UNIVERSITÀ DEGLI STUDI DI MILANO

Université Grenoble Alpes

Corso di dottorato in Filosofia e scienze dell'uomo

Ciclo XXXVI

École doctorale en Philosophie : histoire, représentation, création

Dipartimento di Filosofia "Piero Martinetti"

Institut de Philosophie de Grenoble

Beyond Desert and Moral Responsibility: Rethinking Justice

M-FIL/03 Filosofia morale

Marina BERAHA

Tutor:

Prof. Gianfranco MORMINO

Prof. Philippe SALTEL

Coordinatore del corso di dottorato: Prof. Niccolò GUICCIARDINI

Directeur de l'école doctorale : Prof. Thierry MÉNISSIER

Anno accademico 2022/2023

This thesis is dedicated to the men incarcerated in the AS1 circuit of the Milan-Opera prison, who, with immense generosity, shared their stories with me and made me realise that I needed to radically rethink my conceptions of justice.

Abstract

English Version

This research conducts a comprehensive critique of retributive justice, understood pragmatically as a system for addressing the complex social challenges stemming from reciprocal violence. Starting from an exploration of violence management in pre-State societies, particularly through the scapegoating victimisation mechanisms and the ensuing emergence of religious constructs as elucidated by Girard, the intricate interplay among violence, justice, and religion will be scrutinized. Subsequently, the focus will shift to analysing the modern State's monopolisation of violence and its fundamental role in violence management. Within this context, the traditional justifications for retributive practices will be examined and criticised, alongside an exploration of the relationship between retribution and emotions considered in their adaptive function. Furthermore, the conceptual evolution of retributive justice, in its connection to the notions of free will, desert, and individual moral responsibility, will be traced, commencing from the analysis conducted by 17th-century philosophers. Having ascertained the moral inadequacy of the retributive conception of justice, the possibility of replacing the retributive paradigm with a transformative and restorative justice system, underpinned by utilitarianism and a redefined conception of shared moral responsibility, will be investigated.

Versione italiana

Questo lavoro si propone di svolgere un'analisi critica del concetto di giustizia retributiva, intesa pragmaticamente come sistema per affrontare i problemi sociali legati alla violenza reciproca. A partire da un'indagine sulle modalità di gestione della violenza nelle società pre-statuali, in particolare attraverso i meccanismi di vittimizzazione di capri espiatori e la conseguente nascita della religione, come teorizzato da Girard, verrà indagato il rapporto fra violenza, giustizia e religione. Si passerà, quindi, all'indagine della monopolizzazione della violenza da parte dello Stato moderno e il ruolo che conseguentemente assumerà nella gestione della violenza interna alla società. In questa cornice, si analizzeranno e criticheranno le principali giustificazioni delle pratiche retributive, se ne indagheranno le relazioni con le emozioni e la loro funzione adattiva e se ne esaminerà la storia concettuale, in rapporto alle

nozioni di libero arbitrio, merito e responsabilità morale individuale, a partire dalle analisi condotte dai filosofi del Seicento. Accertata l'inadeguatezza morale della concezione retributiva di giustizia, si indagherà, infine, la possibilità di sostituire il paradigma retributivo con un sistema di giustizia trasformativa e riparativa, fondato sull'utilitarismo e su una riformulazione del concetto di responsabilità morale in chiave sociale.

Version française

Cette étude vise à effectuer une analyse critique de la justice rétributive, conçue de manière pragmatique comme un système permettant de relever les défis sociaux liés à la violence réciproque. En commençant par une exploration de la gestion de la violence dans les sociétés pré-étatiques, en particulier à travers les mécanismes de victimisation des boucs émissaires et l'émergence subséquente de la religion, comme l'a élucidé Girard, l'interaction entre la violence, la justice et la religion sera examinée de près. Par la suite, l'attention se portera sur l'analyse de la monopolisation de la violence par l'État moderne et sur son rôle fondamental dans la gestion de la violence à l'intérieur de la société. Dans ce contexte, les justifications traditionnelles des pratiques rétributives seront examinées et critiquées, ainsi que la relation entre la rétribution et les émotions avec leur fonction adaptative. En outre, l'évolution conceptuelle de la justice rétributive, dans son lien avec les notions de libre arbitre, de mérite et de responsabilité morale individuelle, sera retracée à partir de l'analyse menée par les philosophes du XVIIe siècle. Après avoir constaté l'inadéquation morale de la conception rétributive de la justice, la possibilité de remplacer le paradigme rétributif par un système de justice transformatrice et réparatrice, fondé sur l'utilitarisme et sur une conception redéfinie de la responsabilité morale partagée, sera étudiée.

Table of Contents

Introduction 8

1 The Past: The Origins of Justice 16

1.1 The Problem of Human Violence 16

1.2 Earthly Justice 21

1.2.1 The Sacrificial System 21

1.2.2 The Vindicatory System 41

1.2.3 The Transition to the State System 46

1.3 Otherworldly Justice 49

1.3.1 The Evolution of Christian Justice: From Mercy to Perfect Retribution 52

1.3.2 The Evolution of Protestant Justice: From Damnation to Meritocracy 57

1.3.3 Turning Retaliation into Justice 64

2 The Present: Theories and Practices of Justice 67

2.1 The State, Justice, and Their Victims 67

2.1.1 Retribution and Rationality 70

2.1.2 Meritocracy and its Losers: A Digression 78

2.1.3 Violence by Omission, Violence by Execution 87

2.2 Retributive Emotions: A Moral Delusion 95

2.2.1 Moral Intuitionism 95

2.2.2 The Mirage of Moral Emotions: The Case of Justice 101

2.2.3 Moral Realism and the Right to Be Sure 113

2.3 Punishment: A Philosophical Analysis 116

2.3.1 Between Future and Past: The Traditional Justifications for Punishment 116

2.3.2 Retributive Fitness 128
2.3.3 Building the Conceptual Framework: Justice and Responsibility in Modern Philosophy 131
2.3.4 The Conceptual Framework in Contemporary Philosophy 172
3 The Future: Normative Perspectives 178
3.1 Reshaping Moral Responsibility 178
3.1.1 Interconnectedness 178
3.1.2 Co-responsibility 192
3.2 A Social Justice Approach to Crime 202
3.2.1 Solidarity 203
3.2.2 Social Justice: More Than an Economic Issue 211
3.3 Utilitarianism and Prison Abolition: Rethinking Justice 232
3.3.1 The Missing Match 232
3.3.2 Transforming Conceptions of Justice 254
Conclusion 269
Bibliography 281
Résumé en français 305

Introduction

I first entered a prison in 2018 as a student of a philosophy workshop that combined incarcerated and non-incarcerated students for a semester. It then became two semesters; then I started tutoring incarcerated students; then coordinating the university programme for incarcerated students. Basically, I went to prison every week until March 2020, when access to volunteers and non-essential workers was denied because of the rise of the COVID-19 pandemic. Mentally, however, I remained somewhat behind those bars ever since, as I kept working in my mind on what it means to incarcerate people, on why we do that and, especially, why we so strenuously want it to be done when an outrageous crime makes it to the headlines, why we cherish punishment as "justice".

I had met there some of the people who had been in those headlines or who had not been for accidental publishing circumstances, their crimes being no less heinous. Probably, people I had wished would be sent to prison when I heard of them in the news. Concretely seeing the realisation of what I had wished for – if not for any of them specifically, undoubtedly for others in similar situations – however, made me question if what I was witnessing really was *justice*, the justice I thought we would have achieved once we would put those people behind bars, the justice that we demand from the State when a crime is committed. This is, in the end, what ministries of justice do everywhere in the world: among other tasks, they manage their citizens' punishments. Although I recognised a certain longing for punishing people within myself – for having offenders punished by the State, but also for punishing my loved ones myself, even if only by sulking at them when they would hurt me – I started to struggle at identifying this impulse as something morally worthy, a value, as I would consider justice to be.

As I was getting to know the stories of these incarcerated people, I started thinking that the problem with this conception of justice, that is, retributive justice, was free will. Of course, I would think, if a free and morally responsible person voluntarily and consciously hurt another one, they would *deserve* punishment. However, free will is an illusion; hence, there is no individual moral responsibility, no space for desert, and retribution cannot be fairly applied. But as I delved more into these reasonings, I started wondering: suppose that we were perfectly free, that we possessed the libertarian free will that makes us the sole originators of our own actions; still, why would we achieve

justice through punishment? What is in the voluntary infliction of suffering on someone who has done harm that is intrinsically good?

I also started realising that, no matter how security-centred part of the public discussion around punishment was, the real reason behind this practice was not dangerousness. Even among volunteers and social workers I encountered in prison, people who sincerely believed that those they were meeting there were "rehabilitated", the talk often concerned the need for them to "pay their debt to society". What debt, however? A debt of suffering? The workings of suffering seemed to me utterly different from those of money, however. If I lend some money, once the debt is repaid, I can use the money that is returned to me to buy something, or I can invest it and generate more money, or I can feel safer because I know I have the financial capacity to face emergencies, etc. If I have been harmed, if I suffered, however, what do I do with my offender's suffering? What use is it to me – or to society, if we wish to consider society as the victim of the breaking of the law? Although the idea of debt repayment certainly spoke to me on an emotional level, on a rational level it did not make sense to me anymore.

From this, the two research questions that I will investigate in this thesis were born. First, on the descriptive level, where does this idea of retributive justice come from? Why do we believe that harming someone who harmed us would generate something intrinsically morally valuable? Second, on the normative level, can we do without this conception of justice? What would a society that dispenses with retributive justice look like? I had, in fact, noticed that those who questioned the intrinsic rightness of the retributive paradigm would then remain anchored to its practices, and those who sought to unhinge practices were doing so on a critique not of the theoretical paradigm behind it but of its shortcomings. For example, as I will show, determinists and utilitarians, both rejecting the value of retribution, would end up subscribing to the punitive carceral system, although justifying it in other terms, and those proposing restorative justice approaches would consider the aim of the restorative process as the satisfaction of the victim – left behind by the traditional criminal system – in full retributive style. On the other hand, prison abolitionists often founded their attacks on the current system on a critique of prison as a racist institution rather than on the inherent moral impropriety of punishment. With this thesis, I aim to fill these gaps.

To tackle these questions, I will take a critical theory approach, aiming at unmasking oppressive structures and proposing liberating alternatives for the actors

involved, resisting the idea that justice must be understood in the form of a zero-sum game. I will situate this approach within two conceptual frameworks: on the one hand, the mimetic and sacrificial theory of Girard, which I will introduce in the very first pages and which will serve to understand the mechanisms of violence and the way human societies have dealt and are dealing with it. On the other hand, the normative proposal will be framed within the utilitarian normative theory, understood as a moral realist framework based on the shared experiences of pleasure and pain.

The topic of justice lies at the intersection of many different fields of study. My own perspective will be eminently moral, drawing on the moral literature on the subject (Bruce Waller, Thomas Nagel, Gary Watson, Michel Foucault, Jean-Marie Guyau, J. J. C. Smart, Saul Smilansky among others). However, it would be impossible to deal with this topic without touching on the different disciplines that constitute it, especially legal philosophy and theory (through the works of H. L. A. Hart, Herbert Morris, Ronald Dworkin, Nils Christie) and political philosophy (John Rawls, Martha Nussbaum, Angela Davis, Iris Marion Young, Elisabeth Anderson, Paul Dumouchel, Michael Sandel, Charles Taylor), given the eminently institutional focus of this work – dictated by reasons of space and time and not by the belief that the institutional level is essentially different from that of individual relations. Anthropology (René Girard in the first place, Raymond Verdier, Lucien Scubla) and its intersections with sociology (Émile Durkheim, Michael Young, Max Weber) and psychology (Fritz Heider, Melvin Lerner) will also play a crucial role in my analysis. Finally, much space will be dedicated to prominent authors of modern philosophy (Luther, Hobbes, Descartes, Spinoza, Malebranche, Leibniz, Kant, Bentham), who have drawn the boundaries and main paths of the philosophical debate on justice and the associated philosophical concepts.

The breadth of my perspective in tackling this issue should already be evident from the table of contents, which I will now illustrate in more detail. The thesis is divided into three chapters addressing respectively *The Past*, *The Present*, and *The Future*, sketching out a journey that, from *The Origins of Justice*, that is, from an analysis of the systems put in place in archaic human societies to solve the fundamental problem of human cohabitation, namely, violence, will pass through contemporary *Theories and Practices of Justice*, analysed in their historical evolution, to finally elaborate innovative *Normative Perspectives* that aim to tackle that same problem

unhinging the oppressive system that has been deployed, disrupting retribution and resting on the principle of suffering minimisation.

Each chapter will be divided into three sections. Concerning the first chapter, *The Past: The Origins of Justice*, its first section, *The Problem of Human Violence*, will set the foundations of my approach to the topic of justice presented as a pragmatic solution, before being a moral value, to the problem of violence. The second section, *Earthly Justice*, will introduce the Girardian framework of this work. I will illustrate Girard's mimetic theory and the mechanisms leading to crises of generalised violence in human groups. I will then explain the solution theorised by Girard, that is the spontaneous emergence of religion as a system for canalising violence and preventing its outbursts. To the exploration of the sacrificial system, the analysis of Verdier's vindicatory system will follow, illustrated as a system based on solidarity bonds that I will interpret as a stage in the path that will eventually lead to the State's monopolisation of legitimate violence. Finally, through Dumouchel's analysis, I will focus on the role of Christianity in the erosion of the solidarity bonds that kept the vindicatory system together, leading to the establishment of the State system. In the third and last section, *Otherworldly Justice*, I will dwell precisely on religious conceptions of justice, highlighting how they always take a retributive form. I will show, in fact, how the attempts of both early Christianity (read through a Girardian lens) and Protestantism (in the works of Luther and Calvin) to emancipate divine justice from retribution have failed, and I will interpret this fact as a sign of the deep-rootedness of the retributive impulse. I will argue that these divine conceptions of retributive justice have played a crucial role in transforming such impulse into the moral value of justice.

I will open the second chapter, *The Present: Theories and Practices of Justice*, with a section titled *The State, Justice, and Their Victims* where I will address the State's practices of justice and only then move on to theoretical justifications because I believe that, here too, practice comes before theory, pragmatic solutions before moral values, the latter often being the result of our quest for justifications for what we are already doing for practical reasons. I will show how the State's justice system is a form of retribution that tries to differentiate itself from vengeance through a pretence of rationalisation without challenging its fundamental principle in any way. The theme of retribution in the contemporary State will lead me to analyse – alas too briefly – the theme of meritocracy, that is, the ideal of the just society in Western culture. Through

the works of Michael Young and Michael Sandel, I will expose its shortcomings, to then present Dumouchel's sacrificial interpretation of meritocratic mechanisms. However, I will oppose it to my idea that the real sacrificial victims of the contemporary State system are to be found behind prisons' walls, making the carceral system a preventive – in addition to, obviously, repressive – system not in terms of deterrence, as it is typically portrayed in the debate around it, but in sacrificial terms.

The second section, *Retributive Emotions: a Moral Delusion*, will be dedicated to a critique of moral intuitionism, a realist moral framework that, I will argue, plays a crucial role in our elaboration of the concept of justice. I will show that intuitive moral judgements are not "guided" by our reactive emotions; they are rather polluted by them. Such emotions, in fact, introduce self-centred biases in our moral reasoning and often constitute the core of a so-called moral intuition. I will focus especially on the case of anger and its influence on our ideas of justice, proposing, within this framework, a critique of Kant's theory of retributive justice. I will then introduce utilitarianism as an alternative realist normative framework that can overcome the problems outlined regarding intuitionism.

The third section, *Punishment: a Philosophical Analysis*, will analyse punishment and its supposed vital role in the achievement of justice from the retributive perspective. Other than illustrating and criticising the traditional backwards- and forward-looking justifications that constitute much of the contemporary philosophical debate on punishment, I will introduce my idea of the fundamental fitness of the retributive paradigm to the contemporary Western world. I will show how punishment fits our system on the economic level (through the work of Angela Davis), the political level (Michel Foucault), the emotional-psychological level, the religious level, and, finally, the conceptual level. The rest of the chapter will be dedicated to the investigation of this conceptual fitness, that is, the conceptual system that was developed over the course of the 17th-century philosophical debate through the works of Hobbes, Descartes, Spinoza, Malebranche and Leibniz, to occupy then the central place in contemporary reflections and practices of justice. It is the conceptual framework that links justice, individual moral responsibility, desert, and guilt.

The third and last chapter, *The Future: Normative Perspectives*, will be dedicated to my proposal for an alternative form of management of the problem of interpersonal violence in the spirit of liberation from oppression. In the first section, *Reshaping Moral Responsibility*, I will question the concept of individual moral

responsibility, with reference to the philosophical works of Waller, Beardsley, and Smilansky, as well as of Milgram and Heider on the psychological side. Through these studies, I will argue that the recognition of the fundamental interconnectedness of individuals among each other and their environment should lead to a reformulation of the concept of moral responsibility that drops the essentially individual component and opens the space for a shared kind of responsibility, that I will elaborate starting from Taylor's concept of co-responsibility and Waller's concept of taken responsibility.

In the second section, *A Social Justice Approach to Crime*, I will argue that a different conception of justice as a system for crime management is needed once it is recognised that moral responsibility is shared. I will contend that the revolution brought about in the concept of justice by the introduction of social justice can, and should, be extended to the penal system instead of being complementary to it. In particular, I will argue in favour of substituting the dyad of individual moral responsibility and retributive justice with that of shared responsibility and social justice. To this end, Iris Marion Young's social connection model of responsibility will come to the aid.

How to practically rethink the justice system will be the object of the third and last section, *Utilitarianism and Prison Abolition: Rethinking Justice*. After positioning my proposal within the utilitarian framework and rejecting some common objections to utilitarian conceptions of justice, I will explain, referring extensively to Bentham's texts, why I think that the focus of many utilitarians (e.g., J. J. C. Smart) on the utilitarian justifications for punishment has been misleading and simplistic. I will oppose it to Bentham's systemic approach to utilitarianism, which I will couple with the prison abolition movement. Prison abolitionists have ignored utilitarianism as a possible moral framework for their proposal, which they based instead on fragile moral grounds, as I will show. I aim to fix this problem precisely through the utilitarian conceptualisation of the prison abolition project. Finally, in this framework, I will endorse a transformative justice approach to crime prevention and a restorative justice approach to respond to crimes that could not be prevented.

It is evident that such a broad and multidisciplinary posture will not allow me to go into the details of each of the elements I will touch on. Although this goes against the current trends in research, I have deemed that the topic of justice necessitated a broader gaze, as it pervades so many areas of our public and private life, being rooted in our psychological structure before our practices and institutions and occupying a central place in common sense, philosophical thought, and even political propaganda. It is by

virtue of this greater generality that we can grasp the relevance as well as the uniqueness of this subject. Each element of this larger picture will be here understood not in itself but through its relationships within the broader context that will be progressively built. Hopefully, this work may provide the basis for further, more detailed analysis of each individual component of the picture I am about to outline.

1 The Past: The Origins of Justice

1.1 The Problem of Human Violence

The problem of the management of violence is probably the most fundamental in human relations. It need not be ascribed to intrinsic human cruelty but, as Hobbes has stated,

> from […] equality of ability, ariseth equality of hope in the attaining of our Ends. And therefore if any two men desire the same thing, which nevertheless they cannot both enjoy, they become enemies; and in the way to their End, (which is principally their owne conservation, and sometimes their delectation only,) endeavour to destroy, or subdue one an other.[1]

Proximity eventually entails competition. Still, humans cannot live without one another: being mammals, they are highly dependent on their mothers after their birth; also, human infants are born at an earlier developmental stage compared to other mammals, requiring more attention from their caregivers for a more extended period. Parents often profit from the support of a wider community of family members and neighbours during this caregiving period, implying the need for social bonds and the organisation of life in groups. Vicinity is, therefore, both the cause of survival through cooperation and the cause of death through violence.

Until the 1990s, most anthropologists and ethnologists would deny this tendency to violence in human behaviour, having inherited the Rousseauian conception of "the noble savage" and attributing to the complexity of some ancient and modern societies the root of violence, both within and between groups[2]. Despite the observations of early anthropologists, and even some of Rousseau's disciples themselves, tended to

[1] Thomas Hobbes, "Leviathan," in *Three-Text Edition of Thomas Hobbes's Political Theory. The Elements of Law, De Cive and Leviathan*, ed. Deborah Baumgold (Cambridge: Cambridge University Press, 2017), 135.

[2] Claude Lévi-Strauss observes that it is a mistake to attribute to Rousseau the naive glorification of a peaceful state of nature, since Rousseau is well aware of the fact that "society is inherent to man", who cannot be analysed as an isolated individual (it is rather Diderot who developed Rousseauvian theory by exalting a solitary state of nature) Claude Lévi-Strauss, *Tristes Tropiques*, trans. John Russell (New York: Criterion Books, 1961), 389.

Rousseau's aim is to analyse the life of people in simpler societies in order to understand the evils inherent in complex societies that can be overcome in order to find "[t]he natural man [who] is not prior to, nor external to society. It is up to us to rediscover his form, immanent to the social state outside of which the human condition is inconceivable [...]." Lévi-Strauss, 470, my translation.

disconfirm the preconception of the peaceful savage[3], as Lawrence Keeley claimed in his 1996 revolutionary book *War Before Civilisation*, anthropology has been a victim of moralising myths that prevented the acknowledgement that, in Steven LeBlanc's words, "[l]ives in the past were those of fear, war, worry and hunger"[4] – it is impossible, in these words, not to hear the echo of Hobbes's famous account of human life in a state of nature: "solitary, poore, nasty, brutish, and short"[5].

Keeley's 1996 work opened the path to the study of the history and prehistory of violence. As for all prehistory, studies are troublesome, and results are hardly ever conclusive. We have little or no evidence for long periods, for tools were likely made of perishable materials, and animals or other natural events would often dislocate bones. We only have scant fragments as the basis for studying the prehistoric past. Therefore, scholars disagree as to whether Keeley and LeBlanc are right or wrong in their account of prehistoric life, whether war was more or less frequent than what has been experienced in historic times. Still, a certain non-negligible degree of inter- and intra-group violence seems to be confirmed by archaeology[6] and osteoarchaeology (i.e., the study of prehistoric bones). Specifically, the survey of blunt-force trauma on skulls and the presence of projectiles embedded in bones (though representing a minority of wounds, likely because they hardly ever reach the bones and tend to cause death through wounds in soft tissues) provides relevant information regarding violent causes

[3] Lawrence H. Keeley, *War before Civilization: The Myth of the Peaceful Savage* (Oxford: Oxford University Press, 1996), 7; Mark W. Allen, "Hunter-Gatherer Conflict: The Last Bastion of the Pacified Past?," in *Violence and Warfare Among Hunter-Gartherers*, ed. Mark W. Allen and Terry L. Jones (Walnut Creek: Left Coast Press, 2014), 15–25.

[4] Steven LeBlanc, "The Origins of Warfare and Violence," in *The Cambridge World History of Violence. Volume 1: The Prehistoric and Ancient Worlds*, ed. Garrett G. Fagan et al. (Cambridge: Cambridge University Press, 2020), 42.

[5] Hobbes, "Leviathan," 138. Here, Hobbes refers to what he considers the historical reality of the isolated life humans lived before the establishment of the political state. As should be clear, I tend to agree more with Pierre Gassendi's position that the possibility of isolated humans is only fictional, since they are "social animals", in the sense specified above concerning the need for the other because of the frailty inherent to human life. Pierre Gassendi, "Syntagma Philosophicum," in *Petri Gassendi Diniensis Ecclesiae Praepositi, et in Academia Parisiensi Matheseos Regii Professoris Opera Omnia*, ed. Friedrich Frommann (Stuttgart-Bad Canstatt, 1964), 755a.

Therefore, the state of generalised war that Hobbes situates in the isolated state of nature is displaced by Gassendi, following Lucretius, to the pre-political state of human cohabitation, that generates the social pact. For a comparative analysis of Lucretius', Hobbes', and Gassendi's social contract theories, see: Gianfranco Mormino, "Il Triplo Contratto Sociale in Lucrezio e in Gassendi," in *Lucrezio e La Modernità: I Secoli XV-XVII: Atti Del Convegno Internazionale, Università Di Milano-Bicocca, 13-14 Dicembre 2007*, 2011, 141–62.

[6] Terry L. Jones and Mark W. Allen, "The Prehistory of Violence and Warfare among Hunter-Gatherers," in *Violence and Warfare Among Hunter-Gartherers*, ed. Mark W. Allen and Terry L. Jones (Walnut Creek: Left Coast Press, 2014), 354–55.

of death[7]. Prehistoric massacre sites are also identifiable where the remains of many – as compared to the small groups in which humans lived at the time – are left unburied or buried in common graves without any *post-mortem* treatment[8].

Once the "long chronology" of violence is endorsed, and the myth of the peaceful savage is set aside, we can proceed to consider how human groups have dealt with the problem of reciprocal violence, which, as briefly noted, stems from the closeness that humans cannot do without. Understanding how ancient communities have survived, therefore managing to regulate violence and avoid mutual extermination, may help us better understand present methods of violence regulation that go by the name of "justice", a very general term by which people often actually refer to the more specific domain of retributive justice.

Retributive justice is generally considered as the moral value achieved when the action of morally responsible agents is reciprocated on the basis of desert. Grounding it in the management of violence, however, I interpret it as a practice originating from vengeance that served the aim of protecting human societies from self-destruction through internecine violence – something of vital practical value but not inherently moral. I will argue, in fact, that the very concept of justice, and the related concepts of desert and individual moral responsibility, serve as an ex-post justification for institutionalised retaliation in order to distinguish it from private vengeance and identify it as a good form of violence, when in fact it is not intrinsically good. As I will show, the murkiness of the concepts of desert and, to some extent, moral responsibility seems to indicate that this kind of justice was not born as a moral theory but as a practice that was then given a theoretical cover-up in moral terms.

To understand this process, however, it is necessary that we first delve into the origins of these practices, to only then analyse the emotional and rational instances they

[7] LeBlanc, "The Origins of Warfare and Violence"; Steven LeBlanc, "Forager Warfare and Our Evolutionary Past," in *Violence and Warfare Among Hunter-Gartherers*, ed. Mark W. Allen and Terry L. Jones (Walnut Creek: Left Coast Press, 2014), 26–46; Jorg Orschiedt, "Violence in Palaeolithic and Mesolithic Hunter-Gatherer Communities," in *The Cambridge World History of Violence. Volume 1: The Prehistoric and Ancient Worlds*, ed. Garrett G. Fagan et al. (Cambridge: Cambridge University Press, 2020), 58–78.

[8] Christian Meyer, "Early Massacres: Mass Violence in Neolithic Europe," in *The Cambridge World History of Violence. Volume 1: The Prehistoric and Ancient Worlds*, ed. Garrett G. Fagan et al. (Cambridge: Cambridge University Press, 2020), 299; Martin J. Smith, Rick J. Schulting, and Linda Fibiger, "Settled Lives, Unsettled Times: Neolithic Violence in Europe," in *The Cambridge World History of Violence. Volume 1: The Prehistoric and Ancient Worlds*, ed. Garrett G. Fagan et al. (Cambridge: Cambridge University Press, 2020), 88.

appeal to and their limits. For this reason, they will be the object of the first part of this work.

Having cited Hobbes at the very beginning of this chapter, it comes naturally to think about the modern conceptualisation of the problem of the management of violence through social contract theories. Social contract theories' starting point is always that of a "state of nature", which is a pre-political condition of human life. According to Hobbes's premises cited above, his state of nature cannot be but one of war of all against all. In Hobbes's theory, the misery of life in a state of nature causes humans to create the State – an absolute monarchy – through a social contract, a rational agreement between isolated individuals, that enables them to live in a condition of peace.

The idea of a social contract subsequent to a state of nature, that is introduced in modern thought[9] by Hobbes and has since had, and still have, a great fortune among philosophers. Different thinkers theorised different states of nature according to their own anthropology and conceptions of natural laws. Locke, for example, considered the state of nature to be pre-political but not pre-moral and conceived the institution of the political State as the means to enforce the respect of these natural rights. Famously overturning Hobbes's assumptions, Rousseau, in his *Discourse on the Origin and Foundations of Inequality Among Men*, considered the state of nature as a state of peace and cooperation, ruined by the establishment of private property, and the subsequent birth of inequality, greed, competition, and most vices. The social contract is therefore established to protect private property, thus fossilising inequalities and ending the graceful state of nature for good. Rousseau means this account to be historical, descriptive of how things have gone; however, a few years later, he wrote *The Social Contract*, theorising a new social pact, this time with a normative purpose to fix the flaws of the historical one and aiming at a fairer society.

[9] Though having introduced the social contract theory in the modern philosophical debate, Hobbes takes up this theory from Lucretius' *De rerum natura*, where the state of nature is described as a "life [led] in the wide-roaming way of wild beasts" ("*more ferarum*"). Lucretius, *De Rerum Natura*, ed. Monica R. Gale (Oxford: Oxford University Press, 2009), 74–75.

Lucretius then elaborates the passage from this "wild life", where individuals live isolated lives and have more contacts with wild animals than with other humans, to a social life, where contacts between humans intensify and, as we shall see, so does conflict, generating the need for the establishment of a social pact to regulate it. Lucretius, 80–81.

David Hume was the first to point out that a historical pact would not be binding for present-day citizens[10]. This criticism emphasizes an essential feature of social contract theories: their significance does not lie in their being good historical accounts. Were they meant to be accurate historical accounts, Hume's observation would undermine their relevance to contemporary political philosophy. But following Rousseau's path traced in his *Social Contract*, subsequent philosophers, including many present-day thinkers, continue to think of political theory in terms of social contracts by virtue of such theories' solid explanatory power with regard to the role and the function of the State (as is the case for in John Rawls' *A Theory of Justice*, where the so-called "original position" corresponds to the state of nature[11]). Therefore, it does not matter whether the historical account is accurate – which, of course, it is not since the assumption of a rational agreement between previously isolated and mostly hostile individuals is clearly fictional. What remains unaltered in all accounts is the reason the pact is stipulated, hence the fundamental role of the State these theories highlight: the resolution of violent conflicts. According to Lucien Scubla,

> if judicial power comes in the third position, after legislative and executive power, in our conception of right, it is actually the first one. There were judges […]

[10] Hume does not merely criticise social contract theories of the justification of political power, but proposes a positive theory that posits violence and usurpation as the origin of the political regimes of his time. "If the agreement, by which savage men first associated and conjoined their force, be here meant, this is acknowledged to be real; but being so ancient, and being obliterated by a thousand changes of government and princes, it cannot now be supposed to retain any authority. But besides that this supposes the consent of the fathers to bind the children, even to the most remote generations (which republican writers will never allow) besides this, I say, it is not justified by history or experience, in any age or country of the world. Almost all the governments, which exist at present, or of which there remains any record in story, have been founded originally, either on usurpation or conquest, or both, without any pretence of a fair consent, or voluntary subjection of the people. […] By such arts as these, many governments have been established; and this is all the original contract, which they have to boast of." David Hume, "Of the Original Contract," in *Political Essays*, ed. Knud Haakonssen (Cambridge: Cambridge University Press, 1994), 189–90.

In Hume's theory, once the government is violently established, its subjects consent to it tacitly because they are accustomed to it: custom is, in fact, the basis of morality, that founds the duty of allegiance, i.e., the submission to the government that grants the respect for the interest in the stabilisation of property of human societies. Hume, 193–97.

For a detailed analysis of the connection between the violent origin of the government, the duty of allegiance, moral obligation and human interests see: Philippe Saltel, *Hume : Oberver, Philosopher* (Paris: Ellipses, 2023), 81–89.

[11] "In justice as fairness the original position of equality corresponds to the state of nature in the traditional theory of the social contract. This original position is not, of course, thought of as an actual historical state of affairs, much less as a primitive condition of culture. It is understood as a purely hypothetical situation characterized so as to lead to a certain conception of justice." John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1999), 11. Of course, despite this formal similarity, the original position differs in content from the state of nature, as the former aims to account for the choice of specific principles of distributive justice, while the latter aims to explain the origin of the state.

> before there were laws. Likewise, there were judges before there was a public power to enforce their sentences.[12]

If social contract theories are not to be considered historical theories, another theory proposed by René Girard may provide an account closer to how things have gone in the deep past. As previously noted, it is hard to reach definitive conclusions concerning prehistoric facts, but archaeological and ethnographic observations may corroborate certain theories. It seems to be the case for what concerns Girard's victimary theory.

1.2 Earthly Justice

1.2.1 The Sacrificial System

The basic structure of Girard's victimary theory follows that of Hobbes' social contract theory. Girard's starting point, in fact, is comparable to a Hobbesian state of nature of war of all against all[13]. But while Hobbes had stated conditionally that "*if* any two men desire the same thing, which nevertheless they cannot both enjoy, they become enemies"[14], according to Girard, two people necessarily desire the same thing because of how their desiring mechanism works. Desire, as Girard discovers through the analysis of the works of great literary authors such as Cervantes, Dostoevsky, and Proust[15], is fundamentally *mimetic*: we do not desire an object in and for itself but because we think some other desires it[16]. Mimetic desire is also called "triangular

[12] Lucien Scubla, "Sur Une Lacune de La Théorie Mimétique : L' Absence Du Politique Dans Le Système Girardien," *Cités*, no. 53 (2013): 114–15, my translation. Scubla continues: "Among the Nuer, this power or rather function is held by the leopard-skin chiefs, who are ritual mediators, external to the segmental structure constituting their political system. Like our judges, they can only intervene at the request of the parties involved; unlike our judges, they cannot require binding power to impose their decisions, they can only threaten religious sanctions and make expiatory sacrifices." Scubla, 115. The binding force of the judge's authority, in the absence of an executive independent power, only derives from the parties' request for the judge's intervention in the dispute. Their power is entirely symbolic, yet its consequences are real, even if they do not take the form of the direct use of force as in contemporary societies, but rather result in forms of de-solidarisation within the group.

[13] They are comparable but not identical, as observed by Mormino: "The object of [Girard's] reflection is not […] an hypothetical original state of nature but the situation of archaic societies, that fight to survive with instruments already forged by previous generations, the remote origin of which, similar in all the world's populations, is lost in a time that is not documented by any text. It is, however, true that the mechanism by which a society recasts itself after a crisis may shed light on how societies may have first been formed, thus on the absolute birth of the human." Gianfranco Mormino, *René Girard. Il Confronto Con l'altro* (Roma: Carocci, 2012), 143, my translation.

[14] Hobbes, "Leviathan," 135.

[15] René Girard, *Deceit, Desire & the Novel. Self and Other in Literary Structure*, trans. Yvonne Freccero (Baltimore: The Johns Hopkins University Press, 1965).

[16] Mormino criticises Girard's omission of the philosophical contributions to the topic of mimetic desire. Girard, in fact, considers philosophical thought entirely *sacrificial*, in the sense that will be shortly

desire" by Girard, as it does not constitute a linear relation of a subject with an object, but rather a triangular relationship between a desiring subject, the desired object, and what Girard calls the "mediator", the other subject to whom the original desire for the object is attributed. Subjects do not desire the object *per se* but as a means to become like another person that is considered to own that "metaphysical autonomy"[17] that they lack and is, therefore, worth imitating: the mediator, the rival.[18] This mechanism of desire is then abstracted by Girard from this literary context and used as the basis for his anthropology, characterised by the mimetic nature of human beings.

While this explanation of the mimetic nature of desire perfectly fits the Girardian analyses of modern literature, the extension of the mimetic theory to general anthropology requires further generalisation. In *Violence and the Sacred*, Girard connects the human inclination to mimesis to the experience of learning, specifically through the relationship between a master and a disciple[19]. Mormino develops this intuition, considering imitation as an extension to a perceived "other" of the more fundamental learning mechanism of *self-imitation*[20].

Following Hobbes's conceptual path, it is clear that such an anthropology, characterised by this mimetic inclination, will generate diffuse hostility and, eventually, generalised war. This is what Girard calls a mimetic crisis, in which violence becomes itself the object of mimesis, and communities are torn apart, all social bonds being broken. This mimetic character of violence can be traced back to Lucretius, whose state of nature (*more ferarum*) can be distinguished from Hobbes's as a rather peaceful state where humans live an isolated life and only the incidental fact of coming to live in communities led to the explosion of violence and the subsequent need for a social contract.[21] Differently from such theories, according to Girard, a mimetic crisis is

---

defined, fundamentally incapable of recognising the mimetic nature of human desire. Some connections between Girard's and Hobbes' thoughts have already been pointed out, but the great absent in Girard's dissertation on mimetic desire is Spinoza, who treats the topic of imitation and its twofold effects, generating both cohesion and rivalry, at length in his *Ethics*. This failure to meet with the philosophical tradition is ascribed by Mormino to the dangers of the materialistic philosophy of authors such as Hobbes and Spinoza to Girard's theological aspirations. Mormino, *René Girard. Il Confronto Con l'altro*, 31; 101–12.

[17] Mormino, 58.

[18] Girard, *Deceit, Desire & the Novel. Self and Other in Literary Structure*, 1–52; René Girard, *Violence and the Sacred*, trans. Patrick Gregory (Baltimore and London: The Johns Hopkins University Press, 1977), 145–46.

[19] Girard, *Violence and the Sacred*, 146–47.

[20] Gianfranco Mormino, *Per Una Teoria Dell'imitazione* (Milano: Raffaello Cortina Editore, 2016).

[21] Or, in Lucretius' thought, multiple subsequent social contracts. Lucretius, *De Rerum Natura*, 80–81, 86–89.

resolved when the war of all against all is turned into the war of all against one[22]: individuals unite against one single victim, whom Girard calls a "surrogate victim"[23], that is considered the cause of the crisis and collectively killed as if they were the root of all evil: a scapegoat.

Girard corroborates his theory in *Violence and the Sacred*, analysing ancient founding myths, which he considers to be proof of such killings. Of course, since such narratives are given by the killers themselves – the community founded by that very act of killing – they cannot be considered impartial accounts of what happened but are marked by misconceptions that must be unveiled and exposed[24]. The killers cannot tell the story of the arbitrary killing of an innocent victim; they must conceal the happenings with an acceptable disguise. This is not meant to be a deliberate act of deceit, though: the killers deceive themselves, too, and believe their own narratives; this same misconception is a fundamental condition for the functioning of the victimary system. Because they think they have eliminated an evil being that was troubling the community, they can find the community pacified once again after the killing. Therefore, they can kill their victim without bearing the guilt of murder, the killing consisting of the necessary elimination of a demon[25].

According to Girard, this mythical misrepresentation deceives Freud too. In fact, in *Totem and Taboo*, he comes close to the discovery of the victimary system, where he describes the birth of the "first" super-ego in the context of the so-called *primal horde*, "a more or less organized group of limited size"[26], a family group ruled by a tyrannical father who metaphorically castrates the other males through his monopoly of the females of the group. The brothers thus unite and kill the father but, finding themselves deprived of prohibitions to regulate their desires, now risk the disintegration of the group because of sexual rivalries and, hence, establish the law against incest as a substitute for the father's prohibition[27]. Thus, from this killing, the super-ego, sacrificial

[22] René Girard, *I See Satan Fall Like Lightning*, trans. James G. Williams (Maryknoll, New York: Orbis Books, 2001), 22.

[23] Girard, *Violence and the Sacred*, 2.

[24] René Girard, *Things Hidden Since the Foudation of the World*, trans. Stephen Bann and Michael Metteer (Stanford: Stanford University Press, 1987), 148; Girard, *I See Satan Fall Like Lightning*, 110; Lucien Scubla, "L'impératif Sacrificiel et Ses Masques," *L'Homme*, no. 156 (2000): 244.

[25] Scubla, "L'impératif Sacrificiel et Ses Masques," 242.

[26] Sigmund Freud, *Totem and Taboo. Some Points of Agreement between the Mental Lives of Savages and Neurotics*, trans. James Strachey (London: Routledge, 2011), 145 n3.

[27] "Sexual desires do not unite men but divide them. Though the brothers had banded together in order to overcome their father, they were all one another's rivals in regard to the women. Each of them

rites, religion, moral laws, and, eventually, all culture originate[28]. However, as Girard notices, Freud cannot make the final step to understanding the secret of the victimary system, i.e. the innocence of the victim: his account is marked by the typical mythical misconception that considers the victim of the founding homicide guilty, an "oppressive monster"[29].

In Girard's theory, the pacification that follows the founding murder leads to the sacralisation of the victim[30]: they were at once the origin of the great evil of the mimetic crisis and the cause for the subsequent reconciliation, which is the greatest good for the community; therefore, they must be an all-powerful entity: a God. Girard can extrapolate this mechanism from many ancient myths, but it is in "the horrible miracle of Apollonius of Tyana"[31], as narrated by Philostratus, that he identifies a "missing link"[32] between ancient incomprehensible mythical transfiguration and later more transparent episodes, such as the transfiguration of witches during the Middle Ages or the Jewish conspiracies, narrated in what Girard calls "text of persecution", that allows to connect the phenomena and decipher the mechanism underlying ancient myths.

According to Philostratus, Apollonius of Tyana, known in the II century A. D. for being able to perform miracles, is called by the city of Ephesus, which is going through a plague epidemic. Epidemics are not only medical emergencies but social ones: when people start to fear death in their daily activities, they fail to fulfil their duties and obligation in pursuit of personal safety, and the ecosystem of mutual cooperation in the city collapses[33]. Apollonius of Tyana must restore the city's unity into a community. He, therefore, chooses a surrogate victim who displays the typical

---

would have wished, like his father, to have all the women to himself. The new organization would have collapsed in a struggle of all against all, for none of them was of such overmastering strength as to be able to take on his father's part with success. Thus the brothers had no alternative, if they were to live together, but – not, perhaps, until they had passed through many dangerous crises – to institute the law against incest, by which they all alike renounced the women whom they desired and who had been their chief motive for despatching their father." Freud, 167.

[28] Freud, 164–70.

[29] Girard, *Violence and the Sacred*, 204.

[30] Whereas in Freud's "mythical" account such sacralisation (the transposition of the father into the totem) derives not from the homicide itself but from the previous admiration for the father. Freud, *Totem and Taboo. Some Points of Agreement between the Mental Lives of Savages and Neurotics*, 166.

[31] Girard, *I See Satan Fall Like Lightning*, 49.

[32] Girard, 73.

[33] I suggest this interpretation of Ephesus' crisis, but it must be pointed out that it differs from Girard's own interpretation. According to him, there is no infectious disease epidemics whatsoever in Ephesus, the term "plague" refers to the mimetic crisis itself. Girard, 52.

scapegoat features[34]: an old blind beggar, thus a weak, lonely, marginalised person whom probably nobody would avenge. Apollonius then tells the Ephesians that the beggar is the origin of their problems: though he presents himself as a poor old man, he is, in truth, a demon they should kill. After a moment of hesitation, the beggar is stoned to death by the whole community of Ephesians; removing the stones after the killing, they see the old man has transformed into a demon. The community is unified against a common enemy and is therefore pacified[35]. This story needs no further explanation: the victimary mechanism is clearly visible and can be taken as a model to analyse other myths as well as sacrificial practices.

The old beggar's weakness and marginalisation are fundamental features of a scapegoat: were the victim integrated into a web of social bonds, the killing could not be unanimous, some of the potential killers would stand in favour of the scapegoat, exacerbating social division, and, should the killing be carried out nonetheless, another cycle of violence would be inaugurated, having the exact opposite outcome from that which was meant to be achieved.

Girard compares this story to the biblical *Pericope Adulterae*, the story of the woman who is about to be stoned because she has committed adultery[36]. In that story, too, a crowd is hesitant about the stoning of a victim, but Jesus, as opposed to Apollonius of Tyana, defuses the victimary system. While Apollonius incites the crowd so that someone would throw the first stone at the beggar – after which all other stones will follow, imitating the first act of violence – Jesus manages to prevent the throwing of the first stone by focusing everyone's attention on it: "let him who is without sin cast the first stone". He thus invests the first one to cast a stone with responsibility for the action instead of trying to dissimulate it in the ensuing frenzy. Contrary to what happens in Ephesus, as a result, no stone is cast, and "those who heard began to go away one at a time"[37]: instead of uniting the group against the victim, the crowd is divided into individuals and dispersed.

[34] For a complete account of scapegoat features, refer to: René Girard, *The Scapegoat*, trans. Yvonne Freccero (Baltimore: The Johns Hopkins University Press, 1986), 12–23.

[35] Girard, *I See Satan Fall Like Lightning*, 49–50.

[36] Girard, 55–61.

[37] Gospel of John, 8:9.

Girard analyses another biblical passage concerning this mechanism of victimisation through imitation: the episode of Peter's denial[38]. Peter finds himself in a group hostile to Jesus after his arrest, breaking their sense of community[39]: when questioned about his relations to Jesus, he denies knowing him three times. To avoid being discriminated against, Peter imitates the others' discrimination; the identification of a common enemy seals the community.

Girard also finds these elements in the Greek myth of Oedipus[40] that, according to his analysis, is to be considered a substitute victim necessary to put an end to the mimetic crisis that is going on in Thebes, here again, afflicted by the plague. Once Oedipus is expelled, Thebans are reconciled, but this is not the end of the story: Oedipus, in fact, goes to Colonus. There, he is considered both terrifying and very precious, his corpse eventually becoming a sort of talisman for the city[41]. Once more, we witness the sacralisation of the victim, who is considered to have the power both to destroy and to save the city.

This latter element is also partly present, as we have seen, in the case of the beggar of Ephesus and again in innumerable myths, such as the death of Romulus, founder of Rome and fratricidal murderer at the same time. Livy presents his death in two different versions: "According to the first, a storm broke out while Romulus was inspecting the army, and he was raised directly to heaven, where he became Quirinus, god of war"[42]: his death is mysterious and followed by a sacralisation.

> According to the second, during the course of the same storm, he was the victim of a collective murder: he was assassinated by the assembly of senators. Already modern, Livy was surprised by the naïveté of his fellow Romans, who so easily accepted the blatant lie invented by the senators in order to hide their crime.[43]

This second version complements and clarifies the mysterious death of the first: Romulus was killed by the senators, who misconceived and misrepresented the

[38] René Girard, "Peter's Denial and the Question of Mimesis," *Notre Dame English Journal* 14, no. 3 (1982): 177–89.

[39] Girard, 180.

[40] Girard, *Violence and the Sacred*, 68–88. Contrary to the myth itself, according to Girard, Sophocles' tragedy introduces elements of doubt concerning the guilt of the characters and therefore their victimisation, constituting a primal element in the interpretation of ancient myths as persecution texts.

[41] "Talisman" in the original text (René Girard, *La Violence et Le Sacré* (Paris: Bernard Grasset, 1972), 125.), rendered as " valuable relic" in the English translation. Girard, *Violence and the Sacred*, 85.

[42] Paul Dumouchel, *The Barren Sacrifice. An Essay on Political Violence* (East Lansing: Michigan State University Press, 2015), 17.

[43] Dumouchel, 17.

homicide, and he was later sacralised by virtue of his peace-bearing function as a scapegoat.

Sacralisation brings us to the last element of Girard's theory that will be analysed herein. We have talked about mimetic desire and mimetic crises, surrogate victims and scapegoats; their sacralisation leads us to consider the link between *violence and the sacred*, that is, sacrifice.

Once communities overcome mimetic crises, they manage to remain stable thanks to the misinterpretation that saw the surrogate victim as supremely guilty and supremely powerful. Not only is the scapegoat transformed into a God, but the act that originally pacified the community, the founding homicide, is (rightly) believed to be the secret for mutual reconciliation and is, therefore, periodically repeated. The origin of sacrifice is nothing other than the periodical repetition of the original killing[44]. Girard thus manages to explain the (virtual) omnipresence of sacrifice in all cultures that developed independently around the world. After the founding homicide, sacrificial violence kept the communities united, once again against a common victim: the scapegoat.

This realisation allowed Girard to extend his theory not only beyond the borders of Western culture but beyond the limits of literary texts, moving on to the analysis of practices observed by anthropologists and ethnologists all over the world. The analysis of the sacrificial practices of the most different societies led him to the finding of the common elements of the scapegoating mechanism everywhere. In this way, Girard gives an explanation in terms of usefulness – or, rather, necessity – for sacrifices and violent rites in general, which are typically dismissed by modern thought as criminal aberrations[45], in a unitary theory that comprises modern western literature, ancient myths, biblical accounts, and rites observed in the most diverse societies.

Just like the founding homicide is misconceived and therefore misrepresented in its mythical account, so are the sacrificial rites that re-enact it. Thus, despite sacrificial communities tend to represent their rites as "offerings" or "communions" with the Gods, as Scubla maintains, "a good theory must not explain the practices of men by the beliefs that come with them, it must go back to the common causes of them both"[46].

[44] Girard, *Violence and the Sacred*, 95.
[45] Girard, 26.
[46] Lucien Scubla, "Le Sacrifice A-t-Il Une Fonction Sociale ?," *Pardès* 39, no. 2 (2005): 145, my translation.

