment and consequent remorse, would this be a function of little importance to Smith?

In the remainder of this section, I intend to answer these two questions by arguing two theses.

First, Smith does not believe that strengthening our motives to respect justice is the only function that this concept performs. In fact, the direct appeal to utility in some cases generates duties of justice, more precisely duties to punish injustice, which would not exist if the conception of harm and resentment were not combined with an aggregative conception (*TMS* II.II.III.11). Second, Smith believes that strengthening our natural motives against injustice is by no means a trivial function. It is indispensable in the administration of justice, because in many cases the appeal to utility is the only way to counter the influence of sympathy in favour of withholding appropriate punishment. Let us consider the two cases separately. Let us start with the latter.

Smith argues that the appeal to utility plays a key role in the administration of punishment because in some cases it allows the execution of a sentence that, if justified on the basis of impartial sympathy for the victim's resentment, would be in danger of not being carried out because of sympathy for the condemned man's fate. In other words, Smith argues that a murderer in chains is seen as a harmless person, and this arouses feelings of sympathy for him. In these circumstances, only considerations appealing to the usefulness of punishment to humanity as a whole will ensure that it is actually carried out, so that justice is done.

> […] we frequently have occasion to confirm our natural sense of the propriety and fitness of punishment, by reflecting how necessary it is for preserving the order of society. When the guilty is about to suffer that just retaliation, which the natural indignation of mankind

> tells them is due to his crimes; […] when he ceases to be an object of fear, with the generous and humane he begins to be an object of pity. The thought of what he is about to suffer extinguishes their resentment for the sufferings of others to which he has given occasion. They are disposed to pardon and forgive him, and to save him from that punishment, which in all their cool hours they had considered as the retribution due to such crimes. Here, therefore, they have occasion to call to their assistance the consideration of the general interest of society. They counterbalance the impulse of this weak and partial humanity, by the dictates of a humanity that is more generous and comprehensive. They reflect that mercy to the guilty is cruelty to the innocent, and oppose to the emotions of compassion which they feel for a particular person, a more enlarged compassion which they feel for mankind. (*TMS* II.II.III.7)

Let us now consider the first point. Smith claims that there are cases in the administration of justice, such as that of a military sentinel who falls asleep on duty and is executed for this offence, where the punishment is not justified by the resentment of the victim. According to Smith's hypothesis, the sentinel's behaviour causes no real individual harm, and therefore no resentment on the part of the victim, but only a possible future inconvenience to society. Smith claims:

> Upon some occasions, indeed, we both punish and approve of punishment, merely from a view to the general interest of society, which, we imagine, cannot otherwise be secured. Of this kind are all the punishments inflicted for breaches of what is called either civil police, or military discipline. Such crimes do not immediately or directly hurt any particular person; but their remote consequences, it is supposed, do produce, or might produce, either a considerable inconveniency, or a great disorder in the society. A sentinel, for example, who falls asleep upon his watch, suffers death by the laws of war, because such carelessness might endanger the whole army. This severity may, upon many occasions, appear necessary, and, for that reason, just and proper. When the preservation

> of an individual is inconsistent with the safety of a multitude, nothing can be more just than that the many should be preferred to the one. (*TMS* II.II.III.11)

Smith does not specify exactly what inconvenience is caused by the sentinel's behaviour. However, it is plausible that he has in mind the fact that if this behaviour were not punished by death, it could spread among the soldiers and, in the long run, become extremely harmful to society, which, deprived of effective defences, would run the risk of becoming easy prey for its enemies. The punishment of the guard is therefore justified in so far as it is an effective, i.e. useful, means of preventing the suffering of fellow citizens, although it would be incomprehensible if we were to base it solely on the principle of harm and the resulting resentment of the victim.

At this point we might try to probe Smith's response more deeply. What is the appeal to utility that justifies extending injustice to the sleeping sentinel's behaviour? Smith tells us two things. First, the sentiment of approval for the sentinel's punishment is very different from the one we have for the punishment of a murderer, and this is evidence that the 'approbation of the one is far from being founded upon the same principles with that of the other' (*TMS* II.II.III.11). Second, the 'Sentinel case' is governed by an aggregative principle, according to which if the sacrifice of one person's life is a condition for the lives of many, then we must sacrifice that person. Smith claims:

> When the preservation of the individual is incompatible with the salvation of the multitude, nothing can be more just than to prefer the many to the one (*TMS* II.II.III.11).

The question at this point is: is Smith's conception of moral evaluation capable of supporting this principle? And again, if moral evaluation is based on sympathy, is Smith's

sympathy capable of performing the function it is supposed to perform in order to make moral evaluation work to justify this principle?

Let us begin with sympathy. As we saw in the first section, sympathy is based on an operation of the imagination, typically concerned with the construction of the circumstances in which an individual operates, and not with an abstract entity such as the welfare of society. This seems at first sight to prove that moral approval of the punishment we are discussing, whatever it may be, cannot be based on sympathy. This would also imply that the moral evaluation in question must be of a different kind from the two modes we considered in Sections 4.2-3, which are based precisely on the principle of sympathy.

This reading has found a place not only among Smith scholars, but also in contemporary ethical reflection. According to Voorhoeve, the fact that Smithian sympathy typically concerns individual emotions shows that his theory is able to accommodate morally the importance of the separateness of persons, i.e. the fact that an action that sacrifices the interests of the few for the benefit of the many is not automatically a morally justified action. According to Voorhoeve, the function of the principle of sympathy is precisely to restrain an aggregative conception of value that would lead to the rights of individuals being violated in order to promote the happiness of the many.[158] Indeed, this aggregative view could be countered by arguing that it produces non-meritorious actions, since an impartial bystander would sympathise with the sacrificial victim's resentment of his value-aggregating abusers. But is it not precisely this sacrifice that Smith is justifying here? Is he not arguing that the sentinel's right to protect his own life must be sacrificed when the security of society is at stake? Are we then to conclude that Smith is using a new conception of sympathy here different from that described in the first part of *TMS*? Is he using a new, a third, mode of moral evaluation?

[158] See Voorhoeve 2014: 73.

I think the answer to this question should be in the negative. Although Smith argues that the sentinel's punishment is based on an aggregative principle that is not at work in typical cases of injustice, his conception of sympathy and moral evaluation is perfectly capable of working with it. This becomes clear in the section immediately preceding the one in which the Sentinel example is discussed, where in *TMS* II.II.9 Smith touches in passing on the question of our concern for the general welfare of society. The interest in the happiness and unhappiness of the multitude is given by the composition of our interest in the individuals who make it up. We do not take an interest in the individuals out of an interest in the multitude; on the contrary, we can only take an interest in the multitude because it is composed of individual elements in which we take an interest. Smith claims:

> The concern which we take in the fortune and happiness of individuals, does not, in common cases, arise from that which we take in the fortune and happiness of society. We are no more concerned for the destruction or loss of a single man, because this man is a member or part of society, and because we should be concerned for the destruction of society, than we are concerned for the loss of a single guinea, because this guinea is part of a thousand guineas, and because we should be concerned for the loss of the whole sum. In neither case does our regard for the individuals arise from our regard for the multitude; *but in both cases our regard for the multitude is compounded and made up of the particular regards which we feel for the different individuals of which it is composed*. (*TMS* II.II.III.10, my italics)

A little further on he adds an important clarification: this interest is not to be confused with feelings of love, esteem and affection - which we feel only for particular friends or acquaintances - but rather indicates a general feeling of sympathy that we feel for everyone as our fellow human beings. So when Smith points out that our interest in

the multitude depends on our interest in the individual, he is talking about sympathy. Thus, although Smithian sympathy typically has an individual person as its object, it can also have a multitude as its object, as long as this operation is the result of the multiple operations of sympathy directed at its components.

This gives us a key to interpreting how the appeal to utility - that is, the aggregative principle regarding the value of human life - can be accommodated within Smith's theory of the moral evaluation of justice. First, from the perspective of the impartial observer, the sentinel's behaviour can be seen as inappropriate to the circumstances. He is not exercising control over his own inclinations in a circumstance in which, given the loss of life, he should have been in control. The Sentinel's behaviour is therefore disapproved as inappropriate to the circumstances. However, in what way can this behaviour also be considered non-meritorious?

Since there is no real harm, this judgement would seem to be ruled out. However, although Smith does not say so explicitly, it can be argued that human beings know from their past experience that the long-term tendency of this kind of action is to cause great harm in terms of human lives. We can imagine and sympathise with the outrage that many people feel at such behaviour. On the other hand, we can also sympathise with the resentment that the guard might feel towards those who want to put him to death, even though he has not actually killed anyone. But here the aggregative principle comes into play, which tilts our sympathy towards the social resentment rather than that of the sentinel, leading us to regard his condemnation as just.

In conclusion, Smith's theory of justice is fully capable of accommodating the role that the principle of utility plays in our practice of justice, both when it comes to reinforcing our judgments in order to make punishment effective, and when, even in the absence of actual harm, we administer justice on the basis of aggregative considerations.

Chapter 5

Justice, Genealogy and Utility in J.S. Mill

In this final chapter I examine Mill's theory of justice. As in the previous chapter, Mill's theory will be examined against the background of Hume's conception of justice. I will argue that, although Mill's conception of the content of justice is much broader than Hume's, he takes up three fundamental aspects of it. The first concerns the explanation of the emergence of the concept of justice, which takes up some crucial aspects of the genealogy examined in Chapter 2. The second concerns the role of utility, which, less explicitly present in Smith's conception of justice, now returns to play a crucial role in explaining the absolute character of the rules of justice. The third, finally, is about the motive of justice, which Mill identifies with a specific version of duty.

5.1 Our Common-sense Idea of Justice. Why it is an Issue for Mill

Chapter 5 of *Utilitarianism* sets out an examination of justice that is structured on five levels of analysis concerning respectively the phenomenology of the sentiment of justice, the different uses of 'just' and 'unjust', the etymology of the term 'just', justice as a set of perfect obligations, and finally, the analysis of the psychological components making up the sentiment of justice. Although shorter than the other four, the first level of analysis plays a crucial role in the complex argumentative strategy of Chapter 5 because it highlights the fundamental problem of justice that Mill seeks to solve in *Utilitarianism*.

This depends on the combination of two theses. The first, which Mill traces back to our 'common idea of justice', is that the value of just actions is independent of their consequences. The second, which depends on his utilitarian conception of morality, is that the value of just actions ultimately depends on their consequences for the well-being of human beings. These two theses are inconsistent, i.e. it is not possible for both to be true.

Why is this a problem for Mill? After all, the two theses are placed at a different level of inquiry. The first is a philosophical thesis which aims to identify the true justification of the value of justice. The second, on the other hand, is presented by Mill as merely describing the phenomenology of our common experience of justice, that is, a pre-reflective description that is not mediated by philosophical concepts. Why, then, is the contrast between these two theses a problem, in fact the very problem with which his treatment of justice begins?

In this section, I argue that there are at least three interrelated considerations in answering this question, which shed light not only on the nature of the problem Mill raises, but also, and more importantly, on the methodological assumptions and philosophical aims of his theory of justice.

A first general point to make is that the confrontation with what Mill calls our 'common conception of justice' is necessitated by his basic methodological commitment to a conception of ethics based on observation and experience. This approach was clearly at the heart of his attempt to prove his moral theory from the desiderability of happiness, as set out in Chapter 4 of *Utilitarianism*. It is at work again in Chapter 5, where Mill seeks to measure the acceptability of utilitarianism, this time by its ability to fit or explain our common-sense conception of justice.

Then there is a second consideration, which concerns the content of our common conception of justice. This idea, once articulated, seems to be explicable only by moral

intuitionism, that is, by an antagonistic metaethical conception of utilitarianism. The common idea of justice thus poses not only a general challenge to the observational method of utilitarianism, but also a very specific challenge to find an explanation of the phenomenology of justice that is better than what seems to be its most plausible explanation. Let us examine this aspect in more detail.

Mill argues that (i) our common idea of justice never fails to be associated in our minds with a peculiar sentiment. The phenomenology of this sentiment highlights two components: (i) a feeling, which Mill characterizes as having greater binding force than that aroused by mere utility, and (ii) a clear and immediate perception concerning the distinction between just and unjust. Regarding (i), Mill claims

> […] inasmuch as the subjective mental feeling of Justice is different from that which commonly attaches to simple expediency, and, […] is far more imperative in its demands, people […] think that its superior binding force requires totally different origin. (*UT* 5.2)[159]

In relation to (ii), Mill further argues that the sentiment of justice is similar to an instinct because of the immediacy with which we are to distinguish just from unjust. Focusing this time on what appears to be the cognitive component of this sentiment, he adds

> The powerful sentiment and the apparently clear perception, which that word [Justice] recalls with a rapidity and certainty resembling an instinct, have seemed to the majority of thinker to point to an inherent quality in things […]. (*UT* 5.1)

[159] The text I use is J.S. Mill, Oxford Philosophical Text of *Utilitarianism*, edited by Roger Crisp, Oxford and New York, Oxford University Press, 1998. This text is also that used in the *Collected Works* (Mill 1961-1991). All self-standing references (e.g. 5.2) are to chapters and paragraphs of *Utilitarianism*.

Although not explicitly expressed, Mill's idea seems to be that if it depended on other characteristics, such as its consequences, then our knowledge of justice would be mediated by a calculation of its consequences and would not be immediate.

Mill argues that the combination of (i) and (ii) leads common sense to believe that (iii) the feeling of justice corresponds to an objective and absolute property. He then writes:

> Mankind are always predisposed to believe that any subjective feeling, not otherwise accounted for, is a revelation of some objective reality. (*UT* 5.2)

Examining the possible consequences that this inclination of the human mind has on our common-sense conception of justice, Mill adds that 'justice or injustice of an action' can be regarded as 'a thing intrinsically peculiar, and distinct from all its [the action's] other qualities', and thus a *fortiori* distinct from the quality of being useful to something else.

> […] Just must have an existence in Nature as something absolute, generically distinct from every variety of Expedient, and, in idea, opposed to it […]. (*UT* 5.1)

The belief that there is an objective property of justice that is revealed by the feeling of justice is, in turn, the basis for the further belief that (vi) the judgments associated with the feeling of justice are 'infallible' and need not be controlled by our intellect (*UT* 5.2).

The combination of (i), (ii), (iii) and (iv) constitutes what we can call the realist common sense conception of justice. This holds that justice is a normative property of

actions that is not only distinct from the property of promoting pleasure, but also cannot be explained in terms of any other property. If we ask 'why does justice have value?', we cannot give an answer. So justice is an intrinsic property[160] of actions. This metaphysical view leads to an epistemological view according to which this kind of entity cannot be known in any other way than by intuition. Our experience of justice therefore seems to be consistent with the propositions of moral intuitionism, the main metaethical opponent of utilitarianism, which Mill had addressed in Chapter 1 of *Utilitarianism* and in the essays *Whewell on Moral Philosophy* and *Sedgwick's discourse*. In contrast to intuitionism, Mill had already argued that our ability to distinguish what is right from what is wrong can be explained without assuming a special faculty, i.e. 'moral sense' (SD 10.51), but by referring exclusively to our intellect and our senses.[161] Our common idea of justice, however, reactivates the conflict between intuitionism and utilitarianism. Although common-sense is willing to grant that the observance of justice promotes the well-being of human beings, this fact does not seem to be able to explain the way in which human beings experience its obligations, which seems instead to be more easily explained by intuitionism, which hypothesizes the existence of a sui generis moral faculty. While not explicitly mentioning moral intuitionism, Mill claims:

> For the purpose of this enquiry, it is practically important to consider whether the feeling itself, of justice and injustice, is *sui generis* like our sensations of colour and taste, or a derivative feeling, formed by a combination of others. *And it is the more essential to examine* [italics mine], as people are in general willing enough to allow, that objectively the dictates of justice coincide with a part of the field of General Expediency; but inasmuch as the subjective feeling of justice is different from that which commonly

[160]For this interpretation of the distinction between 'intrinsically valuable' and 'extrinsically valuable' as depending on whether or not it is possible to explain why an entity has value, see C. Korsgaard 1996: 108-109.

[161] On this aspect, see See Crisp 1997: 8. See also Cremaschi and Marcello 2006: 45 ff.

> attaches to simple expediency, and, except in extreme cases of the latter, is far more imperative in its demands, people find it difficult to see, in justice, only a particular kind or branch of general utility, and think its superior binding force requires a totally different origin. (*UT* 5.2)[162]

In conclusion, initial contradiction constitutes a problem for Mill not only because it creates a potential and generic obstacle to his conception of an experience-based ethics, but more precisely because it reignites the metaethical contrast between utilitarianism and intuitionism, apparently already resolved in the early parts of *Utilitarianism*. Mill rightly poses this problem at the opening of Chapter 5 since this highlights the strategic importance of the discussion of justice for the defence of his utilitarian metaethical conception.

Finally, there is a third consideration to be made about the nature of the problem under discussion. Even assuming that utilitarianism offers a better explanation than intuitionism of the value of justice there would remain the problem of the potential negative effects that the spread of this explanation would have on compliance with its rules. This potential risk, though not explicitly mentioned by Mill, is implied in the passage quoted a few pages above, when he argues that although 'there is no necessary connection between the question of its origin, and that of its binding force ... these two opinions are very closely connected in point of fact'. The idea is that human beings tend to believe that the 'binding force' of the feeling of justice as well as its ability to readily and reliably distinguish right from wrong depends on the fact that this feeling is aroused by a peculiar objective property of actions that is irreducible to others. i.e., the property of justice. The existence of this property is exactly what utilitarianism questions since it

[162] On the connection between 'general expediency' and respect for justice see also *UT* 5.1, particularly when Mill writes that people tend to believe that '[justice is] never, in the long run, disjoined from it [every variety of expedient] in fact'.

identifies justice with a quality that is instead reducible and explicable from other qualities of actions. This means that the utilitarian conception of justice, once it becomes public knowledge, could undermine both epistemic confidence in our ability to perceive justice and the overwhelmingness of its duties.

For these reasons, the idea of justice that emerges from our first-level common-sense judgments poses a problem for Mill's conception of ethics. Note the similarity between this starting point and Hume's. In T 3.2.1, Hume uses the Circle argument to show that our common-sense conception of justice as a virtue conflicts with part of his meta-ethical conception, i.e. his philosophical theory of virtue. Indeed, the combination of the thesis that justice is a virtue with an unquestionable maxim creates a circular argument. In this chapter, I will argue that the two authors share more than just the starting point of the investigation of justice. Like Hume, Mill shows how the problem can be solved by a genealogical explanation of the emergence of justice. Moreover, like Hume, Mill argues that the common-sense notion of justice is justified precisely by his philosophical conception of ethics as a set of virtues and norms that promote human well-being.

5.2 The Elements of Justice

The philosophical problem posed in *UT* 5.1-2 is approached through different kinds of inquiry, first identifying different kinds of just actions and circumstances, then the etymology of the term just, and finally a new analysis of the sense of justice, this time aimed at discovering its primary components. Mill intersperses the second and third types of inquiry with a brief discussion of the distinction between perfect and imperfect obligation, which is not clearly related to the other two. Let us therefore consider the various aspects of Mill's enquiry separately, beginning with an examination of the

'various modes of action' and 'arrangements of human affairs' to which the terms 'just' and 'unjust'(*UT* 5.3)) are applied in common parlance, and see whether there is a common meaning to the different uses.

Using contemporary terminology, we might characterise this operation as one in which, starting from an extensional definition of justice that lists just objects, we seek to identify properties common to these objects that would allow us to provide an intensional definition of justice, i.e. conditions under which an utterance containing this concept is true. However, as Mill himself acknowledges, this first attempt to solve the initial problem of justice comes to nothing, because the various kinds of objects we call just do not turn out to share any fundamental property governing the uses of 'justice' that is capable of resolving the internal instability of the common-sense idea of it.

However, this survey is interesting for two main reasons. First, it highlights the breadth of this notion through the identification of what Roger Crisp has called the different ' "spheres" of justice' [163]. In particular, in relation to the limited content of justice as described by Hume in the *Treatise of Human Nature*, Mill's 'spheres' show that the term 'injustice' turns out to have a variety of reference, which he identifies in turn with the violation of legal rights (such as property and personal liberty), the disregard of moral rights, the failure to give us what we deserve, the violation of an expectation that we have also implicitly created, partiality in contexts that require impartiality, and, finally, inequality.[164] Second, the survey is instructive because, although it does not reveal the common property that Mill is seeking, it brings in some features of justice that phenomenological inquiry had not captured. Mill makes it clear that the property underlying justice must be something that gives rise to duties that may not be codifiable

[163] See Crisp 1997: 157.