Thus, Scubla follows Girard in taking the functionalistic approach inaugurated by Durkheim, synthesised in the following passage from *The Elementary Forms of Religious Life*:

> No doubt, sometimes an error does indeed perpetuate itself in history. But barring an altogether unusual conjunction of circumstances, it cannot maintain itself […] unless it proves to be *practically true* – that is to say, if, while not giving us a correct theoretical idea of the things to which it is related, it expresses correctly enough the manner in which those things affect us, for better or for worse.[47]

Epicurus had already taken apart the idea of the offering to the Gods, pointing out that being immortals and perfectly happy, Gods would not need any human offering whatsoever. On the contrary, they are already masters of everything, and therefore nothing can be offered to them that is not theirs already[48]. Still, such misconceived interpretation captures a critical element of the victimary theory, namely, the feeling of indebtedness towards the Gods, that, according to Scubla, derives from the fact that

> men do not owe their collective salvation to their will or their reason, like the philosophers of the Enlightenment would think, but to a mechanism that transcends them and to the religious institutions that stem from it. Institutions that made men much more than they have made the institutions themselves: a principle that constitutes the common denominator of all religions.[49]

Since the very function of the rite is misconceived by those who perform it, and since rites undergo the normal progressive transformation of cultural practices as time passes, some elements that were initially fundamental might disappear in certain practices, and others that are accessories to the primary function may become the central feature of the rite. Still, without aspiring to find exact replicas in sacrificial rites around the world and during different ages, some common features can be identified that testify to a common functional origin[50].

Sacrificial victims are typically situated between the inside and the outside of the community, just like the old beggar in Apollonius' tale, an Ephesian, and yet a marginalised one. Such a feature is valid for both human and animal sacrificial victims. As for the humans, one of Girard's most interesting examples is that of the Tupinamba, an Amazonian population that finds its victims among prisoners of war, hence

[47] Emile Durkheim, *The Elementary Forms of Religious Life*, trans. Karen E. Fields (New York: The Free Press, 1995), 77.
[48] Scubla, "Le Sacrifice A-t-Il Une Fonction Sociale ?," 146.
[49] Scubla, 154, my translation.
[50] Girard, *Violence and the Sacred*, 103–4.

individuals external to the community. Still, after their capture, these prisoners live in their captors' community as integrated members for months, even marrying local women, thus being partly internalised by the community. Only then are they sacrificed, killed, and eaten in a cannibalistic rite[51].

Not being considered fully-fledged community members, women and children can also be good sacrificial victims if no man is believed to have reason to avenge them[52]. This was the plan behind Iphigenia's sacrifice at the hands of her father Agamemnon, who, according to the mythological account, needs to kill her to appease Artemis, that would not otherwise let his ships set sail to Troy. Being both a woman and a child under his jurisdiction, nobody would seem apt to avenge her, but, according to Sophocles' *Oresteia* tragic trilogy, this belief is proven wrong once he arrives back in Mycenae after the end of the war. Clytemnestra, his wife, proves full agency and organises his homicide, avenging their daughter, giving rise to a long cycle of blood feuds within the family. She will be assassinated by her son Oreste, who is, in his turn, persecuted by the Erinyes, the divinities of vengeance.

As for the animal sacrifices, they must be regarded as an evolution from a more original practice of human sacrifice, which more closely mimics the founding murder. If one manages to internalise the animal victim enough, animals provide an excellent substitute for human victims, for they surely cannot retaliate[53]. Girard talks about the bovine sacrifices of the Nuer, an African population that chooses its sacrificial victims among its farmed cattle. The Nuer can recognise each individual bovine, they know their ancestry up to the fifth generation, and they create an imaginary bovine society parallel to their own, with which they live in symbiosis[54]. The animals are, therefore, internalised in the community through the analogy of their social structures. The same can be said for the Dinka, who sacrificed animals after having isolated them from their herd and moved them closer to the community's homes, thereby internalising them[55].

---

[51] Girard, 274.

[52] Girard, 12–13. In *The Scapegoat* "the weakness of women" is considered one of the "signs of victims" (*signes victimaires*), see: Girard, *The Scapegoat*, 21.

[53] The need to have a "reservoir" of possible sacrificial victims is, according to Girard, at the origin of animal domestication. René Girard, Pierpaolo Antonello, and João Cezar de Castro Rocha, *Evolution and Conversion. Dialogues on the Origins of Culture* (London and New York: Continuum, 2008), 116–17.

[54] Girard, *Violence and the Sacred*, 3.

[55] Girard, 272–73.

Practices that proceed in the opposite direction are found by Girard in African sacred monarchies[56]: there, the sacrificial victim is internal and must be externalised. The victim is the sacred king, even though the king himself is often spared, and a surrogate of his is killed instead. He is part of the community, and the externalisation process works through his performance of heinous crimes: acts of violence, breaking of taboos, notably committing incest. According to Girard, these are all acts meant to make the king "other", therefore suitable for sacrifice.

Apart from their duplicity consisting of being both internal and external to the community, sacrificial victims often share the duplicity of being both honoured and scorned. It is the case of the rituals of the Dinka[57], the Tupinamba[58], the sacred monarchies, but also, as suggested by Girard at the end of his analysis of Euripides' *The Bacchae*, of the role of women in Greek mythology[59].

The essential feature for sacrifices like for founding murders, identified by Girard in a wide variety of rites and myths, is the *unanimity* of violence against the sacrificial victim: as for the myths, it is evident in the Hindu Yadjur-Veda[60], the Indonesian Hainuwele[61], the blinding of the cyclops in the *Odyssey*[62], the previously cited biblical episode of Peter's denial[63], Euripides' *The Bacchae*[64], among others. Concerning the rites, this feature is evident in so many "that there's no need to list them all"[65]. Some examples are the Ngadju-Dayak (Borneo) sacrificial rites[66], the Kaingang's (Amazon)[67], the Swaziland's sacred monarchy's (Africa)[68], the Dinka's (Africa)[69], the Arabic sacrifice of the camel[70] and the Dionysian *sparagmos* (Ancient Greece and Rome)[71].

---

[56] Girard, 104–18.
[57] Girard, 98.
[58] Girard, 274–75.
[59] Girard, 141–42.
[60] Girard, 100.
[61] Girard, 100.
[62] Girard, 100.
[63] Girard, "Peter's Denial and the Question of Mimesis."
[64] Girard, *Violence and the Sacred*, 131.
[65] Girard, *La Violence et Le Sacré*, 145. Passage omitted in the English translation, see: Girard, *Violence and the Sacred*, 100.
[66] Girard, *Violence and the Sacred*, 100.
[67] Girard, 101.
[68] Girard, 110.
[69] Girard, 110.
[70] Girard, 100.
[71] Girard, 100, 131.

In this ritualised sacrificial repetition lies the birth of religion. The very etymology of the word "sacrifice", from "*sacer*" and *"facere*", means "creating the sacred", the elementary object of religion. In his *Dictionary of Indo-European Concepts and Society*, the French linguist Émile Benveniste refers to another word that was used in Latin for the verb "to sacrifice", that is, *mactare*[72]. Interestingly, this word remained in Latin languages with the simple meaning of "to kill" as in Spanish (*matar*) or in the Italian form *mattatoio* ("slaughterhouse"). Benveniste follows Hubert and Mauss in explaining this connection between the making of the sacred and the killing of a living being:

> It shows that the sacrifice takes place so that the profane world can communicate with the divine world through the priest and by means of the rites. To make the animal 'sacred', it must be cut off from the world of the living, it has to cross the threshold which separates these two universes; this is the point of putting it to death.[73]

However, Girard's connection between the two seems more straightforward, explaining their *intrinsic* rather than accidental connection. Girard's account also explains the original meaning of the German word *heilig*, English *holy*, from the Gothic *hails*: "safety, health, physical and corporal integrity"[74]: in fact, according to Girard's theory, safety is precisely what is brought to the community after the mimetic crisis by the founding homicide and the subsequent sacrifices.

Finally, Burkert refers the etymology of the German word for "sacrifice", i.e. *Opfer* to the Latin *operari*, simply "to act", therefore identifying sacrificial killing as the primary form of human action: "Sacrificial killing is the basic experience of the 'sacred'. *Homo religiosus* acts and attains self-awareness as *homo necans*. Indeed, this is what it means 'to act', *ρεζειν*, *operari* (whence 'sacrifice' is *Opfer* in German) […]."[75]

This fundamentally agential character of religion, reflected in Burkert's etymological analysis, is consistent with Scubla's functionalistic approach that, evoking Durkheim, defines religion as "a system of prohibitions and collective rites that implies a net separation between profane things and sacred things, that, in various forms, seems

[72] Émile Benveniste, *Dictionary of Indo-European Concepts and Society*, trans. Elizabeth Palmer (Chicago: HAU Books, 2016), 490.

[73] Benveniste, 461.

[74] Benveniste, 459.

[75] Walter Burkert, *Homo Necans. The Anthropology of Ancient Greek Sacrificial Ritual and Myth*, trans. Peter Bing (Berkeley and Los Angeles: University of California Press, 1983), 3.

to have universal extension"[76]. These practical features are quite more fundamental than the theoretical and theological elaborations that make religions, in common thought, primarily systems of beliefs. This is why the Enlightenment approach that we have inherited, which criticises religion on the intellectual level of truth versus falsity, proves to be arid: it doesn't understand religion's fundamental function[77].

What is the sacred that is created by the sacrifice? After the analysis conducted so far, we can now understand that what is projected on the sacrificial victim, in a repetition of the original projection on the founding homicide's victim, is the community's violence itself, "externalised, reified, and kept at a distance"[78]. According to Girard and Scubla, this sheds new light on the biblical story of Cain and Abel. Of the two brothers, the latter gives God well-accepted animal sacrifices, while the former gives him some vegetable offers that God refuses. Cain, contrary to his brother, is not canalising his violence anywhere, and this is dangerous: missing a sacrificial victim, it will be unleashed on a member of the community, risking the ignition of a mimetic crisis; that is why God refuses his non-bloody offerings, and why Cain, transported by anger, ends up killing his brother[79]. It is also why God protects him thereafter, establishing the law against murder[80]: the murder that is now forbidden is that of Cain, who might be a victim of revenge. Finally, it is also why his children are all to be civilising heroes, creating farming, music, and metallurgy, as well as great founders of cities[81].

---

[76] Lucien Scubla, "Les Hommes Peuvent-Ils Se Passer de Toute Religion ? Coup d'oeil Sur Les Tribulations Du Religieux En Occident Depuis Trois Siècles," *Revue Du MAUSS* 22, no. 2 (2003): 91. However, Benveniste remarks that the linguistic distinction between sacred and profane is introduced by Latin and is not present in ancient Greek, ancient Germanic languages: Benveniste, *Dictionary of Indo-European Concepts and Society*, 459, 460, 469.

[77] Scubla, "Les Hommes Peuvent-Ils Se Passer de Toute Religion ? Coup d'oeil Sur Les Tribulations Du Religieux En Occident Depuis Trois Siècles," 95.

[78] Scubla, "Le Sacrifice A-t-Il Une Fonction Sociale ?," 154, my translation.

[79] It is worth pointing out here that Scubla also remarks that two Hellenistic cults, the Orphic and the Dionysian one, apparently situated on the opposite sides of the spectrum of violence, since Orphic communities are non-violent and vegetarian, whereas the Dionysian rite, the *sparagmos*, is extremely violent, can be considered the two sides of the same coin. Not only do they both reject the official sacrificial rites of the city, situating themselves outside of it, but they address the same divinity, Dionysus: "as if, in the final analysis, the only possible choice was between a violence restrained and civilised by ritual and a savage, naked violence." Lucien Scubla, "Fuir Vers Une Terre sans Mal et sans Violence ?," *Revue Du MAUSS* 55, no. 1 (2020): 181, my translation.

[80] Girard, *I See Satan Fall Like Lightning*, 84. A similar path is described in the aforementioned *Oresteia*: the revenge cycle is ended by the institution of a Greek criminal court.

[81] Scubla, "L'impératif Sacrificiel et Ses Masques," 242; Scubla, "Le Sacrifice A-t-Il Une Fonction Sociale ?," 157.

It is interesting noticing here, with Mark Anspach, that the Moussey, a population of central Africa, interrupts the cycles of vengeance that undermine their communities with an animal sacrifice offered to the god of Vengeance[82]. This practice plainly illustrates the process that generates divinity and what such divinity consists of: the externalisation and reification of the origin of social disorder[83]. Vengeance is, as was evident in the previous account of the *Oresteia*, the most dangerous form of violence in a community, for it extends in space and time, infecting other community members through mimesis and persisting over multiple generations on feud cycles, possibly driving the community to self-destruction[84], "as a slight cut can prove fatal to a haemophiliac"[85]. Sacrifice, on the other hand, is "violence without risk of revenge"[86].

It is no accident that the Greek divinities of Vengeance, the aforementioned Erinyes, or Furies, are also called "Eumenides", i.e., "the benevolent ones". This double identity should not be surprising now: the divinities of vengeance are benevolent because they are the externalisation of vengeance itself, its reification, and expulsion from the community. Their creation is the peacemaker, without which societies are bound to succumb to reciprocal vengeful violence.

The topic of vengeance in myths, religion, and sacrificial rites, is, in fact, recurrent. Rites are due to the Gods, or they will seek revenge; rites are also performed to interrupt divine violence as it is taking revenge for some human fault. But in most sacrifices, both animal and human (cannibalistic), the most precious element of the sacrificed victim, namely, its flesh, is not actually given to the Gods but consumed by the community instead[87]. As Scubla remarks in his analysis of the Greek sacrifice of the ox, this fact shows the real function of the sacrifice: "The ceremony periodically restores the great division that opposes the immortals embalmed with sacrificial

[82] Mark Rogin Anspach, *À Charge de Revanche. Figures Élémentaires de La Réciprocité* (Paris: Seuil, 2002), 18.

[83] Vengeance is certainly the main and most dangerous element of social disorder, but not the only one. As Girard notes "It should be no cause for surprise […] that the august inhabitants of Olympus have been charged with a fair number of rapes, murders, parricides, and incestuous acts, as well as innumerable incidents of savagery and madness." Girard, *Violence and the Sacred*, 253.

[84] As a concrete example of this process of self-destruction, Girard cites a group of the Amazonian Kaingàng people that, as witnessed by the American Anthropologist Jules Henry, after the Brazilian intrusion lost its capacity to regulate internal vengeance and became incapable of polarising violence against external enemies. Girard, 52–53.

[85] Girard, 18.

[86] Girard, 13.

[87] This element was already observed by Epicurus, as remembered by Scubla: Scubla, "Le Sacrifice A-t-Il Une Fonction Sociale ?," 146.

fragrances, to the mortal meat-eaters. The fellow citizens are fellow sacrificers"[88]. It is the community, not the Gods, which is revealed to be "blood-thirsty", and which periodically needs to kill a surrogate victim to defuse reciprocal violence and feud cycles, eating their flesh in a shared meal, thus sealing its unity.

Given what has been illustrated so far, we can now assert that way before the establishment of States, before any imaginary subscription to a social contract, violence was dealt with through religion. This is the heart of the victimary theory.

Recent archaeological discoveries in present-day Turkey, analysed by Girard, Jean-Pierre Dupuy, Paul Gifford, and Pierpaolo Antonello, among others, seem to corroborate this thesis of the fundamental function of religion in archaic societies. Girard analysed some murals found in the site of Çatalhöyük[89], considered one of the biggest (hosting up to 8000 people) and best-preserved Neolithic settlements, inhabited between 7500 BC and 6400 BC. During the Neolithic, human groups started to live sedentarily, as opposed to the previous migrant lifestyle of foragers. Of course, the process that led to the spread of sedentism is long and non-linear, and not all human groups engaged in it. But according to the archaeological evidence, it was during the Neolithic that some groups of humans undertook it, consequently developing agriculture and animal domestication. The remains of Çatalhöyük are the remains of one such group that settled in Anatolia.

The murals there found, visible in the reconstructions displayed below, represent the killing of some animals. According to Girard, what might at first be interpreted as hunting scenes is revealed to represent some sacrificial killings. Girard infers it from some peculiarities of these drawings, such as the erect penis of the animals[90], the presence of women (in Figure 2 in an explicitly provocative pose), the almost complete absence of weapons, except for a few bows, but without any arrow, that would probably have been useless against the thick skin and fur of those animals anyway[91]. Other remarkable elements are the presence of severe injuries, like amputated limbs and heads, that deer would hardly cause[92] since they tend to flee rather than engage in fights,

[88] Scubla, "Fuir Vers Une Terre sans Mal et sans Violence ?," 180, my translation.

[89] René Girard, "Animal Scapegoating at Çatalhöyük," in *How We Became Human. Mimetic Theory and the Science of Evolutionary Origins*, ed. Pierpaolo Antonello and Paul Gifford (East Lansing: Michigan State University Press, 2015), 217–31.

[90] Girard, 221.

[91] Girard, 225.

[92] Girard, 226.

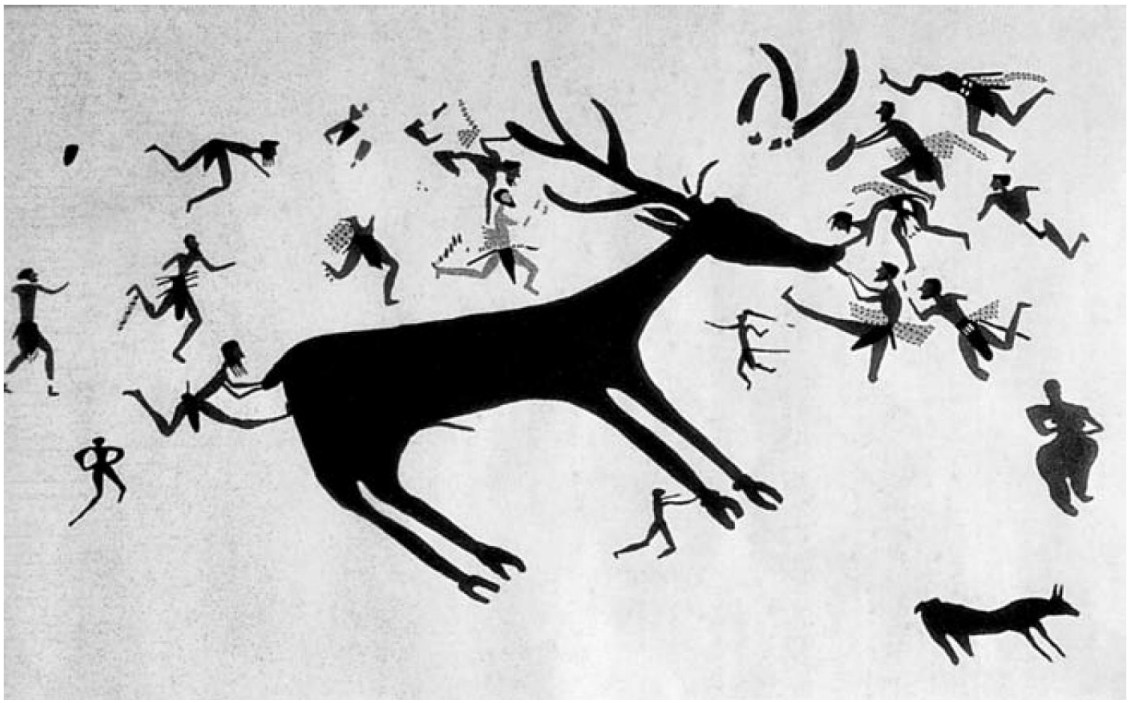

Figure 1: Reconstruction of a Neolithic mural from Çatalhöyük showing a stag surrounded by sacrificers. © Alan Mellaart.

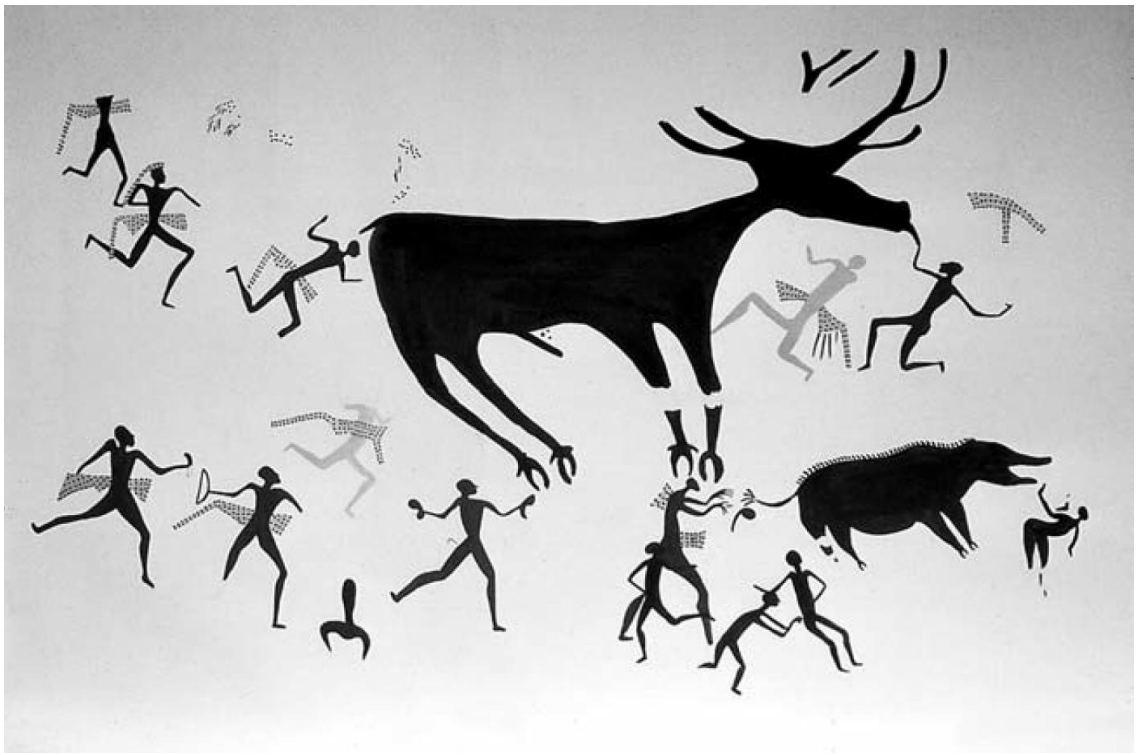

Figure 2: Copy of a Çatalhöyük mural showing a boar and a deer surrounded by sacrificers. Photograph: Omar Hoftun. © Alan Mellaart.

and their exaggerated size in comparison to the size of humans. Both these latter features, according to Girard, serve the purpose of depicting the animals as more threatening than they really are. This menacing trait is crucial to the sacrificial interpretation of the drawings, for they transform a peaceful animal into a good scapegoat that can take upon itself the violence of the community, pacifying it by means of its destruction. Finally, in both representations, the animal is attacked on its front legs, therefore showing a form of ritualised killing whose procedure is repeated among different occurrences[93]. The attack on the legs makes the animal fall down, facilitating its lynching, one of the basic forms of collective killings, attested in many other sacrificial rites mentioned before, such as the Arabic sacrifice of the camel or the Dionysian *sparagmos.*

A further element that bolsters the sacrificial interpretation of these representations is the presence of horns in the tombs of the ancestors found in Çatalhöyük: from menacing entities, after the sacrifice is complete and the community is once again pacified through the collective killing, these animals become protectors, through the typical sacrificial transfiguration discussed above[94]. Ian Hodder, the archaeologist who led the excavation and the research on the site, confirms that the community that inhabited the area was, in fact, peaceful and mostly nonviolent: few violent marks were identified on the human bones there found, in comparison to other sites of comparable size, as cited by Jean-Pierre Dupuy in his essay on Çatalhöyük[95]. According to the Girardian interpretation of the site, this is due to the effectiveness of the sacrificial rites performed and testified by the murals.

The other archaeological site previously mentioned is that of Göbekli Tepe, dating back to 3000 years before Çatalhöyük, between 9600 BC and 8200 BC[96], at the beginning of the Neolithic period. But contrary to Çatalhöyük, the remains of Göbekli Tepe do not belong to those first Neolithic settled groups: despite being "the oldest known example of monumental architecture – the first structure human beings put

[93] Girard, 223–24.
[94] Girard, 226.
[95] Jean-Pierre Dupuy, "Self-Transcendence and Tangled Hierarchies in Çatalhöyük," in *How We Became Human. Mimetic Theory and the Science of Evolutionary Origins*, ed. Pierpaolo Antonello and Paul Gifford (East Lansing: Michigan State University Press, 2015), 246.
[96] Paul Gifford and Pierpaolo Antonello, "Rethinking the Neolithic Revolution: Symbolism and Sacrifice at Göbekli Tepe," in *How We Became Human. Mimetic Theory and the Science of Evolutionary Origins*, ed. Paul Gifford and Pierpaolo Antonello (East Lansing: Michigan State University Press, 2015), 261.

together that was bigger and more complicated than a hut"[97], no trace of permanent human settlement was found, nor any evidence of local agriculture or farming[98]. This means that Göbekli Tepe was a gathering place for the many nomadic hunter-gatherer communities in the area. But what did they gather for? Let's turn to the site's architecture to try to understand its function.

The site comprises at least twenty rings of stone pillars connected by a shorter wall, facing two bigger T-shaped menhirs at the centre of each ring. The stones are carved and decorated with human anatomical parts (arms, hands, legs), clothes, and wild animal figures[99]. The rings were built progressively, one after the other, over a span of 1400 years, while the previous and dismissed ones were filled with debris[100]. Other noteworthy elements are the vast amount of flint weapons found on site[101] and the finding of unusual quantities of human and non-human bones that underwent ritualistic handling[102].

It seems coherent – as hard as it is to find solutions to such ancient dilemmas – to say that at Göbekli Tepe a lot of people would meet to perform shared rites, and the presence of so many weapons, likely manufactured on-site, speaks of violent rituals: sacrifices. Such a hypothesis is sustained by other elements concerning the site's architecture: for example, the walled ring, which would serve as an arena for the community to witness the sacrifice from above, is also a trap for the victim, who cannot escape since there are no breaches in the surrounding walls[103].

Lastly, this Girardian interpretation provides an explanation for the constant rebuilding of the rings: the function of sacrifices lies in their pacifying and unifying powers, yet the misconception that is part of their very nature implies their reconciling potential cannot but be limited, for they do not address the real causes of social tension, that will cyclically come again. Therefore, the feeling of inadequacy of the previous remedy could lead the sacrificers to try performing it again but changing it slightly in the hope of making it more effective[104]. Such a change could be the change of arena,

[97] Gifford and Antonello, 267.
[98] Gifford and Antonello, 269.
[99] Gifford and Antonello, 265.
[100] Gifford and Antonello, 264.
[101] Gifford and Antonello, 267.
[102] Gifford and Antonello, 281.
[103] Gifford and Antonello, 282.
[104] Gifford and Antonello, 283–84.

thus explaining the short life (70 years on average[105]) of each ring, despite the enormous effort necessary to build each of them with the limited technology of the time.

The discovery of Göbekli Tepe, together with its sacrificial interpretation, is so revolutionary in religious anthropology and the history of religion in general to justify the title of Gifford and Antonello's paper that has served as the primary source for these pages: *Rethinking the Neolithic Revolution*. In fact, what they call "the traditional view"[106], born in the 19th century and taken on by much of the subsequent anthropologists, assumes that religion was born after sedentism to encourage social cooperation. The view held by Girard and the authors analysed so far is opposite to it, believing in a much more fundamental role of religion in human history and in hominization itself. According to Girard, it is because humans gathered to perform sacrificial rites that they domesticated plants and animals, built long-lasting constructions, and thus developed the knowledge, technology, and cooperation skills consequently needed, laying the foundations for human settlement. In a word, it is out of religion that not only sedentism, but the whole of culture was born[107]. Göbekli Tepe, having been built and used by forager, non-settled communities on the edge between Palaeolithic and Neolithic, points fiercely in the direction of such an original function of religion, proving the so-called traditional view wrong.

Based on this Girardian thesis, we can now go back to the beginning of this chapter, to the archaeological and ethnological studies debating whether archaic societies were highly violent or whether noble savages populated them. The answer is neither: just like all subsequent societies, ancient communities had found their way to deal with the inclination to reciprocal violence that develops spontaneously in human relations due to the aforementioned dynamics. All societies, or rather, all societies that survived, succeeding in not succumbing to internecine violence, originally found their way out from the cycles of reciprocal violence through religion, which played a more fundamental role in the development of human culture than it was previously realised.

The victimary theory hereto exposed does not only account for the birth of religion but also for the birth of morality[108], the first collective attribution of good and bad qualities to actions. Specifically, the violence of the founding homicide and of the

[105] Gifford and Antonello, 264.
[106] Gifford and Antonello, 270.
[107] Girard, Antonello, and de Castro Rocha, *Evolution and Conversion. Dialogues on the Origins of Culture*.
[108] Mormino, *René Girard. Il Confronto Con l'altro*, 229.

sacrifice emerges as *good violence* thanks to the unanimity of its performance; a difference in meaning is therefore established between violent actions: the form of the action is the same, namely, the killing of someone, but a different moral judgement is attached to each. Collective violence is good, thus prescribed, whereas private violence is bad, thus forbidden[109].

This difference in meaning, despite being culturally determined, must not be considered a misrepresentation of the respective state of things: the victims of both acts of violence might be innocent, but unanimous violence *does* bring peace and is therefore good for the community, whereas private violence initiates or carries on vengeance cycles that, eventually, destroy entire families and communities. Sacrificial violence is good insofar as it prevents further violence: "Only through the mediation of the scapegoat mechanism can violence become its own remedy"[110].

This idea should not sound novel to the attentive reader: it is, in fact, probably the most influential justification to present criminal justice penal systems, that of the deterrent power of prison, and punishment in general, invoked to advocate for longer sentences and more brutal sanctions, as will be analysed at length in the next chapter. The violence of the modern State administering punishment is *good violence* because it is meant to be unanimous: in Italy, magistrates sentence defendants "in nome del Popolo italiano", in France "au nom du Peuple français" – on behalf of their respective peoples. Of course, being the State a superindividual entity whose strength is incomparable to an individual's or a group's retaliatory capacity, it does not need to reconstruct the perfect unanimity of sacrificial violence, for it can tolerate a certain degree of dissent: it is not put in danger by the risk of vengeance. Still, the retention of such a formula at the end of the criminal trial, when the sentence is declared and the punishment begins, directly connects the criminal system to its antecedent, the sacrificial system. They share the same aim: the safeguarding of the community from internecine violence; partly, they also share the same method: the use of violence to prevent further disorders, which makes this violence legitimate.

Of course, there is a fundamental difference between the two cases, namely, that the sacrificial victim is innocent, while the sentenced offender is guilty. But guilt has not always been of paramount importance in criminal justice systems as it is today. In

[109] Girard, *Things Hidden Since the Foudation of the World*, 128.
[110] Girard, 195.

the first section of his history of the prison, dedicated to the *ancien régime*'s torture system (*supplice*), Foucault observes that, before the criminal reform of the 18th century, guilt was not considered in a binary way. The torturing inquiry was both a first punishment and a means to discover the truth through confession. Before the person under investigation was ascertained to be culpable, they already deserved some degree of punishment by virtue of the mere fact of being inspected[111]. The principle of culpability, so crucial nowadays, will only be adopted after the aforementioned criminal reform, which will restrict the administration of punishment by the establishment of "judicial truth" – the truth that emerges from the trial.

The criminal reform also concealed the administration of punishment from the public eye. This is a major change from the sacrificial and the torture systems. Sacrifices needed to be collective events because their victims are proxy victims for everyone in the community, who, either directly participating in the killing, or supporting it by being non-intervening witnesses, can project their own rage, frustration, and hatred onto the sacrificial victim. The same goes for the torture victims in the Ancien Régime, who are not placed on an altar but on the public square, where violence can be once again the focal point of society's gaze.

Nowadays, punishment is hidden by prison walls, and prisons are generally located in isolated places, mostly quite far from city centres and hard to reach. Yet, the sacrificial character of this institution is exposed by the spectacularisation of criminal events in the black chronicle sections of newspapers and television news that follow investigations step by step, frequently pronouncing guilty verdicts well before trials have even started – let alone established a judicial truth – together with the subsequent public interest for the sentences, often indignantly considered too mild. Finally, it is not so uncommon for courts to condemn innocents in order to provide the public with a guilty verdict for heinous crimes.

The public is still "blood-thirsty" and frustrated by the impossibility of expressing its own need for violence, and the media provide them with their surrogate victims. However, their hatred is perceived as legitimate because it points towards a felon, a villain: we victimise victimisers[112]. Offenders are demonised, like the old beggar of Ephesus, but are not sacralised anymore; as Girard said: "We remain capable,

[111] Michel Foucault, *Discipline and Punish. The Birth of the Prison*, trans. Alan Sheridan (New York: Vintage Books, 1995), 42.

[112] Girard, *I See Satan Fall Like Lightning*, 158.

in other words, of hating our victims; we are no longer capable of worshipping them"[113]. Each individual carries out the process on their own, in front of the TV or the newspaper, expressing their anger and hatred with family and friends, but not managing to transform the exercise of public violence into a unifying force. The result of the process remains shallow: an all too ephemeral satisfaction for the suffering of another with whom they have nothing to do, a catharsis that only satisfies a momentary appetite for violence. Girard questionably attributes this jam in the sacrificial mechanism to the evangelical revelation of the victim's innocence, as will shortly be addressed.

We can now aim at the criminals instead of killing weak and innocent victims in sacrificial rites because the State can avenge the community through its sentences without generating feud cycles: the State is too powerful for individual citizens to attack it back in revenge.

> [O]ur judicial system […] serves to deflect the menace of vengeance. The system does not suppress vengeance; rather, it effectively limits it to a single act of reprisal, enacted by a sovereign authority specializing in this particular function. The decisions of the judiciary are invariably presented as the final word on vengeance.[114]

This unbalance of power allows for a shift from the completely preventive vocation of victimary systems to the partially repressive (although also partially sacrificially preventive, as we have just seen) vocation of penal systems. Girard refers to it in a few pages[115], to which we shall later return, but does not speak of the transition from one system to the other. While claiming to prove the Girardian victimary theory wrong, the French anthropologist Raymond Verdier offers an interesting articulation of this passage through the notion of vindicatory system.

1.2.2 The Vindicatory System

Verdier directed a multidisciplinary research group that published between 1980 and 1984 four volumes on vengeance in Non-Western countries[116], ancient

[113] Girard, *Things Hidden Since the Foudation of the World*, 37.
[114] Girard, *Violence and the Sacred*, 15.
[115] Girard, 14–26.
[116] Raymond Verdier, ed., *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume I. Vengeance et Pouvoir Dans Quelques Sociétés Extra-Occidentales* (Paris: Éditions Cujas, 1980); Raymond Verdier, ed., *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume II. Vengeance et Pouvoir Dans Quelques Sociétés Extra-Occidentales* (Paris: Éditions Cujas, 1980).

civilizations[117], and Western thought[118], in an effort to comprehensively investigate practices and theories of vengeance across space and time. In his introductory essay to the first volume[119], he presents the vindicatory system as the social system underlying most ancient civilisations. The vindicatory system is based not on the banishment of vengeance but on its regulation. While modern thought tends to think of vengeance from a psychological point of view as an individual phenomenon arising from anger and resentment, regulated vengeance can actually play an ethical role, as it has been acknowledged by some Greek and even Christian philosophers[120]. So Aristotle approves of vengeance inasmuch as it derives from anger, which is a natural passion, and, if not excessive, it can be well-balanced and effective; Thomas Aquinas approves of it when it is conformed to reason, deriving from the wish to improve one's fellow, and preserve *justice* – we shall return on this later on.

Vengeance can therefore be rational, nothing like Girard's unleashed violence that destroys families and entire communities. Following the path traced by these philosophers, Verdier identifies the three principles that transform the vindicative passion into the reasonable basis for the social organisation of vindicatory systems[121]:

1) *Principle of reciprocity*: sanctioned by the first law ever proved in human culture, the *lex talionis* of the Babylonian Code of Hammurabi, also present in the Bible in the form of the "eye for an eye". Despite appearing crude at first look, it is a first limitation to vengeance that cannot be disproportionate but must be accurately restricted to mirror the first offence, as in the payment of a debt[122].

[117] Raymond Verdier and Jean-Pierre Poly, eds., *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume III. Vengeance, Pouvoirs et Idéologies Dans Quelques Civilisations de l'Antiquité* (Paris: Éditions Cujas, 1981).

[118] Gérard Courtois, ed., *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume IV. La Vengeance Dans La Pensée Occidentale* (Paris: Éditions Cujas, 1984).

[119] Raymond Verdier, "Introduction : Le Système Vindicatoire," in *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume I. Vengeance et Pouvoir Dans Quelques Sociétés Extra-Occidentales*, ed. Raymond Verdier (Paris: Éditions Cujas, 1980).

[120] Raymond Verdier, "Le Désir, Le Devoir et l'interdit : Masques et Visages de La Vengeance," *Déviance et Société* 8, no. 2 (1984): 182, https://doi.org/10.3406/ds.1984.1406.

[121] Verdier, 184–87.

[122] Verdier remarks that the vocabulary of many societies links the notion of vengeance and that of debt: it is the case for the Bulsa, the Kabiyè, the Maengue, the Kanchin, the Mongols. Verdier, 185. Many societies allow for material compensations as a substitute for retaliations, as is the case for the Amazonian Achuar: Raymond Verdier, "Droits Pré-Étatiques et Théorie Du Droit," *Revue Interdisciplinaire d'études Juridiques* 21, no. 2 (1988): 193.

2) *Principle of solidarity*: vengeance is not conceived as an individual phenomenon but as a group affair. The familial group is considered an inseparable unit that makes the offence to one member an affront to the existence and the identity of the whole unit and therefore deserves a collective response.

3) *Principle of social distance*: vengeance is regulated differently according to the relations binding different groups.

This last principle leads us to the heart of Verdier's theory of the vindicatory system. This system is, in fact, organised on the basis of three "social spaces of violence"[123]: the peripheral space characterised by hostile relations, reciprocal misrecognition, and proneness to mutual elimination: the area of war. The central space that groups the family and is characterised by a relation of identification: here, vengeance is forbidden, for it would lead to the group's self-destruction[124], and infractions are treated by penal or proto-penal (ritual) sanctions that aim at reconciliation. Finally, an intermediate space, where adversity relations prevail, and groups recognise each other as autonomous, equal, and possibly rival. Here, vengeance is prescribed to preserve the group itself if one of its members is offended – according to the principle of solidarity[125].

This tripartition sheds light on the double bind that links vengeance and solidarity: on the one hand, solidarity forbids revenge inside the group; on the other, it requires it on the outside. It also shows that there is a "fake diachronic opposition"[126] between vengeance and institutionalised punishment, making vengeance punishment's antecedent, while the two coexist synchronically, only in two different spheres.

Therefore, Verdier claims to have refuted Girard, showing that vengeance is not an infective virus that spreads, contaminating entire social groups, but a valuable and effective instrument for regulating groups' relations. However, Girard does not exclude

[123] Verdier, "Le Désir, Le Devoir et l'interdit : Masques et Visages de La Vengeance," 187.

[124] Anspach remarks that, in the aforementioned tragedies of the *Oresteia*, the tragic mechanism is triggered by the infraction of this prohibition of vengeance inside of the inner circle of the family, when Clytemnestra decides to avenge the death of her daughter Iphigenia, a sacrificial victim, killing her husband Agamemnon. Mark Rogin Anspach, "Vengeance and the Gift," in *Palgrave Handbook of Mimetic Theory and Religion*, ed. James Alison and Wolfgang Palaver (New York: Palgrave Macmillan, 2017), 55.

[125] Verdier, "Introduction : Le Système Vindicatoire," 24; Verdier, "Le Désir, Le Devoir et l'interdit : Masques et Visages de La Vengeance," 187.

[126] Verdier, "Le Désir, Le Devoir et l'interdit : Masques et Visages de La Vengeance," 184, my translation.

the possibility of *regulated*, or better, *ritualised*[127] forms of vengeance that have lost their spontaneous character: in the vindicatory system, in fact, vengeance is not *tolerated*; it is *prescribed.* Girard devotes part of his work to the analysis of how such a destructive form of violence could be domesticated, and he proposes a sacrificial thesis that Verdier does not refute: in fact, Verdier himself notices the close connection between vengeance – and violence in general – and religious rites, acknowledging that "in many past and present societies, law is inseparable from religion, and their respective domains are more or less tightly nested"[128]. He notes how often counteroffensives and retaliations can be ended by shared religious rites of reconciliation[129], how vengeance is generally considered a sacred duty[130] , and, where vengeance is forbidden inside the group, how reintegration requires collective religious rituals of purification and restoration[131].

This last element might be the best to understand the close relation of Verdier's theory to Girard's. This community's inner circle is the one where violence is more likely to spread because of mimetic reasons[132]. Here, as Verdier remarks, vengeance is forbidden: the vindicatory system might as well be put in place to offer violence an outlet from the inner circle, therefore avoiding mimetic crises. The violence that needs to be diverted from the inside must be exercised on the outside to variable degrees proportional to the associated risk. As Paul Dumouchel remarked:

> The sacrificial character of this structuration of social space is readily visible in that the degree of violence permitted augments as the social distance between individuals becomes greater. Like other sacrificial institutions, this organization of the

[127] Verdier, "Introduction : Le Système Vindicatoire," 25.

[128] Verdier, "Droits Pré-Étatiques et Théorie Du Droit," 197, my translation.

[129] Verdier, "Le Désir, Le Devoir et l'interdit : Masques et Visages de La Vengeance," 191–92; Raymond Verdier, "Violence et Intégration Socio-Culturelle : Vengeance et Serment," *Revue Européenne Des Sciences Sociales* 30, no. 94, Les formes de la violence : IXe colloque annuel du Groupe d'Etude "Pratiques Sociales et Théories" (1992): 71.

[130] Verdier, "Le Désir, Le Devoir et l'interdit : Masques et Visages de La Vengeance," 192.

[131] Verdier, "Introduction : Le Système Vindicatoire," 22.

[132] To better understand why, the distinction between internal and external mediation in Girard's thought should be introduced: "We shall speak of external mediation when the distance is sufficient to eliminate any contact between the two spheres of possibilities of which the mediator and the subject occupy the respective centers. We shall speak of internal mediation when this same distance is sufficiently reduced to allow these two spheres to penetrate each other more or less profoundly." Girard, *Deceit, Desire & the Novel. Self and Other in Literary Structure*, 9. Only internal mediation generates rivalry and, although the distance Girard here refers to is not to be understood in geographical terms, physical proximity and kinship ties that characterize the central space of the vindicatory system do contribute to the internalisation of mediation, therefore generating conflict.

> space of solidarity and hostility aims to drive violence away from the center of the community toward "acceptable victims.[133]

For external yet proximate communities, regulated and ritualised vengeance allows for the preservation of balanced and peaceful relations, while communities situated at a greater degree of distance, with which it is less likely to initiate a mimetic crisis, can be the object of unleashed violence. The sacrificial character of such war is made evident by the Tupinamba's custom of using war prisoners as sacrificial victims[134] after their aforementioned internalisation. Quoting Anspach: "[it is possible that] the risk of a savage vengeance is primary and the vindicatory system mainly constitutes a set of ritual devices […] that aim at pushing back this threat"[135]. Instead of refuting Girard's thesis, Verdier seems to corroborate and enrich it, further substantiating the distinction between legitimate and illegitimate violence, and vengeance in particular.

Evidently, Verdier and his research group cannot acknowledge this continuity, for their work is strongly conditioned by an ideological position[136], that of right-wing anarchism, that situates them in opposition to Girard's conceptual path. In fact, Girard, reconducting criminal justice to the practice of vengeance, morally undermines it, whereas, on the contrary, Verdier's group aims precisely at rehabilitating the practice of revenge not only as a form of conflict resolution but as a fully-fledged practice of "justice"[137] in opposition to the impersonal monopoly of State courts. Verdier ends his introductory essay to the first volume by denouncing the condition of "crisis of the State and decline of the law"[138]; the French philosopher Gérard Courtois begins his introductory essay to the fourth and last volume by lamenting the exclusion of vengeance from moral and juridical discourse in favour of the "anonymous justice of

---

[133] Paul Dumouchel, "The Barren Sacrifice," in *Palgrave Handbook of Mimetic Theory and Religion*, ed. James Alison and Wolfgang Palaver (New York: Palgrave Macmillan, 2017), 283.

[134] After their capture, such prisoners are abandoned by their tribe, that does not try to save them, rather, it awaits their death in order to avenge them afterwards: in other words, the tribe accepts their member has become another's tribe sacrificial victim, in order to subsequently capture themselves their own sacrificial victim from the other tribe. Mark Rogin Anspach, "Apprivoiser La Violence. Le Sacrifice, La Vengeance, Le Don," *Revue Du MAUSS* 55, no. 1 (2020): 132.

[135] Mark Rogin Anspach, "Penser La Vengeance," *Esprit* 128, no. 7 (1987): 105, my translation. This is explicitly recognised by Verdier, when he says: "In the vindicatory system, vengeance is *ritualised* in many ways: it is not exercised on whoever, whenever and wherever. The consideration of the need of appeasement and reconciliation leads to the canalisation of aggressivity and its containment into certain limits." Verdier, "Le Désir, Le Devoir et l'interdit : Masques et Visages de La Vengeance," 190, my translation.

[136] Mormino, *René Girard. Il Confronto Con l'altro*, 159.

[137] Gérard Courtois, "Introduction : La Vengeance, Du Désir Aux Institutions," in *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume IV. La Vengeance Dans La Pensée Occidentale*, ed. Gérard Courois (Paris: Éditions Cujas, 1984), 11.

[138] Verdier, "Introduction : Le Système Vindicatoire," 36, my translation.

the State"[139]. Against this detached and bureaucratic system, as is repeatedly stressed by Courtois in his essay on vengeance in the thought of Aristotle and Seneca, the resumption of the practice of vengeance can shift the focus to the victim and their needs[140] as well as to the bonds of solidarity that bind the group[141], placing justice in the hands of civil society as opposed to the State and providing a valid alternative to the systems in force in the contemporary world.

1.2.3 The Transition to the State System

Putting this debate aside, from what has been said about the sacrificial and the vindicatory system, the close connection between the management of violence and the birth and development of religion should be clear to the reader. The theory exposed so far should also enlighten the role played by vengeance and retribution in ancient religions, with their choleric and vengeful gods. This is why, according to Girard, Christianity represents an anomaly in the panorama of religion. The Gospels expose the sacrificial mechanism and its "horrible" secret: the unleashing of unanimous violence against a helpless and innocent victim[142]. Of course, the victim here is Jesus, who, aware of the mimetic nature of desire and action, foresees that all his companions would deny him and, as we have seen in the episode of Peter's denial, was right.

Yet, albeit late, a minority manages to break the mimetic unanimity and declare the victim's innocence[143]: Jesus is the real *Agnus Dei*, the lamb of God, whose pure blood washes sins away. He matches the exact definition of a scapegoat: the goat that, in the Jewish ritual, is charged with all the sins of Israel and then expelled[144]. We can here clearly appreciate the victimary mechanism because, unlike the ancient myths mentioned before, the Gospels avoid the mythical logic that misconceives the events

[139] Courtois, "Introduction : La Vengeance, Du Désir Aux Institutions," 9, my translation.

[140] "It [justice] is not concerned with chastening or correcting a wrong but, like revenge, with restoring something of value to a victim." Gérard Courtois, "Le Sens et La Valeur de La Vengeance, Chez Aristote et Sénèque," in *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume IV. La Vengeance Dans La Pensée Occidentale*, ed. Gérard Courtois (Paris: Éditions Cujas, 1984), 102, my translation. "[…] the essence of revenge is less the repetition of the same evil than the satisfaction of the victims." Courtois, 104, my translation.

[141] "Whereas Aristotle was concerned with thinking of an order among the citizens, in the hands of the citizens, Seneca puts reason above the social body and heralds the modern state with its split between the state and civil society." Courtois, "Le Sens et La Valeur de La Vengeance, Chez Aristote et Sénèque," 114, my tranlsation.

[142] Girard, *I See Satan Fall Like Lightning*, 142.

[143] Girard, 123.

[144] Girard, 144–45.

and represents them distorted to such an extent that they are barely recognisable: they oppose a biblical truth to a mythological lie[145]. The Gospels directly expose the truth of the innocence of the sacrificial victim because they are the account of that repented minority that, although not strong enough to avert the mimetic mechanism, eventually recognised its mistake and could present the point of view of the victim instead of that of the killers.