[164] Note that Mill switches several times from examining the meaning of unjust to that of just, although at the beginning of his survey of spheres of justice he explicitly states his preference for investigating the negative notion of injustice because it is easier to examine (*UT* 5.3).

by precise rules. In fact, justice encompasses an area of behaviour that is *not always* governed by rules. This is made clear by examining the use of the terms 'just' and 'unjust' in the sphere of justice as impartiality. After discussing cases of duties of justice that are codified (the judge who applies the laws of the state, an evaluator of an open competition, etc.), Mill argues for other cases in which the duty of impartiality typically concerns circumstances, such as the relationship between parents and children, in which the idea of following rules seems out of place (*UT* 5.2): in these cases, the duty of impartiality has to do with the exercise of a set of capacities, such as the scrupulous examination of circumstances, reflection on the welfare of one's offspring, and doing one's best, which cannot be reduced to the application of codified rules.

Having further characterised the meaning of justice at the level of our common language, let us now examine the third level of Mill's investigation, namely what he calls the etymology of 'justice'.

5.3 Is Mill's Account of Justice a Sort of Genealogy?

Mill continues his account of justice through three further levels of inquiry. First, he examines the etymology of 'justice' (*UT* 5.11-13). He then characterises the duties of justice in terms of perfect duties and rights (*UT* 5.14-15). Finally, he returns to the sense of justice by examining its various components (*UT* 5.16-23). In this section, I intend to argue that each of these three elements are aspects of a single explanation that can be described as a form of genealogy of justice that places Mill in the wake of Hume. In the first part of the section, I will look specifically at the etymology of the idea of justice, while in the second part I will concentrate on examining the sentiment that Mill associates with this idea.

5.3.1 Mill's Etymology of Justice

The purpose of Mill's etymology is to identify the quality common to different modes of just behaviour, starting from the assumed original meaning of 'justice'. Mill's hypothesis seems to be that if we understand the original idea of justice, then, by constructing a narrative sensitive to historical circumstances, we can follow the 'progressive growth' of that 'idea' to what is believed in advanced societies. This would allow us to identify what Mill calls the 'idea of justice' (*UT* 5.16), that is, the quality common to the different spheres of justice described in *UT* 5.4-10. Having articulated the stages that make up the etymology, I argue that the distinction between perfect and imperfect rights that concludes this inquiry is not an alien concern, but is integral to the quasi-historical narrative that makes up the etymology.

Mill divides his narrative into three stages. The first identifies the original idea of justice as that which conforms to the law and places it in the historical interval from Judaism to early Christianity. In this stage, law is seen as (a) a set of rules coming directly from God, which means that there can be no unjust laws, and (b) law is the criterion of justice. A second stage, which Mill places in the period from Hellenistic Greece to Roman times, is characterised by a new belief that contradicts (a): law is a human product and as such fallible. Following this new belief, the concept gradually evolves and justice becomes that which conforms to 'the laws as they ought to be' (*UT* 5.12). Finally, the third stage, which describes the concept of justice in advanced societies, is one in which conformity to the law 'as it ought to exist' is present in a 'modified shape' (*UT* 5.13): there may be instances in which it is preferable that this law not become the law of the state, even if its violation results in justified punishment.

Through a historically sensitive narrative of the conceptual transformations of the notions of law and punishment, Mill offers an initial characterisation of the quality of

justice, which he identifies with the domain of obligatory action. Note that this is not simply a restatement of the argument made at the beginning of the chapter (*UT* 5.1-3), since it is now supplemented by the notion of punishment. Just actions are those that society can require of people who deserve legal or social punishment if they do not behave as they should. But this characterisation is still partial. This becomes clear in the light of the new comparison that Mill draws in the next paragraph between the spheres of justice and morality, specifically moral duty, which he defines as (a) that which can be required of persons who (b) deserve the punishment of public disapproval (and their conscience) if they do not behave as they ought. Mill observes that if the common quality of just action were merely obligatory, there would be no way of distinguishing moral duty from the duty of justice. The point is instructive: the common quality that Mill seeks is not simply something that must allow one to distinguish justice from the general sphere of expediency, as was the case in the opening sections of Chapter 5, but must also allow one to identify the specific difference between justice and moral duty.

To overcome this difficulty, Mill uses the familiar distinction between imperfect duties and perfect duties, identifying moral duties with the former and duties of justice with the latter. The idea is that duties of justice are perfect because, unlike moral duties, they do not leave the agent free to choose the circumstances and the way in which they are to be fulfilled, but specify certain kinds of action towards certain persons who have a claim or, more precisely, a right to demand that the agent should take that action. We might now ask: is this argument something that depends on the etymological investigation that Mill has made so far, or is it instead an independent philosophical argument that Mill merely derives from natural law theory?

On first reading, one might be tempted to argue for the latter. Indeed, Mill makes the distinction by referring it back to the language of 'ethical writers' (*UT* 5.15) or the precise language of 'philosophical jurists' (*UT* 5.15). This suggests that the distinction

between duties depends on a philosophical conception and cannot be traced from etymology to the terms of our common language. There are, however, two passages that stand in the way of this interpretation.

In the first, Mill argues that the distinction between simple moral wrongness and injustice, which starts from the notion of law, is part of the popular conception of justice and not merely the technical knowledge of jurists or philosophers. Shortly after introducing the distinction between perfect and imperfect rights, Mill claims:

> In our survey of the various *popular* acceptations of justice, the term appeared generally to involve the idea of a *personal right* - a claim on the part of one or more individuals, like that which the law gives when it confers a proprietary or other legal right. Whether the injustice consists in depriving a person of a possession, or in breaking faith with him, or in treating him worse than he deserves, or worse than other people who have no greater claims, in each case the supposition implies two things- a wrong done, and some assignable person who is wronged. (*UT* 5.15, my italics)

The connection between justice and rights is actually at the level of our common language, and takes the form of the belief that in order to identify injustice, as opposed to what is required to identify immorality in general, it is necessary to identify not only the content of a duty (a particular behaviour, such as the fulfilment of a promise), but also the subject who has a claim, a right to have a particular person or group of persons (duty bearers) fulfil that duty. As Sumner has argued, this kind of right, which we might call a claim-right, implies a normative relationship between A's claim to receive X and B's duty to give X to A.[165] This structure emerges clearly from Mill's example that the injustice of treating someone 'worse than he deserves' consists not only in the duty to treat someone

[165] On this aspect, see Sumner 2005:184-198.

'as he deserves' but also in the idea that a subject, the holder of the right, has a justified claim to demand punishment from the one who has violated the duty to treat him as he deserves.

The second relevant passage, however, concerns the connection between moral right and unjust laws and is prior to the distinction between perfect and imperfect rights. Mill argues that it is part of the 'universal or widely spread opinion' on the uses of 'Just' and 'Unjust' (*EU* 5.4) to hold that when we consider a state law unjust, we think it is so because it violates some right of someone which, since it cannot be legal, is called a 'moral right'. He claims:

> When, however, a law is thought to be unjust, it seems always to be regarded as being so in the same way in which a breach of law is unjust, namely, by infringing somebody's right; which, as it cannot in this case be a legal right, receives a different appellation, and is called a moral right. We may say, therefore, that a second case of injustice consists in taking or withholding from any person that to which he has a *moral right*. (*UT* 5.6, my italics)

Taken together, the two passages help us resolve the question posed just above because they indicate that the characterization of justice in terms of rights is dependent on etymology. Mill argues that the notion of right, construed as an individual's claim to receive a certain treatment that another person has a duty to perform, is a constitutive part of our popular notion of justice. This means that it is part of the meaning of our common notion of injustice that it consists in someone's violation of an individual claim who deserves punishment because of the violation of that claim.

What kind of punishment does the offender deserve? In some circumstances, of course, it will be a legal sanction, since the right is established by state law. But in other

circumstances, and this is Mill's interesting point, it will not. In those situations, either because there is no state law yet protecting that right, or because it is good that the matter is not regulated by law, the violated right justifies only public disapproval. This means that, in order to give a true account of the common conception of justice, we need the notion of moral right, which has a precise location in Mill's etymology: it is intelligible only from the second stage of his narrative, in which justice is associated with the notion of 'laws as ought to exist'. The staged narrative of which the etymology is composed thus makes it possible not only to characterise just actions as dutiful, but also to characterise those duties as corresponding to individual claims on the part of individuals, the violation of which is linked to a sanction even in the absence of civil laws.

Having characterized the Millian investigation of the *idea* of justice, let us now examine the second aspect of the notion of justice, namely the *feeling* associated with that idea.

5.3.2 The Sentiment of Justice and its Components

The investigation of the sentiment of justice has not received much attention in Millian scholarship. In his seminal article on justice in Mill, for example, David Lyons barely mentions it, arguing that the philosophical core of Mill's theory lies in etymology, which addresses the crucial question of the relationship between rights and justice.[166] This approach is biased, to say the least, because it tends to obscure the naturalistic (and Humean) dimension of Mill's account of justice. I argue that this aspect becomes visible

[166] See Lyons 1994: 67 f. In contrast to this line, though only incidentally, Barry S. Clark and John E. Elliott mentioned the importance that the feeling of justice, especially when supported by public education, has in the development of high human capacities. See Clark and Elliott 2001: 488.

once we highlight the methodological similarities between etymology and the analysis of the sentiment of justice.

Mill begins by asking whether this sentiment is provided by nature or develops from other principles. He claims:

> Having thus endeavoured to determine the distinctive elements which enter into the composition of the idea of justice, we are ready to enter on the inquiry, whether the feeling, which accompanies the idea, is attached to it by a special dispensation of nature, or whether it could have grown up, by any known laws, out of the idea itself; and in particular, whether it can have originated in considerations of general expediency. (*UT* 5.16)

The expression 'we are willing to' is not accidental and is intended to tie this part back to the question left unanswered in *UT* 5.3. At the beginning of Chapter 5, Mill had already examined the feeling of justice with the intention of explaining its phenomenology so that it would not be an obstacle to his utilitarian conception of morality. The investigation had reached an impasse because he had been unable to decide whether it was a peculiar feeling, similar to that of secondary qualities, or whether it was instead a derivation of that of utility. Mill had therefore undertaken the examination of the quality of justice with the intention of returning to the original problem. He now resumed the dilemma in the same terms as he had left it in *UT* 5.3, asking whether the sense of justice depends on a 'special dispensation of nature', that is, on a first principle analogous to that which explains the perception of secondary qualities, or whether it depends instead on a more general principle such as that of general expediency. However, while the terms of the problem remain the same, the object of inquiry changes significantly from *UT* 5.1-3.

The sense of justice now under discussion is no longer a sense of a particular binding force, but a desire to punish the perpetrator of an injustice. Mill claims:

> We have seen that the two essential ingredients in the sentiment of justice are, the *desire to punish a person who has done harm*, and the knowledge or belief that there is some *definite individual* or individuals to whom harm has been done. (*UT* 5.18)

The result of the etymology determines the point of view from which the feeling associated with the idea of justice is observed. Now the perspective is no longer that of the agent describing the phenomenology of the feeling that accompanies his or her perception of justice, but that of the observer describing the reactive feeling about the injustice before him or her. Note the effect of discussing injustice in terms of the violation of a perfect right in the passage quoted. Mill does not associate harm with the simple violation of a general moral obligation (imperfect right), but with an obligation to act in a certain way towards certain people (perfect right). The belief that there is harm or injury, as Mill will later say, echoing Hume, to certain persons is indeed a condition for the disapproval of injustice to be aroused in the mind of the spectator.

Let us now consider the naturalistic aspect of this enquiry. Mill tries to explain the origin of this feeling in terms of more basic components of our psychology, which are common to the animal world. He claims:

> Now it appears to me, that the desire to punish a person who has done harm to some individual is a spontaneous outgrowth from two sentiments, both in the highest degree natural, and which either are or resemble instincts; the impulse of self-defence, and the feeling of sympathy.
>
> It is natural to resent, and to repel or retaliate, any harm done or attempted against ourselves, or against those with whom we sympathise. The *origin of this sentiment it is*

> *not necessary here* to discuss. Whether it be an instinct or a result of intelligence, it is, we know, common to all animal nature; for every animal tries to hurt those who have hurt, or who it thinks are about to hurt, itself or its young. (*UT* 5.19-20, my italics)

The passage is very dense and contains several lines that need to be distinguished. Mill's starting hypothesis is that the desire to punish a person who has violated someone's right, i.e. the sense of justice, can be explained by the desire to take revenge for some harm done to us or to those with whom we sympathise. The idea is based on the plausible assumption that a desire caused by a narrow class of actions, i.e. the violation of an interest in having a certain right respected, can be explained by a qualitatively similar desire caused by a broader class of actions, ideally including the narrow one. If this assumption is correct, and Mill believes it is, then the more general desire for revenge takes explanatory priority over the specific desire for revenge caused by the perception of unjust acts. This is an example of Mill's first naturalistic commitment. Following Hume's methodological naturalism, Mill shows how the feeling of justice does not require an ad hoc explanatory principle, but can be explained in terms of a more primitive passion that is intelligible independently of it. To this naturalistic commitment is added a second, more substantial one, which depends on the empirical thesis, introduced in the passage quoted above, that primitive resentment is a passion common to the animal kingdom. Not only can the feeling of justice then be explained in terms of other, simpler psychological principles, but these explain human behaviour in continuity with the rest of nature.

The function of the naturalist explanation is thus to rule out as implausible the first option of the dilemma about the origin of the feeling of justice, thereby making the second option, which connects this feeling to utility, viable. Note that in order to fulfil this function, Mill does not have to commit himself to solving the further question of whether or not the original desire for revenge is an instinct. The important point is that the

sentiment of justice is not an instinct, i.e. that unlike what we believe happens with our experience of secondary qualities, that sentiment is explicable out of other, more basic principles.

At this stage, Mill is finally ready to explore in greater detail the connection between justice and utility, hitherto only mentioned unsystematically. This is done by examining the transition between natural resentment and moral resentment, which is facilitated by the ability to progressively expand the scope of our self-interest. In particular, the transition is determined by two psychological features that distinguish humans from non-human animals. Mill claims:

> Human beings, on this point [the capacity to resent], only differ from other animals in two particulars. First, in being capable of sympathising, not solely with their offspring, or, like some of the more noble animals, with some superior animal who is kind to them, but with all human, and even with all sentient, beings. Secondly, in having a more developed intelligence, which gives a wider range to the whole of their sentiments, whether self-regarding or sympathetic. By virtue of his superior intelligence, *even* apart from his superior range of sympathy, a human being is capable of apprehending a community of interest between himself and the human society of which he forms a part, such that any conduct which threatens the security of the society generally, is threatening to his own, and calls forth his instinct (if instinct it be) of self-defence. (*UT* 5.20)

Because of their increased capacity for sympathy, humans are able to feel vicarious pain not only for the harm suffered by their loved ones or those who have benefited them, but also for any sentient being capable of feeling pain. Human sympathy is therefore a development of animal sympathy, in that it incorporates its characteristics and transcends its limits. But it is still not enough to moralise animal resentment. On its own, it can only generate an extended desire for revenge, which, unlike animal resentment,

is sensitive to the harm done to people far away from us, but it cannot transform this desire into the selective passion that reacts only to the harm caused by the violation of moral rights. For extended resentment to become moral, it requires the intervention of a second quality, which Mill calls intelligence, and which he describes as the capacity to grasp the common interests between ourselves and those who constitute the community of which we are a part. Like sympathy, this capacity does not mark a qualitative separation from the rest of the animal world, and is human only insofar as it is 'superior', that is, more developed than our non-human fellows. As a result of this rational capacity, we gradually come to understand that there are common interests between members of the same community. This leads to a broadening of the sources of our self-interest: we become aware that it is in our interest to act not only for the desires we have as individuals and separate from others, but also for those we have as part of a community. Through intelligence, we realise that we have reasons to avoid acting in ways that threaten the security of society, and in accordance with this understanding, we transfer the desire for self-defence and revenge also to those behaviours that harm us indirectly through the damage they cause to society.

The emphasis on intelligence could give the impression that it plays an explanatory role over and above that of sympathy. In fact, the latter seems to be limited to making animal resentment less partial, which nevertheless remains activated by any kind of harm. Intelligence, on the other hand, makes it possible, even without sympathy, to grasp those common interests to which the desire for revenge is linked, and to turn them into a sense of justice. In this way, the moralisation of revenge would depend on purely prudential considerations. But this hypothesis is implausible. In fact, shortly afterwards, Mill argues:

> The same superiority of intelligence joined to the power of sympathising with human beings generally, enables him to attach himself to the collective idea of his tribe, his country, or mankind, in such a manner that any act hurtful to them, raises his instinct of sympathy, and urges him to resistance. (*UT* 5.20)

The mere knowledge of the existence of an instrumental link between the interests of society and my own self-interest is not enough to activate the feeling of moral blame when the rules of justice are violated. Rather, this feeling arises from our sympathy for the painful consequences for society of the violation of the rule. Mill, like Hume, formulates an explanation of the emergence of justice that uses animal passions and feelings as explanatory factors. Although he does not distinguish, as Hume clearly does, between the natural and moral obligations of justice, he does not believe that either self-interest or natural resentment can guarantee respect for justice. What guarantees that respect is the understanding that justice is the set of rules that protect what Mill calls 'the essentials of human well-being' (*UT* 5.33), towards which our impartial sympathy is directed. It is precisely impartial sympathy for this well-being that enables us to develop a 'sense of repugnance' towards injustice, which is the surest safeguard for our stable adherence to the rules of justice.

This concludes and resolves the problem that Mill raises in the opening sections of Chapter 5. Not only does utilitarianism not contradict the phenomenology of our experience of justice, it explains its main features. Moreover, the very way in which it explains them cannot but in turn recommend the rules of justice anew to impartial human sympathy.

5.4 The Indefeasible Character of the Rules of Justice and the Corrective Role of the Principle of Utility

We have examined how Mill's rules of justice secure the possession of goods which are the very conditions of human well-being. They stand at the highest level of the scale of social utility and are therefore, as Hume famously argued (*T* 3.2.1.15/SBN 482-3)[167], inviolable. Mill's position, however, is not as straightforward as it might at first appear. Indeed, it needs to be reconciled with two additional aspects of his theory of justice. The first concerns the fact that Mill believes that there are conflicts between principles of justice and that their resolution may lead to the misapplication of one of them (*UT* 5.28-31). The second, however, concerns conflicts between justice and other moral duties which, in extreme cases, require the triumph of the latter over the duties derived from the principles of justice (*UT* 5.37). In this and the next section I will show how Mill uses the principle of utility to resolve these conflicts. This function of utility is not without precedent. He had clearly argued in *Utilitarianism*, Book 2 that 'if utility is the ultimate source of moral duties, it may be invoked to decide between them when their demands are incompatible' (*UT* 2.25). Moreover, he had illustrated this by the relationship between utility and secondary principles of morality. However, it is the fact that Hume uses this function of utility specifically for justice that makes these cases interesting, since this is intertwined with the theme of the inviolability of the rules of justice. In this section I will examine how Mill uses the first principle of utility to resolve these conflicts, and examine the sense in which the rules of justice are nonetheless inviolable. Examining this issue is important not only for a clearer understanding of Mill's theory, but also because it opens

[167] Hume describes them more precisely as 'inviolably by the laws of society' (*T* 3.2.1.15/SBN 482-83). However, as I will argue further there is not much difference on this point between Hume and Mill.

up important further ground for comparison with Hume's theory of justice, which is the main focus of this chapter.

5.4.1 Internal Conflicts within the Principles of Justice

Mill introduces the topic of internal conflicts in justice in *UT* 5.27, where he claims that this is composed of different principles from which conflicting practical maxims can be deduced. He claims:

> We are continually informed that Utility is an uncertain standard, which every different person interprets differently, and that there is no safety but in the immutable, ineffaceable, and unmistakable dictates of justice, which carry their evidence in themselves, and are independent of the fluctuations of opinion. […] Not only have different nations and individuals different notions of justice, but in the mind of one and the same individual, justice *is not some one rule, principle, or maxim, but many*, which do not always coincide in their dictates, and in choosing between which, he is guided either by some extraneous standard, or by his own personal predilections (*UT* 5.27, my italics).