> To represent collective violence accurately, as the Gospels do, is to deny it the positive religious value that myths give it, it is to contemplate it in its purely human, morally culpable horror, it is to free oneself from the mythical illusion that either transforms violence into a praiseworthy, sacred action, because it is useful to the community, or it removes it altogether, as scientific research on mythology does nowadays.[146]

Jesus, had it not been for the minority who eventually followed him, would have made the perfect victim, for he was perfectly innocent and, therefore, apparently harmless. But his very weakness, revealed by his refusal to respond to aggression in any way, the refusal to succumb to the sacrificial logic, proves revolutionary and plays a major role in the erosion of the vindicatory system. It is the theory elaborated by Paul Dumouchel, according to whom forgiveness and charity play a crucial role in dismantling the vindicatory system. "Forgiving teaches us to not make our foes into our victims"[147], therefore denying the principle of reciprocity that governs group relationships in the intermediate space theorised by Verdier, while charity "teaches us that absence of reciprocity does not free us of all duties toward others, that it does not deliver the Other to us like prey into our hand."[148]

Dumouchel remarks that, for them to have a subversive effect on social organisation, forgiveness and charity do not need an ideal adhesion: people do not need to assume them as guiding values in their life, and they do not need to be spiritually charitable and forgiving. There only needs to be a practical and opportunistic adherence that modifies behaviour rather than thought[149]. These Christian principles, in fact, can prove to be quite advantageous in avoiding solidarity obligations that are often onerous on the individual member of the group: participation in punitive raids, in fact, is

[145] Girard, 113.

[146] René Girard, *Je Vois Satan Tomber Comme l'éclair* (Paris: Grasset & Fasquelle, 1999), 17–18, my translation. This passage is taken from the introduction of the original French edition because in the English edition this introduction has been replaced by an alternative text.

[147] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 121.

[148] Dumouchel, 119.

[149] Dumouchel, 121.

dangerous and burdensome, exposing the individual to risks that they can now avoid appealing to forgiveness and charity. This leads to the progressive erosion of the bonds of solidarity that are structured around the double bind that ties solidarity to vengeance: subsequently, the regulated dimension of revenge disappears, and vengeance is hence downgraded to an individual passion. It loses its function as the group's glue, the aspect of unanimity that made its violence legitimate and good. In the absence of the bonds of solidarity, revenge is 'naturalised' and, losing its sacred element, delegitimised.

This interpretation of the revolutionary power of Christianity provided by Dumouchel turns out to be more compelling than Girard's, for the latter, as Mormino has remarked, deserts the materialist terrain that characterises the rest of his work, here describing in terms of truth and falsity dynamics that can be more simply explained on the basis of the power relations that characterise modern societies, which struggle to close in on victimising unanimity because of their complexity and heterogeneity[150]. Girard seems to forget that truths can be concealed for self-interested reasons and thus that such a discovery does not provide a satisfactory account for large-scale behavioural transformation.

The very history of the unveiling of the sacrificial truth by the Gospels supports this materialistic interpretation. The apostles could expose it because they lived in a complex and broad enough society to tolerate such dissent, nothing like the small groups or even the Greek *poleis* where the foundational myths and sacrificial rites were first forged. Dumouchel's argument proves more persuasive specifically for its reference to the "opportunistic" adhesion to Christian principles, which need not be recognised as morally superior to the ones previously adopted, only more advantageous on the individual level.

Of course, this is only possible within a political framework, that of the progressive establishment of the State as the monopoliser of violence, and vengeance in particular, that implies the overcoming of the need for the vindicatory system as the primary device to manage violence in the community. Being religious at its origins, this political framework could only develop by intermingling with religious institutions and practices, thus bearing their signs. Therefore, before turning to the analysis of justice systems in the State system, a digression on forms of justice in religion is in order. This will be the topic of the next section.

[150] Mormino, *René Girard. Il Confronto Con l'altro*, 243, 288–95.

1.3 Otherworldly Justice

Once the original practical function of religion as a system for the management of conflicts through the distinction between good and bad violence is acknowledged, it is not surprising to realise that reflections on justice occupy such a big part of religious thinking. Even though, as we have seen at length, religion is not originally a theory and must not be fundamentally identified in a theology, these later theorisations bear the sign of that primary role and reflect the subsequent need for other forms of conflict management.

Once human groups have grown, making it hard to close in on the sacrificial victim and establish the sacrificial unity of all against one, other ways to deal with violence had to be found. It seems that virtually in all cultures around the world this practice has taken the form of retribution, and it is attested in theological and philosophical writings concerning the topic of otherworldly life. Life after death is always characterised by an idea of justice that "rights the wrongs" endured in one's earthly life, rewarding those who suffered and punishing those who undeservingly flourished.

In the first pages of his book *The Birth of Purgatory*, the French historian Jacques Le Goff provides the reader with an overview of ancient conceptions of life in the afterlife. From the Greek and Roman dualistic conception of Hades and the Elysian Fields[151], which seems to foreshadow the Christian Hell and Heaven, to the Vedic reincarnation cycles that can eventually lead both to Paradise for some and to Hell for the incorrigibles[152]; from the terrifying hell imagined by ancient Egyptians[153] to the Platonic retributivist reincarnations as narrated in the *Republic*[154] and the *Phaedrus*[155]; from the Babylonian underworld of dust and darkness – from which some can be freed[156] – to the Jewish tripartition, developed at the beginning of the Christian era, of Gehenna (hell), Eden, and Sheol[157], an intermediate place where those who are "neither

[151] Jaques Le Goff, *The Birth of Purgatory* (Aldershot: Scolar Press, 1990), 2.
[152] Le Goff, 18.
[153] Le Goff, 20.
[154] Le Goff, 21–22; Plato, *The Republic*, trans. Desmond Lee (London: Penguin Classics, 2007), 10:615 a-e.
[155] Plato, *Phaedrus*, trans. Robin Waterfield (Oxford: Oxford University Press, 2009), 248a-249e.
[156] Le Goff, *The Birth of Purgatory*, 25.
[157] In the Old Testament, Gehenna is just a valley close to Jerusalem where child sacrifices devoted to the Molech cult were performed: Chronicles 28,1-3; 33,1-6; Jeremiah 7,31-32; 32,35. This practice is

entirely good nor entirely bad, […] will be punished for a time after their death and then go to Eden"[158]. All these conceptions share a retributive notion of justice. It is not surprising: the inclination to seek revenge is so ingrained in human behaviour[159] that it is naturally reflected in conceptions of the afterlife.

When one is attacked, their spontaneous reaction is often to strike back: this is the root of revenge. This impulse to retaliate has obvious evolutionary advantages: if one would just stand still in the face of aggression, they would be too easily killed. Of course, sometimes one has the alternative of flight, but this works well mostly for non-social animals, not for those living in groups. In human communities, expulsion has been, in fact, tantamount to a death penalty: a human being could hardly survive alone in the wild. We are thus justified in believing that humans have the propensity[160] to strike back. In the previous discussion on the dangers of vengeance, we have noted how dangerous retaliation can be for small societies, to the point of leading to their possible self-destruction in feud cycles. This is why archaic societies forbade revenge and often placed it in the realm of the sacred, hence untouchable. But some ways can be found to circumvent such issues – otherworldly retaliation is one such way.

We have mentioned that one of the main justifications for punishment is its deterrent power. According to Foucault's "punitive semio-technique"[161], i.e., the technique of punitive signs, sentences do not need to be actually administered; they only need to be imagined, "represented", for them to be effective. This "*rule of sufficient ideality*"[162] is clearly put into practice by these imageries, where the violence of retaliation is only imagined in expectations and can be exploited by those societies where striking back the attacker is too dangerous: it is a prospect of retaliation that cannot be avenged at its turn. This imagery can, thus, serve the purpose of

---

ended by King Josiah: 2 Kings 23:10. The only underworld there cited is the Sheol, and it is uncertain whether it should be considered an infernal place or rather just the equivalent of the tomb, where the dead corpse is laid down, therefore denying the existence of an afterlife. See for example: Bart D. Ehrman, *Heaven and Hell. A History of the Afterlife* (New York: Simon & Schuster, 2020), chap. V. Death After Death in the Hebrew Bible.

[158] Le Goff, *The Birth of Purgatory*, 40.

[159] We have started investigating this subject in the previous section and we shall further analyse it in the next chapter, when we'll discuss retributive emotions.

[160] By "propensity" I do not mean "instinct", i.e. a behaviour inscribed in a being's nature that does not need any further explanation or cause and that is supposed to happen without exception. This propensity to strike back finds its reasons in human sociability, as has been noted, as well as in the mimetic attitude that has been investigated in the previous section. It is susceptible to exceptions and can be eradicated from individual behaviour.

[161] Foucault, *Discipline and Punish*, 94.

[162] Foucault, 94.

complementing sacrificial practices where the unifying mechanism is jammed because societies are broadening and comprising more people and different cultures. However, if more complex societies have the problem of being incapable of re-establishing the sacrificial unity, in virtue of that same complexity they might also be able to treat this-world vengeance with more ease.

We have noticed that the fundamental moral distinction between good and bad violence relates to this violence's consequences on the group: whether it is a pacifying violence or a warmongering one. The same goes for revenge: as has been briefly noted before, while private vengeance is the most dangerous form of violence and is, thus, always bad, institutional revenge can be good because individuals or even family groups are not strong enough to retaliate against the institution[163]. This institution will eventually become the political state, at least partly separated from religious institutions. But as we have seen, the first forms of authority seem to have been religious: it is religion that first took over the duty that social contract theories ascribe primarily to the state, namely the management of violence.

While we now live in bureaucratic states that have gathered the means to enforce their retaliatory measures, to have sufficiently detailed and specific legislation, and be capable of collecting complaints, verifying them, identifying the culprits, putting them on trial, and punishing them according to the law if necessary requires a level of technological advancement that was reached only recently. The alliance between the state and religion helped overcome these issues. On one side, religious conceptions of otherworldly justice provided significant aid in preventing misconduct; on the other, for a long time, crimes corresponded with sins, and investigations, as well as punishment administration, were carried out by clerics – just think of the Catholic Inquisition – as the Church had a well-established presence and authority all over the territory.

While theologically these practices can be justified and idealised in many ways, for example through the purifying conception of suffering that would thus aim at the

[163] In Sicily, between the end of the 70s and the beginning of the 90s, Cosa Nostra, the main mafia organisation of the region, undertook a conflict with the Italian state that is commonly referred to as a war between the state and the mafia. Having the state historically displayed a weaker presence in Southern Italy, it gave local groups the opportunity to gain power over the population and become so strong as to have the ability to retaliate against the State itself. The rise of armed groups rival to the state is typical of situations where the state is particularly weak and cannot monopolise violence on the territory, as we are set to investigate later.

victim's own redemption[164], their material reality consists of the concrete exercise of violence, often in the form of torture, on individuals. As Montaigne has said, "[h]uman reason is a dye spread more or less equally through all the opinions and all the manners of us humans, which are infinite in matter and infinite in diversity"[165]: any practice may be justified in one way or another, it is, hence, pointless to try understanding them solely through their rational justification; the power relations and the needs of society that underlie them should instead be investigated. Still, this does not mean that such arguments are assumed by people in bad faith, with the conscious intent to deceive: they themselves are deceived by these narratives.

These justifications that transformed domination dynamics into moral values via their connection with the fundamental goodness of God are relevant and worth investigating, for they are still operating in common – even lay – thought regarding justice. The association of retribution with a perfect world that rights all wrongs helped transform retribution into an ideal that may provide a model not only for the world of the dead (that serves the purpose of eradicating dissent against the oppression and abuse suffered in life, which is disvalued in the face of the eternal life after death where compensation is expected) but also for that of the living, substantiating a retributivist conception of justice and a meritocratic ideal of society that can be internalised even by its victims, who believe they deserve the suffering they endure.

1.3.1 The Evolution of Christian Justice: From Mercy to Perfect Retribution

The case of the Medieval transformation of the Christian conception of the afterlife is especially interesting. While, at the time of Jesus, millenarian expectations that predicted the Apocalypse and universal judgement to be near implied that life after death would take place either in heaven or in hell, in the centuries that followed, this

[164] It is the conception of divine chastisement elaborated by the theologians Clement and Origen of Alexandria in the II and III centuries CE: Le Goff, *The Birth of Purgatory*, 52–53. During the Middle Ages this conception will be adopted to justify the practices of the Inquisition, aimed at the heretic's repentance and salvation. In fact, even once condemned to the stake, heretics could be saved if they repented, as is attested by the episode of the Italian heretic Pietro da Martinengo, who repented when the pyre had already begun to burn and was rescued "*semicombustus*" to be cured at the inquisitor's expense. Yet, as Marina Benedetti shows, Christian charity was not the only motivation for the inquisitor's solicitude in Pietro's cures, as he also served as a spy to identify other heretics before he died. Marina Benedetti, "Forme Inquisitoriali Di Polizia Nel Medioevo," in *Le Polizie Informali*, ed. Livio Antonielli (Messina: Rubbettino, 2003), 28.

[165] Michel de Montaigne, *The Complete Essays*, ed. and trans. M. A. Screech (London: Penguin Books, 1991), 126.

bipartite representation is developed into a tripartite conception of the afterlife with the introduction of purgatory. The presence of hell in early Christian conceptions is already startling since, as remarked earlier discussing the Girardian interpretation of the Gospels, Jesus rejects violent reciprocity, and Christianity is set to be a non-retributive religion. Indeed, many excerpts of the Gospels support this interpretation, from the *Pericope Adulterae* aforementioned to the famous exhortation to turn the other cheek to Luke's Gospel's clearly anti-retributivist passage:

> If you love only someone who loves you, will God praise you for that? Even sinners love people who love them. If you are kind only to someone who is kind to you, will God be pleased with you for that? Even sinners are kind to people who are kind to them. If you lend money only to someone you think will pay you back, will God be pleased with you for that? Even sinners lend to sinners because they think they will get it all back. But love your enemies and be good to them. Lend without expecting to be paid back. Then you will get a great reward, and you will be the true children of God in heaven. He is good even to people who are unthankful and cruel. Have pity on others, just as your Father has pity on you.[166]

Girard thus interprets the Apocalyptic passages of the Gospels as a description of the eventual mimetic crisis, resulting from purely human dynamics and attributed to God due to the mythical misconception characteristic of the sacred, where humans externalise their own violence and attribute it to God[167]. However forced Girard's interpretation may sound, it cannot be denied that an anti-retributive thrust is fiercely present in the Gospels and that the Christian God is indeed represented as the God of mercy, in contrast with the Old Testament's vengeful and choleric God. The subsequent focus of theologians and believers on retribution in the afterlife points in the direction of a fundamental struggle to emancipate ourselves not only from the sacrificial dynamics already analysed but also from retributive dynamics.

Some discomfort towards the concept of hell, i.e., infinite suffering as a punishment for finite fault, can be observed early in Christian theology. Starting from the Greek theologian Origen of Alexandria (3rd century), who believed hell to be only temporary and aimed at the sinner's purification through suffering – therefore

[166] Luke 6:32-36. For all the quotations from the Bible I have used the Contemporary English Version (CEV).

[167] Mormino, *René Girard. Il Confronto Con l'altro*, 270; Girard, *Things Hidden Since the Foudation of the World*, 186; René Girard, *Battling to the End. Conversations with Benoît Chantre*, trans. Mary Baker (East Lansing: Michigan State University Press, 2010), xvi; René Girard, *Job: The Victim of His People*, trans. Yvonne Freccero (Stanford: Stanford University Press, 1987), 25.

eliminating hell altogether and making an *ante litteram* purgatory of it[168] – to Ambrose of Milan (4th century), who maintained that for "minor" sinners otherworldly fire would be excruciating only until it purifies their soul[169], to Augustine of Hippo (4th-5th century) who believed that his prayer could "hasten the day when the soul enters heaven"[170], anticipating the importance of the suffrages to shorten purgatorial time, that will eventually result in the sale of indulgences.

While Origen's conception is interesting as an antecedent of purgatory, which will only be formulated in the Christian doctrine during the course of the 12th Century[171], to be later established by the Second Council of Lyon (1271) and then reaffirmed by the Council of Trent (1563), Origen's emphasis on salvation for all will be dropped and become emblematic of heretical currents. It is Ambrose's and Augustine's conceptions that will evolve into the later conception of purgatory: "[t]he ideas that living human beings formed about the other world were inspired, I think, more by a need for *justice* than by a yearning for salvation"[172], said Le Goff. While talking generically about "justice", Le Goff clearly refers to retribution and retribution alone.

This need for retribution led to the formulation of a mathematical correspondence between faults and punishments that establishes "[a] proportion between the time spent in sin on earth and the time spent in torment in Purgatory […] and eventually led to the institution of a complex system of bookkeeping associated with the world to come"[173]. This practice takes up and develops that of the penitential books, initially created in Ireland in the 6th century and propagated around Europe in the following centuries[174]. And in fact, the two are strictly related, for, according to Thomas Aquinas, earthly penance was considered the beginning of the expiation process that would be completed in purgatorial punishment[175].

This latter element brings us back to the connection between these imageries and real-life power dynamics. In fact, as remarked by Le Goff, during the 12th Century, not

[168] Le Goff, *The Birth of Purgatory*, 55.
[169] Le Goff, 59.
[170] Le Goff, 66.
[171] Le Goff, 133.
[172] Le Goff, 210, my emphasis.
[173] Le Goff, 227–28.
[174] Alfredo Lucioni, "Monachesimo Occidentale e Monaci Missionari," in *Storia Del Cristianesimo II. L'età Medievale (Secoli VIII-XV)*, ed. Marina Benedetti (Roma: Carocci, 2015), 68.
[175] Le Goff, *The Birth of Purgatory*, 272.

only was the doctrine concerning purgatory clarified but also canonical law was established in opposition to the attempts of feudal lords, kings, and territorial princes to monopolise the administration of justice[176]. Moreover, the sale of indulgences, stemming from the newly-formed conception of retribution in the afterlife, helped consolidate the clergy's economic and political power "until it became a dangerous weapon that was ultimately turned back against the Church"[177]. Le Goff refers here, of course, to the Lutheran Reformation, to which we should shortly turn.

We have cited above the power of conformity to the *status quo* of the concept of otherworldly life and justice, making this-world life vacuous and inane in the face of eternity, but such theories of retribution pull this conforming power even further. They make suffering a source of merit that can be used both in the afterlife to pay back previous sins and crimes and in this life to secure a meritorious capital that will grant salvation after death. While this mechanism has been well established in Christian culture since its beginnings through the sanctification of the martyrs, the Medieval Church had expanded this belief by spiritualising poverty, transforming it into the "preferred condition of Christian life […] a kind of spiritual capital"[178]. This view that informs much of Christian Medieval and Modern thought – just think of Francis of Assisi and the popularity of the Franciscan movement – is far from abandoned in contemporary times; a prominent example is that of Mother Theresa of Calcutta, who, in the words of her famous critic, the British journalist and author Christopher Hitchens, "was not a friend of the poor. She was a friend of poverty. She said that suffering was a gift from God"[179], and "wasn't working to alleviate poverty, […] she was working to expand the number of Catholics"[180]. But the link between suffering and merit has entered common lay thought as well: we tend to attribute more merit to those who

[176] Le Goff, 211.

[177] Le Goff, 249.

[178] Carter Lindberg, "Luther's Struggle with Social-Ethical Issues," in *The Cambridge Companion to Martin Luther*, ed. Donald K. McKim (Cambridge: Cambridge University Press, 2003), 171.

[179] Christopher Hitchens, "Mommie Dearest. The Pope Beatifies Mother Teresa, a Fanatic, a Fundamentalist, and a Fraud," Slate, 2003, https://slate.com/news-and-politics/2003/10/the-fanatic-fraudulent-mother-teresa.html, accessed on March 9th 2023, 11:30. Hitchens is famous for his radical anti-religious stance, being a major figure in the New Atheism movement, together with Richard Dawkins, Daniel Dennett, Sam Harris and Ayaan Hirsi Ali. He developed his critique of Mother Teresa in his book *The Missionary Position. Mother Teresa in Theory and Practice* (London: Verso, 1995).

[180] Rebecca Leung, "The Debate Over Sainthood," CBS News, October 19, 2003, https://www.cbsnews.com/news/the-debate-over-sainthood/. Accessed on 29/04/2024, 10:08.

achieve a certain result by overcoming more obstacles, with more effort, by suffering more: this is what Bruce Waller has called "effort-deserving"[181].

Montaigne, in his essay *On Cruelty,* addressed this issue, noticing that virtue is linked to struggle and hardship: "virtue rejects ease as a companion"[182], and, in his typical ironic and stinging style, wonders whether the one who, having grown accustomed to enduring hardship, and having become able to face with ease the obstacles imposed by virtuous action, his virtue thus no longer being "painful"[183], would still be considered virtuous: "this […] seems to produce an innocent man but not a virtuous one; a man exempt from doing evil but not one apt for doing good"[184].Without explicitly refuting such a conception of virtue, he reveals its absurdity, pushing it even further by pointing out that God himself could not be said to be virtuous since effort is unknown to him: "That is doubtless why we can call God good, mighty, bountiful and just, but we cannot call him virtuous: his works are his properties and cost him no struggle."[185].

The need to be recognised as deserving for one's struggles and pains seems to me to be fundamentally a way of consoling oneself over the unfairness of having to face hardships spared to others. Of course, this merit is often projected into the afterlife, as one is rarely rewarded in this life for the suffering one has endured, and the idea that this-world suffering is part of a divine plan animated by justice ensures that people do not interpret their painful fate as a source of misery.

In the seventies, social psychologists Ronnie Janoff-Bulman and Camille Wortman conducted research on paraplegic and tetraplegic victims of accidents that, providing a divine broader context to the accident, eventually considered it a positive event[186]. Lerner comments:

> The conclusion from these findings, generated in diverse situations, is that people are impelled, possibly by habit and certainly by a strongly felt need, to perceive

[181] Bruce N. Waller, "Just and Nonjust Deserts," *The Southern Journal of Philosophy* 25, no. 2 (1987): 230–31.

[182] Montaigne, *The Complete Essays*, 474.

[183] Montaigne, 476.

[184] Montaigne, 477.

[185] Montaigne, 472.

[186] Ronnie Janoff-Bulman and Camille B. Wortman, "Attributions of Blame and Coping in the 'Real World': Severe Accident Victims React to Their Lot," *Journal of Personality and Social Psychology* 35, no. 5 (1977): 351–63. These same experimental subjects also showed high degree of self-blame.

> what happens to themselves and others in their world as manifestations of a 'just world'.[187]

While from a psychological point of view, and for the analysed case specifically, this may be considered a useful, if not entirely positive, mechanism, for it alleviates the suffering of the individuals concerned, in the context of poverty or other social injustices, this attitude hinders societal improvement that, it must be noted, always derives from the active struggles of the victims. Justice as proportionality, both between guilt and punishment and between suffering and merit, apart from being an instrument of power to maintain the *status quo*, constitutes a framework of meaning for human life that helps accept everyday struggles and is thus so difficult to unhinge[188].

The historical development of Christian thought on justice in the afterlife seems to corroborate this thesis: from the figure of Jesus, who turns the other cheek and opposes all forms of retaliation and retribution, to one of the most complex systems of retributive proportionality that perfectly mirrors real-life criminal retributive justice.

1.3.2 The Evolution of Protestant Justice: From Damnation to Meritocracy

A similar path can be observed in Protestantism, albeit starting from the opposite extreme. As was noted above, Luther's controversy sparked from the sale of indulgences, thus precisely from the Christian conception of justice in the afterlife and its connection to worldly life. According to Luther, humans are so radically distant from God that not only it is impossible for them, lay or clerics, to save others through their doing – as was implied by the practice of indulgences – but also to save themselves and to be recognised for any merit whatsoever.

As the Augustinian monk that he was, Luther was greatly influenced by Augustine's ideas on free will and divine grace. While in the first phase of Augustine's thought after his conversion, ample space is devoted to the subject of free will as the source of human merit and responsibility[189], especially in his *De libero arbitrio*, in his later works, written during his controversy with the Pelagian heresy, accused of

[187] Melvin J. Lerner, *The Belief in a Just World: A Fundamental Delusion* (New York: Springer, 1980), 125.

[188] Lerner, vii, 125.

[189] Augustine, "On the Free Choice of the Will," in *On the Free Choice of the Will, On Grace and Free Choice, and Other Writings*, ed. and trans. Peter King (Cambridge: Cambridge University Press, 2010), 3–126.

deifying humanity through its almighty free will, Augustine's focus shifts to the gratuitous grace of God, the only means of salvation for humans whose nature has been corrupted by the original sin[190].

While Pelagius preached that the original sin is replicated through the imitation of Adam, according to Augustine's late thought, after Adam's fall, humanity has become a *massa damnata*, that "bear[s] inborn defect from our mother's womb"[191], that can only be redeemed by divine intervention. It is impossible for humans to emancipate themselves on their own nor to understand the divine plan; there is no space for free will, responsibility, and merit: everything is in the hands of God, to whose unknown will the Christian must submit. This conception will be taken to the extreme by John Calvin, who considers humans to be so detached from God that they cannot even aspire to understand the workings of his choice and justice[192], thus emancipating him from the obligations of proportionality that retributive justice entails and that had generated the idea of purgatory[193].

Augustine's quarrel with Pelagius is reflected in Luther's diatribe with Erasmus of Rotterdam. Luther wrote his *De servo arbitrio* in response to Erasmus' *De libero arbitrio*, which Luther accused of "outstrip[ping] Pelagians"[194] in its defence of free will, considered by Erasmus to be "wounded through [original] sin, […] not extinct"[195]. Erasmus, in truth, does not rule out theological determinism, as he claims that "in all things must we believe that whatever delightful or sad happens to us during life, God

[190] Augustine, "Quaestiones VII Ad Simplicianum," in *Mélanges Doctrinaux; Quaestiones 83; Quaestiones VII Ad Simplicianum; Quaestiones VIII Dulcitii; De Divinatione Daemonum*, ed. Gustave Bardy, Jérôme André Beckaert, and Joseph Boutet (Paris: Desclée De Brouwer, 1952); Augustine, *Confessions*, ed. and trans. Thomas Williams (Indianapolis & Cambridge: Hackett Publishing Company, 2019), 119–38.

[191] John Calvin, *Institutes of the Christian Religion. Volume 1*, ed. John T. McNeill, trans. Ford Lewis Battles (Louisville: Westminster John Knox Press, 2006), 247.

[192] Calvin, 211–13, 743–44.

[193] "For as Augustine truly contends, they who measure divine justice by the standard of human justice are acting perversely." John Calvin, *Institutes of the Christian Religion. Volume 2*, ed. John T. McNeill, trans. Ford Lewis Battles (Louisville: Westminster John Knox Press, 2006), 987.

[194] Martin Luther, "The Bondage of the Will," in *Discourse on Free Will*, ed. Ernst F. Winter (New York: Continuum, 2002), 122.

[195] Desiderius Erasmus, "A Diatribe or Sermon Concerning Free Will," in *Discourse on Free Will*, ed. Ernst F. Winter (New York: Continuum, 2002), 26. Erasmus, as a defender of the Church, cannot deny the importance of divine Grace in salvation, but for this very reason his theory on free will is incomprehensible, even contradictory: "You plainly assert that the will is effective in things pertaining to eternal salvation, when you speak of its striving. And again you assert that it is passive, when saying that without the mercy of God it is ineffective. […] You side with neither party and escape safely through Scylla and Charybdis, in order that coming into open sea, overwhelmed and confounded by the waves, you can then assert all that you now deny, and deny all that you now assert!" Luther, "The Bondage of the Will," 105.

has caused it for our salvation"[196], and yet, talking about Augustine's, Wycliffe's, and Luther's theory of divine necessity, asks: "[w]hat could be more useless than to publish this paradox to the world? […] What a loophole the publication of this opinion would open to godlessness among innumerable people?"[197].

Erasmus' practical defence of free will, which is assimilated to those "errors which it is better to ignore than to eliminate"[198], is echoed in some contemporary compatibilist theses against hard determinism. The most prominent example is that of Saul Smilansky, who, despite acknowledging the nonexistence of free will, considers it a "necessary illusion"[199] in order to maintain a desirable social system based on desert and individual moral responsibility[200]: it is, ultimately, an appreciation for retribution in society that grounds his desire to preserve the illusion of free will. The same could be said of Erasmus himself, who aims at protecting the moral space of guilt and virtue[201], punishment and reward[202]:

> Why, you ask, is anything attributed to the freedom of the will, then? It is in order to justify blaming the godless ones who resist spitefully the grace of God; to prevent calumnies attributing cruelty and injustice to God; to prevent despair in us; to prevent a false sense of security; to stimulate our efforts. For these reasons the freedom of the will is asserted by all.[203]

As all determinist theories, Protestantism does undermine the common concept of retributive justice, so much so that the philosopher Michael Sandel argued that "[t]he Protestant Reformation was born as an argument against merit"[204]. Yet, the interpretation of the Protestant conception of justice in retributive terms remains fairly ambiguous. On the one hand, retribution is somewhat denied, for one can be damned even if they have accomplished good works in their worldly life[205], and another can be saved even if they have done nothing good; on the other hand, retribution is to some extent affirmed, as it is claimed that all are fundamentally evil sinners, for Adam's sin

[196] Erasmus, "A Diatribe or Sermon Concerning Free Will," 9.
[197] Erasmus, 11.
[198] Erasmus, 11.
[199] Saul Smilansky, *Free Will and Illusion* (Oxford: Oxford University Press, 2000), 165.
[200] Smilansky, chap. "The Positive Importance of Illusion", p. 145-191.
[201] Erasmus, "A Diatribe or Sermon Concerning Free Will," 53.
[202] Erasmus, 59.
[203] Erasmus, 93.
[204] Michael Sandel, *The Tyranny of Merit* (New York: Farrar, Straus and Giroux, 2020), 38.
[205] It must be noted that, according to Luther, "when God is not present to work in us, everything we do is evil, and […] we necessarily act in a way not availing unto our salvation". Luther, "The Bondage of the Will," 111.

resides in us independently of our own action, and hence deserve God's wrath and damnation[206]; still, once again retribution is contravened because God arbitrarily chooses to save some, though not deserving, and in this justification by Grace alone lies the inscrutable divine justice.

Thus, Protestantism tries to redefine justice, exempting God from the logic of retribution[207], but the result remains muddled, no less than Erasmus' attempt to save free will, as is evident in Calvin's work. There, at times he tries to justify the damnation of the non-elect in the ordinary sense of retributive justice[208], while at other times he relates the concept of justice to God's will to save humans without reference to human merit or guilt[209]. In this perspective, God's act of gratuitous salvation is upheld as the supreme act of justice, even to the point of identifying justice with the remission of sin granted to some. This ambiguity is unsatisfactorily resolved by recurring to the fundamental inscrutability of divine plans that a finite being cannot aspire to understand[210].

Still, Protestant authors have to face the problem of the normativity inherent to Christian ethics, which Erasmus deemed "disturbing"[211] if divine necessity were to determine every human action: "Why do you [God] blame me, when all my works, good or bad, are accomplished by you, and I am only your tool? […] Why do you curse me when I have merely sinned through necessity?"[212]. Luther replies that law is not a means of empowerment but a means of knowledge: it makes sin manifest, together with one's helplessness against it[213]. But does this mean that there is not "any room for

[206] Calvin, *Institutes of the Christian Religion. Volume 1*, 251, 318. Thus, Protestant authors abolish *individual* responsibility, but not *moral* responsibility *tout court*, for it is considered in a collective form: the sin has been committed by Adam, but all humanity bears collectively its responsibility.

[207] "For even though in our miseries our sins ought always to come to mind, that punishment itself may incite us to repentance, yet we see how Christ claims for the Father's secret plan a broader justice than simply punishing each one as he deserves." Calvin, 211.

[208] Calvin, 248, 317.

[209] Calvin, 744, 751, 763, 765.

[210] An author contemporary with the Protestant Reformation who, although definitely not adhering to the Reformation, shares with it the idea of the incommensurable distance between the human and the divine and the consequent unknowability of divine will and justice is the aforementioned Montaigne. Montaigne's case is interesting because, although starting from similar considerations, he reaches the opposite conclusion, that is, the liberation of humans from the oppression of divine plans, which are so detached from them that there is no point in being concerned about them in this life. Montaigne, *The Complete Essays*, chap. An Apology for Raymond Sebond, 489-683.

[211] Erasmus, "A Diatribe or Sermon Concerning Free Will," 81.

[212] Erasmus, 35.

[213] "By the words of law man is admonished and taught what he ought to do, and not what he can do." Luther, "The Bondage of the Will," 126; Calvin, *Institutes of the Christian Religion. Volume 1*, 324–28. Knowledge is very important in Calvin's conception of guilt: being able to discern good from bad

prayer or our own striving?"[214]. It is the problem of the so-called *ignava ratio*, the laziness that derives from the conviction that everything has been already fatalistically decided, that makes one's efforts futile and pointless, for they cannot change their fate.

To solve this problem, we must briefly turn to the Lutheran conception of "calling" and its successive elaboration in Protestant theology and common belief. As noted by Max Weber, the present-day conception of "calling" or "vocation", "in the sense of a life-task, a definite field in which to work"[215], derives from the Lutheran translation of the Bible, and it is only in Protestantism that worldly activity takes on religious and ethical value[216]. For now, the problem of the *ignava ratio* remains: while mundane activity might have spiritual significance, how could one be convinced to struggle if such struggles do not contribute to their salvation?

As Max Weber sharply remarks, the answer to this question requires another fundamental question that must have haunted believers "in an age to which the after-life was not only more important, but in many ways also more certain, than all the interests of life in this world"[217]. Such question is: "Am I one of the elect?"[218] and has been chiefly addressed by Calvinism and English Puritanism. While Calvin himself did not believe that an answer could be found in one's conduct, for he had built such a distant and impenetrable image of God that any attempt to investigate his decrees is to be considered presumptuous and vain[219], ministers had to address the anxieties of the mass of the followers and did so by exhorting them to dedicate themselves to mundane work, evoking the Lutheran conception of the calling. Calling is a gift of God and is the proof that God works in us and through us; thus, success in mundane activity could be interpreted as a sign of one's election: "But God maintains his order by finally making

---

neutralises ignorance as an excuse for sin, that is committed voluntarily, albeit necessarily. Calvin, 263–64, 282, 296, 309.

[214] Erasmus, "A Diatribe or Sermon Concerning Free Will," 15.

[215] Max Weber, *The Protestant Ethic and the Spirit of Capitalism*, trans. Talcott Persons (London and New York: Routledge, 1992), 39.

[216] Weber, 40. "[T]he fulfilment of worldly duties is under all circumstances the only way to live acceptably to God". Weber, 41.

[217] Weber, *The Protestant Ethic and the Spirit of Capitalism*, 65.

[218] Weber, 65.

[219] Still, in the *Institutio christianae religionis*, he identifies the ability to love as a sign that follows redemption: "Now this is an argument a posteriori, by which something [redemption] is proved from the evidences that follow. [...] By faith, therefore, we gain forgiveness; by love we give thanks and testify to the Lord's kindness." Calvin, *Institutes of the Christian Religion. Volume 1*, 667.

clear with his vocation the grace he previously kept hidden within himself. It can therefore be said that, by calling, he attests his election"[220]. In Weber's words:

> Thus, however useless good works might be as a means of attaining salvation, for even the elect remain beings of the flesh, and everything they do falls infinitely short of divine standards, nevertheless, they are indispensable as a sign of election.[221]

So, one should work hard in order to discover through their success that they are among the elect. However, such a conception eventually re-established the connection between salvation and good works:

> They are the technical means, not of purchasing salvation, but of getting rid of the fear of damnation. In this sense they are occasionally referred to as directly necessary for salvation or the *possessio salutis* is made conditional on them.[222]

Once the link between one's own worldly merits and their otherworldly fate is restored, the gate to the full recovery of afterlife retribution is open and works easily become once again the sources of salvation and damnation[223].

Whereas Luther's original conception promoted an ethics of humility, where everyone is fundamentally a sinner, and all that is earned is an undeserved gift, subsequent Protestant elaborations, especially in Puritanism, endorsed an ethics of mastery that ended up completely reversing the Lutheran original attitude towards oneself and one's neighbour. While Luther inspired the first social welfare policies and institutions all around Germany in the 1520s[224], the Puritan "attitude toward the sin of one's neighbour [is] not of sympathetic understanding based on the consciousness of one's own weakness, but of hatred and contempt for him as an enemy of God bearing the signs of eternal damnation"[225] and if God helps those who help themselves, the successful ones are earning their own salvation and do not owe anything to anybody else.

Here, we can clearly see the roots of the individualistic values that drive today's Protestant countries, now evolved into the "providentialism without God"[226] that constitute meritocracy according to Sandel. In these countries, the ideology of the self-

[220] Calvin, *Institutes of the Christian Religion. Volume 2*, 964.
[221] Weber, *The Protestant Ethic and the Spirit of Capitalism*, 69.
[222] Weber, 69.
[223] Sandel, *The Tyranny of Merit*, 40.
[224] Lindberg, "Luther's Struggle with Social-Ethical Issues," 172.
[225] Weber, *The Protestant Ethic and the Spirit of Capitalism*, 75.
[226] Sandel, *The Tyranny of Merit*, 42.

made man blinds citizens to the unfairness of the social structure and helps to dismantle those social welfare institutions that Protestantism itself contributed to creating in the first place. Pragmatically, though not theologically, Erasmus' effort to save some space for individual responsibility and for an intuitive, proportional, retributive systemic justice prevailed over the Protestants' attempt to glorify God's limitless power that unhinged retribution.

Of course, Weber analyses the development of the Protestant ethics as a sociologist: he is interested in understanding the effects of religion on the development of society. Such views, marked by a pragmatic approach, are not shared by many Protestant theologians, who remained faithful to the original interpretation of Luther and Calvin of the Bible, characterised by a strong Augustinian undertone. The Swiss Protestant theologian Karl Barth, for example, in his works on the *Epistle to the Romans*, repeatedly stresses both the anger of God towards men[227] and, more importantly, their complete passivity regarding their salvation[228], insisting multiple times on the condemnation of anyone who believes that their salvation depends on their own striving[229]. Naturally, the presence of such a view in Protestant theology does not undermine Weber's sociological analysis.

---

[227] "[T]he condemnation of man, preached by the Gospel, does in fact apply to all men. Everyone has every reason to apply to himself God's wrath […]." Karl Barth, *A Shorter Commentary on Romans*, ed. Maico M. Michielin, trans. D.H. van Daalen (Aldershot: Ashgate, 2007), 17. See also: Barth, 14–17.

[228] "Sanctification is entirely God's grace. It is not man's affair, but God's – the affair of the God who works for man in Jesus Christ. No man can take it for himself." Barth, *A Shorter Commentary on Romans*, 39. See also: Barth, 27, 28, 33, 55, 59.

[229] "For this is the disobedience, this is the strong, living sin, compared to which the others are only dummies: the contempt of God's grace, the human effort to lay our hands on that which God wants to be and to do for us, the endeavor to save, safeguard and exalt ourselves, while he wants to be our sole salvation, safeguard and exaltation." Barth, *A Shorter Commentary on Romans*, 50. See also: Barth, 51, 52, 55.

Add to this that whereas for Luther, and Augustine before him, predestination has a consoling value for the believer, who, through their faith, can acquire awareness of their salvation and, hence, pacify the anguish connected to the lack of agency in the economy of salvation, Barth, on the other hand, distances himself from the father of Protestantism, taking up the Kierkegaardian theme of the "silence of God" and, thus, of the anguish of the believer, who cannot have any *ex ante* insight concerning their salvation *gratia sola*. Kierkegaard, in fact, rejects Luther's reassuring conception of religion, considering "Augustine's strong doctrine of predestination as a[n] […] illegitimate coping mechanism for the anxiety of Christian faith." Jack Jr. Mulder, "Grace and Rigor in Kierkegaard's Reception of the Church Fathers," in *A Companion to Kierkegaard*, ed. Jon Stewart (Hoboken, New Jersey: Wiley-Blackwell, 2015), 161. See also: Lee C. Barrett, "Kierkegaard's Appropriation and Critique of Luther and Lutheranism," in *A Companion to Kierkegaard*, ed. Jon Stewart (Hoboken, New Jersey: Wiley-Blackwell, 2015), 182–86.

As Kierkegaard notices in *Fear and Trembling*, the awareness of predestination only comes *ex post*, while life is constellated by the question "how can I be sure?". Søren Kierkegaard, "Fear and Trembling. Dialectical Lyric by Johannes de Silentio," in *Fear and Trembling. Repetition.*, ed. and trans. Howard V. Hong and Edna H. Hong (Princeton: Princeton University Press, 1983), 61–62. This is clearly exemplified by the story of Abraham, the "knight of fate" (Kierkegaard, 38.) who, until the very last,

This digression on Protestant conceptions of divine justice aims at showing the common path that it shares, though starting from the opposite starting point, with early Christianity. Jesus preached an all-loving, graceful and nonviolent God, yet later his Church imagined his justice to be retributive, and purgatory was created. Luther preached a distant God whose just wrath and punishment could hit anybody, independently of their action. As we have seen, he is, in some sense, retributivist because humans bear the guilt of the original sin; however, this attitude cannot be properly assimilated with the common understanding of retribution, which is essentially linked to individual responsibility, whereas for Protestant authors humanity is condemnable because of the collective responsibility for Adam's sin.

In this early Lutheran conception, similarly to the indiscriminate grace dispensed by the early Christian God, damnation is indiscriminately dispensed, and only a few will be gratuitously saved. However, this position will prove to be psychologically untenable for believers[230], and Protestant pastors will emphasise the importance of works in the economy of salvation, first as signs of deliverance and eventually as its causes. Protestant countries will become the most tenacious defenders of the ideal of meritocracy, which is a form of retributivism, as we shall see in greater detail in the next chapter.

1.3.3 Turning Retaliation into Justice

This discussion seems to prove that the connection between religion and retribution is hard to break, but the point might be slightly different. In fact, once religion is removed from the picture in secular contemporary Western states, retribution remains firmly in place. We have not analysed religious conceptions to prove a point about religion itself but to investigate justice. As shown in the second section of this chapter, religion is the first tool humans have discovered in their quest to manage violent outbreaks in their community through sacrificial practices. In more complex societies, sacrifice has been abandoned or emptied of its function, and direct retribution

---

could be revealed as a simple murderer (Kierkegaard, 66.). The leap of faith that Kierkegaard requires, and that Barth accepts, is one deprived of any safety net. In Barth's words: "we cannot, believing in the Gospel, hope to find anything in the field of psychology – certainly no grace and life, not even the knowledge of our real sin." Barth, *A Shorter Commentary on Romans*, 46. In Kierkegaard and Barth, merit becomes a concept even more obscure than in was in Luther's and Calvin's writings, for it cannot be derived from any external manifestation, nor from any internal feeling.

[230] Sandel, *The Tyranny of Merit*, 40.

has become the preferred device to address the problem of reciprocal violence, yet not wholly deprived of a sacrificial function.

While nowadays, in that part of the globe that we call "the West", this task, which goes by the name of "justice", is undertaken by the secular State, for most of our history the State has been strictly intertwined with the Church, and the office of justice has been typically religious. This marks a continuity between pre-retributive, i.e., sacrificial, conflict management and subsequent retributive practices. Once it has lost its primary practical function in preventing violent outbreaks through sacrifice, religion has not only participated in repressive practices but has also created narratives that served the purpose of prevention, namely otherworldly retribution.

Theories of the afterlife complement sacrificial and retributive practices, for they have the deterrent function of punishment, circumventing the practical problems for the pre-bureaucratic state in the implementation of punishment. Not only that, but the creation of a divine plan of retributive justice serves as a mirror, a model, and a justification for the management of repression in this world. Moreover, this fictional reality provides a framework of meaning to withstand the suffering and unfairness endured in this life. Finally, the association of repressive practices with the divine, albeit with all the difficulties and confusions highlighted in these pages, sacralises these practices, that is, it makes them immutable and, crucially, good. We are back to the distinction between good and bad violence that we have previously identified as the act of birth of morality in archaic societies. This means that retribution is not just retribution, i.e., a form of vengeful reciprocity, but justice. Simple retaliation is transformed into a value, and its concrete practice, that is, a form of violence perpetrated against an individual, is so well hidden by its transcendence[231] that it is hardly perceived as violence at all.

Durkheim has superbly put in words the connection between justice and transcendence in criminal prosecution:

> When we demand the repression of crime it is not because we are seeking a personal vengeance, but rather vengeance for something sacred which we vaguely feel is more or less outside and above us. Depending upon time and place, we conceive of

[231] "If the transcendence of the judicial institution is no longer there, if the institution loses its efficacy or becomes incapable of commanding respect, the imitative and repetitious character of violence becomes manifest once more; the imitative character of violence is in fact most manifest in explicit violence, where it acquires a formal perfection it had not previously possessed." Girard, *Things Hidden Since the Foudation of the World*, 12.

> this object in different ways. Occasionally it is a simple idea, such as morality or duty. Very often we represent it to ourselves in the form of one or several concrete beings: ancestors, or a divinity. This is why penal law is not only of essentially religious origin, but continues always to bear a certain stamp of religiosity. This is because the acts that it punishes always appear as attacks upon something which is transcendent, whether this is a being or a concept.[232]

Susan Jacoby offers an accurate representation of Durkheim's argument at the beginning of her essay *Wild Justice*, where she describes the 1972 trial against Hermine Ryan, a then-German-American housewife who had secretly served as a dreaded SS in several concentration camps during World War II. The witnesses were mostly survivors of those concentration camps that, when asked if they were seeking revenge, "replied in controlled, dispassionate tones that they only wanted justice"[233].

Now it should come as no surprise that, even in secular States, retributive justice remains strongly linked to the imagery of value and religion, still often represented as the blindfolded *goddess* holding the scale, symbol of proportionality, in one hand, and the sword, symbol of violent vengeance, in the other, perfectly showing the foundational elements that the discourse around justice tends to hide and mystify. Its sacrificial aspects have been outlined in the previous section; in this one, we have begun to discern its ties with vengeance, which we shall further explore in the next chapter.

[232] Emile Durkheim, *The Division of Labour in Society*, trans. W. D. Halls (London: The Macmillan Press, 1984), 56.

[233] Susan Jacoby, *Wild Justice. The Evolution of Revenge* (New York: Harper & Row, 1983), 2.

2 The Present: Theories and Practices of Justice

In the second section of the previous chapter, we have seen that the vindicatory system proposed by Verdier might be an evolution of a more original victimary system theorised by Girard, both being immersed in the sacrificial logic that uses the transfer of aggression towards substitutive victims in order to preserve the social system. In this chapter, we will move on from the vindicatory system to the State and its justice systems, analysing the transition from the former to the latter, addressing the social and moral changes that caused it and that it required, taking up the work that was left at the end of that section. We were then discussing the topic of religion as a means of conflict management, and we moved on to analyse it with respect to Christianity, a religion embedded in a State context, postponing the investigation of this passage in order to grant conceptual unity. We now pick up what was left behind.

2.1 The State, Justice, and Their Victims

As should be clear, religion and the State as means of managing violence in the community are not mutually exclusive. We have seen how the Christian religion proved to be a strong ally to Medieval and Early Modern States in their quest to mitigate violence in their societies, both pragmatically, taking care of the investigation and administration of penalties, and ideologically, through the elaboration of the afterlife deterrent scenarios. Even today, only a few political systems in the world are secularised, and religion keeps supporting (and sometimes determining) most States' justice systems.

The process that led to the affirmation of the State as the only tool for the management of violence is that of its progressive monopolisation of violence. We have seen that violence has always been dealt with through violence (the Girardian interpretation of Matthew 12:26, "Satan fights against himself"), whether by that of standard sacrifice, or by the sacrificial retaliatory practices in the vindicatory system, or by the threat of violence in the afterlife, or, finally, by the Christian penitential system. The founding homicide designates good violence, i.e., useful and peace-bearing, as opposed to bad, private violence, that is dangerous and warmongering. The progressive establishment of the State centralises the moral authority that defines good and bad

violence, whereas it was previously decentralised in the solidarity groups described by Verdier.

According to Paul Dumouchel, the spreading of the Christian religion played a major role in this process. While the vindicatory system, as we have seen, imposes burdensome duties of vengeance towards equal and rival groups in order to preserve the group's identity, a religion such as Christianity, with its universal scope, helped the process of erosion of solidarity obligations. It might sound surprising: solidarity seems to be one of the backbones of Christianism. However, the injunction to love *anyone* as one's brother provides a precious tool for the dismantling of these duties:

> (…) the practice of charity introduces a new recommendation into our interactions with those toward whom we have no reciprocal relations of solidarity. It teaches us that absence of reciprocity does not free us of all duties toward others, that it does not deliver the Other to us like prey into our hand. Charity disconnects obligation from reciprocity. It is an unconditional "obligation" that upsets the structure of the hostility/solidarity space based on social distance. *It brings close what was distant*.[234]

This eventually "separates help from hostility"[235]: groups are not held together by the sharing of a common enemy, solidarity need not entail retaliation. Even more, forgiveness "transforms a transgression, a rejection or avoidance of one's duty to commit revenge, into the ideal of excellence and perfection."[236]

Of course, Dumouchel does not mean that the authentic, spiritual adhesion to such principles caused the disruption of the vindicatory system, he does not even need to assume that such adhesion took place. Instead, forgiveness and charity give the agent acceptable excuses, even of supererogatory nature, that allow them to opportunistically avoid those dangerous and onerous[237] retributive expeditions their group requires. Going further, we can claim that the very possibility for the affirmation of such values depends on the transformation that took place on the social, economic, and political levels, with the broadening of communities and the intensification of the exchanges that made the vindicatory system obsolete, unable to frame the new relationship within and among societies. Thus, the first two circles of Verdier's system collapse on one

[234] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 119, my emphasis.
[235] Dumouchel, 120.
[236] Dumouchel, 120.
[237] Not only from a practical, but also from a psychological point of view. As Martha Nussbaum put it "Revenge made it impossible for anyone to love anyone" (Martha C. Nussbaum, *Anger and Forgiveness. Resentment, Generosity, Justice* (Oxford: Oxford University Press, 2016), 4.), for one was always "burdened by the continual need to avenge something for someone" (Nussbaum, 4.).

another[238] to create one single community, the national community, defined by geographic (or ethnic) criteria, where, in Hobbesian terms, the right to violence is abdicated in favour of a third entity: the State.

We are now back to where this work began: the origin of the State in social contract theories. We have said that they constitute a philosophical fiction yet that they illuminate the primary function and reason of being of the State: the management of social conflict. We can now add that this is accomplished through the monopolisation of violence. That the main characteristic of the State is "*the monopoly of legitimate physical violence* within a particular territory"[239] is, famously, a thesis advanced by Weber in his lecture on *Politics as a Vocation*, but, of course, though in other terms, Weber here repeats the Hobbesian theory of the alienation of personal rights in favour of the Leviathan.

We are left with a world where solidarity bonds are replaced mainly by mutual indifference, to be understood as a sociological phenomenon, "a new institutional arrangement"[240], rather than a psychological disposition, that "limits violence's capacity to turn into a frenzy" [241], but also "isolates and separates agents, it abandons them to their fates and turns each into a potential sacrificial victim"[242] – but only to be sacrificed by the legitimate means of the State.

While it is commonly said that the State takes up the duty to defend its citizens or subjects[243] , the right to violence must not be confused with the right to self-defence, that is, in fact, retained by the individual in contemporary States. In the vindicatory system, retaliatory expeditions are not only means of self-defence in the sense of the immediate need to assure the personal integrity of the group's members; rather, they are part of a long-term strategy of group conservation. Retaliation is not an immediate defence, although it is a response to previous aggression: there is, in fact, a chronological gap between the violent action that is suffered and the one that is

[238] Dumouchel, "The Barren Sacrifice," 283.

[239] Max Weber, "Politics as a Vocation," in *The Vocation Lectures*, ed. David Owen and Tracy B. Strong, trans. Rodney Livingstone (Indianapolis & Cambridge: Hackett Publishing Company, 2004), 33.

[240] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 138.

[241] Dumouchel, xviii.

[242] Dumouchel, xviii.

[243] For example see: Dumouchel, xi. This conception is brought to the extreme by Nozick, that in his social contract theory imagines a minimal legitimate state (legitimate *as* minimal) emergent from the free competition of different protective agencies in a state of nature, converging into one single agency, i.e., the State itself, for reasons of efficiency. Robert Nozick, *Anarchy, State, and Utopia* (Cambridge: Blackwell Publishers, 2001).

inflicted. Vengeance can be inflicted even when no immediate defence is needed, as is made clear by the common saying, "Revenge is a dish best served cold". This fact is important because such temporal disparity introduces the possibility of a different reaction to injury in the long-term conservation strategy, as opposed to the case in which one is experiencing immediate danger and reacts immediately to aggression. The same is true for the State's conflict management through the criminal system.

2.1.1 Retribution and Rationality

We are used to the common rational justifications of punishment in the form of deterrence and incapacitation: we will discuss them in detail later. For now, it will suffice to recall that the institution of prison as the sole means of criminal punishment is the outcome of the criminal reform of the 18th century that did not question the practice of punishment itself; it only changed its form. Of course, the torture system that was in force before the reform, while it might have had deterrent effects, was not aimed at incapacitating (it might only incapacitate in the extreme case of the death of the accused). The prison system inherited the practice of punishment from this former system, and if incapacitation can be an outcome of imprisonment[244], it is not punishment's first purpose. Offenders are, in fact, detained even if no reasonable risk of them reiterating the crime is conceivable[245].

Violence in the community was first dealt with through religion and its preventive, sacrificial system, that developed into the vindicatory system (sacrificial, as we have seen, at its turn), to eventually be handled by the State. The State shared this task with religious institutions, as we have seen, as well as with private citizens: during the Middle Ages, for example, all across Europe, strengthening States tolerated the feuds originated by offences, though trying to persuade victims to rather accept

[244] Though whether incapacitation really happens is debated, for prisons are often considered "schools of crimes", where juveniles enter for petty crimes and during their imprisonment are steered towards more serious crimes, both because of contact with other criminals and because of the difficulty of rebuilding a life in lawfulness once out of prison. We shall see this in greater detail in the last section of this chapter.

[245] Jacoby cites the case of a woman who killed her lover: "Under pressure of circumstances and her own tangled emotions, Harris acted in a way she never had before and – insofar as it is humanly possible to judge – is unlikely ever to act again. […] But none of these facts is relevant to the question of retribution. A man is dead; Harris killed him. Are we to say that no further punishment is required because she is sorry?" Jacoby, *Wild Justice*, 247. A more extreme example of this procedure, this time ending up in death penalty, is the Eichmann trial, that "present[s] an unusual example of a 'pure' retributive rationale" Jacoby, 247.

economic compensation instead of pursuing direct vengeance[246]. It was a process of State-building:

> […] state-building means, among other things, the attempt to monopolize the exercise of violence; that is to say *replacing private vengeance with state vengeance*. […] Generally, the state-building process, when confronted with a culture of vengeance, is successful where it manages to channel, not suppress, the desire for vengeance.[247]

This process proved to be very long and arduous, as is testified by the difficulty States encountered in eradicating the duel, which was banished in many European States in the 16th century but frequently practised up to the 20th century.

This history of the management of violence before and after the centralisation of the authority in the modern State aims at proving that what is commonly called the administration of justice by the State is actually a form of public vengeance that, as Durkheim said, has not changed in nature in comparison to private vengeance, it is only "better directed nowadays than in the past"[248]. In the words of Girard:

> Our penal system operates according to principles of justice that are in no real conflict with the concept of revenge. The same principle is at work in all systems of violent retribution. Either the principle is just, and justice is therefore inherent in the idea of vengeance, or there is no justice to be found anywhere. He who exacts his own vengeance is said to "take the law into his own hands". There is no difference of principle between private and public vengeance; but on the social level, the difference is enormous. Under the public system, an act of vengeance is no longer avenged; the process is terminated, the danger of escalation averted.[249]

While now punishment has taken on the aseptic – yet inhumane – form of incarceration and the aspect of the repetition of violence has been somewhat concealed, like the Erinyes hidden under the Areopagus at the end of the *Oresteia*, as Girard points out, the logic that underpins it is the same as that of the *lex talionis*: the infliction of suffering as a response to those who caused suffering first. This is the heart of the

---

[246] Daniel Lord Smail and Kelly Gibson, eds., *Vengeance in Medieval Europe: A Reader* (Toronto: University of Toronto Press, 2009), xvii–xviii. See also: "The most infamous form of dispute was undoubtedly the feud or blood-feud. This was common in both continental and insular Europe, and had been described by the first-century Roman author Tacitus in his Germania. The feud can be defi ned as a conflict or series of conflicts between individuals or groups of individuals – usually kindred groups – resolved either through private vengeance or by means of the payment of compensation for the initial offence, which in many cases was a homicide. The killing of the slayer by the kinsmen of the victim should not necessarily be regarded as being symptomatic of a lawless society, though the payment of compensation was presumably a more socially acceptable means of resolution." David E. Thornton, "Communities and Kinship," in *A Companion to the Early Middle Ages: Britain and Ireland, c.500-c.110* (Hoboken, New Jersey: Wiley-Blackwell, 2009), 100.

[247] Smail and Gibson, *Vengeance in Medieval Europe: A Reader*, xvii, my emphasis.

[248] Durkheim, *The Division of Labour in Society*, 48.

[249] Girard, *Violence and the Sacred*, 16.

criminal justice system: it is a concession to the vengeful desires of the citizens, though also imposing a limitation on those passions through the respect of proportionality and of guilt.

Literature, being more concerned than philosophy with the investigation of the characters' motives than with moral ideals and justifications, has always understood that. For example, as Jacoby says, in the Elizabethan revenge tragedies, "[r]evenge and justice were not seen as antipodes but as regions of the same moral territory"[250], the tragedy being caused not by the fact of revenge itself but by the usurpation of the right to revenge from the State into the hands of the private individual, with all the violent consequences mentioned above[251].

Remarkably, in today's superhero fiction genre, some of the best-loved superheroes, including Captain America, Hulk, Spider-Man, Thor, and Iron Man, are precisely named *Avengers*. Superheroes take the protection of citizens into their hands when the State fails, as in the crime-plagued Gotham City (inspired by real-world Chicago, the city with the highest rate of violent crime in the U.S.), the setting for the *Batman* saga. The topic of the failure of institutions and its consequences on the escalation of vengeance-related violence is brilliantly addressed by another popular entertainment production, the American crime television series *The Wire* (2002-2008), which chronicles the drug dealing market and its rules in the housing projects of Baltimore, high-poverty areas afflicted by exceptionally high violent crime rates. Remarkably, *The Wire* has been cleverly juxtaposed to the *Njála*, a Medieval Icelandic saga centred on the spiralling destructive force of blood feuds, in the comparative analysis by John Hudson and Mark West[252].

European jurists may oppose that the rehabilitative aim of punishment is clearly integrated in the European legislation, as in the European Prison Rules that, albeit not legally binding, set the standards for the administration of the carceral system in the European Union. Terms like "rehabilitation", "reintegration", and "education" are recurring in their statements[253]. In Italy, this principle was already firmly stated in the

[250] Jacoby, *Wild Justice*, 48.

[251] Jacoby, 34–48.

[252] John Hudson and Mark D. West, "Stringer's Saga: Njal and The Wire," in *Emotion, Violence, Vengeance and Law in the Middle Ages. Essays in Honour of William Ian Miller*, ed. Kate Gilbert and Stephen D. White (Leiden and Boston: Brill, 2018), 271–95.

[253] Committee of Ministers, "Recommendation Rec(2006)2-Rev of the Committee of Ministers to Member States on the European Prison Rules" (Strasbourg: Council of Europe, 2020).

1947 constitution, article 27 of which reads: "[…] Punishments may not consist of treatment contrary to a sense of humanity and must tend towards the re-education of the sentenced person […]."[254] Because of these general guidelines, opposers of my view insist that rehabilitation is to be considered the foundation of punishment; but it does not seem to me that either European rules or the Italian constitution actually say this. In both cases, punishment is presupposed, and re-education is an addition to it: punishments should *tend* towards re-education, but first and foremost, they are punishments, i.e., "intentional infliction[s] of suffering"[255]. When we want to educate someone, our first attempt does not involve intentionally inflicting suffering on them. In the past, when children would fail at something, schools and parents often resorted to punishments, but this practice has been fortunately abandoned, proven ineffective besides being cruel. If the point of criminal justice were forward-looking, focused on education and integration of the offender, we would hardly recur to punishment, and more specifically to a form of punishment that promotes isolation and stigma, *as a first approach*.

If the imposition of the limit of proportionality serves the purpose of distinguishing wild, private vengeance, often disproportionate because driven by the feeling of anger instead of the cold evaluation of facts and motives, the fact of putting a limit to the outcome of a passion does not distinguish the two in principle. While anger, as we shall shortly see in detail, generates the moral intuition (and illusion) of the intrinsic goodness of payback, proportionality satisfies another moral delusion of an aesthetic nature: that some sort of cosmic balance is restored – whence the representation of the divinity of justice holding the scale.

In the *Poetics*, Aristotle gives an example of the best kind of plot, that he identifies in that of the statue of Mitys. In the story, Mitys was killed by a man who, in turn, died crushed by the accidental fall of Mitys' statue[256]. Stories like this give us an intimate satisfaction, a feeling of restoration of the *status quo ante*. Of course, this is a delusion: the killing of the murderer of Mitys does not bring Mitys back to life; it does not actually restore anything, nor does any other form of punishment. As Girard aptly

[254] "[…] Le pene non possono consistere in trattamenti contrari al senso di umanità e devono tendere alla rieducazione del condannato […]", my translation.

[255] Ted Honderich, *Punishment. The Supposed Justifications Revisited* (London: Pluto Press, 2006), 4.

[256] Aristotle, *Poetics*, ed. and trans. Malcolm Heath (London: Penguin Books, 1996), 17.

puts it: "The law against murder involves nothing other than the repetition of murder."[257]

While retributivist thinkers identify an intrinsic good in the proportional suffering of the guilty, it is not enough to state that something is an intrinsic good; the intuition must be supported by something else for the assertion to be convincing, we will see in the next section if retributivist thinkers have provided satisfactory evidence in support of their thesis. For now, we have nothing more than the satisfaction of both retributive emotions – on which we should later return – and aesthetic desires: apparently, not much to fill the moral gap between vengeance and the value of justice.

What this appeal to proportionality does, however, is to give an appearance of rationality to the practice. The State itself, as shown by social contract theories, is presented as a rational entity: the outcome of the rational agreement between individuals[258]. This rational – fictional – process that grounds modern States, this narrative that I use here as a symptom of our perception of the function and characteristics of the State, can be contrasted with the previously-mentioned African sacred monarchies. There, as we have seen, the king is the depositary of the community's violence but is also its victim through the sacrificial rites described. On the other hand, the State is such a powerful entity that it can manipulate violence and remain untouched by it. Being apparently uncontaminated, it can hide it.

The progressive centralisation of moral authority, the monopolisation of violence, is sustained by the severe condemnation of private vengeance that, from a moral obligation in the vindicatory system, is, as Dumouchel put it, de-ritualised and *naturalised*: it is now conceived as a spontaneous reaction[259], not only deprived of moral value but deplored. As is evident in the revenge tragedies abovementioned, and as we have said in the first chapter, the form of the action is the same, but its meaning is different: the unanimity of the institutionalised vengeance that does not initiate blood feuds makes it good and worth pursuing, whereas vengeance exercised by the private individual, or the non-institutionalised group, is bad and banished as an irrational emotion-driven destructive force. A taboo is raised on vengeance, so strong that the connection between the State's punishment and private retaliation is forgotten, hidden

[257] Girard, *I See Satan Fall Like Lightning*, 84.

[258] Or the rational establishment of a single protective agency in Nozick's anarcho-capitalist proposal.

[259] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 120.

behind the *ex-post* rationalisations of deterrence, incapacitation, re-education, but also proportionality and, in a word, *justice*.

Another constraint that the rationality of the State imposes on vengeance is the prohibition of hitting innocent victims: suffering can only be imposed on the defendant proven guilty. At first glance, one could be surprised: vengeance seems to be aimed at culprits by definition; thus, the State would not add anything to simple revenge concerning this aspect. However, one must remember that vengeance outside of the context of the modern State is a group phenomenon. As we have seen, it is a duty relative to that intermediate space where groups recognise each other as equal and rival, among which an offence to the individual corresponds to an offence to the entire group and constitutes an existential threat to it. Just as the distinction between the material victim and their group vanishes in the vindicatory system, so does the distinction between the material offender and their group: members of the group are interchangeable. Thus, retaliation could hit innocent victims, as it often did: in fact, the material offender is likely a strong and violent person – for they have already hurt someone – therefore, they are also a dangerous target, one hard to reach and to defeat: for the sake of the retaliating group's integrity, another member of the rival group, one weaker and more vulnerable, could be chosen[260]. Of course, the designation of such individuals as *innocent* is the judgement of someone (me) external to the vindicatory culture and immersed in a culture of individualised responsibility. In the vindicatory system described by Verdier, we have, thus, reason to believe members of the group shared the culpability because of their being members of that group.

Innocents, or so now considered, were also intentionally made to suffer in the torture system that preceded the criminal reform of the 18th century through judicial torture, as reported by Foucault[261]. The torture system that he describes in the first part of *Discipline and Punish* is a State system that does not aim at hiding the vengeful nature of its justice. Torture, in fact, is the punishment for breaking the law, which is the enactment of the king's will; this makes any crime a form of lese-majesty. Through torture, the king takes his revenge against the guilty party, a revenge that must be

[260] For example, to avoid retaliation, the Siberian Chukchi kill an innocent member of their own group, refraining from hitting the guilty party. Girard, *Violence and the Sacred*, 25–26.

[261] Foucault, *Discipline and Punish*, 39–42.

disproportionate – another distinction from current systems – to reaffirm the disproportion between the king himself and his subjects[262].

But in that system, torture was not only the culprit's punishment; it was also a means to discover the truth about the defendant's guilt. Yet, during the investigation, torture was not used as one would nowadays think, as "a way of obtaining truth at all costs; it was not the unrestrained torture of modern interrogations; it was certainly cruel, but it was not savage. It was a regulated practice, obeying a well-defined procedure"[263]. It was a "judicial game"[264], as in the old ordeals: a competition between the defendant ("patient"[265]) and the judge, the former needing to resist the procedure in order to prove their innocence.

Still, torture was also a form of punishment, not only in the sense that, after the guilty verdict, other tortures would be imposed on the offender but also meaning that the same judicial torture the defendant went through before the verdict was *already* a way of punishing them: "it was a penalty so grave that, in the hierarchy of punishments, the ordinance of 1706 placed it immediately after death"[266]. Punishment could be imposed before the guilty verdict once again because such culture did not share the present conception of responsibility: this time, contrary to the case of the vindicatory system analysed before, the problem does not lie in the *individuality* of responsibility but in its binary character. We are used to considering guilt as either present or absent, but the torture system did not work this way: even someone who would eventually receive a verdict of innocence, by the very fact of being investigated, of having raised suspicions upon themselves, was already considered partly guilty and deserved punishment[267].

Finally, the last practice of voluntary infliction of suffering on an innocent that is abandoned by the modern State is, of course, that of the sacrifice, which, as we have seen, even more than vengeance or torture, is relegated to the realm of the incomprehensible – so much so that the extraordinary phenomenon of the omnipresence

[262] Foucault, 47–49.
[263] Foucault, 40.
[264] Foucault, 40.
[265] Foucault, 40.
[266] Foucault, 41–42.
[267] "The suspect, as such, always deserved a certain punishment; one could not be the object of suspicion and be completely innocent. Suspicion implied an element of demonstration as regards the judge, the mark of a certain degree of guilt as regards the suspect and a limited form of penalty as regards punishment. A suspect, who remained a suspect, was not for all that declared innocent, but was partially punished." Foucault, 42.

of sacrifice in independently developed cultures has long been left unexplained and, surprisingly, dismissed as a purely irrational occurrence to be better forgotten rather than studied. Although sacrifice is undoubtedly not a form of punishment, it must not be forgotten that its links with the world of retributive justice are deep: from its origin as a repetition of the founding homicide, in itself interpreted as a justified retaliation against the cause of the mimetic crisis[268], to its role in preventing violence through violence, as punishments are supposed to do, though using partly different mechanisms, in their alleged deterrent function, to the treatment of sacrificial victims sometimes forced to commit crimes in preparation for the sacrificial rite, as if they were to be made guilty to "deserve" the violent treatment of the sacrifice[269].

So the rational State only victimises victimisers, as we have previously seen, and, as we then remarked, thereby conceals the irrationality of the practice of punishment itself. As noted by Girard, "[i]f our system seems more rational, it is because it conforms more strictly to the principle of vengeance. Its insistence on the punishment of the guilty party underlines this fact."[270] In this way, it also manages to hide the sacrificial character of the institution that we mentioned: the victims of the legitimate violence of the State thus serve as surrogate victims to vent the citizens' anger and frustration.

Nancy Scheper-Hughes, who agrees with a sacrificial interpretation of this form of institutionalised violence towards offenders, compares it to other forms of legitimate violence in society, resuming Franco Basaglia's concept of "peace-time crimes"[271]. Specifically, she investigates the violence exercised towards the elderly in retirement homes, where overworked and underpaid staff turn "the persons and bodies under their protection into things, into bulky objects that can be dealt with in shorter and shorter

[268] The founding homicide is certainly interpreted as the elimination of the cause of the crisis, but it is also a *deserved* death of that who is responsible for it: it is not considered on a par with the elimination of the cause of a natural disaster, the scapegoat is considered *evil* and as such it must be treated, as is clear in the "horrible miracle of Apollonius of Tyana" discussed above, where the beggar's eyes are described as "full of fire" and the old man is eventually uncovered to be "a Molosian dog, but was in size equal of the largest lion […] and vomiting foam as mad dogs do." (Girard, *I See Satan Fall Like Lightning*, 50.) On the guilt of the surrogate victim see also: Girard, *Violence and the Sacred*, 83.

[269] It is the case of the African sacred monarchies described in the previous chapter. See: Girard, *Violence and the Sacred*, 104.

[270] Girard, 22.

[271] "Crimini di pace", a term that gathers "all institutionalised violence, which serves as a preservation strategy for our social system". Franco Basaglia and Franca Basaglia Ongaro, "Premessa," in *Crimini Di Pace. Ricerche Sugli Intellettuali e Sui Tecnici Come Addetti All'oppressione*, ed. Franco Basaglia and Franca Basaglia Ongaro (Torino: Einaudi, 1975), my translation.

intervals."[272] Dumouchel agrees with Scheper-Hughes' comparison of the two phenomena and traces them back to the indifference that governs the contemporary world[273]:

> What links these different forms of indifference is the social mechanism that makes them possible. What allows us to be indifferent to the fate of another is that we are neither obliged to intervene in that person's favor nor to hate that person. In other words, the important factor here is not the physical distance or the difficulty in knowing or imagining what is happening to the other; it is not even the social distance […]. It is what we can call the *moral distance* between the agents.[274]

Whereas I agree that our treatment of the elderly is attributable to such indifference, resulting, as we have seen, from the erosion of the solidarity bonds of the vindicatory system, and I would also relate it to the topic of "violence by omission" that Dumouchel investigates later in his book[275], the case of the State's violence towards offenders in the criminal justice systems seems to me to be different, though in part related.

In the category of "violence by omission" Dumouchel includes all the suffering endured by the outcast of the society, "who are responsible for their misfortune and to whom no one owes anything"[276]. Such thinking is the product of an individualistic culture that has banned the concept of luck from its moral thought, believing in the myth of the self-made man, the ideal of the European Renaissance (most famously in Pico della Miradola's *Oration on the Dignity of Man*) that was brought to its extreme consequences by the present-day meritocratic ideal.

2.1.2 Meritocracy and its Losers: A Digression

The subject of meritocracy warrants a digression of its own, though the topic, which is addressed in issues of distributive justice, is broad, and there is no pretension

---

[272] Nancy Scheper-Hughes, "Coming to Our Senses: Anthropology and Genocide," in *Annihilating Difference. The Anthropology of Genocide*, ed. Alexander Laban Hinton (Berkeley, Los Angeles, and London: University of California Press, 2002), 373.

Scheper-Hughes also addresses the topic of the Brazilian "street children", the children of the poor that "pollute" the middle-class space when they get out of their favelas: "'Street children' are often described as 'dirty vermin' so that unofficial policies of 'street cleaning', 'trash removal', 'fly swatting', and 'pest removal' are invoked in garnering broad-based public support for their extermination." Scheper-Hughes, 372. This example seems to me to be closely comparable to that of the treatment of offenders and I will not explore it further.

[273] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 25–26.

[274] Dumouchel, 26.

[275] Dumouchel, 137–39.

[276] Dumouchel, 139.

here to exhaust it. Its connection with retributivism, which has already been mentioned, stems from the fact that both are based on the idea that justice consists in everyone receiving what they deserve and, in fact, I previously claimed meritocracy to be a form of retributivism. It is worth remembering that the term "meritocracy" was coined in 1958 by sociologist Michael Young in his *The Rise of the Meritocracy*. Surprisingly enough, his use of the term was not intended to be positive: rather, it was meant to recall the concept of "aristocracy" and to show that the appeal on merit would not eventually differ so much from the old appeal on birth rights once it came to limit one's opportunity of development.

Young sets his tale in 2034, a dystopian future where intelligence, in the form of the productive intellectual contribution to the country's economic growth in the context of aggressive international competition, has become the sole criterion for the establishment of social classes and for the distribution of power and economic benefits in society. Although, apparently, by the affirmation of the meritocracy, "[a]greement replaced disagreement, and merit was recognized as the principle which should guide economic, as well as educational, reform"[277], the book is set in a time of crisis. It is, in fact, the analysis conducted by a fictional sociologist in the face of the riots that the so-called "inferiors", allied to the feminist movement, are carrying out. The narrator aims to investigate the causes of his present-day protests and goes through the reforms that, from 1870 to 2033, have brought to the establishment of the meritocracy. I will sketch here the most important steps.

The narrator starts criticising the old aristocratic system that divided society into immutable castes that determined one's destiny according to their birthright. This system was not only unfair but also highly inefficient, for it did not select leaders based on their capacity[278]. It was thanks to the socialists that eventually everyone was granted "equality of opportunity"[279], i.e., access to education independently of one's family's

[277] Michael Young, *The Rise of the Meritocracy 1870-2033. An Essay on Education and Equality* (London: Penguin Books, 1961), 161.

[278] "Inheritance was denounced for the simple reason that a developing industrial country in competition with others could no longer afford second-rate leaders; the needs of the economy reshaped society." Young, 87.

[279] Young, 37.

wealth. This is still nowadays an ideal pursued and worth pursuing that many Western countries are – albeit slowly – approaching, although still with many failures[280].

The interesting question that Young asks is: what happens when the much-desired equality of opportunity is reached? Most likely, differences in capacity will not be levelled out, but they will be highlighted precisely by this equality of opportunity, which will prevent them from being traced back to extrinsic factors. In this fictional history, egalitarians then started to demand that such differences be ignored and that everyone be educated equally[281]. However, whereas the narrator agrees that in an ideal world with unlimited resources, everyone should get the best possible education, the only way to resist the pressure from the global competition was to give the finest education to the best students and to direct the worst students towards manual work[282], accepting individual differences and making intelligence the only criterion for social discrimination.

Young makes this process as fair as possible: not only were students paid to go to school so that their parents' wealth or preference would not condition their access to education[283], but they were also tested multiple times so that "all persons, however humble, know they have had every chance"[284] and that their poor performances were not due to some external accidental event. The resulting so-called "inferiors" did not have access to the highest degrees of education nor to positions of political or economic power. Nevertheless, the state-sponsored narrative did not promote contempt for them: on the ideological level, the "Mythos of Muscularity" served as an appreciation for manual work, to be valued almost as intellectual work[285]; on the economic level, much effort was made to narrow the differences in compensation and justify them not on the basis of merit but for reasons of efficiency. In fact, in his subtle analysis, Young envisages the social criticism that would arise out of a too significant difference in wages – like the one we witness today – and thus imagines everyone to receive the same salary, "the Equal". The difference in compensation for intellectual and more productive

[280] These are evident particularly in Anglo-Saxon countries where most high-quality higher education institutions are extremely expensive. Such failures, that our States imperatively need to address, must not hide the fact that much progress has been made in that direction as compared to the past.

[281] Young, *The Rise of the Meritocracy 1870-2033. An Essay on Education and Equality*, 41.

[282] "In an ideal world, not hampered by shortage of resources, the unfortunate could have large sums spent on them too. But it was not, has not been, nor ever will be, an ideal world. The choice was between priorities, and there was no doubt how the decision had to go." Young, 61.

[283] Young, 58.

[284] Young, 107.

[285] Young, 109.

work, then, consisted of other economic benefits, such as the payment of domestic workers and other "varying expenses as could be justified by the needs of efficiency."[286]

Finally, a eugenic programme encouraged citizens to choose their partner among people with a similar I.Q., therefore basing their relationship on intellectual affinity – rather than other irrational criteria, such as attractiveness, as was done in the past – so that the chance to generate more brilliant children would increase, thus also increasing parents' happiness and satisfaction[287].

Despite the fairness of the selection process, the general agreement on the meritocratic principle, and the equitable distribution of wealth, one cannot but have the feeling of a dystopian rather than a utopian future, disturbingly recalling the Platonic ideal of the *Republic*. There, as well, power was distributed according to the criterion of intellectual superiority, the meritocracy corresponding to the government of the philosophers. There, too, the author expects that citizens of all classes – philosophers, guardians, and producers – would have nothing to complain about, their roles being determined by their nature, their aptitudes, therefore maximising their individual happiness along with the general good of the system. However, in Young's world, riots are just around the corner: the optimistic sociologist, who concludes his essay with the prediction that such turmoil will be "at best an 1848"[288] with no actual revolution in sight, is killed during those very protests he dismissed[289].

So why is there so much discontent in the meritocratic society? The imaginary sociologist quotes in his essay the dissidents' "Chelsea Manifesto"; here is an excerpt:

> *Were we to evaluate people, not only according to their intelligence and their education, their occupation, and their power, but according to their kindliness and their courage, their imagination and sensitivity, their sympathy and generosity, there could be no classes.*[290]

Young aptly suggests that the problem lies in the moral arbitrariness of intelligence as a criterion for social evaluation that divides citizens into "superiors" and "inferiors". After all, he manages to show in his essay, perfect meritocracy does not differ so much from the old aristocracy that it replaced, which was based on birth

[286] Young, 159.
[287] Young, 174–75.
[288] Young, 188.
[289] Young, 190.
[290] Young, 169.

privileges[291]. He remarks how intelligence itself consists of a "natural ability" greatly influenced, though not entirely determined, by genetic inheritance, that is not under the individual's control: some individuals possess it, and others do not. Of course, as is noted by Sandel, there is nothing right or wrong in a certain natural distribution of talents[292]; the problem lies in its moralisation.

Assuredly, talent, "merit", is the best criterion for the assignment of jobs – it is evident that in a competitive job market, the job should be taken by those who are better at doing it – and it could also be an efficient criterion for a certain distribution of wealth, as an incentive to take certain jobs that require hard work, specialised education, and much responsibility. However, all this regards economic efficiency, not moral value: in fact, those considered "more deserving" in our economic system – where the distribution of wealth is so unequal as to lose its efficiency justification anyway – are hardly ever the moral actors on the market.

In Young's dystopia, every effort is made to hide the moral character of the meritocracy from the equal compensation to the Mythos of Muscularity, yet the meritocracy cannot avoid discontent. As Sandel remarked discussing Rawls' distinction between desert and entitlement, the latter being deprived of moral value and depending only on the benefits that the distribution produces on society, "[s]ocial esteem flows, almost ineluctably, to those who enjoy economic and educational advantages, especially if they earn those advantages under fair terms of social cooperation."[293] Notwithstanding the noble intentions or the palliative remedies put in place, the division between the successful and the losers of the system is perceived as a moral distinction, which corresponds, at least in part, to the distinction between the best and the worst.

Add to this that in our ideal conceptions of meritocracy, differently from Young's, such efforts to mask the moral character of the evaluation of individuals are absent, and meritocracy is not generally understood simply as the most efficient system but as the most just, where people get what they deserve. *Desert* is a moral concept that

[291] As noted by Sandel, Thomas Jefferson, who wanted to create a public education system to find and cultivate the smartest among the poor, in one of his letters specifically contrasted "the artificial aristocracy founded on wealth and birth" to "the natural aristocracy [based on] virtue and talents". Thomas Jefferson, "Jefferson to Adams. Monticello, October 28 1813," in *The Adams-Jefferson Letters. The Complete Correspondence between Thomas Jefferson and Abigail and John Adams*, ed. Lester J. Cappon (Chapel Hill and London: The University of North Carolina Press, 1987), 388. See also: Sandel, *The Tyranny of Merit*, 160.

[292] Michael Sandel, *Justice. What's the Right Thing to Do?* (London: Penguin Books, 2009), 165.

[293] Sandel, *The Tyranny of Merit*, 145.

concerns what is *owed* to someone. It can be declined both positively ("they deserve their happiness") and negatively ("they deserve that fine"). On the other hand, *merit* only covers the positive side of the spectrum: someone *meritorious* is someone to be praised. Both appeal to the retributive sentiments that generate our feeling of justice when things go wrong in some way for someone who has acted wrongly, as well as when things go right for someone who has behaved properly.

When we try to move away from the emotional level, however, we struggle to understand what desert and merit actually mean: although we feel that it is *just* for someone who has been morally bad or good to be punished or rewarded, and we call this intuitive connection, this "mysterious relationship"[294], "desert", to explain the concept rationally proves challenging. To make sense of why it is so hard to grasp it rationally, David Miller inscribes it within the participant perspective theorised by Peter Strawson: like reactive attitudes and emotions, we cannot appreciate it from an objective perspective[295]. As I will show at length in the following section on moral intuitionism, I find this to be extremely troublesome for a moral theory; for now, with respect to meritocracy, I will simply note that these conceptions of desert and merit, which are fundamentally, albeit obscurely, moral, have nothing to do with the efficient distribution of wealth that promotes economic development: their use in this context is an unjustified moralisation of the economic system.

The meritocratic ideal as it is conceived in common thought unveils the retributive character of common conceptions of distributive justice: we do not only believe that the assignment of jobs should reflect the talents of each individual so as to have the best candidate for each task – however utopian the very concept of having one person who is the best fit for the role may be. We believe that the people who possess the specific talents that are valued in our society are justified in benefiting from more than what is simply justified in terms of job allocation – not only in economic terms but also in terms of prestige, with all the consequences in power relations that derive from it. The fact that a job that is very demanding on the person performing it but also extremely valuable for society requires incentives – economic or even in terms of prestige – to make sure that a suitable candidate is found is not questioned; the problem

[294] David Miller, "The Concept of Desert," in *Social Justice*, ed. Matthew Clayton and Andrew Williams (Malden, MA: Blackwell Publishing, 2004), 189.
[295] Miller, 189–90.

lies in all that exceeds these practical questions that regard the efficient allocation of resources in the job market for socially desirable purposes.

Why should someone who works more or harder or even better than others but whose work does not have socially desirable outcomes *deserve* more? Of course, the *reason* why techno magnates become billionaires and acquire enormous political power is not because the people believe it to be right; it rather depends on complex economic mechanisms. However, at least part of the reason why *politically* and even *morally* (i.e., in thinking that it is right for those people to possess an immense amount of wealth and power) such inequalities are accepted is that those who do exceptional work – despite it being morally neutral, if not negative – are considered in some way *better* than the others. Of course, "better" is a generic term: one can be better than another at washing the dishes, or resolving mathematical problems, or painting. However, its basic adjective is *good*, the fundamental concept of morality. Although the moral meaning of the word *good* is clearly separated from its neutral connotation upon reflection, I believe this ambivalence generates confusion and misunderstanding in common thought and talk. Hence, those people who are simply *better* at some task that is valued by the contemporary economic system are somewhat perceived as *better* overall – hence, morally better and, thus, deserving more. This then justifies *any* inequality in wealth and power.

In the previous chapter, we have seen that, for the Protestants, economic success could be considered a sign of salvation (and eventually its cause) in what Weber calls "inner-worldly asceticism"[296]. In common secular thought, economic success may not be related to spiritual salvation, but it is certainly related to social appraisal, and, even more so, economic failure is connected to overall failure in life. Nevertheless, as the fictional authors of the Chelsea Manifesto note, the criteria chosen by the economic system for one's success or failure are arbitrary; there is nothing inherently moral in them – if anything, they are often rather immoral, for it is not only intelligence that is rewarded by the system but also aggression, the ability to lie, the lack of scruples. Both the possession of talent and the valorisation of that specific talent by society are morally arbitrary. Moreover, capacities only become talents in the relationship between the person who has the capacity and the society that exploits it: being tall and good at

[296] Max Weber, *Economy and Society. An Outline of Interpretive Sociology*, ed. Guenther Roth and Claus Wittich, trans. Ephraim Fischoff et al. (Berkeley, Los Angeles, and London: University of California Press, 1978), 542.

jumping is only relevant in a society where basketball has been invented, and many people enjoy watching it; being strong, aggressive, and insensitive can be valorised in a society based on war or honour; ease in solving mathematical problems is a talent in a world dominated by the economy and I.T.. So, talents not only do not possess an intrinsic moral value, but even their morally neutral instrumental value depends on arbitrary circumstances.

Hence, we see that there is nothing that one deserves in itself, but one can come to acquire entitlement, in Rawls' terminology, because of the rules discretionarily chosen for a particular institution or competition: if it is decided that a race is won by the fastest, then, according to the rules subscribed by the participants, the one who takes the shortest time to cover the set distance is entitled to win; if it is decided that the job will be assigned to the person who possesses the best mathematical skills, the best mathematician is entitled to get the job.

Such selection processes necessarily create losers as by-products of competition. Whereas in Young's perfect meritocracy there might not even be competition anymore, for the highly efficient system created is able to assign anyone to the specific role corresponding to their ability, our imperfect aspirant meritocracy is based on competition. Both Bernard Perret[297] and Jean-Pierre Dupuy[298] have proposed a mimetic and sacrificial reading of economic competition. Perret has highlighted its inherent violence, especially on the international level, where it is "never totally free of political and even military implications"[299]. On the other hand, on the individual level, the competition promotes the internal mediation that segmentary societies tended to block[300].

A segmentary society, like the medieval feudal society, partly prevents the conflict that derives from the mimetic competition by dividing its population into insurmountable castes: it, thus, prevents rivalry, because the acquisition of the other's object of desire – if ever materially possible – cannot put one in the other's place, for their place is not determined by their possession but by their birth status. Thus, the

[297] Bernard Perret, "The Economy as the Opium of the People," in *Palgrave Handbook of Mimetic Theory and Religion*, ed. James Alison and Wolfgang Palaver (New York: Palgrave Macmillan, 2017), 287–94.

[298] Jean-Pierre Dupuy, "Secularization," in *Palgrave Handbook of Mimetic Theory and Religion*, ed. James Alison and Wolfgang Palaver (New York: Palgrave Macmillan, 2017), 271–77.

[299] Perret, "The Economy as the Opium of the People," 288.

[300] See *supra*, footnote 129.

mediation of the "superior other" was external, the mediation of someone with whom we could not possibly compete because we could never take their place.

Things are radically different in our more egalitarian societies, where status depends mainly on the acquisition of goods – what the sociologist Thorstein Veblen called "conspicuous consumption"[301] – and mediation can always become internal, especially where competitiveness is so valorised as in meritocratic societies. Nonetheless, analogous forms of consumption have been testified even in ancient cultures. A prominent antecedent is, evidently, the *potlatch*: the traditional feast practised among the indigenous peoples of Northwest America, in which the host must give away or even destroy valuable goods in a ceremony that aims at affirming their power. The other participants in the ceremony are set to host a *potlatch* in their turn, in a wealth competition to establish power hierarchies based not on the possession of goods itself but on their distribution or destruction. According to Girard, "the *potlatch* […] is nothing but a ritualised form of mimetic rivalry on the social scale: what is important is not the object that you offer, but the humiliation that you want to inflict on the rival tribe […]."[302] The violent character of the practice is further emphasised by Anspach, who remarks that even slaves were eventually considered goods that could be destroyed in the *potlatch*, "resurrect[ing] a form of human sacrifice" in the course of the ceremony.[303]

Of course, such ceremonies represent paroxysmal isolated events of extreme intensity. In contemporary life, on the other hand, if it does not reach that supreme intensity, conspicuous consumption is normalised and widespread among all social strata as a part of everyday life. So much so that Dupuy was brought to characterise the contemporary world by a constant "low-gear mimetic crisis"[304] whose sacrificial character lies in its inevitable production of losers[305]. Much more could be said about the mimetic and sacrificial interpretation of meritocracy; I will not delve further into

[301] Thorstein Veblen, *The Theory of the Leisure Class*, ed. Martha Banta (Oxford: Oxford University Press, 2007), 49–69.

[302] Girard, Antonello, and de Castro Rocha, Evolution and Conversion. Dialogues on the Origins of Culture, 252.

[303] Anspach, "Apprivoiser La Violence. Le Sacrifice, La Vengeance, Le Don," 134, my translation.

[304] Dupuy, "Secularization," 272. In Girard's words a "mimetic conflict […] [that] does not *normally* degenerate into a fight to the death". Girard, *Things Hidden Since the Foudation of the World*, 93, my emphasis. For the continuity between war and economics see: Perret, "The Economy as the Opium of the People," 288.

[305] Perret, "The Economy as the Opium of the People," 191.

this discussion, for which I refer the reader to the works cited. For our present purposes, a few more words on the topic of meritocracy will suffice.

While a system that explicitly acknowledges the role of luck and arbitrariness in social competition might mitigate the frustration arising from the sense of failure – as Sandel proposes to do at Harvard University, introducing a lottery as part of the admission process[306] – common retributive thinking encourages the victimisation of losers: we are back to Dumouchel's and Scheper-Hughes' sacrificial reading of the losers in the system. As Dumouchel remarks:

> Such people are not simply abandoned to their fates; they are also the objects, in discourse and in reality, of a form of exclusion that reveals that they also have the value of substitute victims. […] Since evil never occurs without reason, and no one has wronged them, it seems more probable that the poverty they suffer comes either from a moral failing, or from a form of incompetency […][307]. They are victims only of themselves, of their incompetency or their vices, and no one has any special responsibility with regard to them.
>
> This twofold discourse, which expresses the guilt of the victims and the absence of everyone else's responsibility, transforms the misfortune that strikes them into justified punishment. It reflects, reveals, though not explicitly, their role as substitute victims.[308]

Meritocracy, perfect or imperfect as it may be, generates victims we tend to blame for their situation. As I said before, I tend to agree with Dumouchel that a severe victimising factor exists in the thinking and acting towards "losers of the system", but after this digression on meritocracy, I would like to draw some distinctions on the three phenomena here discussed under the category of "peace-time crimes": the blaming of the poor, the disinterest for the fate of the elderly, and the punishment of offenders.

2.1.3 Violence by Omission, Violence by Execution

The blaming of the poor perfectly fits into Dumouchel's category of "violence by omission": the poor are not only isolated, left behind, their condition of suffering and deprivation accepted, but they are also the target of the social discourse on personal

[306] Sandel, *The Tyranny of Merit*, 185–86.

[307] The passage continues "that [the incompetency] makes them unfit for the market system, rather than from them being excluded." I omitted the end of the sentence, which I quote here for completeness, because it makes one believe that unfair exclusion is the only reason for the loss in market competition, as if those incompetencies were not real but only attributed *after* and *because of* the exclusion. While sometimes these forms of exclusion do happen, I believe we should also recognise that forms of incompetency, or unfitness to the market logic, exist. The point is that they should not lead to blame and victimisation.

[308] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 138.

responsibility and deserved fate. It is the result of the *just-world fallacy*, the cognitive bias that was formalised as a result of Lerner's seminal work that makes people blame those in a disadvantaged position. Mentioned in the previous chapter while speaking of otherworldly justice, it is interesting to notice that this bias affects lay and atheist people, too, though a strong correlation with religious beliefs has also been found[309]. As psychologists Leo Montada and Melvin Lerner explain it: "When restoration of injustice is costly", as indeed is the case for systemic economic injustice, "people tend to deny injustice by blaming the victims or by minimizing their hardships and disadvantages. In this manner, BJW[Belief in a Just World]-based motivation merges with people's self-interest."[310]

This mechanism is also manifest in the conception of the "natural state", i.e., the *status quo*, as morally prescriptive. Aristotle was probably the first philosopher to consider nature as intrinsically normative, particularly in his analysis of power relations, which he considered to be morally good insofar as they were "natural" – think of his (in)famous description of the "slave by nature" or of the subordination of women, and in general of the superiority of the Greek man, deduced from the observation of factual power relations. Similarly, when "losers" emerge in a liberal system "naturally", that is, as a consequence of the market mechanism net of State influence, their subordination, or in general their suffering, is justified because they are part of a "natural", and therefore intrinsically good, state of affairs.

Concerning the situation of the elderly in retirement homes, they too are treated as system waste and are the victims of the dissolution of solidarity bonds; however, it is clear that they could not bear any responsibility for their situation – and that most will end up in a similar condition eventually. Just-world biases do not seem to intervene here. The problem lies in the staff's indifference towards the patients – generated, as

[309] See for example: James Benjamin Schuurmans-Stekhoven, "Just World Beliefs Mediate the Well-Being Effects of Spiritual/Afterlife Beliefs among Older Australians," *Journal of Religion, Spirituality & Aging* 33, no. 3 (2020): 332–49.

[310] Leo Montada and Melvin J. Lerner, "Preface," in *Responses to Victimizations and Belief in a Just World*, ed. Leo Montada and Melvin J. Lerner (New York: Springer Science+Business Media, 1998), vii. A correlation between the blaming of the poor and the belief in a just world has been recently demonstrated by Efraín Garcia-Sánchez and colleagues: Efraín García-Sánchez et al., "How Fair Is Economic Inequality? Belief in a Just World and the Legitimation of Economic Disparities in 27 European Countries," *Personality and Social Psychology Bulletin* 48, no. 3 (March 1, 2022): 382–95. Similar results have also been found with respect to global poverty and inequality between countries, see for example: David J. Harper et al., "Lay Causal Perceptions of Third World Poverty and the Just World Theory," *Social Behavior and Personality* 18, no. 2 (1990): 235–38. The case of Brazilian street children analysed by Scheper-Hughes could be considered akin to that of poverty in general discussed thus far.

Scheper-Hughes remarks, by "[e]conomic pressures […] [that] bear down on staff to minimize the personal care and attention given to the residents"[311] – as well as in the indifference of the families of the elderly that, in their turn, are often overwhelmed by their own daily lives. Hence, the decision not to further investigate the situation of old parents or grandparents, however tragic it may be, is the result of the perceived absence of viable alternatives and a way of coping with the resulting sense of guilt. Dumouchel cites the case at the beginning of his book, referring to Scheper-Hughes' reworking of Basaglia's category of peace-time crimes, and does not bring it up again when he elaborates on his own conception of violence by omission. Although the elderly are the victims of widespread indifference and the suffering they endure is a consequence of omission, the lack of the victimising trait subtracts the case from Dumouchel's category of violence by omission.

The situation of offenders punished by the State seems to me utterly different. If, on the one hand, they can be compared to the poor and the elderly in that they, too, are left behind by the dissolution of solidarity bonds – in the vindicatory system, in fact, they are defended by their group from retaliation and are ritually reintegrated if their action affects the inner circle they belong to – on the other hand their social treatment, as a result of their moral status, is incomparable. The violence they are subjected to is not violence by omission; it is intentionally inflicted upon them: it is violence by execution. This is true not only for the officers and the magistrates involved in criminal enforcement and for those who condemned the offender and decided their sentence but for the whole of civil society. Contrary to the elderly, towards whom we close our eyes, and to the poor, whom we blame as a consequence of their disadvantaged position because of the just-world fallacy, offenders are the direct target of our blame, anger, and violence. Their suffering is created by the deliberate infliction of their punishment; it is not pre-existent and subsequently justified through blaming mechanisms.

Moreover, concerning the neglect of the elderly, everyone would agree that it is morally wrong. Concerning the blaming of the poor, luck is a major topic in distributive justice thinking, as testified by the current of luck egalitarianism among political philosophers, which can be traced back to Rawls' idea of the "natural lottery"[312] and,

[311] Scheper-Hughes, "Coming to Our Senses: Anthropology and Genocide," 373.

[312] "Within the limits allowed by the background arrangements, distributive shares are decided by the outcome of the natural lottery; and this outcome is arbitrary from a moral perspective." Rawls, *A Theory of Justice*, 64.

later, to Ronald Dworkin's notion of "equality of resources"[313]. The relevance of luck in matters of economic distribution is also widely accepted by common thinking, especially in the light of the acknowledgement that our meritocratic systems are imperfect and, in the absence of equality of opportunity, individual achievements depend only partly on the effort or ability of the individual. If unconsciously we may still tend to blame the poor, more careful consideration of the importance of luck in wealth distribution brings many – at least momentarily – to agree on the injustice of their suffering (the solution for which, though, is hard to conceive for the ordinary citizen, having to do with the hyper-complex global economic system, hence justifying the just-world fallacy).

On the contrary, concerning punishment, the infliction of suffering is not only not considered an injustice but is positively regarded as a fundamental constituent of justice itself. As we have seen, the victimisation of victimisers is considered a rational function of the rational State. Contrary to the other "losers of the system" analysed here, our moral system allows for and justifies their persecution. Their victimary sign[314] is their guilt. Whereas luck is accepted in matters of distributive justice, where the poor – albeit with all the inconsistency we have seen – are often regarded as victims of unfortunate circumstances, the same cannot be said for retributive justice.

The concept of moral luck was introduced by Bernard Williams in 1976[315] and has been famously developed by Thomas Nagel[316], especially in what concerns "constitutive luck"[317], mentioned by Williams but not further elaborated on in his original paper. Constitutive luck undermines the concepts of guilt and individual moral responsibility in general: "Justice requires not merely that *something I am* should be beyond luck, but that *what I most fundamentally am* should be so"[318]. If one comes to

[313] Dworkin has rejected the label of "luck egalitarian" (Ronald Dworkin, "Equality, Luck and Hierarchy," *Philosophy and Public Affairs* 31, no. 2 (2003): 190.), however, for his role in the conceptualisation of luck egalitarianism, see: Carl Knight, *Luck Egalitatianism. Equality, Responsibility, and Justice* (Edimburgh: Edimburgh University Press, 2009), 13–43.