Note that Mill does not simply argue that different moral communities hold different principles of justice. In other words, Mill's view does not fall within what Gilbert Harman has called relativism as a descriptive thesis.[168] His position is that the standard of justice is uncertain because, whatever the differences in moral beliefs between different moral

[168] See Harman and Thomson 1996: chap 1.

communities, within the same community justice is made up of different principles that are potentially in conflict with each other.[169]

Mill illustrates this theme by discussing three applied cases, namely the principles governing just punishment, those governing the just distribution of wages, and finally those regulating the just distribution of taxes. It may be instructive to consider the first of these.

He lists three common views of the circumstances in which punishment is just. The first is that it is just only if it is carried out for the purpose of re-educating the person who has done wrong (*UT* 5.28). The second is that punishment is just only if it minimises the harm to others, that is, prevents them from possible future harm (ibid). The third, finally, is that no punishment is ever just, since we are not free to perform the actions we do, since they are the product of our character, which is never fully under our control (ibid). These views could, of course, give rise to contradictory maxims: the strategy we consider appropriate for re-educating the offender may not be identical to the strategy we believe to be appropriate for preventing him from committing the same act in the future. For example, if I believe that the only purpose of punishment is to ensure that the offender does not repeat behaviour that is harmful to others, I might go so far as to think that a treatment that guarantees this outcome one hundred per cent is appropriate, even at the cost of completely disregarding the offender's quality of life. For example, I might think that the so-called 'Ludovico treatment' that Alex, the protagonist of Burgess's novel *A Clockwork Orange*, has to undergo is a just punishment. Even if the offender has lost the

[169] The claim that justice is a department of morality dealing with potentially conflicting principles is not surprising and can be traced back to the very general notion of morality that Mill sketches in Chapter 1 of *Utilitarianism*, particularly his conception of moral epistemology. Mill argues that our moral faculty is not able to discern between what is right and wrong in particular cases, but guides us through 'general principles of moral judgment' (*EU* 1.3) out of which we infer our moral obligations and judgments concerning the conduct of others. In Chapter 1, Mill had also importantly added that since these conflicts are inevitable no moral theory can avoid identifying a first moral principle, or at any rate a rule capable of establishing an 'order of precedence' among principles, capable of resolving them (*EU* 1.3). With the discussion of principles of justice in Chapter 5, Mill thus illustrates this important issue at the heart of the theoretical project sketched at the beginning of *Utilitarianism*.

ability to choose not to commit evil deeds, but merely refrains from doing them to avoid unbearable physical pain, the very fact that the treatment ensures that he will not commit crime again provides justification that the punitive treatment is just. This view of just punishment obviously contrasts with the view that punishment is only just if it is administered for the purpose of re-educating the offender. In this case, the restriction of liberty will be just if it is necessary to bring the offender to understand why the action for which he is being punished is wrong, so that this understanding will feed into the decision-making process that leads him to refrain from such actions in the future. Both of these ideas, then, contrast with the idea that since the offender is not responsible for his evil character, he is not punishable for what he has done.

Mill's important point is that these three conceptions result from as many principles that are constitutive of justice each of which is true. Mill writes:

> All these opinions are extremely plausible; and so long as the question is argued as one of justice simply, without going down to the principles which lie under justice and are the source of its authority, I am unable to see how any of these reasoners can be refuted. For in truth every one of the three builds upon rules of justice *confessedly true*. […] Each is triumphant so long as he is not compelled to take into consideration any other maxims of justice than the one he has selected; but as soon as their several maxims are brought face to face, each disputant seems to have exactly as much to say for himself as the others. No one of them can carry out his own notion of justice without trampling upon another equally binding. (*UT* 5.28, my italics)

The conflict is thus internal to justice and is generated by the conjunction of some of its secondary principles. But how exactly is utility supposed to make it possible to resolve internal conflicts in justice? By intervening at the level of general rules, refining them or replacing them with others, or by intervening at the level of particular cases,

deciding on a case-by-case basis and allowing us to choose a new maxim that is directly deducible from the principle of utility, bypassing the level of general laws?

In Chapter 5, Mill provides little guidance in this regard, arguing only that 'social utility alone can decide preference' (*UT* 5.30). More guidance can be found, however, if we place this question in the context of his more general discussion of the interplay between secondary principles and the first principle of morality in Chapter 2 of *Utilitarianism*.

Secondary principles are defined by Mill as those rules that allow us to apply the first principle of utility in our everyday choices (*UT* 2.24). This is necessary because if we were to be guided directly by the principle of utility in choosing which actions to take, moral action would require an excessively laborious activity of calculation and comparison, and would be subject to inevitable error. Moral rules make it possible to avoid these difficulties because they are the distillation of shared human experience about the consequences of broad classes of actions for human well-being. Through private and public education, these rules are internalised by people and become simple and immediate guides to behaviour, enabling them to act in accordance with the utilitarian end without making long, complicated and uncertain calculations (*UT* 2.24). Mill is keen to point out that, although we learn from childhood to regard these rules as something 'sacred' (*UT* 2.23), like any other rule they allow for exceptions, the identification of which depends on the principle of utility.

Let us consider how the principle of utility can legitimately fulfil this function. Regarding the internal conflict in justice between the secondary principle of telling the truth and the principle of giving everyone what he or she deserves, Mill argues:

> Yet that even this rule, sacred as it is, admits of possible exceptions, is acknowledged by all moralists; the chief of which is when the withholding of some fact (as of information

> from a malefactor, or of bad news from a person dangerously ill) would preserve some one (especially a person other than oneself) from great and unmerited evil, and when the withholding can only be effected by denial. But in order that the exception may not extend itself beyond the need, and may have the least possible effect in weakening reliance on veracity, *it ought to be recognised, and, if possible, its limits defined*; and if the principle of utility is good for anything, it must be good for weighing these conflicting utilities against one another, and *marking out the region within which one or the other preponderates*. (*UT* 2.23, my italics)

This passage points to a first feature of the way in which the utility principle is authorised to intervene in the resolution of conflicts between conflicting (secondary) principles of justice. The restriction on the application of the secondary principle must never undermine the collective confidence in the moral value of the principle on which the correction operates, in this case that of truthfulness. A plausible interpretation of this constraint might be that not only must the exception be clearly and publicly justified on the basis of the utility principle, but that the exception must be case specific, i.e. it must identify a repeatable general circumstance on the basis of which a sub-rule specifying the exception to the general principle can be constructed (e.g. 'the principle of truthfulness does not apply when lying is the only way to save a human life'), so that the exception is not perceived as a violation of the rule and thus weakens our confidence in its general observance.

The use of the utility standard in conflicts between secondary principles of justice thus seems to involve a reflexive process that concerns the refinement of secondary principles. This highlights a second aspect, which concerns the qualities of the agents engaged in this kind of reflection. Mill argues that the resolution of moral conflicts through the principle of utility involves being vigilant against falling into forms of 'self-deception' (*UT* 2.25) that lead us to exchange general utility for forms of 'partialities'

and ‘personal desires’ (*UT* 2.25), and he characterises these agents as morally responsible and virtuous. These characteristics constitute a psychological prerequisite that enables agents engaged in conflict resolution to resolve it by adopting a general perspective that transcends their particular interests and allows them to imagine solutions that are applicable to future conflicts involving the welfare of others.

Before moving on to Mill’s analysis of justice, a remark on this point is in order. Mill’s reference to virtue cannot fail to evoke Aristotle’s discussion of ‘decency’ in Book V of the *Nicomachean Ethics*, where Aristotle claims, with regard to justice in general, that there is a kind of excellence which goes beyond mere observance of the law and consists in the ability to correct it (*NE* 137 b13-16). Because the law is universal, Aristotle argues, it may not be right in some particular circumstances (NE 137 b13-16). ‘Decency’ allows us to deal with this difficulty excellently, since it is the quality that enables us to imagine what the legislator would have chosen if he had been faced with the particular circumstance in which we find ourselves, and from this knowledge to correct the general law (*NE* 1137 b22-24). Needless to say, the meaning of virtue in Aristotle’s discussion of justice is much broader than that of the mental quality of ‘decency’. Like Aristotle, however, Mill argues that the application of justice requires a virtuous quality that allows us to adopt a general perspective, similar to that of the Aristotelian legislator who aims at the good of the citizens, from which to correct the principles of justice by adapting them to particular circumstances.

Chapter 2 provides a framework for integrating the brief remarks in Chapter 5 on how the principle of utility intervenes to resolve the inevitable conflicts that characterise the everyday practice of justice. As Chapter 2 shows, Mill's idea is not to prefer one principle to another merely on the basis of the consequences of adopting that principle in a particular circumstance; more precisely, to examine the consequences of the particular action that principle leads us to take. Nor is it to ignore conflicting principles and decide

what action to take by directly examining the consequences of possible courses of action. Rather, Mill seems to be describing the role of the utility principle as one that leads to refinement of the rules in play, taking care not to undermine confidence in the authority of those rules and in their being followed by all. This implies both that the utility principle intervenes in the rules themselves, and that the agents involved in this process do so in a virtuous and morally responsible way.

5.4.2 The Conflicts Between Obligations of Justice and Imperfect Moral Obligations

So far, we have looked at Mill's description of the nature of conflicts within justice and how the principle of utility can help resolve them. But what happens when the conflict is between the rules of justice on the one hand and the imperfect obligations of morality on the other? Mill addresses this issue in passing in the conclusion of Chapter 5, posing the question of whether it is permissible to steal or deprive someone of liberty by kidnapping him for the purpose of saving someone's life. After reiterating that the rules of justice are placed at the highest level of social utility, Mill poses the following problem:

> […] particular cases may occur in which some other *social duty* is so important, as to *overrule* any one of the *general maxims of justice*. Thus, to save a life, it may not only be allowable, but a *duty*, to steal, or take by force, the necessary food or medicine, or to kidnap, and compel to officiate, the only qualified medical practitioner. In such cases, as we do not call anything justice which is not a virtue, we usually say, not that justice must give way to some other moral principle, but that what is just in ordinary cases is, by reason of that other principle, not just in the particular case. By this useful accommodation of language, the character of *indefeasibility* attributed to justice is kept up, and we are saved from the necessity of maintaining that there can be laudable injustice. (*UT* 5.37)

Unlike the conflict involving principles of justice, Mill now explicitly argues that in these cases there is a duty to obey the moral obligation versus what justice would require one to do. This is probably because of the urgency of saving a human life versus, for example, the conflict between ways of justifying punishment for an offender. My focus, however, is on the similarity between these cases and cases of conflict within the justice system.