[314] *Signe victimaire*, Yvonne Freccero translates it as "sign of a victim": Girard, *The Scapegoat*, 18.

[315] Bernard Williams, "Moral Luck", *Proceedings of the Aristotelian Society*, supplementary volume L 1976): 115–35, revised and reprinted in Bernard Williams, "Moral Luck," in *Moral Luck. Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981), 20–39.

[316] Thomas Nagel, "Moral Luck," in *Free Will*, ed. Gary Watson (Oxford: Oxford Readings in Philosophy, 1983), 174–86.

[317] "[…] the phenomenon of constitutive luck – the kind of person you are, where this is not just a question of what you deliberately do, but of your inclinations, capacities, and temperament." Nagel, 177.

[318] Williams, "Moral Luck," 38. Hobbes, in the *Elements*, had mentioned a similar kind of "fortune" that is necessary for obtaining and developing mind and body faculties. However, in one of his efforts to make his theory compatible with religion, he claims that this "good fortune […] is really nothing else but

believe that what one most fundamentally is – one's character, capacities, desires – does not depend on one's self, the idea of desert is shaken and retributive justice loses its grounds. If one's ability to follow or even understand ethical principles is essentially a matter of luck – one being unlucky in feeling enjoyment in the face of violence, in having been raised in a cruel environment, even in having a specific brain conformation – what does it mean for someone to *deserve* their punishment? From this point of view, the criminal is a social burden, that is, someone who needs special assistance from the welfare state, comparable to the poor, the elderly or any other vulnerable member of the society.

Classifying violent offenders as vulnerable people who need protection can appear surprising at best, if not outrageous. The way we think we should treat offenders is opposite to the way we think we should treat the elderly or the poor, even according to the most progressive people: we aim at punishing the formers and protecting – up to a certain extent – the latter. It seems wrong to equate offenders and victims. However, another comparison with vulnerable members of society might make my point more straightforward, although maybe not less outrageous on the surface. I ask the reader to bear with me for a moment and wait for the end of the reasoning before drawing their conclusions.

Offenders hurt others, which apparently gives them a different moral status that forbids associating them with victims. However, there is another category of people that occasionally proves dangerous for its fellow citizens: the mentally ill. This comparison might seem outrageous because much work has been needed to decriminalise mental illness; however, the fact that "the fool" and "the criminal" have been long connected is not accidental, and the fact that this association has been (rightly) overcome should not prevent us from investigating the connection that has long been standing. In both cases, we face a non-standard mode of social interaction. Society is kept together by a set of explicit and implicit rules that an individual internalises during their development

---

the favour of God Almighty." Thomas Hobbes, "The Elements of Law," in *Three-Text Edition of Thomas Hobbes's Political Theory. The Elements of Law, De Cive and Leviathan*, ed. Deborah Baumgold (Cambridge: Cambridge University Press, 2017), 66.

By so doing he disrupts the concept of fortune, which is by definition entirely disconnected from that of desert and is united to that of randomness, of chance. If we equate good fortune with God's favour – and therefore misfortune with God's disfavour – we are, even loosely, reintroducing desert in the picture. Even if not entirely speaking of Providence, it is evident that God acts within a scheme of reasons, although unknowable for us. Therefore, our good or bad fortune may be connected to what we have done, hence respecting our deserts.

process – what Freud has called "the super-ego" that corresponds to the "moral conscience" in everyday talk. Tragically, in many cases, mental illness disrupts this internalisation process, and one finds that one cannot make reasonable expectations about the behaviour of someone suffering from certain kinds of mental illness (empathy disorders, for example[319]). The American Psychiatric Association defines a mental disorder as "a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underling mental functioning"[320]. Given such conditions, it is not surprising that people suffering from it often inadvertently unhinge minor habits of social interaction, and this generates the reasonable expectation that other, more important, social rules – including laws – might be broken as well, as a consequence of an unpredictable reasoning process. Regrettably, statistics on the mental health of the incarcerated population show that this happens repeatedly[321].

[319] Simon Baron-Cohen, *The Science of Evil. On Empathy and the Origins of Cruelty* (New York: Basic Books, 2011).

[320] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. (Washington, DC, 2013), 20.

[321] For example, whereas in the general population three percent of males and one percent of females suffer from antisocial disorders, half of the male prison population and a quarter of the female prison population suffer from it. Baron-Cohen, *The Science of Evil. On Empathy and the Origins of Cruelty*, 67; Seena Fazel and John Danesh, "Serious Mental Disorder in 23 000 Prisoners: A Systematic Review of 62 Surveys," *The Lancet* 359 (2002): 345–50. Whereas the case of antisocial disorder might be extreme, other studies confirm that higher rates of mental health disorders are found in prison population compared to general population, even when studies only take into account recently admitted prison population to control for prison-generated mental disorders. A 2020 study, for example, found that the recently admitted prison population in Northern France is about three times more likely to suffer from mental disorders than the general population (drug addiction excluded, which is 7.8 more likely in the prison than in the general population). Thomas Fovet et al., "Mental Disorders on Admission to Jail : A Study of Prevalence and a Comparison with a Community Sample in the North of France," *European Psychiatry* 63, no. 1 (2020): 1–7. A 2007 study on recently admitted prison population in Connecticut's prisons confirmed the trend: "More than two of three inmates met the criteria for at least one lifetime psychiatric disorder, almost half for an anxiety disorder, and more than one-third for an affective disorder, substantially higher levels than those reported in community prevalence studies. These rates are comparable both to levels reported in clinical samples and to those reported in a recent BJS study. Not surprisingly, levels of lifetime PTSD and antisocial personality disorder (ASPD) were markedly higher than in community studies. The levels of current affective and anxiety disorders and PTSD also were 5 to 10 times higher than those reported (based on past year prevalence) in community studies." Robert L. Trestman et al., "Current and Lifetime Psychiatric Illness Among Inmates Not Identified as Acutely Mentally Ill at Intake in Connecticut's Jails," *The Journal of the American Academy of Psychiatry and the Law* 35, no. 4 (2007): 496. In the US prison population, roughly 37% has been told by a mental health professional they had a mental disorder. The percentage remains unchanged if one considers only prisoners admitted since less than one month, thus, attributing the diagnosis to their pre-incarceration period and therefore controlling for prison-generated illness. Jennifer Bronson and Marcus Berzofsky, "Indicators of Mental Health Problems Reported by Prisoners and Jail Inmates, 2011-12," *Us Department of Justice, Office of Justice Programs, Bureau of Justice Statistics*, 2017, 6.

On the other hand, the fact of committing an offence is also in itself an indicator that something in the process of internalisation of the rules that implicitly govern social interaction has gone wrong. I do not think that offenders differ from people who do not commit crimes in that they perform utility calculations differently, as the deterrent conception of punishment would have us believe. Advocates of the justification of punishment based on deterrence, as we should investigate further in a following section, maintain that the reason why we need institutionalised forms of punishment is that people need disincentives in order to avoid committing crimes as if one would normally ponder whether to kill or harm another who hinders them in their aims and then refrain from that because of the risk of being caught and sent to prison. But if I were to tell someone that I was considering poisoning my competitor for a post-doc position but that I then considered it too risky for myself as there were good chances of me being caught and punished so that I probably would not do it in the end, my interlocutor would think that there is something wrong, and maybe that I should be reported to the police or to sanitary authorities before I harm someone. The thing is, that is just not how standard[322] interaction works. The deterrent conception of punishment provides a misleading understanding of the process that leads to criminal actions[323].

This last remark should deepen the understanding of the connection between the mentally ill and the offender. It is no coincidence that social expectations about reasonable behaviour are disrupted in both cases: it is the result of an equally disrupted mental process that leads to action. This finally brings us back to constitutive moral luck. I argue that to be the kind of person who seriously considers the idea of killing or hurting another to achieve their goals is a severe form of constitutive bad luck. It is a sign that something has gone terribly wrong in the process of internalisation of social rules, not differently – if not in degree – from what generated an antisocial personality disorder or psychopathy[324].

The (much needed) decriminalisation of mental disorders brought to the abolition of asylums in favour of alternative institutions that care not only about

[322] I will not delve into the discussion around what can be considered "normal" or whether there is some kind of intrinsic normativity in the way the majority behaves. I simply say that social interaction is based on the reasonable predictions one can make about the actions of others on the basis of implicit norms of behaviour. If these predictions, which are based on the behaviour of the majority, are disappointed, a social interaction problem occurs.

[323] Of course, as we should later see, this does not mean that punishment has zero deterrent power.

[324] Baron-Cohen, *The Science of Evil. On Empathy and the Origins of Cruelty*, 64–84.

protecting society and the institution itself from the sick person[325] but, first and foremost, about improving the person's living conditions[326]. Equating the mentally ill and the offender through the concept of constitutive luck is not meant to bring the criminalisation of illness back; instead, it aims at helping liberate offenders from the social stigma – one that is much more ingrained and socially accepted than any other in contemporary society – and support the prison abolition movement in favour of alternative institutions that foreground the needs and shortcomings of the people they host. Once again, it is no coincidence that the prison abolition movement gained momentum together with the anti-psychiatry movement of the 1960s and 70s as part of an overall social tendency to deinstitutionalisation[327], although only the latter succeeded.

Once we have mined the concept of desert through luck, the association between the offender, the mentally ill, the abandoned elderly, and the poor should be transparent. They are all cases of victims, in one way or another, of circumstances that put them in a disadvantaged position for social interaction, that make them vulnerable to the indifference generated by the erosion of social bonds. Offenders, however, are the victims of that specific kind of violence that I have called violence by execution: the violence one exercises towards them, contrary to other cases analysed, is fundamentally shameless. It is probably the only case in which one can unabashedly publicly state that another should be put to death, put in isolation for the rest of their life, castrated, even raped and battered by fellow inmates after incarceration. Not only that, but all this violence is covered with the moral value of justice. It is not simply accepted because

[325] Sociologist Erving Goffman, one of the central figures of the anti-psychiatry movement in the 1960s, wrote on the purpose of asylums: "part of the official mandate of the public mental hospital is to protect the community from the danger and nuisance of certain kinds of misconduct. In terms of the law and of the public pressures to which the mental hospital is sensitive, this custodial function is of major importance. Within the institution surprisingly little explicit reference is made to it, however, the focus being on the medical-like therapeutic services the hospital supplies patients." Erving Goffman, *Asylums. Essays on the Social Situation of Mental Patients and Other Inmates* (Garden City: Anchor Books, 1961), 352.

[326] One need not idealise this process either, for, as psychiatrist E. Fuller Torrey pointed out, "Jails and prisons have increasingly become surrogate mental hospitals for many people with serious mental illness." Edwin Fuller Torrey, *Nowhere to Go: The Tragic Odyssey of the Homeless Mentally Ill* (New York: Harper & Row, 1988), 42. Such a consideration might seem to cast a shadow on my argument. However, I should have made abundantly clear that my argument is one for liberation for all, not of repression.

[327] For a historical account of the intertwining of these movements: Anne E. Parson, *From Asylum to Prison. Deinstitutionalization and the Rise of Mass Incarceration after 1945* (Chapel Hill: The University of North Carolina Press, 2018), 69–97.

anger is understandable in the face of tragedy, nor is it just legitimate; it is *morally right*.

Williams ended his paper with the hope that such considerations on moral luck would bring about a change in moral conceptions:

> Scepticism about the freedom of morality from luck cannot leave the concept of morality where it was, any more than it can remain undisturbed by scepticism about the very closely related image we have of there being a moral order, within which our actions have a significance which may not be accorded to them by mere social recognition.[328]

However, retributive justice thinking has not proved as permeable to the concept of luck as distributive justice thinking. Given all this, offenders make the perfect sacrificial victims. The need for surrogate victims adds to the need to vent retributive emotions and generates a very solid institution of public punishment, resulting from the intertwining of rational justifications, retributive emotions, and sacrificial outlets for social violence. This latter element has been treated at length; I shall now turn to the first two.

2.2 Retributive Emotions: A Moral Delusion

2.2.1 Moral Intuitionism

Emotions play a crucial role in our day-to-day moral experience, not only in the sense that we feel outraged, sad, or resentful when a moral principle we care about is violated and happy, grateful, or inspired when it is defended and supported, but that we implicitly take them to be our moral guides in many situations. We do hold pre-reflexive moral standards, some of which we may be able to rationally justify *ex-post*, while others might struggle to pass the rational analysis test. The sympathy we *feel* towards a defenceless baby tells us that we *should* (normatively) protect and care for them; the revulsion we *feel* towards the suffering of an innocent tells us, at the very least, that we *should* not inflict gratuitous pain on them, and maybe even that we *should* intervene were we to witness someone abusing them; the gratitude we *feel* towards someone who has done us good feels *deserved*, meaning that we pre-reflexively believe that the one who has done us good *should* be treated in a special positive way – as a minimum, they should be thanked as a reward for their action. If someone were to react

[328] Williams, "Moral Luck," 39.

to an act of kindness with complete indifference, let alone disdain, we would consider it *wrong*. Similarly, but undoubtedly more strongly, if someone hurts us, we *feel* outraged in a way that makes us *feel* entitled to react with a similar offence, even just a rude response if our interlocutor has been disrespectful towards us.

In the situations described, we *feel* that those are the appropriate reactions, we do not *think* so. Most metaethical theories should easily and convincingly justify the moral beliefs that emerge from the feelings felt in the first two situations: all other things being equal, we should do good to the helpless, and we should not harm the innocent. It is accepted by deontologists, who take these to be basic rules of morality; by consequentialists, for, given the premise "all other things being equal", following these feelings produces good consequences (whether in terms of general utility for utilitarians or in terms of other criteria for other types of consequentialism); by virtue ethicists, who can take these feelings to be good manifestations of a person's virtue.

As for the latter two situations, things get more complicated. These feelings prescribe a retributive conception of justice, reminiscent of Polemarch's definition at the beginning of Plato's *Republic*: "justice is to benefit one's friends and harm one's enemies"[329]. It is, in fact, the conventional view of justice that Socrates is set to refute in his enquiry concerning the good arrangement of the city. While we would not express it nowadays in terms of "friends" and "enemies", the meaning of Polemarch's definition coincides with the principle that can be induced from the description of the two situations above: justice is to distribute benefits and harms according to desert, that is, to benefit those who create value and harm those who create disvalue. As Socrates in *The Republic* shows, things get more complicated once we try to justify these principles rationally. While such a view can be justified in deontological and virtue ethics terms, a consequentialist justification is not so straightforward. The difference, I believe, lies in the different role that moral intuition, and thus, as we shall see, the emotions, play in these normative ethical theories.

Moral intuitionism is an epistemological approach to ethics according to which "normal human beings have an immediate awareness of moral values"[330]. This epistemological approach also has ontological implications: it implies, in fact, that there are such things as moral values that exist in the world, independently from the

[329] Plato, *The Republic*, 1:332 d.
[330] W. D. Hudson, *Ethical Intuitionism* (London, Melbourne, Toronto: MacMillan, 1967), 1.

perceiving subject; it is a realist theory. Not only that, but it also implies that such moral real facts are of a metaphysical kind, in need of their own faculty in order to be perceived. Intuitionists believe moral knowledge can be objectively reached thanks to the so-called faculty of intuition that reveals moral truth to the human mind in the form of clear and simple ideas that cannot be further analysed in other terms. As Peter Strawson put it, it is the only "analysis of moral judgement *in terms of feeling* ever been suggested which was not monstrously paradoxical or artificial[...]."[331]

Intuition can be defined as an "initial intellectual appearance"[332] and has been conceived in both emotional and intellectual senses. The moral sense theorists to whom the idea of intuitionism may first be ascribed, like Hutcheson and Shaftesbury, believed it to be analogous to perception, whence the idea of a sort of moral blindness of those who do not "see" true moral principles[333]. The rationalists (like Price[334] and Clarke[335]), on the other hand, equated moral intuition to the cartesian clear and distinct ideas, making reason, or the intellect, the faculty that grasps them, and equating intellectual stupidity with the moral one. In both cases, there is moral knowledge that can be accessed by some, whereas others, blind or stupid, are left outside the realm of moral truth[336]. Obviously, in both cases, the problem is how to distinguish moral truth from

[331] Peter Strawson, "Ethical Intuitionism," *Philosophy* 24, no. 88 (1949): 31, my emphasis.

[332] Michael Huemer, "Revisionary Intuitionism," *Social Philosophy and Policy* 25, no. 1 (January 2008): 370.

[333] Shaftesubury maintains that humans share a "Sense of Right and Wrong", that is "a Sentiment or Judgment of what is done, thro' just, equal, and good Affection, or the contrary" (Anthony Third Earl of Shaftesbury, *Characteristicks of Men, Manners, Opinions, Times. Vol. II. An Inquiry Concerning Virtue and Merit. The Moralists; a Philosohical Rhapsody*, ed. Douglas Den Uyl (Indianapolis: Liberty Fund, 2001), 18.) "a Liking or Dislike of moral Actions" (Shaftesbury, 30.).

According to Hutcheson, through the moral sense "we perceive *Virtue*, or *Vice*, in our selves, or others." Francis Hutcheson, *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*, ed. Aaron Garrett (Indianapolis: Liberty Fund, 2002), 17.

[334] "The [moral] principles themselves, it should be remembered, are self-evident; and to conclude the contrary […] because of the obscurity attending several cases wherein a competition arises between the several principles of morality, is very unreasonable." Richard Price, *A Review of the Principal Questions in Morals*, ed. D. D. Raphael (Oxford: Oxford University Press, 1974), 168.

[335] "There is a fitness or suitableness of certain circumstances to certain persons and an unsuitableness of others founded in the nature of things and in the qualifications of persons, antecedent to will and to all arbitrary or positive appointment whatsoever, must unavoidably be acknowledged by everyone who will not affirm that it is equally fit and suitable in the nature and reason of things that an innocent being should be extremely and eternally miserable as that it should be free from such misery. There is, therefore, such a thing as fitness and unfitness, eternally, necessarily, and unchangeably in the nature and reason of things." Samuel Clarke, "A Demonstration of the Being and Attributes of God," in *A Demonstration of the Being and Attributes of God and Other Writings*, ed. Ezio Vailati (Cambridge: Cambridge University Press, 2010), 83.

[336] Hudson, *Ethical Intuitionism*, 57.

moral illusion and, given the high variability in moral beliefs, to decide *who* is the repository of moral knowledge.

What role are the so-called "moral emotions" or "moral sentiments" to play in this process? On the one hand, an intuitionist theory such as W. D. Ross's is based on the self-evident character of what he calls the *prima facie* duties that one just *knows* to be moral duties, like the duty to keep a promise and, more in general, "[t]he main moral convictions of the plain man", that are to him "not opinions which is for philosophy to prove or disprove, but knowledge from the start"[337]. These intuitions, which are thus purely intellectual and are not further explained by Ross[338], are accompanied by a "feeling of approval or interest for what is good"[339], but such feelings are only derivative of the judgement of goodness, which is primal.

On the other hand, according to Michael Moore, who is probably referring to sentimental intuitionism, "[a]n intuitionist […] believes that the emotions stand to moral judgement in a relationship exactly analogous to the relation between perceptual experience and scientific judgement"[340], whereas for Peter Strawson "[e]motion may be the gatekeeper to the moral world; but intuition is the gate"[341]. In both conceptions, emotion plays a crucial epistemic value in morals, providing a fundamental guide to

[337] David Ross, *The Right and the Good* (New York: Oxford University Press, 1930), 21. Among these convictions, we find the right to punish wrongdoers, however, instead of simply justifying it on the basis of its self-evidence as many intuitionists would do, Ross does question it, and only validates it once coupled with the duty to produce as much good as one can (Ross, 58.) and with the duty to keep promises: the promise the State has made to the victim, their relatives, and society on the whole that offenders would be punished (Ross, 63–64.): "Most intuitionists would perhaps take the view that there is a fundamental and underivative duty to reward the virtuous and to punish the vicious. I am inclined to diverge from this view. Two things seem to me to be clear: that we have *a prima facie* duty to do this, and that a state of affairs in which the good are happy and the bad unhappy is better than one in which the good are unhappy and the bad happy." Ross, 57–58. "Thus to our sense that *prima facie* the state has a right to punish the guilty, over and above the right which it has, in the last resort, of inflicting injury on any of its members when the public interest sufficiently demands it, there is added the sense that promises should *prima facie* be kept; and it is the combination of these considerations that accounts for the moral satisfaction that is felt by the community when the guilty are punished, and the moral indignation that is felt when the guilty are not punished, and still more when the innocent are. There may be cases *in* which the *prima facie* duty of punishing the guilty, and even that of not punishing the innocent, may have to give way to that of promoting the public interest. But these are not cases of a wider expediency overriding a narrower, but *of one prima facie* duty being more obligatory than two others different in kind from it and from one another." Ross, 64.

[338] "Many readers will perhaps say that they do *not* know this to be true. If so, I certainly cannot prove it to them; I can only ask them to reflect again, in the hope that they will ultimately agree that they also know it to be true." Ross, *The Right and the Good*, 21.

[339] Ross, 91.

[340] Michael S. Moore, "The Moral Worth of Retribution," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987), 200.

[341] Strawson, "Ethical Intuitionism," 24.

moral judgement and action. This is why some emotions are also called "moral sentiments", and many philosophers have believed them to be essential to a healthy moral life. R. Jay Wallace identified these emotions with "resentment, indignation, and guilt"[342], picking up Peter Strawson's reactive attitudes[343], attitudes Strawson considers so fundamental that, being incompatible with determinism, they prove its falsity. Even John Stuart Mill spoke of the "feeling of justice"[344] and the "sentiment of justice"[345], referring to "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"[346] – something very similar to the moral emotions and reactive attitudes identified by Wallace and Strawson.

I will follow Martha Nussbaum in comprising resentment, indignation, and in general, all emotions that involve both the perception that something wrong has been done and the wish for payback under the more general category of anger.[347] Guilt, although more indirectly, can also be included under the category of anger: it is, in fact, anger directed towards one's self, composed of the belief that one has done something wrong and that it is just for them to suffer for it, that the self-inflicted suffering from guilt is deserved.[348] As Wallace pointed out, anger, and reactive attitudes in general, are generally taken to demonstrate "our commitment to certain moral standards, as regulative of social life"[349], and, according to Moore, "[l]ack of anger at criminals, if it does not represent simple indifference to the suffering of others, may represent our self-deception about the potential for evil in humanity"[350], as a symptom of an effort to deny the offender's responsibility in order to refuse the possibility of a will to harm.

Moore then turns to the epistemic value of emotions. He distinguishes virtuous and non-virtuous emotions, finding guilt in the first category and resentment,

[342] R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge/London: Harvard University Press, 1996), 2.

[343] Peter Strawson, "Freedom and Resentment," in *Freedom and Resentment and Other Essays* (London: Routledge, 2008), 1–28.

[344] John Stuart Mill, "Utilitarianism," in *Utilitarianism and On Liberty: Including "Essay on Bentham" and Selections from the Writings of Jeremy Bentham and John Austin*, ed. Mary Warnock (Hoboken, New Jersey: Blackwell Publishing, 2003), 216.

[345] Mill, 224.

[346] Mill, 224.

[347] Nussbaum, *Anger and Forgiveness*, 262.

[348] Nussbaum, 130.

[349] Wallace, *Responsibility and the Moral Sentiments*, 69.

[350] Moore, "The Moral Worth of Retribution," 211.

indifference to others, self-deception, cowardice and pity[351] in the latter[352]. He then moves on to investigate the judgements that such emotions generate: concerning the non-virtuous emotions, resentment generates retributive judgements, whereas the other emotions listed generate non-retributive judgements. Given the non-virtuous nature of these emotions, none should be given epistemic value in matters of morality, and they should be discarded as moral guides. Guilt, on the other hand, is, according to Moore, a trustworthy moral guide: "[t]he alternative, of not crying over spilt milk (or blood), is truly indecent. A moral being *feels* guilty when he or she *is* guilty of past wrongs."[353] Even though we might excuse others, sympathise with them and withhold anger towards them, were we to commit the crime, we would never excuse ourselves in the same way, and we would deem it appropriate to feel that form of self-anger that is guilt.

> What would you feel like if it was you who had intentionally smashed open the skull of a 23-year-old woman with a claw hammer while she was asleep, a woman whose fatal defect was a desire to free herself from your too clinging embrace? My own response, I hope, would be that I would feel guilt unto death. I couldn't imagine any suffering that could be imposed upon me that would be unfair because it exceeded what I deserved.
>
> Is that virtuous? Such deep feelings of guilt seem to me the only tolerable response of a moral being.[354]

Here is a perfect example of sentimental intuitionism: Moore, with a somewhat populist move, describes a heinous crime, one that would typically generate anger and calls for retribution against the offender, and then asks the reader to direct such emotions towards themselves. Having given epistemic credentials in moral matters to guilt as a virtuous emotion, he concludes that retribution is the only morally acceptable reaction to wrongdoing. Surprisingly enough, while Moore had listed multiple non-virtuous emotions, he had been restrained in listing virtuous emotions: in fact, we only find guilt – in spite of "an entire tradition that regards [it] as a useless passion"[355] – in

[351] "In pity we do not care about the other any more for his own sake, but only insofar as he allows *us* to become, in our own and others' eyes, better." Moore, 212. It is not clear to me from where he derives such a negative characterisation of pity.

[352] Moore, 212.

[353] Moore, 214.

[354] Moore, 213. Herbert Morris' reasoning works in a similar way: he reaches the paradoxical conclusion that punishment is an offender's right, since it respects the self-determination of their choice. It is implied that, intuitively, someone who has committed a crime would not want to be treated like a dangerous animal or a mentally ill person to be re-educated, for this would be perceived as an attack on their autonomy and, hence, their worth as an independent human being. Punishment, therefore, seems "the only tolerable response [to] a moral being". Herbert Morris, "Persons and Punishment," *The Monist* 52, no. 4 (1986): 475–501.

[355] Moore, "The Moral Worth of Retribution," 214.

his writing. Although he is counting pity as non-virtuous, should compassion not be considered virtuous instead?

Moreover, serious guilt might be felt even in situations that would not justify such a harsh response. Imagine, for example, crushing someone's hand while closing a heavy door, resulting in this person losing two fingers. It was not an act of carelessness; from your perspective, you could not see that someone's hand was on the door's way; it was just an unfortunate event. Still, if you were just to walk away saying that that person's suffering was none of your business, you would probably be blamed, and we would deem it appropriate for you to feel guilty, even if you did not intend to cause harm, nor had you failed at behaving sufficiently carefully in that situation. Intuitively, we would expect you to stop, offer help, call an ambulance, and maybe get in touch to see how things are going in the following days. In a word, an appropriate response would be to suffer to a certain degree for the harm you unintentionally caused, therefore implying a retributive attitude. However, an explicitly retributive stance that asked that you should be made to suffer for the harm caused would hopefully be deemed inappropriate. This example shows that even so-called virtuous emotions can prove unreliable guides to moral judgement and action. This is because emotions' function is far from moral.

2.2.2 The Mirage of Moral Emotions: The Case of Justice

Darwin initiated the study of emotion from a functional viewpoint in his 1872 pioneering book *The Expression of the Emotions in Man and Animals*, originally conceived as a chapter of *The Descent of Man, and Selection in Relation to Sex* (1871)[356]. However, Darwin focused only on the physical expression of emotions, and particularly on facial movements and noticeable physiological changes, famously, for example, on blushing[357]. The scientific study of emotions remained quiet for almost 100 years[358] and was taken up by Paul Ekman and Carroll Izard in the 1960s when their mentor Sylvan Tompkins asked them to search for proof for Darwin's theory of the

[356] Paul Ekman, "Darwin's Contributions to Our Understanding of Emotional Expressions," *Philosophical Transactions of the Royal Society* 364 (2009): 3449.

[357] Charles Darwin, *The Expression of the Emotions in Man and Animals*, ed. Francis Darwin (Cambridge: Cambridge University Press, 2009), chap. XIII. Self-attention-Shame-Shyness-Modesty: Blushing: 328-367.

[358] Michael Hammond, "Evolutionary Theory and Emotions," in *Handbook of the Sociology of Emotions*, ed. Jan E. Stets and Jonathan H. Turner (New York: Springer, 2006), 368.

evolution of emotion[359]. Darwin had based his theory on his own subtle observations of his family's children and pets, of the orangutans at the London Zoo, of hundreds of pictures, as well as on his memories of his encounters with people from different cultures during his travel on the *Beagle*, on his interviews on emotion recognition, on physicians' descriptions of psychiatric patients, and on a questionnaire that he had circulated among travellers directed towards indigenous cultures[360]. Ekman and Izard focussed only on human emotions and demonstrated a high rate of intercultural emotion recognition even among the most diverse cultures. Since then, many have moved on to expand research to non-human animal emotion, most famously the ethologist Frans De Waal[361], substantially confirming Darwin's observations and intuitions. As Ekman noticed, although the methods and samples analysed by Darwin would be considered insufficient in nowadays' scientific research, it was thanks to his evolutionary perspective that he reached so many accurate conclusions[362].

Of course, expression is only one of the elements that constitute an emotion, and research has broadened to include other aspects such as other non-visible, or in general non-communicative, physiological changes (e.g., muscular tension, vessel dilatation or restriction, augmented sensory perceptions, activation of certain regions of the brain, etc.) and reported or inferred psychological changes (e.g., feeling, action readiness, motivation, etc.) reaching what is now known as *basic emotion theory*[363], based on Ekman's and Robert Plutchik's pioneering studies, published in the 1970s and 1980s but not yet superseded and still commonly cited. The key idea of the basic emotion theory is that since emotions' physiological expression has been observed and compared in several species, and since emotions' psychological effects have been reported analogously in humans of virtually all cultures and inferred analogously in several species as well, emotions must serve an adaptive function not limited to the communicative function initially analysed by Darwin. It is now noncontroversial among emotion scholars that, in order to understand an emotion, one needs to understand such

[359] Jessica L. Tracy, "An Evolutionary Approach to Understanding Distinct Emotions," *Emotion Review* 6, no. 4 (2014): 309.

[360] Darwin, *The Expression of the Emotions in Man and Animals*, 16–18.

[361] Frans De Waal, *Mama's Last Hug. Animal Emotions and What They Teach Us about Ourselves* (New York: W. W. Norton & Company, 2019).

[362] Ekman, "Darwin's Contributions to Our Understanding of Emotional Expressions," 3450.

[363] Andrea Scarantino, "The Philosophy of Emotions and Its Impact on Affective Science," in *Handbook of Emotions*, ed. Lisa Feldman Barrett, Michael Lewis, and Jeannette M. Haviland-Jones, Fourth (New York: The Guildford Press, 2016), 20.

function[364] relative to the context in which the specific emotion proved adaptive and was thus selected for. This does not simply refer to the environmental context in which emotions, as well as all other adaptive traits, were selected throughout evolutionary history, long before humans started living in socially and technologically complex societies, but to the specific circumstances in which the aroused emotion would provide an advantage in terms of reproductive fitness (which often, but not always, corresponds to survival).

Plutchik has identified in hierarchy, territoriality, identity, and temporality the four universal problems of adaptation, i.e., problems that any individual would need to tackle in order to survive and reproduce, and associated each to two emotions, that he called basic, thus listing eight fundamental universal emotions, respectively: anger and fear, exploration and surprise, acceptance and rejection, sadness and joy[365]. While the list of universal problems and emotional adaptation varies among scholars[366], the basic idea that specific emotions were selected thanks to their ability to tackle specific problems, a "recurrent evolutionarily significant situation"[367], is unanimous.

From this picture, emotions emerge as "adaptive action patterns"[368], "*mechanisms for motivating or facilitating behavior*"[369], "specialized, domain-specific Darwinian algorithms"[370],

[364] Randolph M. Nesse, "Evolutionary Explanations of Emotions," *Human Nature* 1, no. 3 (1990): 262; Tracy, "An Evolutionary Approach to Understanding Distinct Emotions," 308.

[365] Robert Plutchik, "A General Psychoevolutionary Theory of Emotion," in *Emotion. Theory, Research, and Experience. Volume 1: Theories of Emotion*, ed. Robert Plutchik and Henry Kellerman (New York: Academic Press, 1980), 27–30.

[366] Silvan Tomkins, for example, lists nine basic emotions (Silvan S. Tomkins, "Effect as Amplification: Some Modifications in Theory," in *Emotion. Theory, Research, and Experience. Volume 1: Theories of Emotion*, ed. Robert Plutchik and Henry Kellerman (New York: Academic Press, 1980), 143.), Izard and Buechler ten (Carroll E. Izard and Sandra Buechler, "Aspect of Consciousness and Personality in Terms of Differential Emotions Theory," in *Emotion. Theory, Research, and Experience. Volume 1: Theories of Emotion1*, ed. Robert Plutchik and Henry Kellerman (New York: Academic Press, 1980), 167.).

[367] Tracy, "An Evolutionary Approach to Understanding Distinct Emotions," 308; John Tooby and Leda Cosmides, "The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables," in *Handbook of Emotions*, ed. Michael Lewis, Jeannette M. Habiland-Jones, and Lisa Feldman Barrett, Third (New York: The Guildford Press, 2008), 117.

[368] Michael Lewis, "The Development of Emotional Life," in *After Darwin. Animals, Emotions, and the Mind*, ed. Angelique Richardson (Amsterdam - New York: Rodopi, 2013), 281.

[369] James D. Weinrich, "Toward a Sociobiological Theory of the Emotions," in *Emotion. Theory, Research, and Experience. Volume 1: Theories of Emotion*, ed. Robert Plutchik and Henry Kellerman (New York: Academic Press, 1980), 125.

[370] Leda Cosmides and John Tooby, "Evolutionary Psychology and the Generation of Culture, Part II. Case Study: A Computational Theory of Social Exchange," *Ethology and Sociobiology* 10 (1989): 60.

> specialized modes of operation shaped by natural selection to adjust the physiological, psychological, and behavioral parameters of the organism in ways that increase its capacity and tendency to respond adaptively to the threats and opportunities characteristic of specific kinds of situations.[371]

Let us now go back to the so-called moral emotions and, specifically, to the one that constitutes, as we have seen, the so-called "sentiment of justice", i.e., anger. It appears in every list of the basic emotions, not only because it has been observed in most species and cultures, as well as at early stages in infants, but also because its adaptive role is quite evident: it stimulates aggression in response to attack when other exit strategies are not available, in the popular response scheme of freeze-flight-fight. That of conflict is clearly one of those recurrent evolutionarily relevant situations that have shaped animal behaviour patterns.

The evolutionary role of aggression was the subject of ethologist Konrad Lorenz's 1963 essay *On Aggression*. Richard Dawkins has rightly criticised Lorenz for his insistence on species preservation in the place of individual preservation, which fits better with evolutionary theory[372]. However, Lorenz's work remains notable as an antecedent for more recent studies thanks to its extraordinarily broad perspective. Lorenz's observations range from the behaviour of humans and other mammals to that of fish and birds, depicting an interesting evolutionary pattern, however outdated his theory may now be. Although he does not explicitly talk about emotion, he mentions "life-preserving […] instincts" that elicit aggression[373]. Such "instincts" I interpret to be precisely the emotions we have been talking about, and, in the case of aggression, anger.

The French philosopher Jean-Marie Guyau, writing almost a century before Lorenz, was one of the first thinkers to understand the relevance of Darwin's evolutionary theory for moral philosophy and constituted "one of the first 'evolutionist' ethics"[374]. He identified this same "natural and legitimate instinct" that Lorenz is describing with a "sentiment" that generates a "general law of life", i.e.:

> Every animal […] answers to an attack by a defence, which is itself most often an attack in answer, a kind of shock in return. […] Those beings in which this instinct was keenest and most developed have more easily survived, like the rose-trees provided with thorns. […] In us there is a spring ready to unbend against him who touches it […]. It is,

[371] Nesse, "Evolutionary Explanations of Emotions," 268.
[372] Richard Dawkins, *The Selfish Gene*, 30th anniv (Oxford: Oxford University Press, 2006), 8.
[373] Konrad Lorenz, *On Aggression*, ed. Julian Huxley, trans. Marjorie Kerr Wilson (London and New York: Routledge, 2002), 44.
[374] Philippe Saltel, "Présentation," in *Esquisse d'une Morale sans Obligation Ni Sanction* (Paris: Les Belles Lettres, 2008), 23, my translation.

> in fact, necessary for the life on the individual. In every primitive society one must be able, in order to live, to bite when bitten, to strike when stroke. […] We find the same feeling among animals.[375]

As Guyau well recognised, the function of anger is not inherently moral; to the contrary, it is an emotion of power[376], hierarchy[377], destruction[378]. It is not only evident from the behaviour pattern that normally follows it (i.e., aggression, whether physical or only verbal) but also from the physiological changes that are correlated with it, such as increased muscle tension, heart rate and blood pressure, and release of adrenal hormones[379] – all elements that prepare the body for aggression. Darwin himself had observed that "an indignant man throws himself into a fitting attitude for attacking his opponent, though without any intention of making an actual attack"[380].

It is true that it is also elicited when moral norms are violated, and this generates the illusion that the sentiment is in itself moral. However, this violation, I believe, triggers anger because the moral order becomes part of our identity, and the violation is perceived as an attack. Not only that but, as we have seen, the scapegoat mechanism also comes into play. Plutchik has formalised this tendency to discharge violent emotions on innocent victims as a form of psychological "ego defence":

> Basically, all ego defenses are coping strategies designed to deal with conflicts over particular emotions. If the expression of anger is dangerous, that is, produces threats or invites retaliation, then displacement expresses the anger toward a weaker scapegoat and avoids the danger.[381]

Anger is undoubtedly an indicator that something is wrong, but not necessarily – and not primarily – in a moral sense. It indicates that there is a danger and that aggression might be a solution. But if we realise it is not, we may divert it towards other victims, the scapegoats. The danger the emotion indicates might be moral: we might feel attacked in our moral self, our values, and this often generates outrage. However, remembering the adaptive function of primary emotions should also make us wary of

---

[375] Jean-Marie Guyau, *A Sketch of Morality Independent of Obligation and Sanction* (London: Watts & Co., 1891), 173–74.

[376] Scott Schieman, "Anger," in *Handbook of the Sociology of Emotions*, ed. Jan E. Stets and Jonathan H. Turner (New York: Springer, 2006), 508.

[377] Plutchik, "A General Psychoevolutionary Theory of Emotion," 27.

[378] Plutchik, 16.

[379] Jerry L. Deffenbach et al., "State-Trait Anger Theory and the Utility of the Trait Anger Scale," *Journal of Counseling Psychology* 43, no. 2 (1996): 131–48.

[380] Darwin, *The Expression of the Emotions in Man and Animals*, 369. Remember that indignation has been here included under the broader category of anger.

[381] Plutchik, "A General Psychoevolutionary Theory of Emotion," 25.

anger's moral indications, as it is indeed biased in favour of the subject experiencing it in order to increase their chances of survival and reproduction.

A silly example: when I ride my bike, I mostly feel entitled to all road spaces and sometimes to pavements too. A feeling of anger arises in me when a car honks at me – maybe because I am riding on the wrong side of the road – or when a pedestrian yells at me to get off the pavement. I then recognise that I am at fault, but still, I tend to find such reactions exaggerated. Things are quite different when it is me behind the car's wheel, and cyclists seem to me reckless, irresponsible, and, above all, in the way. Not to mention when I am on foot, and I consider myself the weaker inhabitant on the road and feel that bikes should be my allies against the cars and not steal and endanger the little space I have on the pavement. I will not be talking about what I feel as a pedestrian towards cars and as a car driver towards pedestrians; it should by now be clear.

Despite all this, it cannot be denied that reactive emotions can be demonstrations of "our commitment to certain moral standards"[382], and they might, in this sense, have positive heuristic value – although other emotions, such as grief, could fulfil the same role (*pace* Moore) without anger's troublesome implications – but this does not make them a good guide for subsequent action. As Nussbaum remarked, anger is normatively problematic because it is either justified through "down-ranking" or through payback. In the former case, the offence is considered such in relation to one's status, and thus to offend in return does re-balance the relative status of the two agents.[383] However, it is problematic, to say the least, to consider that harm is a matter of status, and luckily it is an idea mostly relegated to the past. Having set cultures of honour aside, we now focus on the suffering that is caused by wrongdoing, not on its down-ranking effects.

The latter case probably fits present-day conceptions better but is structurally problematic. The concept of payback assumes that the punishment of the guilty may somehow restore what has been damaged by the offence. According to Nussbaum, this idea derives from an implicit notion of the world being ordered through a cosmic balance which should be rebalanced[384]. This conception is not so implicit either: the two-pan scale is the symbol of the Western conception of justice, a simple Google image search of the word "justice" may prove it – of the first fifty images found by the

[382] Wallace, *Responsibility and the Moral Sentiments*, 69.
[383] Nussbaum, *Anger and Forgiveness*, 5–6.
[384] Nussbaum, 5.

search engine, forty are images of two-pan balances. Once the idea of payback is explicitly exposed, it becomes clear that it makes no sense: nothing is brought back once the blow has been returned. As common as it is to speak of offenders as being "indebted to society" and needing to "pay their dues", crimes are really not like debts that are erased when the money or the goods are returned.

While I agree that the idea of payback encompasses the idea of cosmic balance, the derivation relationship identified by Nussbaum seems incorrect. In fact, as she claims a few pages later, the idea of payback is part of the conceptual content of anger[385]. So, the idea of balance derives from it, it is not its origin: in the beginning was emotion, we might say. However, the imagery of the cosmic balance is then used as a justification – irrational but accepted because it appeals to common sentiments – for the practice of retribution. If we accept a retributive theory of punishment, we are sublimating the reactive emotion of anger, transforming it into a moral principle, that of justice. We are performing, in Hart's words, that "mysterious piece of moral alchemy in which the combination of the two evils of moral wickedness and suffering are transmuted into good"[386]. As Albert Camus noticed in his essay on capital punishment,

> Whoever has done me harm must suffer; whoever has put out my eye must lose an eye; and whoever has killed must die. This is an emotion, and a particularly violent one, not a principle. Retaliation is related to nature and instinct, not to law. Law, by definition, cannot obey the same rules as nature. If murder is in the nature of man, the law is not intended to imitate or reproduce that nature. It is intended to correct it. Now, retaliation does no more than ratify and confer status of law on a pure impulse of nature.[387]

This transmutation is brought to its extreme consequences by Kant in his radically retributive doctrine of punishment:

> Even if a Civil Society resolved to dissolve itself with the consent of all its members […] the last Murderer lying in the prison ought to be executed before the resolution was carried out. This ought to be done in order that every one may realize the desert of his deeds, and that bloodguiltiness may not remain upon the people; for

[385] Nussbaum, 15.

[386] Herbert L. A. Hart, *Punishment and Responsibility. Essays in the Philosophy of Law* (New York: Oxford University Press, 2008), 234–35.

[387] Albert Camus, "Reflections on the Guillotine," in *Resistance, Rebellion, and Death*, trans. Justin O'Brien (New York: Knopf, 1961), 198. It is the same position as Guyau, who adds that it is a confusion that also involves language, transferring terms "which belong to the sphere of the passions" on the moral level. Guyau, *A Sketch of Morality Independent of Obligation and Sanction*, 206.

> otherwise they might all be regarded as participators in the murder as a public violation of Justice.[388]

It may be surprising that Kant, the champion of Reason, should be treated in a section on emotions. It is true, in fact, that his theory does not appeal to emotions directly: in order for the execution of an offender to be just there is no need for anybody to be angry with them, according to Kant. However, I still defend the choice to address Kant here because, although he lacks any reference to emotion, I believe emotion to be at the basis of his retributive stance: it is because humans are angry and resentful that one comes to believe that justice consists in the replication of harm. It is the experience of the satisfaction of vengeful desire that makes one consider retribution as a value in itself: retribution, in fact, looks at the past but cannot restore it.

That the basis of retributive justice is found in an evolutionarily adaptive emotion explains its quasi-universality, justifying the theorisation of a natural tendency to retribution by many thinkers: Peter Strawson argues that reactive attitudes such as resentment (i.e., another form of anger) are essential to human life, their suppression being not only undesirable, bringing to an incalculable impoverishment of existence, but impossible[389]. Individual moral responsibility and the retributive system resulting from it are founded on participative-reactive attitudes: the impossibility of abandoning them implies the impossibility of abandoning the paradigm of responsibility and retribution based on it.

Susan Jacoby, to give another example, speaks of an "instinct to make others suffer as we have suffered"[390] and warns her readers about the impossibility to suppress it and the dangers of attempting to do so[391]. Similarly, Girard, as we have seen, conceptualises it in the form of the imitation of violence, that he considers so fundamental to the human being that he who has managed to free himself from mimetic violence cannot but be more than human: the son of God[392]. As Mormino remarked[393],

---

[388] Immanuel Kant, *The Philosophy of Law. An Exposition of the Fundamental Priciples of Jurisprudence Ad the Science of Right*, trans. W. Hastie (Edimburgh: T. & T. Clark, 1887), 198.

[389] Strawson, "Freedom and Resentment," 14. Gary Watson has called Strawson's theory an "expressive theory of responsibility": Gary Watson, "Responsibility and the Limits of Evil. Variations on a Strawsonian Theme," in *Responsibility, Character, and the Emotions. New Essays in Moral Psychology*, ed. Ferdinand Schoeman (Cambridge: Cambridge University Press, 1987), 261.

[390] Jacoby, *Wild Justice*, 333.

[391] "Dismissing the legitimate aspects of the human need for retribution only makes us more vulnerable to the illegitimate, murderous, wild impulses that always lie beneath the surface of civilization […]." Jacoby, 362.

[392] Jean-Michel Oughourlian summarises it in his conversation with Girard: "This unprecedented task of revealing the truth about violence requires a man who is not obliged to violence for anything and

Girard weakly justifies the passages of the Gospels in which Jesus shows a more vengeful conception of divinity – that he calls "minor defects"[394] – with the need for Jesus to "speak their [his listeners'] language up to a certain point and take into account illusions that cannot yet be eradicated"[395]. Furthermore, his divinisation of the ability to escape retributive logic is clearly ideological, without the need to appeal to intimate feelings of compassion and forgiveness; Martha Nussbaum offers three major examples of public figures that in the 20th century taught to eschew violent reciprocity, the three masters of the non-violence movement: Mahatma Gandhi, Martin Luther King, and Nelson Mandela[396].

When Guyau, at the end of the 19th century, remarked that "a justice which proportions good to good and evil to evil […] is but the metaphysical symbol of a deep-rooted physical instinct, which in reality is part of the instinct of the preservation of life"[397], he was laying the foundations for a critique of a retributive conception of justice[398], not for its justification. He was conceiving a path of moral progress that should eventually overcome the concept of punishment[399], whose relationship with demerit he defines as a "mystical alliance"[400]. On the contrary, authors such as Strawson and Jacoby use it to justify retributive justice. Furthermore, the fact that other non-

---

does not think in terms of violence-someone who is capable of talking back to violence while remaining entirely untouched by it.

It is impossible for such a human being to arise in a world completely ruled by violence and the myths based on violence. […] For this reason the Gospels and the whole New Testament, together with the theologians of the first councils, proclaim that Christ is God not because he was crucified, but because he is God born of God from all eternity." Girard, *Things Hidden Since the Foudation of the World*, 218–19.

[393] Mormino, *René Girard. Il Confronto Con l'altro*, 268–69.

[394] Girard, *Things Hidden Since the Foudation of the World*, 188.

[395] Girard, 189.

[396] Nussbaum, *Anger and Forgiveness*, chap. 7. The Political Realm: Revolutionary Justice, 211-246.

[397] Guyau, *A Sketch of Morality Independent of Obligation and Sanction*, 177.

[398] "Let it be noted that the moral notion of justice or of merit is still unknown throughout this whole mechanism. If an animal without its brain bites him who hurts it, there is here no idea of sanction. If you ask a child or an uncultured man why he strikes someone, he will think himself fully justified when he tells you that he himself was first struck. Ask him no more. Really, for him who only considers the general laws of life, this is a sufficient reason.

We are here at the very origin, and almost at the point of physical emergence, of this pretended moral need of sanction, which so far does not offer us anything moral […]." Guyau, 175.

[399] Guyau, 180–86. "We shall wish more to help those who suffer than to reward in a brilliant and superficial way those who have acted well. For distributive justice, which is an entirely individual and personal justice, a justice of privilege (if such words are not a paradox), must, therefore, be substituted a justice of a more absolute character, and which at bottom is really charity. Charity for all men, whatever may be their moral, intellectual, or physical worth, should be the final aim to be pursued even by public opinion." Guyau, 186.

[400] Guyau, *A Sketch of Morality Independent of Obligation and Sanction*, 153.

human animals demonstrate similar reactive mechanisms has led some to argue that they have a "sentiment of justice"[401].