Consistent with the importance of rules in his moral theory, Mill argues that the duty to comply with the rules of justice is an obligation that stands even when we have an obligation to save a human life. This is because there is no real tradeoff between the positive consequences of saving a human life and the negative consequences produced by weakening trust in compliance with a rule of justice. Indeed, stable compliance with these rules generates a kind of social utility that is not commensurable with that generated by the immediate consequences of a single action, no matter how quantitatively large they are. Therefore, although we have an obligation to act to save a human life we must do so with the other kind of obligation in mind as well. How? Mill takes exactly the line already taken for internal conflicts of justice. It is again a matter of producing a sub-rule that circumscribes the scope of the rule of justice, so that action *prima facie* describable as a violation of a rule turns out to be a justifiable exception to that rule.

Having discussed the two types of conflicts involving justice, we are now in a better position to specify the meaning of indefeasibility that Mill considers appropriate to it. Justice obligations are indefeasible in the sense that, considered in relation to the practice they support, they are not violable. Indeed, these practices produce a type of well-being that is placed at the highest level of the social scale (UT 5.37).[170] This, however, is

[170] Mill on this point claims: "It appears from what has been said, that justice is the name for certain moral requirements, which, regarded collectively, stand higher in the scale of social utility, and are therefore of more paramount obligation, than any others […]" (*UT* 5.37).

not incompatible with the fact that these rules can be improved as human experience about the sources of our well-being progresses, and that part of this improvement also consists in identifying new circumstances to which these rules do not apply. The indivisibility of the rules of justice is thus compatible with the fact that some of them must be sacrificed, both when they conflict with other rules of justice and when they conflict with obligations, such as social obligations, that are external to justice.

Is this form of indefisibility different from the inviolability that Hume holds for the rules of justice in Book 3 of the *Treatise*?

Let us briefly recall Hume's view set forth in *T* 3.2.1. Hume argues that the rules dealing with property, i.e. rules of justice, are held to be 'sacred and inviolable' (T 3.2.10.15/SBN 562-3; see also *T* 3.2.2.27/SBN 501 and *T* 3.2.6.10/SBN 533). This claim had been preceded by an argument showing why natural motives are not the characteristic motives of the virtue of justice. Specifically, he had argued that were it so, we would have no reason to act justly in a variety of circumstances, such as demanding repayment of a loan when a person is in great financial difficulty or returning money to a person who would use it to harm himself (*T* 3.2.1.13). We might ask: does this discussion prove that duties of justice, unlike in Mill, must always triumph over considerations concerning the good of particular people? According to Michael Frazer, the answer is positive. The way in which Hume introduces the theme of inviolability serves to illustrate an important feature of the concept. According to Fraser, inviolability means that the impartial duties derived from the rules of justice must never yield to practical considerations that depend on the possession of certain natural virtues, such as benevolence or compassion.[171] This interpretation, in its crudeness, can be misleading because it implies that the duties of justice must always be fulfilled, whatever the cost. First, Hume is not as interested as Mill

[171] See Frazer 2010: chap 3.

in discussing the problem of possible moral conflicts arising from the application of justice. That said, the admiration we pay to the trait of justice does not depend on the fact that it motivates us to obey the rules of justice at all times. We admire it because it motivates us, moved by a non-instrumental respect for justice, to perform acts that support a pattern of actions that is highly beneficial to society. We approve of the character trait of a just person because we impartially sympathise with the general welfare of society, which is indirectly promoted by that trait. This is entirely consistent with the fact that – *pace* Frazer - we do not cease to approve a character as just because he sometimes acts benevolently against a duty of justice. This does not mean, however, that justice, understood as a set of rules, can be sacrificed to any principle about how a natural virtue should be applied to particular circumstances. In this sense, the rules of justice are inviolable: they cannot be violated for the sake of other moral principles.

This way of proceeding is not very different from Mill's. Although he does not discuss the question of individual character in the examples given, he never questions the inviolability of the rules of justice. He does, however, admit that in exceptional circumstances they may be disregarded in favour of duties that are not part of the rules of justice.

5.5 The Virtue of Justice and the Motive of Duty

Moral virtue is identified by Mill with a set of motives that are traceable to concern for the welfare of others (*UT* 2.19) and that have a tendency to cause right actions, i.e. actions that are morally appreciable out of the utilitarian first principle. From the metaethical point of view, the moral value of actions depends not on that of the motives that cause them but on the consequences they bring about for the welfare of the people involved (*UT*

2.19). Actions can thus be morally right or wrong regardless of whether they are caused by virtue and vice, respectively (*UT* 2.20). Mill thus differs from Hume, who gave virtue an axiological primacy over actions instead. An action such as saving someone from drowning - as Mill notes in a famous example in Chapter 2 of *Utilitarianism* - is morally appreciable even if it is performed out of the non-virtuous selfish desire to derive personal benefit from the rescue (*UT* 2.19).

From an epistemological perspective, however, actions – not isolated actions, but the predominant tendencies of behaviour - are the inductive basis for knowing whether or not a person's motives are virtuous (*UT* 2.20). Not isolated actions, but the predominant tendencies of behaviour. Indeed, Mill writes that although 'actions which are blameable often proceed from qualities entitled to praise', it is inconceivable to consider 'any mental disposition as good, of which the predominant tendency is to produce bad conduct' (*UT* 2.20). Virtue is thus identified with a motivational disposition to perform actions of a certain kind, which will not be occasional but will be stable and predictable: whenever the virtuous agent will, for example, be in the presence of characteristics that concern people's pleasure or pain he will tend to give the appropriate virtuous response to those characteristics.

Although reminiscent of the Aristotelian thesis of virtue as a 'habitual state that produces choices' (*NE* II, 6, 1106b36-7a3), Mill differs from Aristotle by not requiring that the full possession of virtue requires the ability to always act under the influence of an appropriate emotion, at the perfect time, toward the right persons, and for the appropriate end (*NE* II.9.1109a24-30). Instead, Mill seems rather to embrace the Humean thesis of 'virtue in rags' (*T* 3.3.1.19/SBN 584), according to which the full possession of virtue is compatible with sometimes acting in a blameworthy way (*UT* 2.20). Significantly, Mill does not specify whether the virtuoso's blameful actions are simply an effect of unforeseeable adverse fate or depend on the virtuous agent's poor cognitive and

predictive ability, and this could therefore mean that less than excellent cognitive ability does not automatically show a defect in virtue.

This said, Mill believed, however, that the possession of virtue is something that admits of degrees of competence. As we saw in the previous section regarding the Millian take on the Aristotelian concept of fairness, there are degrees in the possession of virtue that Mill traces back to the greater or lesser ability to point out exceptions to moral rules or the ability to more or less effectively apply the standard of morality directly to actions whenever we are faced with new circumstances not covered by the rules. Mill further adds in Chapter 2, that degrees of virtue are also related to the extent of our private benevolence and interest in the public good,[172] motives that depend not only on elements originating in our passionate make-up, but also on contingent social-historical factors, such as whether we have the misfortune to live in a corrupt society characterized by unjust laws or whether we carry on an occupation that does or does not allow us to cultivate our affections.[173]

Unlike Hume who lists a list of virtues, Mill does not offer a systematic treatment of the content of virtue, which is to be gleaned from the observations scattered throughout Chapters 2, 4, and 5 of *Utilitarianism*.

---

[172] On the existence of these differences in virtuous dispositions, Mill claims: 'Genuine private affections and a sincere interest in the public good, are possible, though in unequal degrees, to every rightly brought up human being' (*UT* 2.14).

[173] On this point, Mill argues: 'Capacity for the nobler feelings is in most natures a very tender plant, easily killed, not only by hostile influences, but by mere want of sustenance; and in the majority of young persons it speedily dies away if the occupations to which their position in life has devoted them, and the society into which it has thrown them, are not favourable to keeping that higher capacity in exercise. Men lose their high aspirations as they lose their intellectual tastes, because they have not time or opportunity for indulging them; and they addict themselves to inferior pleasures, not because they deliberately prefer them, but because they are either the only ones to which they have access, or the only ones which they are any longer capable of enjoying' (*UT* 2.7). Later in the chapter, he adds: 'Genuine private affections and a sincere interest in the public good, are possible, though in unequal degrees, to every rightly brought up human being. In a world in which there is so much to interest, so much to enjoy, and so much also to correct and improve, every one who has this moderate amount of moral and intellectual requisites is capable of an existence which may be called enviable; and unless such a person, through bad laws, or subjection to the will of others, is denied the liberty to use the sources of happiness within his reach, he will not fail to find this enviable existence, if he escape the positive evils of life, the great sources of physical and mental suffering — such as indigence, disease, and the unkindness, worthlessness, or premature loss of objects of affection' (*UT2*.14).

In Chapter 2, virtue is described as a motivational disposition that has as its object 'the multiplication of happiness' (*UT* 2.29). Despite this broad formulation, in this chapter Mill tends to identify virtue with private benevolence. This thesis rests on two arguments, the first conceptual and the second empirical. The conceptual argument is that we should not confuse the 'motive' of a morally appreciable action with the 'rule' or 'morality' of the action (*UT* 2.19). The latter concerns the ultimate standard of morality that justifies actions in 'proportion' (*UT* 2.2) to what is 'promoting the general interests of society' (*UT* 2.19). Motive, on the other hand, is the psychological inclination that causes the morally appreciable action. The motive for right actions, as we have seen, can also be morally neutral or blameworthy. However, Mill is not so much interested in that aspect here, but in the contrast with the motive of duty, i.e., that which impels us to perform the right action out of respect for the 'Greatest Happiness Principle'. The conceptual argument thus argues that it is not a necessary condition for a motive to be a motive for right action that it impels us to perform that action out of the morality of the action. In contrast, the empirical argument holds that not all human beings are in the condition that their actions typically have large-scale consequences. For most of us, actions only affect the happiness or pain of the people with whom we have relationships. Mill's conclusion from these two arguments is that, in the vast majority of cases, virtue is identical with the possession of a disposition to multiply the happiness or reduce the unhappiness of that small number of people with whom we have social and emotional relations (*UT* 2.19). Mill claims:

> The great majority of good actions are intended, not for the benefit of the world, but for that of individuals, of which the good of the world is made up; and the thought of the most virtuous man need not on these occasions travel beyond the particular persons concerned … The multiplication of happiness is, according to the utilitarian ethics, the

> objects of virtue: *the occasion on which any person* (except one in thousand) *has it in his power to do this on an extended scale, in other words, to be a public benefactor, are but exceptional*; and on *these occasions alone is he called on to consider public utility*; in every other case, *private utility, the interest or happiness of some few persons*, is all he has to attend to. Those alone the influence of whose actions extends to society in general, *need concern themselves habitually about so large an object*. (*UT* 2.19, my italics)

Consistent with these observations, Chapter 2 tends to identify the motive of duty with those motivational dispositions to two types of virtue that apply to only a small number of moral agents: public benefactors and those whose behaviour expresses self-sacrifice for the good of others. The former, as the passage above makes clear, are those who, by their social status or public office, are in a position to benefit a large number of people. The latter are those who express the 'highest virtue', which consists in 'serving the happiness of others by the absolute sacrifice of his own' (*UT* 2. 16). Mill illustrates this virtuous disposition by referring to the figure of the 'hero or martyr', adding that they sacrifice themselves to prevent others from having to 'renounce happiness' (*UT* 2.15). The reference to the martyr shows how, for this type of virtue, the complete renunciation of one's own happiness may involve the sacrifice of one's life. The hero is not necessarily a martyr, although this figure expresses a similar self-sacrifice and courage to sacrifice one's personal interests for the good of others. These references show that the person who exemplifies the 'highest virtue' is not the ordinary individual who, in certain circumstances, is capable of completely sacrificing his or her own particular interests, but rather someone who exemplifies a whole way of life characterised by self-sacrifice.

Far from expressing a concession to the ascetic and Stoic morality of renunciation, Mill's thesis rather grounds this view of virtue on an experience-based utilitarian conception: in the less than ideal conditions in which we live, personal sacrifice is in some

circumstances the only way to avoid major suffering (*UT* 2.16; see also *UT* 2.15) and thus to achieve the utilitarian end.

This conception is partially modified in Chapter 5. Before expounding it, let us summarise what has been said so far. Mill argues that virtue is a motivational disposition that tends to promote the happiness of others. The measure of well-being that an agent must promote in order to be virtuous is contingent on the number of people who typically are affected by the agent's actions. A virtue's threshold thus varies among people. For most of us, since our actions affect only those with whom we have stable social ties and relationships, virtue consists in showing a solicitude toward the welfare of those people who are within those spheres of interaction and does not require being moved by the motive of duty. For others, the few whose actions have large-scale effects, virtue requires taking into account, in a wide variety of situations, a large number of individual interests at stake. For these agents, the motive of virtue coincides with the motive of duty. To these two senses of virtue, Mill adds a third one, which he calls the 'highest virtue', consisting of completely sacrificing one's own interests in favour of those of others. This virtue seems not reserved exclusively for those whose actions have large-scale effects, but can also manifest itself among those who possess exclusively private benevolence and who sacrifice themselves within narrow relational contexts.

In Chapter 5 Mill presents a broader conception of the motive of duty that includes the motivational disposition to perform just actions.[174] Although Mill predominantly describes the feeling of justice as a desire to punish that the spectator of an injustice feels toward the injurer (*UT* 5.18), there are passages in which he insists on 'its binding force'

[174]In this chapter I will not dwell on the important theme examined in *Utilitarianism* chapter 4 of the way Mill explains the transition from the instrumental desire for virtue to the desire for virtue as 'part of happiness', that is 'for its own sake' (*UT* 4.6), which characterises the character of the virtuous person. The examination of this theme is in fact partially distinct from the theme of this chapter. For an examination of this aspect of Mill's theory of virtue I refer to R. Crisp 1996. For an examination of the associationist conception of the mind that supports the transition between these two desires see, W. Donner 1991: 18.

and its being a 'criterion of conduct' (*UT* 5.2), suggesting that the anticipation of the pain we would feel should this feeling be directed toward ourselves would be a strong reason to refrain from unjust actions. If this motive is sufficiently stable to become a motivational disposition, what kind of motive would it be?

In chapter 2, Mill identifies the motive of duty with a virtuous disposition that is characterized by the fact that it impels us to perform actions considering the large-scale negative consequences if we omitted to perform them. Moreover, in chapter 3 he describes the internal sanction of duty as pain that is associated with the violation of moral obligations that, in a cultivated character, can become so intense that we regard these violations as practical impossibilities. This might lean towards the view that the motivational disposition of justice can be seen as a virtue based on the motive of duty. Is this a plausible reading?

At least three considerations support this hypothesis. The first, more general, is that Mill refers to justice as a "virtue" (*UT* 5.37). Since this cannot identify with either private benevolence or self-denial, it will have to identify with the only other remaining virtue, which, according to chapter two, coincides with the motive of duty. It is true that the motive of duty is a virtuous motive operative in only a few people. However, one must bear in mind that this thesis is formulated in a chapter 2, where Mill's aim is to defend himself against the objection that utilitarianism is too demanding a theory. This might underlie Mill's complacent attitude that only a few people are required to act considering the utilitarian rule. A second, stronger reason for this identification is that Mill argues that the motive of justice leads us to avoid performing unjust actions not so much to avoid causing harm to a particular person, but because the unjust action belongs to a class of actions whose compliance ensures the most important part of the general

welfare (*UT* 5.21)[175]. The perspective on the consequences of actions that is connected to the motive of justice is therefore analogous to that of the motive of duty, which motivates us to perform actions by considering the consequences they have not on private well-being, but on large-scale one.

Against this identification, it could be objected that when we comply with the rules of justice, unlike when we are moved by the motive of duty, we do not think about the effects that abstaining from this action has on a large scale. However, as Mill states by imagining this objection himself, although he who is moved by the motive of duty may consider the effects of actions on a large scale at the time he acts, this is by no means a necessary condition for being moved by the motive of duty. Instead, what is necessary is for the person to be able to articulate the reason why he acted by referring to the large-scale consequences of his actions. Mill claims:

It is no objection against this doctrine to say, that when we feel our sentiment of justice outraged, we are not thinking of society at large, or of any collective interest, but only of the individual case. It is common enough certainly, though the reverse of commendable, to feel resentment merely because we have suffered pain; but a *person whose resentment is really a moral feeling*, that is, who considers whether an act is blamable before he allows himself to resent it—such a person, though he may not say expressly to himself that he is standing up for the interest of society, certainly does feel that he is asserting a rule which is for the benefit of others as well as for his own. If he is not feeling this—if he is regarding the act solely as it affects him individually—he is not *consciously just*; he is not concerning himself about the *justice of his actions*.

[175] About the sentiment of disapproval of injustice, Mill characterizes the object of this feeling in this way: 'The sentiment of justice, in that one of its elements which consists of the desire to punish, is thus, I conceive, the natural feeling of retaliation or vengeance, rendered by intellect and sympathy applicable to those injuries, that is, to those hurts, which wound us through, or in common with, society at large. […] When moralised by the social feeling, it only acts in the directions conformable to the general good: just persons resenting a hurt to society, though not otherwise a hurt to themselves, and not resenting a hurt to themselves, however painful, unless it be of the kind which society has a common interest with them in the repression of' (*UT* 5.21).

In a manner reminiscent of the well known Aristotelian distinction between acting in accordance with virtue and acting virtuously (*NE* II 4 1105 a 17-33), Mill argues that what characterizes a just person, i.e. someone who possesses the virtue of justice, as opposed to one whose actions are in accordance with justice is precisely the ability to be able to justify his action by referring to the utilitarian standard.

Bibliography

Primary Texts

Hume, D. 1932. *Letters* [*L*], ed. J. Greig. Oxford: Clarendon Press.

Hume, D. 1987 [1742/1752]. *Essays, Moral, Political, and Literary* [*E*], ed. E.F. Miller, rev. edn. Indianapolis: Liberty Fund.

Hume, D. 1995 [1757]. *The Natural History of Religion* [*NHR*]. Raleigh, N.C.: Alex Catalogue.

Hume, D. 1998 [1751]. *An Enquiry concerning the Principles of Morals* [*EPM*], ed. T. Beauchamp. Oxford: Clarendon Press.

Hume, D. 2007 [1739–40]. *A Treatise of Human Nature* [*T*], ed. D.F. and M.F. Norton. 2 vols. Oxford: Clarendon Press.

Mill, J.S. 1998 [1871; 1st edn. 1863]. *Utilitarianism*, ed. R. Crisp. Oxford and New York: Oxford University Press.

Smith, A. 2002 [1790; 1st edn. 1759]. *The Theory of Moral Sentiments*, ed. K. Haakonssen. Cambridge: Cambridge University Press.

Secondary Texts

Abramson, K. 2002. 'Two Portraits of the Humean Moral Agent', *Pacific philosophical quarterly* 83: 301–334.

Abramson, K. 2008. ‘Sympathy and Hume’s Spectator-Centered, Theory of Virtue’, in A E.S. Radcliffe (ed.), *A Companion to Hume*. Hoboken: Wiley, 240-256.

Abramson, K. 2015. ‘What’s So “Natural” about Hume’s Natural Virtues?’, in D. Ainslie and A. Butler (ed.), *The Cambridge Companion to Hume’s Treatise*. Cambridge: Cambridge University Press, 333-368.

Ainslie, D. 2005. ‘Sympathy and the Unity of Hume’s idea of Self’, in J. Jenkins, J. Whiting, and C. Williams (ed.), *Persons and Passions: Essays in Honor of Annette Baier*. Notre Dame, IN: University of Notre Dame Press, 143-173.

Ardal, P. S. 1966. *Passion and Value in Hume’s Treatise*. Edinburgh: Edinburgh University Press.

Aristotle 2000. *Nicomachean Ethics*, edited and translated by Roger Crisp. Cambridge: Cambridge University Press.

Baier, A. 1979. ‘Good Men‘s Women: Hume on Chastity and Trust’, *Hume studies* 5: 1–19.

Baier, A 1980. ‘Hume on Resentment’, *Hume Studies* 6: 133-49.

Baier, A. 1985. *Postures of the Mind. Essays on Mind and Morals*. Minneapolis: University of Minnesota Press.