That at the heart of retributive laws lies this "alchemic" mechanism that transforms these vengeful impulses into a moral duty is made evident by the distinction in punishment between attempted and accomplished crimes, which Hart himself ascribes to a difference in resentment – another form of anger – felt by the victim[402]. The case of attempted but unrealised crimes is a form of moral luck with respect to the consequences of an action – in Nagel's words, "in the way one's actions and projects turn out"[403]. Nagel's example for this type of moral luck is that of a truck driver who commits minor negligence, like postponing their brakes' check with a mechanic, and cannot thus stop in time when a child unexpectedly crosses the road. Worse cases apply, too, as if the driver ran over a child because he was drunk[404]. In both cases, if the child had not run across the street, the driver would be considered slightly negligent or inconsiderate – but because of accidental circumstances, they are now guilty of manslaughter and considered deserving of some degree of punishment.

> That these are genuine moral judgements rather than expressions of temporary attitude is evident from the fact that one can say in advance how the moral verdict will depend on the results. If one negligently leaves the bath running with the baby in it, one will realize, as one bounds up the stairs toward the bathroom, that if the baby has drowned one has done something awful, whereas if it has not one has merely been careless.[405]

As in the case of the contrast proposed by Hart between attempted and accomplished crimes, what changes here again is the resentment felt by the victims, and by the community on the whole, that "ask for justice" while actually asking for punishment.

Moreover, that this "alchemy" is a form of emotional, irrational, hardly justifiable mental process is further testified by the changing in feelings and related sentiment of justice when we superficially learn about a crime story and when we come

---

[401] Jonathan H. Turner and Jan E. Stets, "Moral Emotions," in *Handbook of the Sociology of Emotions*, ed. Jan E. Stets and Jonathan H. Turner (New York: Springer, 2006), 544–66.

[402] "My own belief is that this form of retributive theory appeals to something with deeper instinctive roots than the last mentioned principle. Certainly the resentment felt by a victim actually injured is normally much greater than that felt by the intended victim who has escaped harm because an attempted crime has failed." Hart, *Punishment and Responsibility*, 131.

[403] Nagel, "Moral Luck," 177.

[404] Nagel, 178.

[405] Nagel, 179.

to know its background story. Gary Watson shows this in his account of the heinous crimes of murderer Robert Harris. In 1978, 25-year-old Robert Harris, according to the account given by his brother Daniel, witness of the scene, laughingly killed two teenagers for no reason. Daniel Harris' account makes the blood run cold, and although an ideological stance against the death penalty may make one hostile to the idea of him being condemned to capital punishment, the feeling that he has to "pay his debt to society", "right his wrongs" through his suffering clearly emerges in most readers. However, Watson then cites another interview, that with Robert's sister Barbara, in which she tells of Robert's childhood, one of unspeakable physical and psychological abuse by both parents. An unwanted and even hated child since pregnancy, neglected and affectively deprived during early childhood, raped in his teenage years during a stay in a youth prison for stealing a car[406], "No wonder Harris is as he is!"[407]. As we get to know his tragic story, our retributive feelings fade away, leaving room for compassion – although, obviously, our judgement on the heinous actions he performed does not change, and we surely believe that something must be done to prevent future damages.

We are facing a case of manifest constitutive luck: unfortunate circumstances during Harris' upbringing explain his deviance. The knowledge of these circumstances silences retributive feelings, and suddenly, the idea of a payback seems unfitting. Nevertheless, although we do not always know the circumstances that brought someone to act heinously, and although sometimes they can never be known, for they are obscure even to the agents themselves, agents always act because of reasons, and whether these appear convincing or not is beyond their control: they ultimately depend on constitutive luck.

In his essay, Nagel only refers constitutive moral luck to the moral judgement concerning one's character and inclination, though seemingly conceding that, according to the Kantian conception, such impulses can always be overcome. Still, he disagrees with Kant that one's character and inclination cannot be the object of moral judgement, and since they are, in fact, the object of such judgements, although depending on fortuitous conditions, he traces these judgements back to constitutive moral luck[408]. In my view, constitutive moral luck goes further also to include the ability or inability to repress or overcome these impulses. So, even if Kant were right in that we could and

[406] Watson, "Responsibility and the Limits of Evil. Variations on a Strawsonian Theme," 272–74.
[407] Watson, 274.
[408] Nagel, "Moral Luck," 181–82.

should abstain from judging a person's character, however insensitive, unfriendly, or arrogant they may be, constitutive luck would still be of paramount moral importance. This concerns the structure of action and its relationship with its reasons.

According to Galen Strawson, reasons are "actually existing psychological states of a certain sort, paradigmatically beliefs and desires"[409]. Beliefs do not depend on the agent: on the contrary, when we act, we hope our beliefs reflect a state of things independent from our believing it. We surely hope the world itself to constitute the content of our perceptual beliefs. Things might seem less straightforward for what concerns normative beliefs, yet, I think that if we do not want to take a moral stand that is radically relativistic – and, ultimately, nihilistic – we must believe that something can be considered objectively right and that we can attain moral progress – and that we did accomplish some of it, for example when we stopped discriminating people based on their gender, race, sexual orientation, and so on.

Although we might be conscious about our own moral limited view, and therefore allow for a certain degree of practical relativism – that is, not imposing our values on different cultures in the assumption of being the repositories of moral truth, especially given the morally catastrophic consequences of such attempts in the past – this does not mean that we believe that we should not struggle to find some kind of moral truth. If we do not think so, the idea of moral progress does not make any sense. A belief in moral progress implies a stance on moral realism. Then, whether our beliefs are true or false depends on our capacity to recognize moral truths – whether we consider them to be external to the individual, like some kind of objective value handed over to humanity by God, or internal, like the objectivity of our primary needs and the related fundamental experience of pleasure and pain, as in the ethical system elaborated by Spinoza and, later, by the utilitarians. Just like it is not in our power to recognise other kinds of truths – they impose themselves on us when we recognise them, and we cannot but accept them, having no choice in believing that two plus two equals four or that the snow that we see is white – so it is for moral truths. The possession of this capacity, like any other capacity, depends on the constitutive luck that determines the elements that concurred in the formation of our character: our genetic heritage and experiences.

[409] Galen Strawson, *Freedom and Belief* (New York: Oxford University Press, 2010), 24.

As for desires, I agree with Hobbes: "I acknowledge this liberty, that I can do if I will; but to say I can will if I will, I take to be an absurd speech"[410]. A simple self-analysis reveals that we have no power over our desires: they just emerge in our consciousness. If we do not like them, we might train them so that, in Harry Frankfurt's terminology, first-order desires correspond to our second-order desires[411], meaning that "we want what we want to want". However, the will to train once again depends on reasons: we must believe training to be good and therefore desire it, and, as we have just seen, this depends on constitutive luck.

This short excursus on the motivational structure of action should corroborate my claim that all crime stories eventually can be explained through a tragic story – in the worst case, just an irresistible impulse that one cannot control and that we are lucky not to have. Virtually all such hateful stories that initially arouse anger and payback feelings can engender compassion instead. This instability makes the retributive emotions even worse guides for actions than they seemed before.

Does this mean that we should use compassion as a moral compass? This would once again lead justice back to one emotion. Access to moral knowledge is undoubtedly problematic, for the moral realm is indeed incorporeal: we cannot "see" or "touch" it; we do not have a specific sense – as was theorised by some intuitionists – to get in contact with it. We could hardly develop a sense for it – and surely emotions, as we have seen, are not this kind of sense – for senses are the result of evolution, and morality does not provide an adaptive advantage oftentimes.

2.2.3 Moral Realism and the Right to Be Sure

Huemer maintains that "all justification for evaluative beliefs derives ultimately from intuition."[412] Though it is true that all moral claims can be traced back to "initial intellectual appearances", as we have defined intuition before, for our moral theory to be solid, we need to be able to elaborate on those intuitions and rationally justify them, corroborating them with convincing and sharable arguments for, given the wide variety of moral beliefs held by different cultures across time and space, it is clear that virtually

[410] Thomas Hobbes and John Bramhall, *Hobbes and Bramhall on Liberty and Necessity*, ed. Vere Chappell (Cambridge: Cambridge University Press, 1999), 16.

[411] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68, no. 1 (1971): 5–20.

[412] Huemer, "Revisionary Intuitionism," 370.

everything can be taken as moral intuition. He then proposes a critical methodology for discerning good intuitions from bad ones by identifying and discounting the biases that affect our moral thought[413]. He pinpoints three types of biases that chiefly affect moral beliefs: those dependent on cultural influence, biological programming, and personal interests[414]. In order to attain moral truth, he adds to these the probatory value of coherence among beliefs that support and explain each other forming a system[415]. This method should grant what Alfred Ayer has called "the right to be sure"[416], that is, the ability to answer satisfactorily to the question: "How do you know it?"[417]. Whereas many intuitionists have maintained that "by intuition" should be enough, for intuitions cannot[418] and should not[419] be further verified, it could hardly be considered a satisfactory answer. To just say, "It's my intuition", is equivalent to saying, "It seems to me", "I believe it". Accepting it as an answer would mean equating belief and knowledge when, in fact, something should be added to the former to speak of the latter: an explanation that could be shared and convince others, some evidence that grounds knowledge outside the subject.

At the beginning of this section, I have mentioned the approach deontologists, consequentialists, and virtue ethicists would have towards some common moral intuitions. It is now time to investigate their ability to justify them. I noticed that deontologists could extrapolate basic rules of morality from the first two examples there mentioned – i.e. that, all other things being equal, the helpless should be helped, and the innocent should not be hurt. Deontologists can extrapolate a moral rule from the last two examples, too, the same rule stated by Polemarch, i.e., that rewarding those who do good and punishing those who do bad is just – in fact, not only that: it is itself justice. However, were we to ask the deontologist why it should be so, it seems that they could only appeal to intuition, to the fact that "it is evident", that in the end only means "it seems to me". They do not seem to have earned their right to be sure.

Virtue ethicists, in their turn, could take all these actions as signs of a virtuous character, of one being benevolent in the first two cases or just in the last two. Yet

[413] Huemer, 370.
[414] Huemer, 374–77.
[415] Huemer, 372, 379.
[416] Alfred Ayer, *The Problem of Knowledge* (London: MacMillan, 1956), 34.
[417] Hudson, *Ethical Intuitionism*, 57.
[418] Strawson, "Ethical Intuitionism," 24.
[419] Ross, *The Right and the Good*, 21.

again, how one would identify virtue and its content – e.g., that being just consists of rewarding those who do good and punishing those who do bad – remains mysterious: virtue ethics does not provide a method to grasp virtues; it only enumerates them. Again, we have seen how such intuitions could be, and probably are, biased by many factors, not least our evolutionary history. Of course, both deontologists and virtue ethicists could reject such a definition of justice were they not to find it a duty in the first case or a virtue in the latter. The problem in both cases is that if one identifies rules and duties or virtues as the most fundamental and irreducible concepts of ethics that cannot be understood by reference to something else, these basic concepts will appear arbitrary, to say the least. They rely on intuition only, without further elaborating with convincing rational arguments.

A form of consequentialism, on the other hand, could break the deadlock. At first sight, consequentialism does not seem to overcome the basic moral problem exposed so far: if a good action is defined in terms of its good consequences, it remains to establish how to discern positive from negative consequences. In the examples that we are using, it remains intuitively good that helpless children or innocent people should not suffer; thus, helping the former and avoiding hurting the latter could be moral actions in a consequentialist view, for they bear good consequences. However, things change once we consider the "justice examples". If the consequentialist maintains that punishing the guilty bears good consequences because it produces justice, they introduce another moral criterion, namely, "justice", according to which one needs to evaluate consequences. The goodness or badness of actions is ultimately traced back to arbitrary values, not different from the ones that constitute the duties or the rules of deontologists or the virtues of virtue ethicists. However, consequentialism has a way out of the impasse: hedonistic utilitarianism.

Utilitarianism will be treated at length in the last chapter of this work; for now, suffice it to say that its ability to overcome this challenge lies in the possibility of identifying a universal criterion for the classification of consequences between good and bad, one that is grounded in the positive and negative qualities of our experience. We are set to delve into this topic later; here, a reference to the experience of pleasure and pain as a fundamental criterion for ethics should be enough. Hedonistic utilitarianism so conceived can also tackle Huemer's challenges, for it requires adherence to a minimal set of coherent principles that are also supported by experience, namely the experience of pleasure and pain. These experiences can be appealed to in answering the question

"How do you know it?" thus granting the hedonistic utilitarian their right to be sure, to claim they have reached moral knowledge in a realist framework. It refutes the claim that moral judgements are corrigible like musical tastes – and not like a doctor's diagnosis[420]. Although it is true that producing evidence in moral argumentation means "using the facts to back our attitudes, to appeal to the capacity of others to feel as we feel, to respond as we respond"[421], the fact of feeling the same as sentient beings grounds objective moral truth. Finally, hedonistic utilitarianism also constitutes a method that virtually allows one to solve any moral dilemma.

All these aspects will be examined in depth later. This short introduction should be enough to see how this approach could provide an alternative way of dealing with the justice examples, one that does not appeal to arbitrary rules or virtues, is not biased by emotion and is rationally defensible. Surprising as it may seem, since legal philosophers usually deploy utilitarianism as a defence strategy for the right to punish, I believe utilitarianism fundamentally undermines not only the retributivist justification for punishment but the practice of punishment in itself. I will propose my conception of the utilitarian approach to violence within society in the last chapter, and I will analyse the supposedly utilitarian justifications of punishment in the next section, where the topic of the justifications for punishment will be addressed.

2.3 Punishment: A Philosophical Analysis

2.3.1 Between Future and Past: The Traditional Justifications for Punishment

Two kinds of justification can be invoked to defend the practice of punishment, either forward-looking or backwards-looking. The former can take several forms: it can aim at deterrence, incapacitation, or rehabilitation – most of the times, it consists of a mix of the three[422]. Deterrence, the main focus of utilitarian justifications of punishment, focuses on exemplary punishment; it has its leverage on fear, which was the central feature of the torture system. In fact, before the industrial revolution,

[420] Strawson, "Ethical Intuitionism," 27.

[421] Strawson, 28.

[422] It can be noted that, although typically referred to as justifications *for punishment*, rehabilitation, and incapacitation even more, are really justifications *for incarceration*: death penalty, for example, for obvious reasons does not rehabilitate; torture does neither (although one could support the practice on the assumption that, like children's chastisements, it should have educational value). However, we are so immersed in the carceral paradigm that punishment and incarceration are often confused.

offenders were hardly caught – due to a lack of infrastructures and technologies in general that allow for the investigation, identification, and prosecution of the culprit – and the power system relied on the fact that, even though the probability of being punished where low, the thought of the terrible punishment administered if one were caught[423] would be so intense to overcome the thought of its unlikelihood and convince potential offenders to abstain from felony. As we have seen, according to Foucault's semio-technique rule of "sufficient ideality", such punishments do not even need to be administered; it is enough for potential offenders to imagine them. The present-day penal system works the other way around: punishments are "mild", to recall Beccaria's terminology in his reform proposal against the torture system[424], but the probability of being caught is high. Thus, the utility calculus suggests potential offenders to abstain from wrongdoing.

Of course, such a theory of punishment poses many well-known moral problems to its supporters, the most relevant being the instrumentalisation of individuals – to punish one to educate others – and the justification for disproportionate punishment, as was done in the torture system. A critique of proportionality has already been moved, so the criticism of disproportion might feel inappropriate; however, one cannot help but feel that, say, the death penalty for theft, despite probably being effective, would be unjust. But as I am trying, through a utilitarian framework, to base moral arguments on rational reasoning rather than on intuition and feeling, a utilitarian justification for avoiding disproportionate punishment is needed.

Bentham, in formulating his utilitarian justifications for punishment, was part of that very reform that introduced penalty proportionality – as opposed to the intrinsically excessive suffering administered by the torture system when crime was chiefly considered as an offence towards the monarch. The softening of penalties is, in fact, advocated by the utilitarians because all suffering that goes beyond what is strictly necessary for the positive consequence of deterrence is *useless*. However, deterrence is also not in itself the only consequence that must be taken into account in the utilitarian calculation: if penalties are too hard, the cost of deterrence – in terms of suffering – can become too high, so that it becomes preferrable to accept a higher degree of crime than

[423] See for example Damiens' torture at the beginning of *Discipline and Punish*: Foucault, *Discipline and Punish*, 3–6.

[424] Cesare Beccaria, *On Crimes and Punishments and Other Writings*, ed. Aaron Thomas, trans. Aaron Thomas and Jeremy Parzen (Toronto, Buffalo, London: University of Toronto Press, 2008), chap. XXVII. Mildness of Punishments, 49-51.

to impose punishments that are too severe. Moreover, if we take into account the bigger picture, i.e., the prospect of living in a society where theft is punished by death, even though crime might be more effectively deterred, we recognise that we would hardly be maximising overall utility where, among the many problem that we can picture such society to have, the population would live in fear and the people close to petty offenders would suffer immeasurably. In Bentham's words, in this case punishment is *unprofitable*[425].

Furthermore, that an implicit or explicit utility calculus moves potential offenders is a rationalist assumption that, although intuitively understandable, still needs to be proven, especially in the face of the statistics on the mental health conditions of offenders. As we have previously seen, numerous studies have found that an enormous portion of the carceral population suffers from mental disorders at the time of their incarceration, specific numbers depending on the disorder and the incarcerated population's specific features, but all studies to my knowledge confirm the trend. Moreover, many violent crimes are explicitly classified as "crimes of passion" – by definition excluded from the rational calculation – and in many countries, for this reason, they benefit from mitigation in court. As I previously mentioned, when deciding their course of action, most people do not even consider a criminal action, especially a violent crime, in their implicit utility calculation. For the great majority of the citizens in democratic States, the reason not to commit crimes lies in a certain form of moral and social education, on which we will dwell later. The ordinary decision process does not contemplate crime and subsequent deterrence from fear of punishment.

These elements seem to undermine the effectiveness of general deterrence and even of specific deterrence – i.e., the deterrence of the person who has been incarcerated after their release. The soundness of forward-looking justifications is of empirical nature and must, therefore, be supported by data on the outcome of the practice. If the data is missing or not considered in policy-making, then the practice is forward-looking only as a façade, to hide other non-forward-looking reasons that one cannot rationally justify.

Data on general deterrence are hard to gather since the counterfactual is missing. However, a 2009 meta-analysis[426] that gathered data from seven hundred studies on

[425] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, ed. J. H. Burns and H. L. A. Hart (Oxford: Oxford University Press, 2015), 163.

[426] Dieter Dölling et al., "Is Deterrence Effective? Results of a Meta-Analysis of Punishment," *European Journal on Criminal Policy and Research* 15, no. 1 (2009): 201–24.

deterrence conducted between 1952 and 2006 showed that, although the deterrence hypothesis is not entirely disproven, meaning that some deterrent effect of the criminal system is, in fact, observed, once things are looked at more closely, the overall picture is more blurred. First, punishment is especially effective as a general deterrent with respect to crimes against property and so-called administrative offences[427], while concerning violent crimes – the ones that primarily concern citizens and for which the longest sentences are meted out – show very little efficacy. Certainty is also more relevant than severity[428], as was hypothesised by Beccaria[429], a result consistent with the fact that the toughest of penalties, the death sentence, shows no deterrent effect[430]. The relevance of certainty with respect to severity, in my opinion, calls into question the need for severity, that is, for punishment, at all, for we cannot discern whether the deterrent effect is due to the fact that the response to the detected offence takes the form of punishment, or whether the mere perspective of being caught and having to face certain consequences, which may take other forms – that will be analysed in the last chapter of this work – could be enough.

Even taking into account the consequences on minor offences, the estimated deterrent effect is still small, classified by the authors of the meta-analysis as "non-significant"[431] – a result that is consistent with behavioural research on criminal conduct, which has shown that people tend to comply with the law because they recognise its legitimacy and authority more than for fear of punishment[432]. Minor deterrent effects radically challenge deterrence as a good utilitarian justification: punishment seems unprofitable.

From a utilitarian point of view, to justify punishment on the basis of deterrence – as Bentham tried to do[433], as we shall see in the next chapter – one needs to show not only that some benefit is generated through punishment, that some individuals are deterred, but also that the benefits obtained are greater than the evil that is inflicted, a

[427] Dölling et al., 210, 216, 223.
[428] Dölling et al., 210.
[429] Beccaria, *On Crimes and Punishments and Other Writings*, 49, 72.
[430] Herbert Morris, "Nonmoral Guilt," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987), 220–40.
[431] Dölling et al., "Is Deterrence Effective? Results of a Meta-Analysis of Punishment," 222.
[432] Tom Tyler, *Why People Obey the Law* (Princeton: Princeton University Press, 2006).
[433] Bentham, *An Introd. to Princ. Morals Legis.*

fact that Bentham summarised in the notion of "profitability"[434]. Given that only a tiny fraction of the population is deterred by the threat of punishment, as I have argued, to consider punishment to be profitable is not as straightforward as it might have seemed. Moreover, one needs also to demonstrate that punishment is the most economical way – in terms of generated suffering – to achieve that benefit. Otherwise, as Bentham claimed, punishment is unjustified because it is "needless"[435]. If other systems were to achieve similar results – or even lower results but compensated by a much lower cost of suffering – then the practice of punishment should be rejected, given that we focus on forward-looking practices and renounce the retributive logic. I will illustrate such systems in the next chapter.

Concerning specific deterrence, discouraging data on recidivism show that re-arrest over a 9-year period for former federal convicts in the U.S. reaches a rate of 83%[436]. This is consistent with data collected since the Seventies always showing poor, if not negative, effects in terms of specific deterrence of state-administered punishment[437]. Prisons have been since identified as criminogenic environments[438], where petty offenders are transformed into hardened criminals, which brings us directly to the second consequentialist justification for punishment: incapacitation.

It must be noticed that incapacitation is not, strictly speaking, a justification for punishment in general; instead, it is a justification for that specific form of punishment which is incarceration (or death penalty)[439]. If incapacitation really were the aim of

[434] Bentham, 288–89.

[435] Bentham, 164.

[436] Denis Yukhnenko, Shivpriya Sridhar, and Seena Fazel, "A Systematic Review of Criminal Recidivism Rates Worldwide: 3-Year Update," *Wellcome Open Research* 4, no. 28 (2019). In Italy, 62% of currently incarcerated people has been to prison at least once before, with 18% having been incarcerated at least 5 times: Associazione Antigone, "Il Carcere Visto Da Dentro. XVIII Rapporto Di Antigone Sulle Condizioni Di Detenzione. Numeri Del Carcere," 2022, https://www.rapportoantigone.it/diciottesimo-rapporto-sulle-condizioni-di-detenzione/numeri-del-carcere/. Accessed on 29/04/2024, 10:08.

[437] Michael Tonry, "Doing Justice in Sentencing," *Crime and Justice* 50, no. 1 (2021): 4; Marina Bell, "Abolition: A New Paradigm for Reform," *Law & Social Inquiry* 46, no. 1 (February 26, 2021): 32; R. Antony Duff, "Responsibility, Restoration, and Retribution," in *Retributivism Has a Past: Has It a Future?*, ed. Michael Tonry (Oxford: Oxford University Press, 2011), 64; Andrew Von Hirsch, *Past or Future Crimes. Deservedness and Dangerousness in the Sentencing of Criminals* (Manchester: Manchester University Press, 1985), 3–18.

[438] Michelle Brown and Judah Schept, "New Abolition, Criminology and a Critical Carceral Studies," *Punishment and Society* 19, no. 4 (2017): 447.

[439] Punishment is not the same as imprisonment: it is the voluntary infliction of suffering as a consequence of wrongdoing, and it is not intrinsically related in any way to incarceration – in fact, we impose and suffer forms of non-carceral punishment in our lives both in hierarchical relationships, where many punish their children or their employees, and in equal relations: what else is it to sulk when our partners or close friends hurt us? One might argue that personal and institutional forms of punishment are

incarceration, if no intrinsically retributive intent was added, then it should be comparable to quarantine. Ferdinand Schoeman[440] and, to some extent, by Derk Pereboom[441] tried to explore this way of punishment justification. Most of us recently experienced some form of quarantine due to the Covid-19 pandemic. Despite being stuck in our homes, with all the comforts of everyday life – surely, some homes more comfortable than others, but the vast majority in Western countries more comfortable than a prison cell –, in most cases with someone taking care of them, having almost immediate access to any book, entertainment, internet, food, often with access to more than enough space, and virtually complete autonomy in the organisation of their day, most found this condition distressing. And we believed that all other restrictions not strictly justified by sanitary risks would not have been acceptable. What we witnessed happening in China, where entire cities were quarantined despite the limited number of contaminations, healthy citizens were prevented from meeting with their families and even doing their grocery shopping or going for a walk, where infected people were taken from their homes and brought to designated places, often dirty, overcrowded, and far from home, appeared to be strictly unfair to us, already struggling to follow our own milder rules. Of course, stricter rules, and stricter controls, would have meant fewer contaminations, but we did believe that a certain degree of risk could be taken in order not to impose too harsh conditions on the people at risk of contamination.

Our attitude to offenders' conditions is the opposite: cyclically, the news that one well-known criminal is granted access to comfortable conditions during their detention raises disapproval; that the supervisory magistrate has given permits to facilitate studying, working, or contact with the family to the incarcerated person generates anger. When it is pointed out that these procedures are permitted by law and fall within the terms of the sentence assigned, the typical response is, "They may have spent many years in prison, but now they regain freedom, whereas their victim will

not comparable, and the former cannot help illuminate the latter. I would tend to disagree (although, of course, I do believe that they should not be equated) since, as it should by now be clear, I believe the practices of justice put in place by States derive from our interpersonal feelings and practices. However, should one reject the comparison, it would suffice to remember that, for most of our history, State punishment has not taken the form of incarceration. If certain forms of physical punishment may have incapacitated offenders – mutilation, for example – that was clearly not the main aim of the practice. Foucault described the punishment system of the *ancien régime* primarily as "a whole quantitative art of pain" (Foucault, *Discipline and Punish*, 34.), where the infliction of pain was justified by deterrent and retributive reasons only.

[440] Ferdinand Schoeman, "On Incapacitating the Dangerous," *American Philosophical Quarterly* 16 (1979): 27–35.

[441] Derk Pereboom, *Living Without Free Will* (Cambridge: Camb, 2001), chaps. 174–178.

never return". A certain indignation is also often felt when the author of a heinous crime is set free at the end of their sentence, let alone granting them all possible comforts during their period of detention, where prisons that are not overcrowded and dilapidated are transformed in public imagination into "five-stars hotels", and citizens wonder why so much money is spent on the offenders' comfort instead of on services for innocent citizens – or even just on lowering the taxes. This shows that incarceration is not fundamentally intended as a means for incapacitation.

Moreover, notes Pereboom, comfortable or not, taken in itself, this justification is insufficient, for either we quarantine offenders for the rest of their lives, or we will be back to square one the day they are released. Within the carceral paradigm, incapacitation lasts only for the duration of the sentence (although sentences tend to get longer and longer, especially in the U.S.). Paradoxically, it has been noticed how supposed incapacitation is actually transformed into "capacitation" within the prisons' walls, hence the discouraging results on specific deterrence mentioned previously. Worse still, in the current system, people often do not even return to square one after a prison stay, whereupon social ties are lost, job opportunities are reduced, and links with criminal organisations are created. It is way too common for incarcerated people to form new networks during their stay in prison; formerly incarcerated people are then pressured into re-entering old or exploiting new networks because of the stigma that accompanies their status, which makes it difficult for them to earn a living within the boundaries of legality. In some cases, during the period of incarceration, resentment towards the system grows and trust in the institutions is so undermined that life in lawfulness is not even considered an option anymore.

Hence, we are shifting into the third and last forward-looking justification for punishment: rehabilitation. Whether punishment works as a rehabilitative method is, yet again, an empirical question and the data disprove it. Not only has punishment been abandoned as an educational method for children as it has been found to be correlated with socioemotional problems and subsequent aggressive behaviour of its victims[442],

[442] Mary Keegan Eamon, "Antecedents and Socioemotional Consequences of Physical Punishment on Children in Two-Parent Families," *Child Abuse and Neglect* 6 (2001): 787–802. A recent meta-analysis found, more specifically, correlation with anxiety and depression, reduced cognitive abilities and, most notably for current purposes, aggressive and antisocial behaviour (in itself often correlated with mental health and cognitive problems): Elizabeth T. Gershoff, "More Harm than Good: A Summary of Scientific Research on the Intended and Unintended Effects of Corporal Punishment on Children," *Law & Contemporary Problems* 73, no. 2 (2021): 31–56. Similar results have been found by other studies, for example: John Clark, "Against the Corporal Punishment of Children," *Cambridge Journal of Education*

but, as we have just seen, recidivism rates of formerly incarcerated people are so incredibly high that one could hardly argue in favour of the rehabilitative function of prison, and we could extend this conclusion to other forms of punishment, too. Even when a formerly incarcerated person does not end up reoffending, their life conditions are frequently appalling; as Marina Bell noticed, if one ends up homeless, chronically ill, suffering from addiction, with no access to healthcare or drug addiction treatment, they would still qualify as "rehabilitated" according to the standard evaluation system[443]. However, a person living in these conditions is not less socially problematic than a recidivist offender in most cases.

Things are different where prisons are not so punitive in spirit as in Scandinavian countries, where recidivism rates are incomparably low. In "open prisons", most internal doors are open, and inmates enjoy a certain degree of freedom of movement; external doors are porous membranes that let external people and activities penetrate, and inmates leave on frequent furloughs. Personal responsibility is also encouraged in the organization of one's activities, in contrast with traditional "closed" prisons where inmates are infantilised. Education, vocational training, and job opportunities are also provided.

Economists Giovanni Mastrobuoni and Daniele Terlizzese argued for the difficulty in comparing traditional Western prisons to Scandinavian open prisons: the two systems are too different in many respects, and one cannot trace the different outcomes to the "openness" of the system. In fact, Scandinavian open prisons host roughly one-tenth of the population of an average American prison, inmates spend an average incarcerated period of one-tenth to one-fifth compared to their American counterparts, and governments spend three times more per inmate[444]. Moreover, open prisons coexist with traditional closed prisons in Scandinavian countries, and, their

---

34, no. 3 (2004): 363–71; Joan McCord, "Unintended Consequences of Punishment," *Pediatrics* 98, no. 4 (1996): 832–34. These studies refer to physical punishment; I could not find studies on other forms of punishment. Anyhow, except maybe for extreme cases of abuse, we could hardly find anything strictly comparable to long sentences of reclusion in interpersonal forms of punishment. Hence, the parallel remains partial and it does not allow to draw conclusions on carceral punishment administered to adults. However, both are (or are justified as) forms of reformatory punishment, i.e., voluntary inflictions of punishment for educational reasons, and the problems noted with respect to one can shed light on possible complications regarding the other. Specific studies on incarceration are needed to reach a definitive conclusion and they are not missing, as we will see in the next few lines.

[443] Bell, "Abolition: A New Paradigm for Reform," 41.

[444] Giovanni Mastrobuoni and Daniele Terlizzese, "Leave the Door Open? Prison Conditions and Recidivism," *American Economic Journal: Applied Economics* 14, no. 4 (2022): 201–2.

population being selected, their recidivism rates cannot be fruitfully compared with a random carceral population[445].

However, Mastrobuoni and Terlizzese found that an Italian open prison carries out strategies similar to Scandinavian ones, has comparable population size, length of sentences, and similar – if not lower – per-inmate expenditure compared to American prisons[446]. It is Bollate prison, on the outskirts of Milan, and its population, although being partly selected, had a significant share of randomly assigned inmates displaced from other Italian prisons because of overcrowding, so that results concerning that part of the population is immune to selection bias[447]. It has been observed that among this non-selected population, "spending more time in an open prison, and correspondingly less time in a traditional, closed one, reduces recidivism by a statistically significant and economically meaningful amount."[448] It must also be noted that since non-selected inmates have spent part of their sentence in a closed prison, often, when they arrive at Bollate, they are not offered many of the educational and working opportunities that are available to selected inmates. Their shorter stay in Bollate does not qualify them for the best training and "educational treatments" – as they are called in the Italian carceral jargon – there available. Selected inmates that access these special measures have even lower recidivism rates[449], but it cannot be known whether this is due to the opportunities provided or their independent propensity not to offend again.

Does all this tell us that proper punishment has rehabilitative effects, as claimed by supporters of this justification for punishment? I do not think so. What it tells us is that it is precisely what *is not punishment* – once again, the voluntary infliction of suffering – in incarceration in open prisons that is effective for rehabilitation and reintegration. It is not even proven that any form of incarceration is necessary for reducing recidivism, only that education and working opportunities do the job.

The backwards-looking justification corresponds to the Kantian stance that the fact of violating a rule is in itself a necessary and sufficient condition to punish, independently of any consequence of the punishment. It is the retributive justification

[445] Mastrobuoni and Terlizzese, 202.
[446] Mastrobuoni and Terlizzese, 202.
[447] Mastrobuoni and Terlizzese, 202.
[448] Mastrobuoni and Terlizzese, 224. More precisely, "an extra year spent in Bollate, as opposed to any of the prisons of origin, reduces recidivism by 10.5 pp, i.e., by about 27 percent of the average recidivism rate (39.6 percent). Mastrobuoni and Terlizzese, 221.
[449] Of 17 pp: Mastrobuoni and Terlizzese, "Leave the Door Open? Prison Conditions and Recidivism," 212.

for punishment that is my main object of criticism here, and that was treated in the previous section on retributive emotions. We have said, there, that the feeling of justice generated by retribution is the satisfaction of vengeful desires spawned by the emotion of anger and its correlates, like resentment. So, suppose the good that is achieved through retributive justice is this kind of satisfaction. In that case, retribution seems to be not so backwards-looking after all: the good that the practice of justice generates lies in its consequences (as is well observed by Girard, who, as we have seen, pragmatically considers the justice/vengeance of the State to be good because it puts an end to vendetta cycles among families). Then, the retributive theory is a consequentialist one. This would explain why the Kantian stance that the last murderer of a dissolving society would still need to be executed feels so inappropriate despite the widespread "feeling of justice" when retribution is applied in normal conditions.

If this is correct, then the theory must be assessed by weighing up the consequences inflicted on the punished party and the satisfaction that is gained by those who are wronged – whether we are referring here to the actual victims or to society as a whole, that is offended by the violation of its law. Without needing to perform a Benthamian *calculus felicificus*, it seems clear that that kind of satisfaction cannot be weighted against a death sentence – or a life sentence, or, more broadly, the acute suffering that is inflicted through incarceration. This is especially true in light of the consideration that nurturing these kinds of emotions cannot but bring misery to our lives, as we will see in the next chapter discussing Martha Nussbaum's *Anger and Forgiveness*. For now, it can be noted that survivors of violence themselves seem not to be very satisfied with the results of retributive justice procedures. Victoria Law, for example, remarks that 90% of survivors who are given the chance by the New York City-based organisation Common Justice to go through a restorative justice process together with their offender instead of subjecting them to traditional criminal proceedings choose to do so[450].

Gérard Courtois identifies another element that makes revenge achieve a positive outcome. According to him – and to Aristotle, to whom he refers – the question of retribution is one of honour[451]. Honour is the currency of exchange in social relations

[450] Victoria Law, *"Prisons Make Us Safer" And 20 Other Myths About Mass Incarceration* (Boston: Beacon Press, 2021), 168.

[451] Courtois, "Le Sens et La Valeur de La Vengeance, Chez Aristote et Sénèque."; Aristotle, *Rhetoric*, 1378a31– 33.

that are considered zero-sum games in which the rising of one implies the falling of another. When one is harmed, more than the physical reality of the damage, one needs to consider its moral dimension: it is one moral – or rather social – status that is primarily damaged. If one does not respond to an injury, they are giving a sign of weakness to the rest of society, and others may feel entitled to abuse them because they do not risk any harsh consequences. Vengeance, hence, is a means to re-establish one's status within society, signalling one's strength and, therefore, "respectability" – that is, nothing other than the ability to induce fear through expected anger and revenge in social relationships.

Of course, this theory loses its strength when retribution is not administered in a vindicatory system (that we discussed in the previous chapter) but in a State system where the State itself takes charge of revenge or "justice" – which motivates Courtois and Verdier's critique of the institution of the State as part of their political framework of right-wing anarchism, on which we shall not dwell. They advocate for citizens to take justice/vengeance back into their hands, disempowering the State. The presence of the State is, in fact, precisely what makes honour cultures obsolete: that action will be taken to respond to an offence does not depend on an individual's ability, connections, or what, in general, other people think of them. The State functions as an equaliser and takes care of everybody's defence, making Courtois' retributive theory anachronistic – if one does not subscribe to his broader political view.

Martha Nussbaum provides another critique of the honour-based justification for retribution (or "road of status", as she calls it). She observes that to consider harm as an offence to one's status implies "the tendency to see everything that happens as about oneself and one's own rank", something that is "very narcissistic, and ill-suited to a society in which reciprocity and justice are important values."[452] Even more importantly, she continues:

> It [this tendency] loses the sense that actions have intrinsic moral worth: that rape is bad because of the suffering it inflicts, and not because of the way it humiliates the friends of the victim. […] If wrongful injuries were primarily down-rankings, they could be rectified by the humiliation of the offender, and many people, certainly, believe something like this. But isn't this thought a red herring, diverting us from the reality of the victim's pain and trauma, which needs to be constructively addressed? All sorts of bad acts – murder, assault, theft – need to be addressed as the specific acts they are, and their victims (or the victims' families) need constructive attention; none of this

[452] Nussbaum, *Anger and Forgiveness*, 28.

will be likely to happen if one thinks of the offense as all about relative status rather than injury and pain.[453]

Despite all this, backwards-looking retributive justification is the justification that has dominated the history of the thinking on punishment. The explicitly consequentialist forward-looking justification, in fact, gained some traction only during the 18th Century criminal reform. However, it was just during the 1950s and the 1960s that it prevailed in the penal debate[454], remaining quite lively up until 1975 when data on the poor deterrent and rehabilitative effects of State-administered punishments generated the idea that "nothing works"[455]. This has led to a partial decline in the interest in utilitarian justifications for punishment, for the utility it earns does not seem sufficiently solid to justify the practice in utilitarian terms[456]. Foucault has remarked how the failure of the carceral system in terms of its effects was denounced since its very first establishment, dating appeals to prison reform back to the very birth of the prison institution.[457] However, this has not induced a questioning of the practice itself.

[453] Nussbaum, 28.

[454] R. Antony Duff and David Garland, "Introduction," in *A Reader on Punishment* (Oxford: Oxford University Press, 1994), 8.

[455] Jesper Ryberg, "Punishment and Desert-Adjusted Utilitarianism," in *Retributivism Has a Past: Has It a Future?*, ed. Michael Tonry (Oxford: Oxford University Press, 2011), 76; Duff and Garland, "Introduction," 8–16.

[456] It is interesting to question why there has been such a surge (before the subsequent decline) of the interest in consequentialist justifications for punishment given that another justification, the retributive one, was already available and has returned to the foreground since, although without completely eliminating the utilitarian justification in common thought. My intuition is that consequentialist thought, and utilitarianism in particular, could cloak a fundamentally emotional practice with an appearance of rationality that fits well with the idea of the State as a rational, efficient entity that manipulates other rational, predictable entities: humans. This effort has also taken the form of an "economisation" of life: through the measurement of market value, economics makes everything comparable and, hence, understandable so, at least apparently, rational. The standard model for the rational individual has become that of the *homo oeconomicus*. On the surface, this seems in line with a utilitarian approach, for utilitarianism seems to aim at putting any experience on a scale, to weight it, measure it, assign it a number. Moreover, a substantial part of political work consists in deciding on the allocation of the public budget, in justifying expenditure. A utilitarian view of punishment provides, on the surface, precise calculations that are vital to these justifications. The problem, as we have seen, arose when the calculations could no longer justify the practice in economic terms. Then, as this justification failed, but since it was not at the origin of the will to punish, it was set aside without questioning the practice itself, showing that utilitarianism, in this case, was only a cover to retributive impulses.

[457] "For the prison, in its reality and visible effects, was denounced at once as the great failure of penal justice. In a very strange way, the history of imprisonment does not obey a chronology in which one sees, in orderly succession, the establishment of a penality of detention, then the recognition of its failure; then the slow rise of projects of reform, seeming to culminate in the more or less coherent definition of penitentiary technique; then the implementation of this project; lastly, the recognition of its successes or its failure. There was in fact a telescoping or in any case a different distribution of these elements. And, just as the project of a corrective technique accompanied the principle of punitive detention, the critique of the prison and its methods appeared very early on, in those same years indeed, it was embodied in a number of formulations which - figures apart - are today repeated almost unchanged." Foucault, *Discipline and Punish*, 264–65.

Rather, appeals to retribution have grown and have unveiled the unchanged retributive core of the practice of punishment.

2.3.2 Retributive Fitness

Foucault identified the fitness of the carceral system with the socio-economic framework as a key motivation for its persistence: developed during the Industrial Revolution when disciplinary power was established, prisons fit perfectly the other disciplining institutions that then emerged – the factory, the school, the barrack, the hospital.[458] Likewise, Angela Davis introduced the concept of the prison-industrial complex – recalling the military-industrial complex – to describe the intertwining of the carceral system with economic interests in the contemporary world, where the building of prisons is not only a source of employment, both in its construction phase and in its operating phase[459], but also a reservoir of cheap labour force[460] as well as a pool of consumers that must be provided with all essential goods without having a choice for their purchases in the free market[461].

Both Foucault and Davis also exposed political reasons why the carceral system remains advantageous in spite of its poor results in terms of deterrence and rehabilitation. Foucault concludes his *Discipline and Punish* by claiming that prisons are delinquency generators that serve the purpose of making illegality manageable, depriving it of its political force, generating a sense of insecurity in the general population so that police surveillance is accepted and, under the guise of protecting against petty delinquents, the State can strike at political crime.[462] Davis highlights the political role that the prison system played in the U.S. in maintaining racialised exploitation after the abolition of slavery since, thanks to the "black codes" – laws that were only applicable to the black population and that replicated the "slaves codes" in force on Southern plantations – formerly enslaved people could be incarcerated even

[458] Foucault, 228.

[459] In Italy, two prison officers are hired every three convicts. Mastrobuoni and Terlizzese, "Leave the Door Open? Prison Conditions and Recidivism," 209.

[460] Although forced labour is not legal anymore, convicts that are allowed to work have no contractual power and face a blackmailing from the institutions, as working is considered as part of the so-called "rehabilitative treatment" and, therefore, refusing to work for "starvation wages" would undermine their possibilities to access furlough opportunities. Moreover, as always in cases of extreme poverty, not having access to other working opportunities, such wages remains better than no wage.

[461] Angela Y. Davis, *Are Prisons Obsolete?* (New York: Seven Stories Press, 2003), 88.

[462] Foucault, *Discipline and Punish*, 273–92.

without committing crimes[463], and then, thanks to the convict leasing system, be used as free labour force. All this has led to the racialisation of crime and, hence, the criminalisation of Afro-American communities placed under constant police surveillance.[464]

So, Foucault and Davis analysed the economic and political fitness of the carceral system. I have emphasised the emotional fitness, if we may call it that, of the carceral (but more generally punitive) system with the human, or animal in general, spontaneous response to harm and aggression. In the first part of this work, we have also stressed the fitness of retribution with the Christian religion, with its emphasis on the individual fate of the soul, its immortal life and, as we discussed at length, a retributive conception of justice to be accomplished in the afterlife[465]. Although philosophers and theologians did struggle on the topic of free will and divine grace, popular conceptions imagine each individual soul as free to choose the good or the bad path and expecting to "get what it deserves" after death, probably as a result of priests' and pastors' simplified preaching. The most prominent proof of the retributive character of Christianity is the invention of purgatory, as we have seen.

There is a fifth and last form of fitness – after the economic, political, emotional and religious ones – that is left to be analysed, which justifies the entrenchment of punitive systems in our societies. It is its fitness with a conceptual system that has been dominant in the West for many centuries, to say the least. It is a conceptual framework composed of the interconnected ideas of free will, merit, moral responsibility, and guilt.

The dominance of these ideas that substantiate the retributive conception of justice, in common as well as in philosophical thought, explain the resistance against pure utilitarian justifications for punishment – repeatedly accused of justifying the punishment of the innocent if that could be demonstrated to have the best consequences

[463] Black people, in many states, as a result of the black codes could be incarcerated for begging – and former slaves could hardly have a home to keep them off the streets.

[464] Davis, *Are Prisons Obsolete?*, 22–39.

[465] I have previously mentioned that ancient Judaism did not subscribe to a similar conception, considering death the end of both body and soul. Eastern religions like Buddhism and Hinduism, on the other hand, have a non-dual conception of the afterlife, the goal of spiritual practices being *nirvana*, that is, the cessation of pain and pleasure through the end of the *samsara*, the cycle of individual rebirth, that also corresponds to the end of individual life in the non-dual perspective. However, this does not mean that these religions avoid the temptation of retributivism altogether. In fact, the cycle of reincarnations is strictly retributive, and the popular notion of *karma*, which in common thought roughly means that one gets what one deserves, indicates precisely the positive or negative reincarnation that follows from a morally good or bad life, thus remaining linked to retributive conceptions of cosmic justice.

in terms of utility[466] – as well as aversion for determinist worldviews. For many, the problem of free will and determinism is not a problem of understanding whether the concept of free will actually indicates a certain feature of our experience, but to accept or reject the consequences of determinism on the ethical level, the rejection of free will having a disruptive power on individual moral responsibility, desert, and, therefore, retributive justice. We have already seen that this was one of the strategies put in place by Erasmus in his debate with Luther, and we have seen how a similar argument was made in recent years by Saul Smilansky[467], for whom hard determinism must be rejected in the moral practice because "we can make sense of the distinction between the guilty and the innocent, and need to follow this distinction. Not to do so would be not to respect persons; indeed it would be barbaric"[468].

Before Smilansky, the arguments already cited by Peter Strawson followed a similar line of thought, although more moderate. He claimed that, since reactive attitudes are intrinsic features of human nature, thus ineradicable, the truth or falsity of determinism – or the absence or existence of free will more in general[469] – should not have consequences on the moral level.[470] Therefore, tracing all actions to constitutive moral luck should not, in Strawson's opinion, undermine our practices of retributive justice grounded in inescapable moral sentiments. I will not resume here my criticism of this kind of moral intuitionism, a subject discussed at length in the dedicated paragraph. I will only add that it is against this hypothesis of Strawson[471] that Martha Nussbaum mentions the non-violent, non-angry, and non-resentful leaders of the revolutionary movements of the 20th century against racism and colonisation: Martin Luther King, Nelson Mandela, and Mahatma Gandhi[472]; this, as Nussbaum remarks, should raise questions on Strawson's empirical claim that resentment – a form of anger – really is an

[466] Henry John McCloskey, "A Note on Utilitarian Punishment," *Mind* 72, no. 288 (1963): 599; J. J. C. Smart, "An Outline of a System of Utilitarian Ethics," in *Utilitarianism. For and Against* (Cambridge/New York: Cambridge University Press, 1973), 71; Duff, "Responsibility, Restoration, and Retribution," 64.

[467] Smilansky, *Free Will and Illusion*.

[468] Saul Smilansky, "On Free Will and Ultimate Injustice," *Iyyun: The Jerusalem Philosophical Quarterly /* 52 *עיון: רבעון פילוסופי* (2003): 53.

[469] To deny the existence of free will, one one needs not engage in a determinist conception of the world, as I suggested in my paper "Libertà situata. La libertà oltre il libero arbitrio" (*Etica & Politica / Ethics & Politics*, forthcoming.) introducing the notion of pseudo-determinism.

[470] Strawson, "Freedom and Resentment."

[471] Nussbaum, *Anger and Forgiveness*, 14–15.

[472] Nussbaum, 211–46.

ineradicable sentiment essential for our moral life, and for "anything that *we* could find intelligible as a system of human relationships, a human society"[473].

However, it must be noted that the connection between free will, moral responsibility, and punishment has not always been so clear and defined. It is well known that Ancient Greeks had a "contaminating" or "polluting" notion of guilt that could be passed on to different generations: Oedipus could be blamed and punished for his father's guilt, with no reference to a supposed free will that he "misused" – as will be said in the Christian tradition – making him deserving of punishment. The two elements that drove this change in perspective are Christianity and the centralised State.

The former has been analysed at length. The establishment of centralised States, on the other hand, decisively strengthened the link between individual responsibility and justice. As we have seen in the first chapter, modern centralised States have less need for sacrifices – that explicitly target innocent victims – to deal with the problem of violence. Or, at least, they join sacrificial practices to practices of vengeance because they have the force to target culprits and make them the outlet of social hate and anger, thus concealing the partly sacrificial nature of the practice. It is no coincidence that the concepts of free will and moral responsibility became central in the philosophical debate of the modern age precisely at the time of the consolidation of the centralised State – i.e., modern times. From that moment on, retributive justice became a fitting part of the conceptual framework of common as well as philosophical thought, making it not only hard to challenge it in our practices but also to critically question it in theory.

2.3.3 Building the Conceptual Framework: Justice and Responsibility in Modern Philosophy

In the following pages, I will give a brief and non-exhaustive overview of the debate on free will, moral responsibility and justice in the works of the most influential philosophers between the 17th century to the very beginning of the 18th century, from Hobbes to Leibniz, passing through Descartes, Spinoza, and Malebranche. I chose to focus on these authors because it seems to me that they have shaped philosophical thinking on the subject, constituting the core arguments that have been either accepted or rejected by subsequent authors but not fundamentally changed. Of course, they were not the first and certainly not the last to address these topics. However, these issues gained unparalleled importance at the time as Europe was shaken by the religious wars

[473] Strawson, "Freedom and Resentment," 26.

and the main differences between the opposing sides, as we have seen in the first chapter discussing Luther and the Reformers, lay precisely in issues of justice and, relatedly, free will. In that context, one's philosophical stand implied a religious identity that enormously impacted one's social, political, and economic life, up to and including the possibility of death. It is unsurprising, therefore, that the philosophers of the time devoted unprecedented attention and care to these questions, which led them to forge the conceptual horizon in which the discussion still takes place today.

Hobbes

Hobbes's social contract theory for the origin and the function of the State as the means for managing human violence has been alluded to since the very beginning of this work. Let us recall it very briefly: its starting point is a state of nature, not to be understood as a historical period, but as a *condition* to which there is always the risk of relapsing – the condition of the absence of a central authority who has the power of coercion. It is the same condition as sovereign States among themselves, and that of the civil war that Hobbes witnessed in 17th-century Britain. Absent this authority, given the fundamental equality of power (or better, weakness) among human beings – everyone being exposed to the risk of being killed by anyone else – as well as their tendency to seek supremacy and the convergence of desire on indivisible goods in a situation of scarcity[474], human life in the state of nature is, famously, "solitary, poore, nasty, brutish, and short"[475]. Without centralised authority, in fact, everyone has a natural right to everything that they consider relevant to their survival, which equals saying that nobody has a right to anything, and humans live in a constant state of war of all against all[476]. On the other hand, the law of nature, which fundamentally prescribes what is apt to self-preservation, requires humans to do all they can to escape this miserable condition. Therefore, to reach peace, they subscribe to the social contract and form a civil State[477].

Through this *covenant*, rational individuals renounce their natural right over everything and delegate authority, that is, the power of establishing civil laws and enforcing them through violence, to a third, superior party, ideally an absolute

[474] Hobbes, "The Elements of Law," 133–35; Thomas Hobbes, "De Cive," in *Three-Text Edition of Thomas Hobbes's Political Theory. The Elements of Law, De Cive and Leviathan*, ed. Deborah Baumgold (Cambridge: Cambridge University Press, 2017), 133–35; Hobbes, "Leviathan," 133–35.

[475] Hobbes, "Leviathan," 138.

[476] Hobbes, "The Elements of Law," 136–38; Hobbes, "De Cive," 136–38; Hobbes, "Leviathan," 137–38.

[477] Hobbes, "The Elements of Law," 144–45; Hobbes, "De Cive," 145; Hobbes, "Leviathan," 145.

monarch[478], who is not bound by any civil law and the related threat of punishment. The monarch is only constrained by natural laws – that is, if their action does not grant their subject self-preservation, they pose themselves at risk of being deposed. In fact, by natural law, subjects are only bound by the covenant as long as this grants them better chances of self-preservation than the state of nature.

From this account, it should be clear that a central element of Hobbes's political proposal is the so-called "administration of justice" that, in the *Elements of Law* and the *De Cive*, he calls "the sword of justice"[479]: the right to punishment. The civil state, thanks to the threatening power of punishment, solves the problem of internecine violence: whereas, in the state of nature, fear elicits attack in anticipation[480] – because everyone is roughly equally powerful and therefore exposed to the risk of being killed by another, who, in turn, if successful, would be spared retaliation – in the civil state fear is an agent of deterrence: once a *super partes* and more powerful entity is introduced, one that would retaliate after the victim is killed, fear of this institutionalised retaliation prevents violence outbursts. Punishment is, therefore, a constitutive element of civil law, that would be "vain" if it simply consisted of a prohibition without violent enforcement[481]. Fear of punishment is what makes a covenant – in this case, the subjection to the monarch's law – binding, "[f]or who would lose the liberty that nature hath given him, of governing himself by his own will and power, if they feared not death in the retaining of it?"[482].

Hobbes subscribes to the idea, which will be strongly supported by Girard, as we shall see in greater detail, that there is no difference in the principle of vengeance and of justice meted out through punishment: as we have just seen, they are both acts of retaliation. In both the *Elements* and the *De Cive*, talk of punishment is infused with terms relating to vengeance – to the point that, in the latter, the law that defines punishments is termed "*vindicative* or *penal*"[483], equating the two. In fact, if this "sword of justice" is not appropriately administered by the magistrates, "people, grieved, by their injustice, should take upon them to make their own revenges, to the disturbance of

[478] Hobbes, "The Elements of Law," 220; Hobbes, "De Cive," 220–23.
[479] Hobbes, "The Elements of Law," 216; Hobbes, "De Cive," 216.
[480] Hobbes, "Leviathan," 134.
[481] Hobbes, "De Cive," 383, 406.
[482] Hobbes, "The Elements of Law," 155.
[483] Hobbes, "De Cive," 382.

the common peace"[484] – I have mentioned Girard's choice of avoiding any reference to Western philosophical thought, but the similarity with Hobbes's position here is so staggering that it cannot be ignored: for both authors, revenge must not be suppressed in the passage to the civil State, but it must be taken away from people's hand for it risks generating an escalation of violence, leading to the much-feared relapse into the war of all against all. Hence, like for Girard, the State's punishment is an act of revenge that puts an end to the cycles of violence, for it cannot be avenged in return – therefore, compared to private vengeance, it is justified on consequential grounds, that is, on its peace-making character.

Hobbes, too, identifies the crucial difference between punishment and vengeance in the former's forward-looking character[485]. In the *Elements*, punishment is defined as a "*revenge […] taken upon the consideration [not] only of the offence past, but of the benefit to come*"[486] . This benefit, however, is not only the disruption of the vendetta cycles Girard focuses on but also the "correction" of both the offender and society as a whole[487], anticipating the utilitarian justifications into which we have delved at length[488]. Only the *Leviathan* will be amended from the lexicon of vengeance, making it clear that "[p]unishment is not a revenge, but terrour"[489], which is why its terms must be precisely established before it is administered, its real function lying in its threat. This is why if a punishment more severe than the one previously known and expected is inflicted, it is no longer punishment: it is an "act of hostility"[490], cruelty.

Given the forward-looking character of punishment, it is clear that the problem of excessively harsh punishments results from their uselessness and not from a violation of the proportionality that should be established through desert – in fact, there is no such thing as intrinsic desert or merit in Hobbes's system. Although Hobbes does refer to the traditional conception of justice as *suum cuique tribuere* when discussing distributive justice, his interpretation is definitely unconventional. In his view, distributive justice is

[484] Hobbes, "The Elements of Law," 364.
[485] Even identified as a law of nature: Hobbes, "De Cive," 174.
[486] Hobbes, "The Elements of Law," 174.
[487] Hobbes, "Leviathan," 362.
[488] I find this passage of the *De Cive* especially reminiscent of the utilitarians' future stances: "We must therefore provide for our security, not by compacts, but by punishments; and there is then sufficient provision made, when there are so great punishments appointed for every injury, as apparently it prove a greater evil to have done it, than not to have done it. For all men, by a necessity of nature, choose that which to them appears to be the less evil." Hobbes, "De Cive," 215–16.
[489] Hobbes, "Leviathan," 383.
[490] Hobbes, 383.

the virtue of the "arbitrator" (i.e., the monarch). The peculiarity of the arbitrator's distribution as compared to the distribution that might take place among the subjects is that when the arbitrator gives each their due, this is not distributed according to a principle that precedes them and binds them: Hobbes's monarch has absolute power; therefore they are not submitted to the rules of desert; rather, through their distribution they *create* desert[491]. It is, thus, conventional with respect to punishment as well as to economic distribution or political rank.

The absence of intrinsic desert is consistent with Hobbes's strict determinism, which is made clear in his quarrel with Bishop Bramhall. I will not go through the many arguments put forward by the two parties in favour of free will or determinism, for it would go beyond the scope of this work. I will limit my remarks to the fact that, here again, a relevant place is dedicated to the practical question concerning punishment and reward, blame and praise, law and exhortation in a determinist world. From Hobbes's point of view, given his forward-looking interpretation of punishment and the related practices, determinism does not pose any problem[492]. According to Bramhall, on the other hand, the fact that Hobbes' denial of free will impinges on the retributive and backwards-looking character of blame and punishment is an argument against such conception.[493]

We have seen an analogous argumentation in Erasmus' diatribe with Luther. Obviously, both Erasmus and Bramhall hold a religious perspective in which the retributive character of punishment and blame is embedded in divine law. Hence, the

[491] "And Distributive Justice, the Justice of an Arbitrator; that is to say, the act of defining what is Just. Wherein, (being trusted by them that make him Arbitrator,) if he performe his Trust, he is said to distribute to every man his own: and this is indeed Just Distribution, and may be called (though improperly) Distributive Justice; but more properly Equity; which also is a Law of Nature, as shall be shewn in due place." Hobbes, 171. Does this mean that their action has no constraint? Not exactly, for the monarch is still subject to the laws of nature, which is why they should respect the natural equality of human through equity in their distribution, in order to avoid too much discontent that would lead to an uprising and, therefore, the loss of peace, the North Star of Hobbes's thought. For a more detailed analysis of Hobbes's conception of distributive justice, see: Johan Olsthoorn, "Hobbes on Justice, Propery Rights and Self-Ownership," *History of Political Thought* 36, no. 3 (2015): 477–79.

[492] See, for example: "The intention of the law is not to grieve the delinquent for that which is past and not to be undone, but to make him and others just that else would not be so; and respects not the evil act past but the good to come, insomuch as without this good intention of the future, no past act of a delinquent could justify his killing in the sight of God. But you will say, how is it just to kill one man to amend another if what were done were necessary? To this I answer that men are justly killed, not for that their actions are not necessitated, but because they are noxious, and those which are not noxious are spared and preserved." Hobbes and Bramhall, *Hobbes and Bramhall on Liberty and Necessity*, 25.

[493] See, for example: "[T]hough there were no other end of penalties inflicted, neither probatory nor castigatory nor exemplary, but only vindicatory, to satisfy the law out of a zeal of justice by giving to every one his own; yet the action is just and warrantable." Hobbes and Bramhall, 52.

respect of this postulate can make a theoretical argument. Although this religious premise has been dropped, theoretical reasoning not depending on truths of fate anymore, we have seen Smilansky take a similar conceptual path. He takes some supposed moral truths about retributive justice – the rejection of which, as we have seen, would be "barbaric" – as reasons not so much to refute the truth of determinism but to hide it and dismiss its moral consequences, revisiting the steps previously traced by Bramhall and Erasmus.

Descartes

Descartes probably constitutes a *unicum* in the philosophical panorama of his time because of his peculiar take on ethics. In the preface to the French edition of the *Principles of Philosophy*, "the highest and most perfect Ethics" is presented as "the final stage of Wisdom"[494]. In the famous analogy of the tree of philosophy, ethics is described as one of the branches:

> Philosophy as a whole is like a tree; of which the roots are Metaphysics, the trunk is Physics, and the branches emerging from this trunk are all the other branches of knowledge. These branches can be reduced to three principal ones, namely, Medicine, Mechanics, and Ethics.[495]

The three branches of Ethics, Medicine and Mechanics need to be rooted in the previous elaboration of metaphysics and physics, which hence constitute Descartes' primary subject of interest. The three branches also give "fruits"[496], namely, practical applications in everyday life, but that of ethics is indeed a special branch: thanks to the connection between ethics and epistemology, the fruits it gives bear the seeds of philosophy itself in the form of the correct use of reason and will – i.e., the highest virtue of the Cartesian system: generosity.

According to these passages, then, the massive amount of work dedicated by Descartes to metaphysics and physics should be understood as preliminary to the elaboration of his "highest and most perfect Ethics"[497]. However, Descartes never built the moral system he had announced in the preface. In fact, in his 1647 letter to Chanut, written in the same year as the French preface, he explicitly states his refusal to write

[494] René Descartes, *Principles of Philosophy*, *The Cambridge Descartes Lexicon* (Dordrecht/Boston/London: Kluwer Academic Publishers, 1982), xxiv.
[495] Descartes, xxiv.
[496] Descartes, xxiv.
[497] Descartes, xxiv.

about morality[498], which is only addressed in letters that he wished to remain private[499]. Two years before, in 1645, he wrote to Elisabeth that all was needed to know about morality was contained in the few rules exposed in the *Discourse on the Method*, written in 1637, ten years before *The Principles of Philosophy*, which should have constituted the foundation for the moral theory, were published. These rules are known as the "*morale par provision*", which essentially recommends following common-sense morality[500].

Although Lisa Shapiro has maintained, consistently with Descartes' claim in this letter to Elisabeth, that we should not interpret "*par provision*" as meaning "temporary" but as a juridical expression that indicates what will not be called into question in the final judgement, i.e., Descartes' definitive moral theory[501], such an interpretation is hard to follow. The reference to the elaboration of a new moral theory, rooted in the achievements in the fields of metaphysics and physics, comes way later than the exposition of the "*morale par provision*" and even of the letter to Elisabeth that refers to it, suggesting that, at the time of the writing of the French preface, Descartes believed the "perfect Ethics" to be achievable, although not yet accessible due to the shortcomings in the other sciences. But these can be expected to be filled and a philosophical moral theory to be developed as a consequence.

Chronological facts aside, Shapiro's suggestion is questionable because the "provisional morality" cannot be identified with "the final stage of Wisdom" rooted in metaphysics and physics. On the contrary, provisional morality is constituted primarily by customs, religion and laws, in how they instruct us on the correct behaviour to be held in society and, although sometimes they may be arbitrary, they are in general "so well established that anyone who follows them honestly […] leads a happier, more

[498] "It is true that normally I refuse to write down my thoughts concerning morality. I have two reasons for this. One is that there is no other subject in which malicious people can so readily find pretexts for vilifying me; and the other is that I believe only sovereigns, or those authorized by them, have the right to concern themselves with regulating the morals of other people." René Descartes, "To Chanut, 20 November 1647," in *The Philosophical Writings of Descartes. Volume III. The Correspondence*, ed. John Cottingham, trans. John Cottingham et al. (Cambridge: Cambridge University Press, 1991), 326.

[499] "But I beg you, if possible, to see that these writings do not fall into other hands." Descartes, 327.

[500] Stéphan Vaquero, "L'unité de La Philosophie Chez Descartes: Métaphysique et Topologie Morale," *Revue Philosophique de La France et de l'étranger* 134, no. 4 (2009): 478.

[501] Lisa Shapiro, "Descartes's Ethics," in *A Companion to Descartes*, ed. Janet Broughton and John Carriero (Hoboken, New Jersey: Wiley-Blackwell, 2010), 450.

secure life"[502] than those who do not. This means that the philosopher finds a ready-made base for their philosophical reflection, not that they should be satisfied with it.

Although, as we have mentioned, the "most perfect Ethics" will not be developed by Descartes, continual references to moral themes, that cannot be assembled into a coherent moral theory, are scattered all over in his works. His continual exploration of the subject might be motivated by his progress in his research in the domain of physics, where, in a letter to Chanut, he had claimed to have reached "sure foundations" for morals[503] and, as Frédéric de Buzon and Denis Kambouchner maintain, if the perfect Ethics, like method, "is concerned more with practice than with theory"[504], then "it is not impossible that Descartes, in his last years, had felt he was bringing morality to its maximum perfection"[505]. Whether this is true or not, the absence of a systematic moral theory testifies to a mistrust of Descartes on the subject, which, by the way, does not prevent us from analysing some of the many fascinating insights in moral philosophy that we find in his work, especially in his theories concerning will and freedom that constitute the foundation of morals.

Once again, I will limit my analysis to the elements that are relevant to this research. Although Descartes does not speak explicitly about justice or retribution, not being interested in a philosophical or moral analysis of the political system, his interest in free will leads him to touch on many of the notions that are conceptually related to retributive justice that I mentioned before, i.e., moral responsibility, praise and blame[506].

---

[502] "Les loix communes de la societé, lesquelles tendent toutes à se faire du bien les uns aux autres, ou du moins à ne point faire de mal, sont, ce me semble, si bien établies, que quiconque les suit franchement, sans aucune dissimulation ny artifice, mene une vie beaucoup plus heureuse et plus assurée, que ceux qui cherchent leur utilité par d'autres voyes […]." René Descartes, "Descartes à Elisabeth. [Egmond, Janvier 1646]," in *Œuvre de Descartes. Correspondance IV. Juillet 1643 - Avril 1647*, ed. Charles Adam and Paul Tannery (Paris: Librairie Philosophique J. Vrin, 1996), 357.

[503] "I must say in confidence that what little knowledge of physics I have tried to acquire has been a great help to me in establishing sure foundations in moral philosophy." René Descartes, "To Chanut, 15 June 1646," in *The Philosophical Writings of Descartes. Volume III. The Correspondence*, ed. John Cottingham, trans. John Cottingham et al. (Cambridge: Cambridge University Press, 1991), 289.

[504] René Descartes, "To Mersenne, 27 February 1637," in *The Philosophical Writings of Descartes. Volume III. The Correspondence*, ed. John Cottingham, trans. John Cottingham et al. (Cambridge: Cambridge University Press, 1991), 53.

[505] Frédéric de Buzon and Denis Kambouchner, "Descartes," in *Le Vocabulaire Des Philosophes. Tome 2: La Philosophie Classique et Moderne (XVIIe-XVIIIe Siècle)*, ed. Jean-Pierre Zarader (Paris: Ellipses, 2016), 53–54.

[506] From the perspective of retributive justice, I believe we can consider praise and blame as proxies for reward and punishment, for they are based on the same backwards-looking mechanism: just like a bad action means the agent deserves punishment, so they deserve blame – expressive retributive theories of punishment considering it precisely a form of public blame. The same can be said for good actions with respect to praise and reward.

Just like in Hobbes, and, as we will see, all modern philosophers, free will holds a central place in Descartes' ethical discussion. The topic is addressed mainly in two places: the Fourth Meditation[507], published in 1641, and the famous letter to Mesland, written in 1645[508]. In the Fourth Meditation, Descartes distinguishes two degrees of freedom: the higher consisting in the choice of what is presented to the will as clearly and distinctively good; the lower in indifference, when no reason drags the will in one particular direction, and one can indifferently follow any of the options. The first corresponds to a form of internal determinism that C. P. Ragland has effectively named "clear and distinct determinism"[509]; the latter seems to be a form of indeterminism that, in contemporary philosophy, is based on the principle of alternative possibilities. However, rather than a perfection, indeterminacy attests to a limit in knowledge, for the knowledge of truth and goodness impedes indifference[510].

The correct exercise of free will in both forms – when one follows what one clearly sees as true and good and when one remains indifferent in the case of epistemic balance – is, according to Descartes, meritorious, praiseworthy: it constitutes the highest virtue, generosity. Nevertheless, free will might be misused: this generates mistakes but also blameworthiness. The cartesian link between epistemology and ethics, in fact, makes our epistemic process the basis for praise and blame. To be mistaken on the epistemic level is equivalent to being guilty on the moral level. Note that Descartes does not simply label these actions as good or bad: the reference to praise and blame implies both an underlying conception of moral responsibility and a retributive attitude that remains not investigated.

Blameworthy mistakes can take the form of hasty judgements when one consents to perceptions that are not clear and distinct or does not recognize the epistemic balance that should lead to indifference. But they may also consist in denying one's assent to clear and distinct perceptions, deciding to follow the worst while one

[507] René Descartes, *Meditations on First Philosophy. With Selections from the Objections and Replies*, ed. and trans. John Cottingham (Cambridge: Cambridge University Press, 2013), 74–87.

[508] René Descartes, "To Mesland On Freedom (February 9, 1645)," in *Philosophical Essays and Correspondence*, ed. Roger Ariew (Cambridge/Indianapolis: Hackett Publishing Company, 2000), 273–74.

[509] C. P. Ragland, "Descartes on the Principle of Alternative Possibilities," *Journal of the History of Philosophy* 44, no. 3 (2006): 379.

[510] It also attests the disproportion between the limitedness of the intellect and the limitlessness of the will, for the latter can be directed towards anything, even what we do not understand: it does not need to be guided by an understanding of the true or the good and therefore one can make choices in situations of balance or indifference.

sees the good[511]. This latter case is discussed in the letter to Mesland. There, Descartes goes back to the concept of indifference, which he now considers twofold: on the one hand, it indicates "strictly speaking, […] that state in which the will finds itself, when it is not carried by the knowledge of what is true or what is good to follow one side rather than the other"[512], as was presented in the Fourth Meditation. But on the other hand, Descartes admits that it might also mean

> that positive faculty we have of determining one or the other of two contraries, that is, of pursuing or fleeing, of affirming or denying a particular thing. […] Indeed, I think it is there, not only whenever it is not at all carried by the weight of any argument to one side rather than to another, but even when this faculty is involved in all the other actions of the will, so that it is never determined without using that faculty. Thus, even when a very evident argument carries us to something, even though, *morally* speaking, it is difficult for us to do the contrary, nevertheless, speaking *absolutely*, we can do it.[513] For we are always free to prevent ourselves from pursuing a good clearly known to us or from admitting an evident truth, *provided only that we think it a good to witness in this way the freedom of our will*.[514]

It is not in one's power to perceive something as good or as true, as Descartes affirms in the Fifth Meditation: as long as one contemplates the nature of the triangle, one cannot but believe that the sum of its angles is equal to 180 degrees[515]: "Admittedly my nature is such that so long as I perceive something very clearly and distinctly I cannot but believe it to be true"[516]. However, our nature is also such that we cannot always focus at the same time on what we have previously perceived clearly and distinctively, so that we may doubt what we previously perceived as a truth – even a

[511] Andreea Mihali, "The Role of Freedom in Descartes' Ethics of Belief," *Pacific Philosophical Quarterly* 95, no. 2 (2014): 218.

[512] Descartes, "To Mesland On Freedom (February 9, 1645)," 273.

[513] The distinction between moral possibility and absolute possibility has caused many difficulties to interpreters. An interesting solution has been proposed by Brian Embry, that traced these concepts back to neo-scholastic terminology, according to which moral impossibility amounts to "astronomically low probability" (Brian Embry, "Descartes on Free Will and Moral Possibility," *Philosophy and Phenomenological Research* 96, no. 2 (2018): 391.). As Embry explains, neo-scholastic philosophy elaborated the concepts of moral and absolute possibility to justify the coexistence of two theories of human sin: the first contemplating the ability of the righteous to always avoid sin – that justifies God's punishment for the guilty and collides with Protestant theories; the second excluding the possibility of always avoiding sin – in contrast with Pelagian heretics that made free will all too powerful. The situation for neo-scholastics is indeed similar to that presented by Descartes, as in the first case they need to respect the principle of alternative possibilities and in the second they must comply with a form of determinism. Through the concept of moral impossibility, the neo-scholastics opt for a form of indeterminism: it is extremely unlikely, but still always possible, never to sin. According to Embry, the same goes for Descartes: it is extremely unlikely, but still always possible, to reject what is clearly perceived as good. The proposal is illuminating, but the final few lines of the passage suggest, in my opinion, a different interpretation.

[514] Descartes, 'To Mesland On Freedom (February 9, 1645)', p. 273, my emphasis.

[515] Descartes, *Meditations on First Philosophy. With Selections from the Objections and Replies*, 45.

[516] Descartes, 48.

mathematical truth – afterwards[517]. According to Descartes, it is the existence of God – proved in the Fifth Meditation through the so-called "Ontological Argument" – that grants the stability of truths, that allow us to continue considering true what we have once seen as clear and distinct[518].

But although clear and distinct ideas impose themselves on our intellect, freedom enters into play when the will gives or withdraws its consent to the representation. Contrary to the intellect, which is limited, meaning that we can only have clear and distinct knowledge of a limited pool of things, the will is infinite – a mark of our resemblance to God. To have an unlimited will means to possess that sort of "metaphysical independence"[519] that allows us to give or withdraw our consent to any representation: clear and distinct perceptions may incline our will, but they can never determine it. So much so that, as we have seen in the letter to Mesland, the will can give its consent to and choose precisely what it knows to be evil. The paradox of this stance is that this can be done only by those who know true goodness – which makes the choice of evil absurd in actual fact but relevant in the form of a possibility, even if abstract.

The final passage I emphasised in the quotation from the letter to Mesland, however, seems to me to turn the matter towards a deterministic interpretation of Descartes' thought. Although the excerpt begins with Descartes speaking of indifference, the clause "provided only that we think it a good to witness in this way the freedom of our will" clearly denies that the free agents find themselves in a state of indifference. On the contrary, agents face two perceived goods, one being the pursuit or affirmation of something clearly known as true or good, the other the proving of their free will against clear and distinct determinism. If one considers the ability to deny what is perceived as good to be a greater good, they are still following the rule of the pursuit of the greatest perceived good. Therefore, they are still deterministically driven, so the agent, even in this case, being able to deny what the intellect presents clearly and distinctly, does not seem to pursue one or the other option indifferently.

However, many passages of Descartes' work contradict this interpretation. A paramount example is found in *The Passions of the Soul*, where Descartes maintains that:

[517] Descartes, 48.
[518] Descartes, 48–49.
[519] de Buzon and Kambouchner, "Descartes," 73, my translation.

> I observe only one thing in us that can give us a good reason to esteem ourselves, namely, the use of our free will, and the mastery we have over our volitions. For we can be justifiably praised or blamed only for those actions that depend on this free will of ours, and it renders us in some sort similar to God by making us masters of ourselves, provided that we do not forfeit, by our own weakness, the rights it confers on us.[520]

This claim is clearly reminiscent of Pico della Mirandola's famous "*homo faber fortunae suae*". But more interestingly, it is somewhat reproposed in very similar terms in contemporary libertarian philosophy, as in Roderick Chisholm:

> If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute *only to God*: each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing – or no one – causes us to cause those events to happen.[521]

I note in passing that Chisholm raises this idea as part of the same argument advanced by Erasmus, Bramhall, and also Smilansky up to a certain extent: individual responsibility is an unquestionable premise, a moral truth, from which the truth of this problematic conception of free will descends. Descartes, on the other hand, does not advance that argument. However, he does not argue for moral responsibility either. Moral responsibility is postulated and taken to be descendent from the notion of free will elaborated, but that notion, as I tried to show, bears traces of inconsistency throughout his work and hardly justifies a strong notion of moral responsibility. Whether this depends on Descartes' decision to abstain, for the most part, from moral subjects or a tension between a philosophical elaboration of ethics and common-sense morality that he does not wish to challenge, the question is beyond the scope of this research.

The fact remains that, in Descartes' work, free will is the sole basis for the moral evaluation of actions and people in terms of blameworthiness and praiseworthiness – the retributive notions found in Descartes' thought – as it is the root of moral responsibility. But its notion is inconsistent: the moral edifice has its foundations in muddy soil. However, given Descartes' refusal to systematize his ethics, this should not come as a surprise. I found it relevant to highlight these tensions because it seems to me to be yet another demonstration of the conceptual difficulty – internal to the philosopher

[520] René Descartes, *The Passions of the Soul and Other Late Philosophical Writings*, ed. and trans. Michael Moriarty (Oxford: Oxford University Press, 2015), 258.

[521] Roderick Chisholm, 'Human Freedom and the Self', in *Free Will*, ed. by Gary Watson (Oxford: Oxford Readings in Philosophy, 1983), pp. 24–35 (p. 32, my emphasis).

or in his relation to the political world – to question the fundamental retributive structure of part of moral thought.

Spinoza

In Spinoza, the question of free will is, once again, a major centre of gravity for morality. Differently from Descartes, the possibility of its existence is ruled out since the very beginning of the *Ethics*, where determinism is clearly exposed, and it is stated that "The Will cannot be called a free cause; but only a necessary cause." (E1p32)[522]. Once again, the undermining of the notion of free will disrupts that of moral responsibility: Spinoza's complete silence on the subject is therefore explained, as well as his criticism of the concepts of merit and sin, just and unjust, praise and blame. These latter are said to be born from the illusion of free will in the appendix to the first part of the *Ethics*[523]. On the other hand, merit, sin, just and unjust are called "extrinsic notions" (E4p37s2)[524], concepts that do not have a meaning in themselves but are only relative to something else. The same goes for the basic moral concepts of good and evil, which are defined in part three as "that which satisfies any sort of desire" and "that which frustrates desire" (E3p39s[525]; E3p9s[526]). Nothing is good or bad in itself: things are defined as such as a consequence of the way they affect people. Although they arise from individual affects, an intersubjective, if not objective, moral dimension can be found through the similarities between people constituting a common human nature that makes humans generally share desires and aversions. On the other hand, the concepts of merit, sin, praise, blame, justice and injustice are grounded neither in some individual affect nor human nature but rather in an intersubjective convention: they exist only in the so-called "civil state".

Just like Hobbes, Spinoza distinguishes the "state of nature", a condition in which humans live fearful solitary lives, struggling for survival in the race for resources, and the "civil state", where humans have agreed to live peacefully together and overcome conflict and fear. However, unlike Hobbes, Spinoza does not consider the two states as opposite, one requiring the complete abandonment of the other since the same self-affirmation mechanism is operating in both. One is always moved by the same

[522] Baruch Spinoza, *Ethics*, ed. Clare Carlisle, trans. George Eliot (Princeton: Princeton University Press, 2020), 99.
[523] Spinoza, 105.
[524] Spinoza, 255.
[525] Spinoza, 191.
[526] Spinoza, 171.

basic impulse, the effort to impose their *conatus*, to persevere in their existence. The creation of the civil state does not change this basic functioning of the natural state; rather, it frames it in a more rational and effective way[527]. For Spinoza, as for Hobbes, peace is the primary necessary condition for human thriving[528]: the passage from the natural to the civil state is precisely one from a state of war to one of peace.

Spinoza swings between a contractarian rational theory of this transition in the *Theological-Political Treatise*, and an emotional irrational theory in the later *Political Treatise*. The two must not necessarily stand in opposition, for, as remarked in the *Ethics*, the same effects might be generated by both passions and reason (E4p59)[529]. While the rational contractarian theory can be considered a justification for the existence and the protection of the civil state, the emotional theory of Spinoza's last work investigates the actual mechanism of human aggregation, for it is rooted in concrete human nature and humans, as finite beings, are slaves of the passions, "envious" and "more prone to revenge than to mercy" (E4 §13[530], TP 1.5[531]).

Humans may look like "natural enemies"[532], but they need to cooperate. Instead of a rational contract, what unifies them is a common fear of loneliness – due to the scarcity of resources – but also, says Spinoza in an enigmatic passage of the *Political Treatise*, a common longing for vengeance (TP 3.9)[533], that we will investigate later in more detail. This evolution of the state of nature into the civil state is necessary[534]: it arises from the material conditions of human existence[535], following the laws of human nature, which hold that humans pursue their utility guided by fear and hope (TP 3.3)[536].

Like all natural laws, the laws of human nature are descriptive. Laws originally indicate the description of something's behaviour: such are the laws of physics, descriptive of the behaviour of bodies; likewise, we can formulate laws descriptive of

[527] Emilia Giancotti, "Introduzione," in *Trattato Teologico-Politico* (Torino: Einaudi, 2007), XIX.
[528] Giancotti, XXXI.
[529] Spinoza, *Ethics*, 268.
[530] Spinoza, 281.
[531] Baruch Spinoza, "The Political Treatise," in *The Collected Works of Spinoza. Volume II.*, ed. and trans. Edwin M. Curley (Princeton and Oxford: Princeton University Press, 2016), 505.
[532] Moira Gatens, "The Condition of Human Nature. Spinoza's Account of the Ground of Human Action in the Tractatus Politicus," in *Spinoza's Political Treatise. A Critical Guide*, ed. Yitzhak Y. Melamed and Hasana Sharp (Cambridge: Cambridge University Press, 2018), 49.
[533] Spinoza, "The Political Treatise," 521.
[534] Giancotti, "Introduzione," XIX; Spinoza, *Ethics*, 253–55, E4p37s2.
[535] Gatens, "The Condition of Human Nature. Spinoza's Account of the Ground of Human Action in the Tractatus Politicus," 49.
[536] Spinoza, "The Political Treatise," 518.

human behaviour[537]. In fact, as we have mentioned, the similarities between individual human beings enable the identification of a common human nature[538] that allows for abstraction and formulation of general descriptions. An example of a natural law of human behaviour is provided in the *Theological-Political Treatise*:

> [I]t's a universal law of human nature that no one neglects to pursue what he judges to be good, unless he hopes for a greater good, or fears a greater harm. Nor does anyone submit to any evil, except to avoid a greater one, or because he hopes for a greater good. Between two goods, each person chooses the one he judges to be greater; between two evils, the one which seems to him lesser.[539]

Of course, the result of the judgement will be different according to whether the judging agent is ignorant or wise: the ignorant might pursue an evil, judging it good (or a lesser good, or a greater evil), whereas the wise has the ability to discern good from evil properly, following the former and avoiding the latter. This is why, in the *Political Treatise*, Spinoza enunciates the "law of reason": "it is also a law of reason that we should choose the lesser of two evils" (TP 3.6)[540]. When the wise acts on the law of nature, that is transformed into a law of reason, for reason is perfectly aligned with nature. Ultimately, these laws do nothing but describe the *conatus* itself.

Divine laws revealed in the Scripture must also be understood in this way, as eternal truths. It was the prophets who misunderstood them and made them into commandments, prescriptions, and dispositions that led humans to imagine God as a king and a lawgiver[541]. On the contrary, they do not prescribe that vices should be punished; they just describe that sinners will suffer for their vices since vices are themselves sources of pain.[542]

Civil laws, on the other hand, are really prescriptions, but they are only introduced once the civil state is established; before that, there are no norms and, therefore, no sin[543], merit, justice or injustice[544], notions on which we shall come back.

---

[537] Baruch Spinoza, "Theological-Political Treatise," in *The Collected Works of Spinoza. Volume II.*, ed. and trans. Edwin M. Curley (Princeton and Oxford: Princeton University Press, 2016), 125.

[538] As noted by Moira Gatens, this is not in contrast with the statement that only individual essences exist, as declared in the *Ethics* (E2p40s1), for "human nature" is not to be considered as an existing being, but as an abstraction that groups individual essences thanks to their similarities. Gatens, "The Condition of Human Nature. Spinoza's Account of the Ground of Human Action in the Tractatus Politicus."

[539] Spinoza, "Theological-Political Treatise," 285.

[540] Spinoza, "The Political Treatise," 519.

[541] Spinoza, "Theological-Political Treatise," 133.

[542] As in the Proverb of Salomon "*the punishment of fools is foolishness*." Spinoza, 138.

[543] Spinoza, 283.

[544] Spinoza, 290. Although there being no prescriptive laws in the natural state, according to Spinoza there is still a "natural right", that consists of the right of everyone to "exist and have effects as it is

The formulation of these laws exploits the fundamental mechanism of human behaviour revealed by natural laws: rewards and punishments are delivered to induce certain actions, exploiting the human inclination to hope and fear and the human determination to follow what is judged good and avoid what is judged evil. We have here a forward-looking consequentialist justification for punishment: through punishment (most) violence can be prevented so that the state of war of the state of nature can be transformed into the state of peace of the civil state. So the primary role of the State is to regulate human interactions to reduce conflicts; the goal of the State is to free people from fear and let them enjoy their natural right to life[545]; the means to achieve this end lie in its

> […] right to decide what's good and what's evil, what's right and what's wrong, i.e., what must be done or omitted by each person, or by all together. We've seen, therefore, that the supreme 'powers alone have the right to make laws, and, when there's a dispute about them, to interpret them in each particular case, and to decide whether a given case was decided contrary to the law or in accordance with it […]. TP 4.1[546]

> […] Right to judge each person's deeds, to require each person to account for what he's done, to punish offenders, to settle disputes between citizens concerning the law, and to set up people knowledgeable in the laws, who will administer them in its place. TP 4.2[547]

The core of the State's action lies in what would now be called "the administration of Justice"[548]. As seen before, for Spinoza, that of justice is a conventional notion with no intrinsic moral significance. In the *Theological- Political Treatise* he gives the following definition: "justice is commonly defined as a constant

---

naturally determined to do" (Spinoza, 283.). It "prohibits nothing except what no one desires and what no one can do: not disputes, not hatreds, not anger, not deception. Without qualification, it is not averse to anything appetite urges." (Spinoza, 284.) – it is therefore not bound by abstract moral notions, like that of justice, but only by one's power. This natural right is not suppressed in the civil state, that is just an evolution of, and does not stand in opposition with, the natural state. The State, being much more powerful than each single individual, can impose its will on them through civil law, and thus contribute to the determination of one's action. However, as Spinoza remarks, obedience remains a choice (Spinoza, 297.) and one continue to "exist and have effects as it is naturally determined to do" (Spinoza, 283.) that is, maintaining their natural right. This means that, even if the State itself is not subjected to civil laws, thus appearing limitless in its exercise of power, limits are provided by the conservation of natural right. When its power is exercised arbitrarily, in fact, people are determined by their nature to rebel against it and the State consequently collapses. In fact, as we have seen, people submit to the State because they judge it a lesser evil than life in the state of nature. When this is not considered true anymore, subjects revolt and overthrow the state. There are, therefore, criteria of effectiveness that must be respect if States are to survive. Hence, natural right prevents States from turning permanently into tyrannies.

[545] Spinoza, "Theological-Political Treatise," 346.

[546] Spinoza, "The Political Treatise," 525.

[547] Spinoza, 525.

[548] In TP 10.9, Spinoza affirms that "the laws are the soul of the state. So if they are preserved, the state itself is necessarily preserved.." Spinoza, 600.

and perpetual will to give to everyone his due (*suum cuique tribuere*)."[549] Later in the same work, he repeats the same formulation of justice, adding a small specification: "Justice is a constancy of mind in apportioning to each person what belongs to him (*suum cuique tribuere*) according to civil law"[550]. Similarly, in the *Political Treatise*, Spinoza affirms:

> But in a state, where it's decided by a common Law what belongs to one person and what to another, a person is called just if he has a constant will to give to each person his own, and unjust if he tries to make his own what belongs to someone else. TP 2.23[551]

That of the *suum cuique tribuere* is a traditional definition of justice. It defines justice in distributive terms: it consists of the allocation of goods and evils in society according to what each deserves. Of course, such a definition is very vague, for it gives no indication as to how desert should be identified: what does one deserve for stealing a loaf of bread? Do they deserve to be charged the price of the load plus an additional "moral cost" that must be paid to the one who suffered the theft and to the society whose rules were broken? But how do we define the amount of the moral cost? Is it economic in nature? Or should it take the form of inflicted suffering? The definition leaves these questions unanswered. Furthermore, in Spinoza's case, the *suum cuique tribuere* is even more problematic for, however obscurely, the concept of desert is typically rooted in that of free will, which, as we know, is excluded from Spinoza's system.

Despite all these difficulties, Spinoza apparently takes this distributive stance for granted and does not specify it any further. This might indicate his indifference towards this notion, which typically belongs to political moralism; however, the recurrence of this definition in several places points to its importance and undermines this hypothesis. The lack of an explanation is problematic given Spinoza's consequentialist and positivistic perspective, in contrast with the traditional conceptions of justice that find intrinsic value in "giving everyone their due". To understand Spinoza's take on this notion, we must turn to his conception of civil law and its limits.

Although the State is not itself under the jurisdiction of civil law, natural right describes its range of possibilities of action if it is not to fall into ruin. Its laws cannot be

[549] Spinoza, "Theological-Political Treatise," 127.
[550] Spinoza, 290.
[551] Spinoza, "The Political Treatise," 516.

arbitrary and must make life in the civil state better than what subjects might expect it to be in the state of nature. So, while its laws are always *just*, some of them are better than others in that they are more efficient in achieving the State's goal and, therefore, keeping it in existence. Hence, the proposed criteria for justice do not classify a law as just; rather, they qualify it as more or less efficient and, thus, more or less just. More efficient and just laws are those that avoid conflicts in society, making it convenient for individuals to subject themselves to them.

André Campos Santos, in a subtle analysis, interprets Spinoza's criterion for justice as a criterion for equality[552] for, although humans are not equal in all respects, perceived inequality sparks conflicts. On this basis, Campos Santos formulates different interpretations of the *suum cuique* according to different theories of equality – an equality that is not observed and protected but rather one that is instituted and enforced. I will report the most relevant ones, i.e., ontological, legal, political, and productive equality.

The fostering of ontological equality is related to the State's duty to guarantee everyone's right to existence, the affirmation of one's own (*suum cuique*) *conatus*. Was this not secured in the civil state, there would be no reason for subjects not to overthrow it.

Legal equality is the equal subjection of everyone to the same sovereign authority. In this case, the *suum cuique* stands for the fact that no one has the right to rise over others, which would imply taking possession of something that is not theirs. When this kind of equality is violated, conflicts generated by the envious nature of man[553] arise, and the existence of the State itself is endangered.

Political equality is the equal participation of all in the legislative procedure so that the same individuals that are subjected to the civil laws (legal equality) are the ones that formulate them. Therefore, they do not need to obey any foreign authority but only themselves. Giving everyone this right as their own (*suum cuique*) makes the State more

[552] Andre Santos Campos, "Spinoza on Justice: Understanding the Suum Cuique," *Epoché* 21, no. 1 (2016): 127–43. Interestingly, equity and inequity are also used as synonyms for justice and injustice in the Theological-Political Treatise. Spinoza, "Theological-Political Treatise," 290.

[553] Spinoza, "The Political Treatise," 505, TP 1.5; Susan James, "Politically Mediated Affects. Envy in Spinoza's Tractatus Politicus," in *Spinoza's Political Treatise. A Critical Guide*, ed. Yitzhak Y. Melamed and Hasana Sharp (Cambridge: Cambridge University Press, 2021), 61–77.

stable, for it represses envy and its destructive effects towards the ruling class. Spinoza is, in fact, a supporter of democratic government[554].

Finally, productive equality establishes that the one who performs a certain action must be the same who undergoes its consequences – both positive and negative. Apparently, what productive equality tells is that everyone should get according to their work or, in other words, according to their merit. Campos Santos, hence, justifies Spinoza's denial of forgiveness and mercy[555] on this basis: a bad action seems to call for bad consequences on the agent, but this productive equality is violated by mercy that discounts the agent from these consequences. However, punishment is not a necessary consequence of a bad action, violated by mercy. As we have seen already, according to Spinoza, vices punish themselves without needing external intervention. Even though humans are "more prone to revenge than to mercy" (E4§13[556]; TP 1.5[557]), when speaking of punishments, we are talking about a convention that, in Spinoza's system, is hard to rationally justify on the retributive grounds proposed by this notion of productive equality.

Moreover, according to Spinoza, a person is only a "proximate cause"[558] of an action, that is, in reality, the result of an infinite chain of causes of which a person's will and decision to act are effects. The notion of productive equality introduces in the theory of the civil State a form of individual responsibility that is incompatible with Spinoza's broader theory of humans as finite modes of the infinite *Deus sive Natura*, in which all modes (finite beings) are ever intertwined. As we have seen, this kind of deterministic argument is often accused of undermining the foundations of moral systems. In fact, it is thought to provide a universal excuse that would make it wrong to punish wrongdoers, undermining the retributivist logic. Therefore, determinist positions are believed to bring along indiscriminate mercy.

Still, Spinoza explicitly rejects mercy, proving these assumptions false. This rejection is motivated by consequential considerations: the reason for punishment lies in public safety[559] and is based on the natural laws that guide human behaviour. The State

[554] Spinoza, "Theological-Political Treatise," 287; Giancotti, "Introduzione," XXIV.
[555] Spinoza, "Theological-Political Treatise," 336–37.
[556] Spinoza, *Ethics*, 281.
[557] Spinoza, "The Political Treatise," 505.
[558] Spinoza, "Theological-Political Treatise," 126.
[559] Spinoza, 336–38. As Edwin Curley remarks, "bad men are no less to be feared when they are bad by necessity. So we may properly defend ourselves against them by taking whatever punitive action is

must exploit the human inclination to fear and hope and the human responsiveness to pain and pleasure to minimize interpersonal conflicts. To forgive is to give up on these mechanisms, risking that a bad action would be easily considered favourable, therefore enhancing the possibility for violence and crime to spread within society[560]: it is the classic deterrence-based justification. Hence, mercy threatens the existence of the civil State and jeopardizes everyone's safety. It would thus be irrational to exercise piety towards the individual while being impious towards the State[561]: real piety and charity are only to be found in what fosters social peace and harmony (TP 3.10[562]). So even if the origin of guilt is just ignorance, the guilty must nonetheless be punished – in fact, the civil State must enforce its law "not by reason, which is unable to restrain passion […] but by threats" (EIVp37s2)[563]. As Spinoza claims in a letter to Oldenburg*: "*Indeed, men can be excusable, and nevertheless lack blessedness and suffer in many ways.*"*[564]

While this is the rational justification proposed by Spinoza for the practice of punishment, Campos Santos is analysing the criteria that make the State more stable and efficient in its administration. It, therefore, needs to respond to the irrational instances of people guided by their passions. As we have seen, both reason and passions can lead to the same effects, so while the rational justification for punishment stands on consequentialist grounds, the passionate justification relies on hate and longing for vengeance[565], passions rooted in sadness, that is, in ignorance, that are therefore based on confused ideas such as that of individual moral responsibility. From this perspective, the retributivist stance proposed by the notion of productive equality makes sense.

In the first chapter of this work, I have already highlighted the importance of the longing for vengeance in the escalation of violence and the genesis of the State. In a brief and undeveloped passage recalled earlier, Spinoza, too, recognizes this role. Revenge is defined in the third part of the *Ethics* as "the effort to return an injury"

---

necessary". Edwin M. Curley, "Spinoza's Moral Philosophy," in *Spinoza. A Collection of Critical Essays*, ed. Marjorie Grene (Garden City: Anchor Press/Doubleday, 1973), 373.

[560] Spinoza, "Theological-Political Treatise," 177–78.

[561] Spinoza, 336.

[562] Spinoza, "The Political Treatise," 521.

[563] Spinoza, *Ethics*, 254.

[564] Baruch Spinoza, "Letter 78 (OP). To the Most Noble and Learned Mr. Henry Oldenburg From B. d. S.," in *The Collected Works of Spinoza. Volume II.*, ed. and trans. Edwin M. Curley (Princeton and Oxford: Princeton University Press, 2016), 480.

[565] See also: Chantal Jaquet, "Longing (Desiderium) for Vengeance as the Foundation of the Commonwealth," in *Spinoza's Political Treatise. A Critical Guide*, ed. Yitzhak Y. Melamed and Hasana Sharp (Cambridge: Cambridge University Press, 2018), 91.

(E3p40c2s)[566], and its common longing due to a common offence is identified in the *Political Treatise* as one of the passions that lead to the formation of the State[567]. The association devoted to avenging a common offence is a league, a criminal conspiracy. As observed by Chantal Jaquet, this comparison between the State and a league[568] makes the umpteenth step forward in the desacralisation of the State in Spinoza's work[569], which definitely distances itself from a rational social contract theory[570].

Spinoza's main explanation for the origin of the State remains that of a spontaneous association due to the reciprocal benefit humans find in association: Spinoza's famous theory of the *homo homini Deus* (EIVp35s)[571], according to which nothing is more useful to man than what is most similar to him, that is, the other man (EIVp31, 35)[572]. Even when considering this retributivist alternative just mentioned, he finally maintains that, should this longing for vengeance be satisfied, what impedes the league-State ever to dissolve remains the fear of loneliness[573] – a justification that complements the one based on the usefulness of association. Association is, in fact, so crucial in Spinoza's system because of the fundamental weakness of the human being (and of any individual whatsoever). This is clearly stated in the axiom of the fourth part of the *Ethics*: "There is no individual thing in nature which may not have another above it stronger and more powerful. Rather, anything whatever being given, there is also given another thing more powerful by which the former may be destroyed." (EIVa)[574]. It is because of this fundamental weakness that humans need association.

Nevertheless, the intuition that a shared longing for vengeance might do the same is illuminating for understanding the core function of the State, which was previously identified in the administration of justice[575]. For, even though the rational

[566] Spinoza, *Ethics*, 193.

[567] Spinoza, "The Political Treatise," 521, TP 3.9.

[568] "[The State] is a complex conspiracy that distinguishes itself from a simple league not by a difference in kind but by a difference of degree." Jaquet, "Longing (Desiderium) for Vengeance as the Foundation of the Commonwealth," 83.

[569] Jaquet, 83.

[570] Jaquet, 82.

[571] Spinoza, *Ethics*, 250.

[572] Spinoza, 246, 250.

[573] Jaquet, "Longing (Desiderium) for Vengeance as the Foundation of the Commonwealth," 83.

[574] Spinoza, *Ethics*, 230.