Baier, A. 1991. *A Progress of Sentiments*. Cambridge, Mass.: Harvard University Press.

Baier, A. 2010. *The Cautious Jealous Virtue. Hume on Justice*. Cambridge and London: Harvard University Press.

Baron, M. 1982. ‘Hume's Noble Lie: An Account of His Artificial Virtues’, *Canadian journal of philosophy* 12: 539–555.

Bell, M. 2013. *Hard Feelings: The Moral Psychology of Contempt*. Oxford: Oxford University Press.

Besser, L. 2018. ‘Empathy, Interdependency, and Morality. Building From Hume’s Account’, in P.A. Reed and R. Vitz (ed.), *Hume's moral philosophy and contemporary psychology*. London: Routledge, 208-225.

Bricke, J 1996. *Mind and Morality: An Examination of Hume's Moral Psychology*. Oxford: Oxford University Press.

Buckle, S. 1993. *Natural Law and the Theory of Property*. Oxford: Oxford University Press.

Butler, J. [1736]. *Fifteen Sermons Preached at the Rolls Chapel Upon the Following Subjects […]*. GALE: Eighteenth Century Collections Online.

Capaldi, N. 1989. *Hume’s Place in Moral Philosophy*. New York: Peter Lang.

Clark, B. S. and Elliott, J. E. 2001. ‘John Stuart Mill’s Theory Of Justice’, *Review of social economy* Lix: 467-490.

Cohon, R. 1997. ‘The Common Point of View in Hume’s Ethics’, *Philosophy and Phenomenological Research* 57: 827–50.

Cohon, R. 2008. *Hume’s Morality: Feeling and Fabrication*. Oxford: Oxford University Press.

Cohon, R. 2020. ‘Virtue as a Means to Happiness in Hume’s Second Enquiry’, in J. Taylor (ed.), *Reading Hume on the Principles of Morals*. Oxford: Oxford University Press, 156-176.

Craig, E. 2007. ‘Genealogies and the State of Nature’, in A. Thomas (ed.), *Bernard Williams*, Cambridge: Cambridge University Press, 181-200.

Cremaschi, S. and Volodia, M. (2006). ‘The Mill-Whewell controversy on ethics and its bequest to analytic philosophy’, in E. Baccarini and S. P. Samaržja (ed.), *Rationality in Belief and Action*. Rijeka, Croatia: University of Rijeka, 45-62.

Crisp, R. 1996. ‘Mill on Virtue as a Part of Happiness’, *British Journal for the History of Philosophy* 4.2: 367–80.

Crisp, R. 1997. *Mill on Utilitarianism*. London: Routledge.

Crisp, R. 2019. *Sacrifice Regained. Morality and Self-interest in British Moral Philosophy from Hobbes to Bentham*. Oxford: Clarendon Press.

Darwall, S. 1993. 'Motive and Obligation in Hume's Ethics', *Noûs* 27: 415–448.

Darwall, S. 1995. *The British Moralists and the Internal 'Ought'*. Cambridge: Cambridge University Press.

Darwall, S. 1998. 'Empathy, Sympathy, Care', *Philosophical Studies* 1998: 261-82.

Donner, W. 1991. *The Liberal Self. John Stuart Mill's Moral and Political Philosophy*. Ithaca: Cornell University Press.

D'Urso, V 2013. *Sulla vendetta e sul perdono*, in *Teoria e clinica del perdono*, B. Barcaccia and F. Mancini (ed.). Milano: Raffaello Cortina Editore, 115-128.

Evnine, S. 1993. 'Hume, Conjectural History, and the Uniformity of Human Nature', *Journal of the History of Philosophy* 31: 589–606.

Fieser, J. 1997. 'Hume's motivational distinction between natural and artificial virtues', *British Journal for the History of Philosophy* 5: 373–388.

Flanagan, O. 2021. *How to Do Things with Emotions. The Morality of Anger and Shame across Cultures*. Princeton and Oxford: Princeton University Press.

Forbes, D. 1975. *Hume's Philosophical Politics*. Cambridge: Cambridge University Press.

Frazer, M. L. 2010. *The Enlightenment of Sympathy: Justice and the Moral Sentiments in the Eighteenth Century and Today*. Oxford: Oxford University Press.

Garrett, D. 2007. 'The First Motive to Justice: Hume's Circle Argument Squared', *Hume studies* 33: 257–288.

Gauthier, D. 1992. 'Artificial Virtues and the Sensible Knave', *Hume Studies* 18: 401–27.

Gill, M. 2000. 'Hume's Progressive View of Human Nature', *Hume Studies* 26: 87–108.

Gill, M. 2006. *The British Moralists on Human Nature and the Birth of Secular Ethics*. Cambridge: Cambridge University Press.

Gill, M. 2014. *Humean Moral Pluralism*. Oxford: Oxford University Press.

Griswold, C. 2007. *Forgiveness. A Philosophical Exploration*. Cambridge: Cambridge University Press.

Haakonssen, K. 1981. *The Science of a Legislator: The Natural Jurisprudence of David Hume and Adam Smith*. Cambridge: Cambridge University Press.

Haakonssen, K. 1993. 'The Structure of Hume's Political Theory', in D.F. Norton (ed.), *The Cambridge Companion to Hume*. Cambridge: Cambridge University Press, 182-221.

Hampton, J. 1988. 'Forgiveness, resentment and hatred', in J. Murphy and J. Hampton (ed.), *Forgiveness and Mercy*. Cambridge: Cambridge University Press, 35-87.

Harman, G. and Thomson, J. J. 1996. *Moral Relativism and Moral Objectivity*. Oxford: Blackwell.

Harris, J. A. 2010. 'Hume on the Moral Obligation to Justice', *Hume studies* 36: 25–50.

Harrison, J 1981. *Hume's Theory of Justice*. Oxford: Clarendon Press.

Hills, A. 2010. *The Beloved Self. Morality and the Challenge form Egoism*. Oxford: Oxford University Press.

Home, H. 1792. *Historical Law-Tracts*, The Fourth Edition with Additions and Corrections. Edinburgh.

Hunter, G. 1962. 'Hume on is and Ought', *Philosophy* 37: 148–152.

Hursthouse, R. (1999). 'Virtue Ethics and Human Nature', *Hume studies* 25: 67–82.

Kail, P. J. E. 2009. 'Naturalism, Method and Genealogy in Beyond Selflessness', *European Journal of Philosophy* 17: 113–120.

Korsgaard, C.M. and O'Neill, O. 1996. *The sources of normativity*. Cambridge and New York: Cambridge University Press.

Locke, J. 1997 (1662). *Second Tract on Government*, in M. Goldie (ed.), *Political Essays*. Cambridge: Cambridge University Press.

Lyons, D. 1994. *Mill's Theory of Justice*, in David Lyons, *Rights, Welfare, and Mill's Moral Theory*. Oxford: Oxford University Press, 67-88.

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: 231-53.

Miller, C. 2018. 'Virtue as a Trait', in N. Snow (ed.), *The Oxford Handbook of Virtue*. Oxford: Oxford University Press, 9-34.

Nagel, T. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.

Nussbaum, M. 2006. *Frontiers of Justice. Disability, Nationality, Species Membership*, Cambridge, Mass and London: Harvard University Press.

Nussbaum, M. 2016. *Anger and Forgiveness. Resentment, Generosity, Justice*. Oxford: Oxford University Press.

Pink, T. 2017. 'Hume, Virtue and Natural Law', in G. Duke and R. P. George (ed.), *The Cambridge Companion to Natural Law Jurisprudence*. Cambridge: Cambridge University Press, 187–215.

Pollock, R. 2016. 'Hume and the Problem of Paternalism: When Is Humanity Sufficient?', *The Southern Journal of Philosophy* 54: 107-28.

Prinz, J. 2012. 'Is Empathy Necessary for Morality', in A. Coplan, and P. Goldie (ed), *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press, 211-229.

Pritchard, M. S. 2008. 'Justice and Resentment in Hume, Reid, and Smith', *Journal of Scottish Philosophy* 6: 59-70.

Queloz, M. 2018. 'Williams's Pragmatic Genealogy and Self-Effacing Functionality', *Philosophers' Imprint* 18: 1-20.

Queloz, M. 2021. *The Practical Origins of Ideas. Genealogy as Conceptual Reverse-engineering*. Oxford: Oxford University Press.

Radcliffe, E. 2018. *Hume, Passion, and Action*. Oxford: Oxford University Press.

Rawls, J. 2020 (1971). *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Reed, P. A. 2012. 'Motivating Hume's natural virtues', *Canadian journal of philosophy* 42:134–147.

Rousseau, Jean-Jacques 2003 [1755]. *A Discourse on Inequality*. London: Penguin Classics.

Sayre-McCord, G. 2016. 'Hume on the Artificial Virtues', in P. Russell (ed.), *The Oxford Handbook of Hume*. Oxford: Oxford University Press, 435–469.

Schwarze, M. 2020. *Recognizing Resentment: Sympathy, Injustice, and Liberal Political Thought*. Cambridge: Cambridge University Press.

Slote, M. 2007. *The Ethics of Care and Empathy*. Abingdon: Routledge.

Stewart, J.B. 1992. *Opinion and reform in Hume's political philosophy*. Princeton: Princeton University Press.

Strawson, P. 2008. *Freedom and Resentment*, in *Freedom and Resentment, and other Essays*. London: Routledge.

Sturgeon, N. L. 2001. 'Moral skepticism and moral naturalism in Hume's Treatise', *Hume studies* 27: 3–83.

Sumner, L. 2005. 'Mill's Theory of Rights', in H. West (ed.), *The Blackwell Guide to Mill's Utilitarianism*. Oxford: Blackwell, 184-198.

Taylor, J. 2015. *Reflecting Subjects: Passion, Sympathy, and Society in Hume's Philosophy*. Oxford: Oxford University Press.

Voorhoeve, A. 2014. 'How Should We Aggregate Competing Claims?', *Ethics* 125: 64-87.

Wallace, R. J. 1994. *Responsibility and the Moral Sentiments*. Cambridge: Cambridge University Press.

Williams, B. 1981. 'Internal and external reasons', in B. Williams (ed.), *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press, 101-113.

Williams, B. 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press.

Zagorac, I. 2015. 'Hume's Humanity and the Protection of the Vulnerable', *Diametros* 44: 189-203.