[575] It is also worth noticing that the opposition between a civil state and a state of nature is fictive, the state of war of all against all in nature is an imaginary construction of the Modern era, natural forms of associations always occur even in the absence of a State. It is therefore not the fear of loneliness that holds people together, rather, it is the fear of the escalation of violence due to private vengeance. This passage of the TP that refers to a longing for vengeance as a dangerous as well as unifying inclination shows that Spinoza already had an intuition in this direction.

justification for the "practices of justice", and for punishment in particular, seems to lie on consequential grounds, as we have seen we may attain the same effects whether moved by reason or by the passions. Whereas the former answers Hart's questions, "What justifies men in punishing? Why is it morally good or morally permissible for them to punish?" the latter rather answers, "Why do men in fact punish?"[576]. So the consequentially justified practice of punishment in the State's system of justice finds its main cause in a shared desire for vengeance: creating the State, every individual gives up their own right to vengeance and delegates it to the State (E4p37s2)[577], which can perform the desired task more effectively.

Jaquet compares this Spinozian theory to Locke's passages on the right to punishment possessed in the state of nature[578]. Contrary to Spinoza, Locke based his analysis on the existence of normative "laws of nature", the violation of which gives the offended the right to respond retributively to the offender. Once the State is established, this right is alienated from the individual and is passed to the State itself, which becomes "the avenger of unpunished offences"[579]. But, Jaquet adds, contrary to Spinoza, "Locke is not attempting to show that political society is founded on a longing for vengeance, but rather on a desire for justice, which is not the same thing."[580] However, she does not elaborate on the difference that should distinguish the two.

A difference might be identified in that revenge is a passion, an affect of sorrow, a derivative of hate (E3def37)[581], while justice, if it is to be rational, is an action that follows a utility calculation. Still, this difference remains subtle, and it is not in itself suitable for application to Locke's theory. In fact, speaking of retributive justice in the state of nature, Locke does not justify retributive practices on consequential grounds or in any other way: his conception of justice does not really seem distinguishable from vengeance. Like Spinoza, Locke, too, degrades the State to an avenger with no real appeal to superior values.

A merit of Spinoza's analysis is that this identity between vengeance and practices of justice is not hidden by moralist talk – whatever Girard might have said

[576] Hart, *Punishment and Responsibility*, 74.
[577] Spinoza, *Ethics*, 254.
[578] John Locke, *Second Treatise of Government; and a Letter Concerning Toleration*, ed. Mark Goldie (Oxford: Oxford University Press, 2016), 6.
[579] Jaquet, "Longing (Desiderium) for Vengeance as the Foundation of the Commonwealth," 81.
[580] Jaquet, 81.
[581] Spinoza, *Ethics*, 218.

about the history of Western philosophy and its sacrificial character. It is also justified in non-retributivist consequential terms. These two features of the discourse on punishment, retribution and utility, remain dominant in the contemporary debate, the justifications for punishment being still far from settled.

Malebranche

Besides being a philosopher, Malebranche was a theologian and a priest: God, grace, divine justice, grace and free will are central themes in Malebranche's moral philosophy, which, on the other hand, is not very concerned with the political realm. His reflection's starting point is his dissatisfaction with Descartes' solution to the mind-body problem through the pineal gland. Malebranche advocated for a stricter form of dualism, establishing a "metaphysical distance between mind and body"[582] that cannot interact. To explain how perception (the impression left by the external world in our mind) and wilful action (the operation in the external world motivated by an act of will) are possible, Malebranche turns, hence, to occasionalism, putting God centre stage.

According to occasionalism, God is the only efficacious being in the world: nothing else is endowed with causal efficacy[583]. When a ball strikes another ball, the former does not really move the latter: its movement is just the *occasion* for God's intervention to communicate motion to the latter ball that will therefore move. This does not mean that everything that happens in nature requires a specific and miraculous intervention, making the laws of nature susceptible to constant annihilation by God's arbitrary action. Admittedly, God almost never[584] acts through particular volitions: his wisdom is expressed by the simplicity of his means and not solely, nor pre-eminently, by the effects of his action. Simplicity is a sign of perfection[585]: particular volitions are not simple; therefore, they are not "worthy of God". To act in the simplest possible way, God must act through general volitions; that is, at the time of creation, he must have established general laws that govern the world, the beings that inhabit it, and their relationships and evolution – both in the realm of nature and in the realm of grace. Once

[582] Susan Pepper-Bates, *Nicolas Malebranche: Freedom in an Occasionalist World* (London: Continuum, 2009), 55.

[583] Nicolas Malebranche, *Treatise on Nature and Grace*, ed. Patrick Riley (Oxford: Oxford University Press, 1992), 138.

[584] Malebranche cannot entirely rule out the possibility of divine miracles, for they are narrated in the Scriptures. Still, every time something inexplicable happens, he suggests to consider that there might be a law that explains the phenomenon, that is unknown to us. Malebranche, 197.

[585] Malebranche, 116.

laws have been established, their unfolding does not require any additional intervention from God.

If everything is the product of this unfolding, then God seems responsible for all the evil in the world – both physical, as in the birth of malformed babies or natural catastrophes, and moral, as in crime, sin, or damnation[586]. Thus, Malebranche must develop what Patrick Riley named a "proto-*Théodicée*"[587] and defend God from these allegations. He does so with the theory of general volitions: in general volitions, God does not will a particular outcome, such as the birth of a monstrous baby; rather, he wills the laws that govern the world because they are fruitful, even though they might have fallouts, negative externalities, undesired bad consequences. As Andrew Pessin puts it, general volitions are incomplete[588]: they do not aim at all the consequences they bear. This does not mean that the all-powerful God does not foresee those effects or *could not* act otherwise to prevent them. He could do anything, but, as Patrick Riley said, "[w]hat he can do is simply a question of power; what he actually wills is a question of wisdom and justice."[589] Hence, he limits his otherwise unlimited power to be consistent with his wisdom, which requires him to act through general volitions, even though this means that evil will be created. The world God creates is the *best*; this does not necessarily make it *good* from a human point of view[590]: he loves his own wisdom more than he loves humanity[591], and his purposes do not lie outside of him since he is self-sufficient; hence human suffering can be justified.

In his *Treatise of Nature and Grace*, Malebranche claims that one of his aims is "to make God lovable to men and to make it understood that there is nothing harsh and unjust in the conduct he follows in the establishment of his Church"[592] – an aim, again, strikingly similar to Leibniz's in his *Theodicy*. However, this goal is hardly obtained: concerning nature, the Malebranchian God allows monstrosity and catastrophes in order to respect his wisdom; concerning grace, the lovable God of Malebranche allegedly wants to save all, but not so much as to give up his *modus operandi*, the simplicity of

[586] Malebranche, 118.
[587] Patrick Riley, "Malebranche's Moral Philosophy: Divine and Human Justice," in *The Cambridge Companion to Malebranche*, ed. Steven Nadler (Cambridge: Cambridge University Press, 2000), 259.
[588] Andrew Pessin, "Malebranche's Doctrine of Freedom/Consent and the Incompleteness of God's Volitions," *British Journal for the History of Philosophy* 8, no. 1 (2000): 22.
[589] Riley, "Malebranche's Moral Philosophy: Divine and Human Justice," 233.
[590] Riley, 247.
[591] Riley, 225.
[592] Malebranche, *Treatise on Nature and Grace*, 110.

his ways, and hence allows for widespread damnation: human salvation is not God's goal – that would be a form of anthropocentrism and anthropomorphising of God. Rather, it is his means for establishing the eternal Church, guided by Christ in the afterlife, that is dedicated to his own glorification.

Does this make God unjust towards men? Justice and order, claims Donald Rutherford, require "a balancing of merit and treatment"[593]: in fact, as stated in the *Treatise on Ethics*, God loves his creatures proportionally to their merits, and distributes rewards and punishments accordingly[594]. This is the essence of justice, which lies in objective ratios of perfection, universally valid and comparable to mathematical ratios[595]. Here again, the retributive conception of justice is assumed and not analysed or justified. So how can God make innocents suffer, violating this fundamental moral law of justice as proportionality, and still be just? To answer this question, the Christian caveat of human unworthiness comes to the aid, and Malebranche repeats the Lutheran justification based on the original sin. Since Adam's sin made all inherently sinners, no one suffers undeservedly; grace is a gratuitous gift of God, and no one has a right to salvation that God would be violating.

Once the responsibilities of God have been ascertained, it is time to determine those of humans. Through the theory of general and incomplete volitions, God has been spared from the responsibility of evil, but this has also led to a strict determinism, somewhat similar to the one proposed by Spinoza, ruling out human freedom and responsibility. However, contrary to Spinoza, Malebranche's starting point is theological and needs to comply with the teachings of the Scriptures. As stated in the *Reflexions sur la prémotion physique*, a theologian must avoid the doctrines that would

> clash with one of these dogmas [the ones established by the Church], above all a dogma as essential as that of freedom. For if freedom is rejected, God is made the author of sin, someone unjust and cruel; in a word, this heresy overturns all religion and all morality […][596].

[593] Donald Rutherford, "Malebranche's Theodicy," in *The Cambridge Companion to Malebranche*, ed. Steven Nadler (Cambridge: Cambridge University Press, 2000), 176.

[594] Nicolas Malebranche, *Treatise on Ethics*, ed. and trans. Craig Walton (Dordrecht: Springer Science+Business Media, 1993), 48.

[595] Malebranche, 47.

[596] Malebranche, *Reflexions sur la prémotion physique*, in *Oeuvres Complètes*, 16:27-8, as quoted in Elmer J. Kremer, "Malebranche on Human Freedom," in *The Cambridge Companion to Malebranche*, ed. Steven Nadler (Cambridge: Cambridge University Press, 2000), 195.

However, the accusation of overturning morality need not be religious: we mentioned it earlier when discussing Spinoza's approach to mercy and forgiveness, and it is still strongly present in the modern debate, as we have seen discussing Smilansky.

Malebranche's starting point is that humans ought to be free: grace, to be efficacious, needs cooperation[597]. Not only that, it is precisely grace that gives freedom back to humans after they lost it because of original sin. The human will has a natural inclination towards the good, impressed by God because he inclined humans towards himself, the greatest good. Malebranche even defines it as the "natural and continual movement of the soul towards the good in general"[598]. However, this inclination predisposes us to be attracted by all kinds of goods, even particular goods that give us immediate pleasure.

Adam, who was perfectly free, meaning that he was not a slave of his desires and attention, needed only the light of his reason to recognise the true good and love it. After the fall, humans lost this power over their attention and became slaves of concupiscence: the attraction to immediate goods made "[t]he heart of man […] always the slave of pleasure"[599]. Malebranche is no stoic: pleasure is good in itself, and it is the constitutive element of happiness. However, the pleasures of the senses can be deceptive, for pleasure always excites love for its cause and bodily pleasures induce humans to love powerless creatures, apparent goods that only make us impermanently happy.[600]

Grace rebalances the human will: it instils into it the internal delectation that makes one love God as a way to counterbalance bodily ephemeral pleasures[601]. As we have seen, this does not mean that God gives each and every individual what they need in order to be freed from the bondage of concupiscence: God's providence unravels according to general laws, not particular volitions, which means that it sometimes falls "on hardened hearts"[602], just like sometimes "rain falls in the sea where it is useless"[603].

[597] Malebranche, *Treatise on Ethics*, 76.
[598] Malebranche, *Treatise on Nature and Grace*, 170.
[599] Malebranche, 150.
[600] Nicolas Malebranche, "Traité de l'amour de Dieu," in *Œuvres T.2*, ed. Geneviève Rodis-Lewis (Paris: Gallimard, 1992), 1051.
[601] Malebranche, *Treatise on Nature and Grace*, 151.
[602] Malebranche, 129.
[603] Malebranche, 129.

However, even when grace reaches "prepared grounds"[604], it does not determine the human will to love God. As we have mentioned, the whole point of the gift of grace is to restore lost freedom to human beings. How are humans, then, free and responsible in an occasionalist world? To solve this problem, Malebranche introduces the notion of "immanent acts"[605] that the human will exercises on itself in order to give or suspend its consent to its inclinations. While the fact of our actual pursuit (or not) of particular goods is determined by the laws established by God, according to Malebranche, we can still be responsible for this conduct, thanks to immanent acts.

With an immanent act, we can recognize the limitedness of particular goods and withhold our will to strive for the general good meritoriously; thanks to the action of grace, our natural inclination is not invincible[606]. Still, this suspension of our will, in which human freedom lies[607], does not imply "anything physical", for, as we have seen, the radical dualism supported by Malebranche impedes any interaction between the mind and the physical world. To physically impede us from pursuing the particular good towards which we have withheld our consent, God must intervene with his general law. Hence, there is a correlation – not causation – between our mental and physical acts established by a supernatural law decreed by God.

When we put this freedom into practice, withholding our consent in front of particular goods, we behave deservingly: this is the essence of merit. When we make this the general rule of our action, we become *just*[608]. Following this path is difficult, even painful[609], but pain, according to Malebranche, is precisely what makes virtue deserving, while pleasure is its reward[610]. So, humans are born sinners due to the original sin, and if they are to be deservingly saved eventually, they will need to suffer.

As stated in the *Treatise on Ethics*: "It is in order to merit heaven, that we are living on earth"[611]. So suffering is administered during earthly life as a punishment for sin[612], but accounts will only be definitely settled in the afterlife, where humans will

[604] Malebranche, 129.
[605] Pepper-Bates, *Nicolas Malebranche: Freedom in an Occasionalist World*, 105.
[606] Malebranche, *Treatise on Ethics*, 89; Malebranche, *Treatise on Nature and Grace*, 170.
[607] Malebranche, *Treatise on Ethics*, 83–85.
[608] Malebranche, 66.
[609] Malebranche, *Treatise on Nature and Grace*, 173.
[610] Malebranche, *Treatise on Ethics*, 51.
[611] Malebranche, 183.
[612] Malebranche, 51.

find their final reward or punishment[613]. Malebranche's theory seems somewhat inconsistent: at the same time, he claims that, because of the original sin, no one is innocent, and so no one deserves salvation; however, salvation can be earned by that minority on whom "the rain of grace" falls profitably. Most individuals, hence, are excluded from the possibility of salvation because they do not have the freedom of the will that was lost with Adam's sin, and they are unlucky in that they do not receive the grace that is needed to soften their "hardened hearts" – for which they bear no guilt, in a form of constitutive (bad) luck. Nonetheless, Malebranche's conception of God's justice consists of the distribution, during this life and in the afterlife, of rewards and punishments according to the merit or demerit of each: what the inherent demerit is of the misfortune of not having the kind of grace needed to restore one's free will because of God's action through general volitions is, evidently, perplexing.

This inconsistency is due to Malebranche's attempt to reconcile the commonsense notion that we are in control of our destiny and that we have the power to change it – what we have previously called "the just-world fallacy" – with the religious claim that we depend on an all-powerful God. The same problem must be addressed by lay philosophy: natural laws described by hard sciences in the contemporary world can replace the eternal laws of the Malebranchian God. The difficulties, in terms of human responsibility and retributive justice, are the same. Malebranche is an example of the failure of this reconciliation: to affirm one requires letting go of the other, as the first Protestants courageously did.

Leibniz

Justice and the correlated themes of free will and responsibility are central topics in Leibniz's thought, recurring in many of his writings from his earliest to his latest works. If, during a juvenile phase, he endorsed Trasymachian conception of justice as the "advantage of the strongest"[614], he rejected it soon after. This conception, embraced by Hobbes, empties the notion of justice from its ordinary moral meaning, making it a convention that might justify the most horrible abuses, he later declared. Moreover, in

[613] Malebranche, 202. According to Malebranche, we should try to imitate God, treating people according to their merit, even if we cannot know it precisely (Malebranche, 176.). This might be a guideline for human justice, that is traditionally inspired by divine justice, and is a rare reference to political matters in his works.

[614] See Gianfranco Mormino, *Determinismo e Utilitarismo Nella Teodicea Di Leibniz* (Milano: FrancoAngeli, 2005), 27.

his *Meditation on the Common Concept of Justice*, he claimed that defining justice in this way is to change the meaning of words:

> [T]o say that "just" is whatever pleases the most powerful is nothing else than saying that there is no certain and determined justice which keeps one from doing whatever he wants to do and can do with impunity, however evil it may be. Thus, treason, assassination, poisonings, torture of the innocent, all will be just, if they succeed. This is, indeed, to change the nature of terms and to speak a language different from that of other men.[615]

Already in his 1670-1671 *Juris et Aequi Elementa*, justice is transformed into an eternal truth comparable to mathematical truths. Justice is, in fact, an element of the theory of right, which, like mathematics, is a science that does not depend on experiments, but on definitions[616]: not empirical but theoretical.

Like Malebranche, Leibniz rejects Descartes' conception of God's omnipotence as unrestricted arbitrariness, for the independence of values is the foundation of merit and moral worth and makes God a good master instead of a tyrant. As written in the *Meditation*, the theory according to which

> whatever God wills is good and just […] because God wills it […] would destroy the justice of God. For why praise him because he acts according to justice, if the notion of justice, in his case, adds nothing to that of action? And to say *stat pro ratione voluntas*, my will takes the place of reason, is properly the motto of a tyrant.[617]

Arbitrariness is one of the great enemies of Leibniz's thought. It is excluded from God's plans, which are the fruit of choices based on values external to God himself, and also from human action that, as we shall see, is never generated by a situation of indifference. According to Leibniz, this does not undermine God's omnipotence. On the contrary, Leibniz endorses a quite radical form of divine omnipotence, according to which, being God the primary cause of everything, he is also the endpoint of the chain of causes that determines human action and will.[618]

All this does not imply a refusal of the existence of free will. Instead, Leibniz embraces a compatibilist (Hobbesian) conception of free will as the ability to do what

---

[615] Gottfried Wilhelm Leibniz, "Meditation on the Common Concept of Justice (c. 1702-3)," in *Political Writings*, ed. Patrick Riley (Cambridge: Cambridge University Press, 1988), 47.

[616] Gottfried Wilhelm Leibniz, "Juris et Aequi Elementa," in *Mittheilungen Aus Leibnizens Ungerdruckten Schriften*, ed. Georg Mollat (Leipzig: H. Haessel, 1893), 21–22.

[617] Leibniz, "Meditation on the Common Concept of Justice (c. 1702-3)," 45–46.

[618] Gottfried Wilhelm Leibniz, "On the Omnipotence and Omniscience of God and the Freedom of Man (1670–71?)," in *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, ed. Robert C. Sleigh (New Haven and London: Yale University Press, 2006), 4–25.

one wills and will what one deems good[619]. Free will is most definitely not a form of indifference: it does not lie in the capacity to will one's own will, something that is not only chimerical but even undesirable, for it would reduce each choice to chance.

Freedom, Leibniz says in the *Theodicy*,

> consists in *intelligence*, which involves a clear knowledge of the object of deliberation, in *spontaneity*, whereby we determine, and in *contingency*, that is, in the exclusion of logical or metaphysical necessity.[620]

Let us briefly analyse each of the three components. The criterion of *intelligence or knowledge* implies, conversely, that there is no freedom in a state of ignorance[621]. The criterion of *spontaneity or self-determination*, as should be clear by now, does not imply a form of self-origination *tout court*: it only requires that the principle of action be in oneself, without requiring that it be generated in the self and by the self out of nothing, without further reasons. It is a criterion against coercion[622]:

> free choice is contrary to constraint. This is why, if someone is pulled by force into a house, we shall not say he went there freely. But when those who are at sea in a storm throw their cargo overboard in order to make the ship lighter and save themselves, one can doubt if they acted freely. And I respond yes, for they were not compelled to save themselves if they did not want to.[623]

The impulse towards saving oneself is sometimes considered a form of compulsion. The interest in our own saving seems to be inscribed in our biology, and it is very hard to fight against it. In contemporary jurisprudence, it is, in fact, often considered a condition for acquitting defendants on trial. A famous case is that of the Mignonette[624], a sailboat that shipwrecked in 1884 with four crewmembers onboard. The four managed to board the lifeboat, but, after nineteen days, they were at risk of

---

[619] Leibniz, 26.

[620] Gottfried Wilhelm Leibniz, *Theodicy. Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*, ed. Austin Farrer, trans. E. M. Huggard (Charleston: BiblioBazaar, 2007), 306, my emphasis.

[621] Leibniz had maintained this idea since at least the beginning of the 1680s, when he had written the short essay *On Free Choice*, where he gives the example of a young man kept in the dark about his own affairs by his guardians: "he is denied free administration of his goods, although they have been placed in his hands.", concludes Leibniz. Gottfried Wilhelm Leibniz, "On Free Choice (Between Summer 1678 and Winter 1680–81?)," in *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, ed. Robert C. Sleigh (New Haven and London: Yale University Press, 2006), 135.

[622] On coercion see also Gottfried Wilhelm Leibniz, "On Freedom, ca. 1679," in *Philosophical Papers and Letters*, ed. Leroy E. Loemker (Dordrecht/Boston/London: Kluwer Academic Publishers, 1989), 263.

[623] Leibniz, "On Free Choice (Between Summer 1678 and Winter 1680–81?)," 135.

[624] For a philosophical anlysis of the Mignonette case, refer to Sandel, *Justice. What's the Right Thing to Do?*, 31–33.

starvation; one had drunk seawater and was sick; two crewmembers decided to kill him and feed on him to avoid death. Once rescued and returned to land, they were trialled and sentenced to death. The judges declared:

> It must not be supposed that in refusing to admit temptation to be an excuse for crime it is forgotten how terrible the temptation was; how awful the suffering; how hard in such trials to keep the judgment straight and the conduct pure. We are often compelled to set up standards we cannot reach ourselves, and to lay down rules which we could not ourselves satisfy. But a man has no right to declare temptation to be an excuse, though he might himself have yielded to it, nor allow compassion for the criminal to change or weaken in any manner the legal definition of the crime.[625]

Yet, the sentence was later commuted to six months of imprisonment: apparently, the fact that unreachable standards were set does, in fact, play a role in judgements of freedom, responsibility, and, as we shall see, duties. But considering the choice between life and death as a form of free and autonomous choice is coherent with Leibniz's thoughts on freedom. Freedom is not about the possibility of doing otherwise that, even in a non-deterministic perspective, would seem negligible in the analysed cases. Whatever the ultimate origin, the causal root, or the invincible strength of the impulse towards saving oneself, the principle of the action lies within the subject. Therefore, spontaneity is attained.

Lastly, the criterion of *contingency or absence of metaphysical necessity*. At first glance, it would appear to contradict what has just been said, namely, the unimportance of alternative possibilities. But, with this criterion, Leibniz does not rule out all kinds of necessity or determination. To understand this point, we must grasp the Leibnizian notions of necessity and contingency. "Necessary" is defined in *The Confession of a Philosopher* as "the opposite of which implies a contradiction or cannot be clearly conceived"[626]. "Contingent" is "not necessary"[627], therefore the opposite of which does not imply a contradiction. For things to be contingent, it is only required that their opposite be conceivable without implying that there is a chance that it might actually exist. Hence, the world, according to Leibniz, is dominated by strict determination despite facts being contingent, for their opposite remains conceivable even though they

[625] David C. Brody, James R. Acker, and Wayne A. Logan, eds., "The Queen v. Dudlet and Stephens, 14 Q.B.D. 273 (1884)," in *Criminal Law* (Gaithersbourg, Maryland: Aspen Publishers Inc., 2001), 23.

[626] Gottfried Wilhelm Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," in *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, ed. Robert C. Sleigh (New Haven and London: Yale University Press, 2006), 53.

[627] Leibniz, 55.

are generated by a chain of causes and effects: "nothing can come to pass unless there be a cause so disposed as to produce the effect, this taking place no less in voluntary than in all other actions"[628], claims Leibniz in the *Theodicy*. In fact, God's prescience is untouched by the denial of metaphysical necessity: its foundation lies in "the nature of things"[629], given which events necessarily unfold. This is true for voluntary actions as well since people are "predetermined by [their] preceding state"[630] to the extent that Leibniz defines them as "spiritual automata"[631].

The mechanism that governs human actions is already described in the early letter to Wedderkopf (1671):

> Consider Pilate, who is damned. Why? Because he lacks faith. Why does he lack it? Because he lacked the will to pay heed. Why does he lack this? Because he did not understand the necessity of the matter (the usefulness of paying heed). Why did he not understand? Because the causes of understanding were lacking. For everything must be reduced to some reason, which process cannot stop until it reaches a primary reason, or it must be admitted that something can exist without a reason for existing; but if this were admitted, the demonstration of the existence of God and of many other philosophical theorems would be destroyed.[632]

Therefore, the criterion of contingency does not in any way transform free will into "freedom of indifference", of which Leibniz remains a firm adversary[633].

The necessity that ties the world into a net of causes and effects is called by Leibniz "hypothetical necessity": what happens in the world is not necessary *per se*, or by its own definition, its opposite (that is, not to exist for something that does exist) being, in fact, conceivable. Yet, it becomes necessary given the hypothesis of the existence of its ultimate cause, namely, God[634], that is metaphysically necessary. Leibniz has, therefore, established a logical difference between these two kinds of determination, which is best understood on the level of the analysis of propositions.

Necessary (absolutely or metaphysically speaking) propositions are demonstrated by an analysis of their terms, and their truth value is based on the

---

[628] Leibniz, *Theodicy*, 127.
[629] Leibniz, 148.
[630] Leibniz, 152.
[631] Leibniz, 369.
[632] Gottfried Wilhelm Leibniz, "Letter to Magnus Wedderkopf, 1671," in *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, ed. Robert C. Sleigh (New Haven and London: Yale University Press, 2006), 3.
[633] In the *Theodicy*, Leibniz compares Descartes' idea of the knowledge of freedom of indifference through inner feelings to a magnetic needle that could feel pleasure in turning towards the North, and therefore thought to do it independently of any causes. Leibniz, *Theodicy*, 154.
[634] Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 57.

principle of contradiction. On the other hand, a contingent proposition can only be understood *a posteriori*, by experience, or by reason if the *principle of sufficient reason* is applied, that is, assuming that what does exist is what had more reasons to exist[635]. In the workings of human will, this means that people always choose on the basis of a prevalent reason, the only way to resist one being to forget it[636], even when we are not aware of them, as it often happens[637].

Contingent truths have reasons that incline without necessitating (absolutely speaking). This only means that reasons do not make the opposite of the contingent truth contradictory[638] , and what will happen can only be known *a posteriori*. It does not mean that, all things being equal, things could have been different. In fact, despite the logical distinction, on practical grounds, "the absolute necessity of the sequence of causes would in this matter add nothing to the infallible certainty of a hypothetical necessity."[639], as remarked by Leibniz himself.

As observed by Mormino, necessity and contingency are, therefore, properties of the object, not of the agent[640]. Contingency, in particular, provides a requirement for human fallibility more than for human freedom[641]. Yet, at the same time, Leibniz states in the *Theodicy*:

> It is true that Adam was determined to sin in consequence of certain prevailing inclinations: but this determination destroys neither contingency nor freedom. Moreover, the certain determination to sin which exists in man does not deprive him of the power to avoid sinning (*absolutely speaking*[642]) or, since he does sin, prevent him from being guilty and deserving punishment.[643]

---

[635] Leibniz, "On Freedom, ca. 1679," 264–65; Gottfried Wilhelm Leibniz, "On Freedom and Possibility (1680-82?)," in *Philosophical Essays*, ed. Roger Ariew and Daniel Garber (Indianapolis & Cambridge: Hackett Publishing Company, 1989), 19; Gottfried Wilhelm Leibniz, "On Contingency (1686?)," in *Philosophical Essays*, ed. Roger Ariew and Daniel Garber (Indianapolis & Cambridge: Hackett Publishing Company, 1989), 28–29.

[636] Leibniz, "On Freedom and Possibility (1680-82?)," 19.

[637] Leibniz, *Theodicy*, 146.

[638] They do not make it contradictory *in itself*, that is, contradictory in the sense of a round square. The opposite of what exist is still conceivable. But in the absence of a prevalent reason, they do become contradictory in the concrete situation, in the real as opposed to possible world.

[639] Leibniz, *Theodicy*, 163.

[640] Mormino, *Determinismo e Utilitarismo Nella Teodicea Di Leibniz*, 109.

[641] Mormino, 111.

[642] The English translation reads "speaking generally", whereas the original French edition reads "*absolument parlant*" (Gottfried Wilhelm Leibniz, *Essais De Théodicée, Sur La Bonté De Dieu, La Liberté De L'Homme, Et L'Origine Du Mal* (Amsterdam: Mortier, 1714), 566.): the choice of the word "absolutely" seems relevant to me, as I make clear in the following paragraph, so I decided to translate literally "absolutely speaking".

[643] Leibniz, *Theodicy*, 350–51, my emphasis.

It is not necessary to sin *absolutely speaking*, meaning there is no absolute/metaphysical/logical necessity, but only *hypothetical*, that is, one based on the hypothesis of the causes: this bond to a cause makes necessity not "absolute", that is, not "unbound". But nothing (except for God) is ever unbound; something considered "absolutely" is a mere abstraction.

Passages like the one just quoted seem to reveal that the true target of Leibniz's discord is not the absolute necessity of Hobbes and Spinoza. Rather, he wishes to refute the Augustinian and Protestant doom to sin. As remarked by Mormino, this latter is much more insidious, for it impedes moral improvement; strict causal necessity, on the other hand, is precisely what is needed for growth and amelioration[644]. Therefore, according to Leibniz, this necessity does not stand in contrast to free will. Rather, it is one of its constitutive elements.

This conception of hypothetical necessity, strikingly similar to that of absolute necessity, forces Leibniz to face the two main problems of divine determination, already seen in Bramhall's criticism of Hobbes and in Malebranche's dealing with occasionalism, namely, the attack on human responsibility and the cruelty of God in creating evil. As to whether people are excused for their actions and denied responsibility, Leibniz's answer is negative. Like Malebranche, and Luther before him, Leibniz endorses a form of compatibilism according to which the origin of responsibility relies on will, whatever its origin. So, even though the will can be traced back to God through a series of reasons[645], people deserve praise and blame, punishment and reward, on the sole basis of their free (i.e., not coerced) will. Actions are determined by the will, which is inclined by "true or apparent reasons of goodness"[646], being therefore governed by a sort of final causality, where the end that moves the will consists in what is perceived as good.

Still, it is undeniable that we often fail to pursue what is best. This, according to Leibniz, happens because what is bad is sometimes hidden behind an apparent good[647]. In *The Confession of a Philosopher*, Leibniz was already claiming that "no one voluntarily makes himself evil"[648], and in the *Theodicy* he affirms:

---

[644] Mormino, *Determinismo e Utilitarismo Nella Teodicea Di Leibniz*, 176.

[645] Leibniz, "On Contingency (1686?)," 28; Leibniz, "Letter to Magnus Wedderkopf, 1671," 3.

[646] Gottfried Wilhelm Leibniz, "De libertate et gratia", cited in Mormino, *Determinismo e Utilitarismo Nella Teodicea Di Leibniz*, 109.

[647] Leibniz, *Theodicy*, 223.

[648] Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 77.

> In the schools they teach constantly that as the true is the object of the understanding, so the good is the object of the will. So likewise they teach that as the understanding can never affirm anything save that which is shown to it under the semblance of truth, the will can never love anything which to it does not appear to be good. One never believes the false as such, and one never loves evil as evil. There is in the understanding a natural determination towards the true in general, and towards each individual truth clearly recognized. There is in the will a natural determination towards good in general; whence many philosophers conclude that from the moment when individual goods are clearly recognized by us we are of necessity compelled to love them. […] A very clear recognition of the best determines the will; but it does not necessitate it, properly speaking.[649]

Hence, sin is a consequence of error, but this does not mean that it can be excused[650]. The opinions that incline the will depend on perception, which in turn depends on both the dispositions of the perceived object and the temperament of the perceiving subject[651]: "But we do not form our ideas because we will to do so, they form themselves within us, they form themselves through us, not in consequence of our will, but in accordance with our nature and that of things."[652] Human nature is indeed limited: as we are limited in our bodies, so we are in our minds; "I am not free to fly like an eagle nor to swim like a dolphin because my body lacks the necessary equipment. Something similar can be said about our mind."[653] But this only means that we are fallible, not doomed.

The mechanisms underlying human will analysed this far contribute to the understanding of God's will as well. He, too, chooses according to "reasons of goodness", but, being perfect, he cannot make any mistakes. Therefore, he is compelled to what is best[654] and, in this very compulsion, lies his perfect freedom[655]. In his intellect, there are infinite possible "series", a concept introduced in the *Confession*[656], that will evolve into that of "world" in the *Theodicy*[657], and that indicates the connection of all things. Since all things are connected, and they all trace back to God, they can really be traced back to one single divine act of volition: to the single choice with which

[649] Leibniz, *Theodicy*, 317.
[650] Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 73.
[651] Leibniz, 39.
[652] Leibniz, *Theodicy*, 369.
[653] Gottfried Wilhelm Leibniz, "Dialogue on Human Freedom and the Origin of Evil (1695)," in *Philosophical Essays*, ed. Roger Ariew and Daniel Garber (Indianapolis & Cambridge: Hackett Publishing Company, 1989), 112.
[654] Leibniz, "Letter to Magnus Wedderkopf, 1671," 3; Gottfried Wilhelm Leibniz, "On the Necessity of Choosing the Best (1677?)," in *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, ed. Robert C. Sleigh (New Haven and London: Yale University Press, 2006), 138–39.
[655] Leibniz, "On the Necessity of Choosing the Best (1677?)."
[656] Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 41.
[657] Leibniz, *Theodicy*, 131.

God chose this series[658], and since he could only choose what is best, it is, therefore, *a posteriori* demonstrated that we live in the best of all possible worlds. This world, or series, more than a choice, is the necessary consequence of the existence of God[659]. And if the world we live in is the best possible world, it is also *a posteriori* demonstrated that no world without sins and evil could ever exist.

This brings us to the second problem of divine omnipotence: if all that exists is the fruit of the will of God, and if sin exists, does it mean that God, willing the existence of sin, is evil? Is God responsible for evil? Again, Leibniz gives a negative answer to this question: God does not want sins and evils in themselves but as a means for a greater good. "Why, therefore, does he permit what he can prevent? Here I do not see a better response than that given by Saint Augustine, that God would not permit evil unless good could be made even out of evil."[660]

This means that sins are not "wanted" or "willed" by God, properly speaking. Rather, they are "permitted": "*To permit* is neither to will in favor nor to will against, and nevertheless to know. *To be the author* is by one's will to be the ground of something else."[661] God knows, he foresees, that sins will be a consequence of his creation, and, in particular, of the creation of any single individual (sinner) whose existence he does will[662]. And if, as previously seen, responsibility takes its roots in the will, by not directly willing sin but only permitting it as a consequence of his creation, God is cleared, with a move reminiscent of the Malebranchian theory of incomplete volitions. On the other hand, humans who want sin, because their will was created with an inclination toward evil[663] or because evil is hidden behind a good, are not. Their will, directed towards sin because of ignorance (that can be traced back to God), makes them the authors of their sin.

Given the assumption of God's existence and supreme goodness and the actual existence of a world where evil exists, it is *a posteriori* demonstrated that the existence of such a world is preferable to non-existence. According to Leibniz, a utilitarian calculus made God create such a world in spite of all the evil that creation brought with

[658] Leibniz, 171.
[659] Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 41.
[660] Gottfried Wilhelm Leibniz, *On Freedom, Fate, and the Grace of God (1686)*, ed. Shane Duarte, n.d., 10.
[661] Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 55.
[662] Leibniz, *On Freedom, Fate, and the Grace of God (1686)*, 10.
[663] Leibniz, *Theodicy*, 184.

itself. In fact, the existence of beatitude ("the chosen happiness […] by which gods were united with mortals"[664]) largely compensates for the existence of misery[665], even though, as he notices in the *Theodicy*, the great majority of humans are destined to damnation[666].

Evils are, thus, part of God's plan as means for a greater good. They might be a direct cause of something good that justifies them, but they might also be "means for a greater good" meaning that they are a necessary consequence of existence and existence is the necessary condition for the greatest of goods, namely, beatitude. To these utilitarian considerations, Leibniz joins some aesthetic considerations:

> It's a bit like what happens in music and painting, for shadows and dissonances truly enhance the other parts, and the wise author of such works derives such a great benefit for the total perfection of the work from these particular imperfections that it is much better to make a place for them than to attempt to do without them. Thus, we must believe that God would not have allowed sin, nor would he have created things he knows will sin, if he could not derive from them a good incomparably greater than the resulting evil.[667]

Like shadows in a painting please our eye when seen in connection with the greater picture, so evils, sins, and suffering must please God, that sees the greatest picture, namely, the whole world. This is a controversial statement that taints Leibniz's theory of divine justice. God's criterion in identifying the perfect world is, in fact, that

[664] Leibniz, *On Freedom, Fate, and the Grace of God (1686)*, 22.

[665] Leibniz, 21–23.

[666] Leibniz, *Theodicy*, 61. Leibniz's God, therefore, assumes a detached look on his creation, and the philosopher imitates him, therefore not understanding the existential scope of the complaint against a miserable existence, as remarked by Mormino (Mormino, *Determinismo e Utilitarismo Nella Teodicea Di Leibniz*, 99.). It is absurd, says Leibniz in the *Confession*, to lament one's destiny of damnation, for it equals a complain about one's overall existence: "For to ask why this soul rather than another is subjected from the beginning to these circumstances of time and place (from which the entire series of life, death, salvation, or damnation arises), and why, consequently, it passes from one set of circumstances to others – the series of things external to itself bringing things forth in this manner – is to ask why this soul is this soul. Imagine that another soul began to exist in *this* same body (that is, a body located at the same time and place) at the same time and place as that in which this one had begun; then this very soul that you call another will not be another but will be *this one*. If someone is indignant that he was not born of a queen, and, on the other hand, that his own mother did not give birth to a king, then he is indignant that he himself is not another. Indeed, he is indignant at nothing at all (…). You see to what extent men vex themselves with vain questions. It is as though someone of half-noble birth were irritated with his father because he had married a woman unequal in rank (although men do not lack similar feelings, or even more foolish ones), not thinking that if his father had married someone else, not he, but some other man, would have come into the world." Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 105.

[667] Leibniz, "Dialogue on Human Freedom and the Origin of Evil (1695)," 115.

of *harmony*[668] , and this notion plays a crucial role in the definition and practice of justice.

An essential element of Leibniz's reflections on justice is, as mentioned above, that of punishment, to the extent that Leibniz defines "injustice" as that which deserves punishment ("whether the pain will actually follow or not"[669]). Before being a human practice, that of punishment is a divine one:

> God hates sins, not in the sense that he cannot bear the sight of them, as we cannot bear the sight of what we detest – otherwise he would eliminate them – but because he punishes them. Taken together with punishment or atonement, sins are good, i.e., harmonious. For there is no harmony except as a result of contraries.[670]

Justice is here considered a form of ethical balance, a ratio[671], reminding us that the realm of justice is one of eternal truths, like mathematical truths. In fact, there is no peculiar divine decree determining specific punishments for individual sinners. Rather, as stated in the *Theodicy*, the balance is maintained by a natural connection between sin and punishment[672], both in the afterlife and earthly life. In fact, the world created by God has a "twist towards the good"[673] that grants harmony between the reign of nature (the work of the God-architect) and the reign of grace (the work of the God-monarch), leading nature towards grace[674].

To the accusation that we hardly ever experience this kind of justice in earthly life since we see sinners prosper and the righteous succumb to their abuse, Leibniz replies by appealing to another natural law. It is a natural law, he says, for the strong to prevail over the weak, like a rock would break a ceramic vase: it is foolish to think that God would change such a law for the pleasure of men[675]. When harmony is not attained in earthly life, it only means that justice will be served after death.

These precepts must not be limited to divine justice: they must guide the administration of human justice, which is different from that of God only in degree[676], to the extent that, in the *Principles of Nature and Grace Based on Reason*, the image of

---

[668] Leibniz, "Letter to Magnus Wedderkopf, 1671," 3.
[669] Leibniz, "Meditation on the Common Concept of Justice (c. 1702-3)," 47.
[670] Leibniz, "Letter to Magnus Wedderkopf, 1671," 5.
[671] Leibniz, "Juris et Aequi Elementa," 22.
[672] Leibniz, *Theodicy*, 166.
[673] Mormino, *Determinismo e Utilitarismo Nella Teodicea Di Leibniz*, 125, my translation.
[674] Gottfried Wilhelm Leibniz, "Principles of Nature and Grace, Based on Reason (1714)," in *Philosophical Essays*, ed. Roger Ariew and Daniel Garber (Indianapolis & Cambridge: Hackett Publishing Company, 1989), 212.
[675] Leibniz, *Theodicy*, 259.
[676] Leibniz, "Meditation on the Common Concept of Justice (c. 1702-3)," 45.

the perfect State is not one where people do not commit any crime, but one in which all crimes are punished[677].

However, and incoherently with what we have seen thus far, the most famous of the Leibnizian definitions of justice is not that of harmony but that of *charity*. The starting point is again God's attitude towards its creatures, his main sentiment being love.

God is just:

> Th: What is to be *just*?
>
> Ph: He is *just* who loves everyone.[678]

God loves everyone, for even in the worst of all (*pessimus*) the potential for goodness can be loved[679]. So must we, in order to be just. This is the most original conception of justice in Leibniz's work. It does not mean that "justice is the greatest of follies", for it induces people to seek the profit of the other at their own expense.[680] On the contrary, the just person pursues the good of others because they are pleased by it. One might be pleased by the pursuit of the good of others because they hope that they will get something in return, live gratitude and favours; hence, the good of others pleases them instrumentally: it is just a means to other ends and, should such a reward fail to happen, then the good of the other would lose its pleasantness, and, therefore, its motivational force. Still, the instrumental value of the good of others can be enormous: if humans realised the material advantages that would come from mutual cooperation, they would become like God for each other (*homo homini deus*) instead of mutual predators (*homo homini lupus*)[681].

But the good of the other can also be pleasurable *per se*: an intrinsic good sought after for itself that nothing could take away from us: to find this kind of pleasure is to love disinterestedly[682]. Note that the notion of "disinterestedness" is not to be understood in a Kantian way as something detached from our own advantage: as previously mentioned, nothing can be pursued if it is alien to our interest. Rather, it

---

[677] Leibniz, "Principles of Nature and Grace, Based on Reason (1714)," 212.

[678] Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 29.

[679] Gottfried Wilhelm Leibniz, "Definitio Iustitae Universalis," in *Historische Beiträge Zur Philosophie. Vol. II*, ed. Adolf Trendelenburg (Berlin: G. Bethge, 1855), 268.

[680] Gottfried Wilhelm Leibniz, "De Justitia," in *Mittheilungen Aus Leibnizens Ungerdruckten Schriften*, ed. Georg Mollat (Leipzig: H. Haessel, 1893), 35.

[681] Leibniz, "Meditation on the Common Concept of Justice (c. 1702-3)," 57.

[682] Leibniz, *Theodicy*, 424.

indicates that something does not have instrumental value but is in itself the object of our interest. Once this is understood, it is easy to see why Leibniz defines justice as the combination of goodness and wisdom, which "is nothing else than the knowledge of the good"[683]. Once one knows where to find their own good, namely, in the good of others, a loving and just attitude follows necessarily.

The possibility of disinterested love links the conception of justice as charity to Leibniz's motivational hedonism[684] and makes justice achievable. For Leibniz, it is, in fact, impossible to demand someone to do something against their own interest and "nobody can be forced to his own damage"[685]. Moreover, he believes ought to imply can, and since "it [is not] possible to renounce, more than verbally, the impulse to one's own good without at the same time renouncing one's own nature"[686], if disinterested love was not possible, then justice as charity could not be a normative stance.

If, in loving disinterestedly, we are imitating God, nonetheless, it must be noted that "because of the universal harmony of things, which sets off a picture by shading, consonance by dissonance"[687], God cannot save everyone. Since our justice should replicate his, we too should be just by loving everyone[688] but be ready to sacrifice some. Given Leibniz's already sketched theory of responsibility, those whose will is inclined towards evil ("*pessimi*"[689]) are the ones we must sacrifice in order to maintain our social order.

In fact, justice is not only defined as the habit of loving but also as the mastery of love and hate[690]: like God "hates sin", meaning that he punishes it, so we must do with sinners or criminals. As stated in the *De Justitia*, justice is a form of regulation of our inclination for pleasing and hurting others[691]: when we are just we are, therefore, not only looking for the good of others, but, in certain conditions, we might be just while hurting them. This inclination must be regulated both according to merit and convenience for the greatest number[692]. In this early text, the two conceptions of justice,

[683] Leibniz, "Meditation on the Common Concept of Justice (c. 1702-3)," 50.
[684] Gregory Brown, "Happiness and Justice," in *The Oxford Handbook of Leibniz*, ed. Maria Rosa Antognazza (Oxford: Oxford University Press, 2018), 625.
[685] Leibniz, "Juris et Aequi Elementa," 24, my translation.
[686] Leibniz,'Preface to the Mantissa', 3, as quoted in Brown, "Happiness and Justice," 625.
[687] Leibniz, "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)," 31.
[688] Leibniz, "Juris et Aequi Elementa," 30.
[689] Leibniz, "Definitio Iustitae Universalis," 268.
[690] Leibniz, "Juris et Aequi Elementa," 26.
[691] Leibniz, "De Justitia," 35.
[692] Leibniz, 36.

a retributive one and a utilitarian one, which will be more explicitly discussed in the *Theodicy*[693], are already sketched.

The utilitarian conception of justice is based on three principles: that of the correction of the offender (rehabilitation in contemporary terms), that of the defence of the potential victims (incapacitation), and that of the exemplary punishment (deterrence). All three, says Leibniz, do not require any kind of metaphysical or absolute freedom: in fact, they can all be imposed on non-human animals as well[694].

"True retributive justice", on the other hand,

> assumes […] intelligence and freedom in him who sins, because the harmony of things demands a satisfaction, or evil in the form of suffering, to make the mind feel its error after the voluntary active evil whereto it has consented. Mr. Hobbes also, who does away with freedom, has rejected retributive justice […].[695]

Such a conception of justice is the natural consequence of the aesthetic considerations previously observed and present in the *Theodicy* as well. There, the concept of wisdom acts as a link between aesthetics and morals. As wise people are pleased by the perception of good music or architecture, so they are by a convenient punishment, for they experience a form of spiritual compensation[696]. Vindictive justice also has a theological explanation, as this kind of satisfaction may be connected to the *satisfactio vicaria Christi*[697]. What still lacks is a specifically moral explanation that Leibniz fails to give.

Retribution and utilitarianism come into conflict in distributive justice. According to Leibniz, distributive justice should be regulated by the traditional principle of the *suum cuique tribuere*[698], distributing goods and evils according to merit but also aiming at the greatest possible common good: evils that would otherwise be shared must fall on those who are responsible (in the sense earlier defined) for them and, conversely, the authors of positive actions must be rewarded.[699] But what happens when such a distribution does not grant "the greatest possible common good"? In this case, the utilitarian principle overwhelms the retributive principle: "the supreme rule of

[693] Leibniz, *Theodicy*, 163–67.
[694] Leibniz, 163.
[695] Leibniz, 425.
[696] Leibniz, 165.
[697] Mormino, *Determinismo e Utilitarismo Nella Teodicea Di Leibniz*, 203.
[698] Gottfried Wilhelm Leibniz, "De Tribus Jusris Naturae et Gentium Gradibus," in *Mittheilungen Aus Leibnizens Ungerdruckten Schriften*, ed. Georg Mollat (Leipzig: H. Haessel, 1893), 8.
[699] Leibniz, 10.

right is: what is useful for the community, that must be done."[700] In the end, Leibniz's justice appears to be utilitarian at its core.

We have covered a varied path into Leibniz's theories of justice. Their guiding principles are *harmony*, which leads to a retributive conception of justice, *common good*, which leads to a utilitarian conception, and *charity*, the most original element, which does not lead to a particular conception, and is complemented by Leibniz with one of the former principles – but that will hopefully be developed in the third part of this work. The utilitarian conception seems predominant, yet Leibniz spends many pages on retributive justice and its requirements, i.e., moral responsibility and contingency. In this account, responsibility is founded on a (questionable) compatibilist theory and the (dubious) difference between hypothetical and absolute necessity. The reasons that bring Leibniz to elaborate such a theory are clear: frequently, he repeats the importance, for moral purposes, of practices based on responsibility, like praise and blame. Still, this retributive account remains unconvincing.

2.3.4 The Conceptual Framework in Contemporary Philosophy

The debate on punishment, responsibility, and free will has not changed very much in the last four centuries, the features of the arguments having been set in modern times, and the conceptual framework having remained more or less stable since then. Religious dogmas (Descartes, Malebranche) have been replaced by moral dogmas (Smilansky, Strawson) – common sense morality having become unquestionable, leaving to philosophy the task of analysing it and justifying its conclusions rather than challenging them. The consequentialist justifications for punishment, advanced by Hobbes, Spinoza, and Leibniz, have become central after the penal reform of the 18th century, thanks to the efforts of the utilitarian reformers, Beccaria and Bentham foremost among all, although retributivist justifications were never abandoned. Aesthetic considerations, like those advanced by Leibniz, not only remain dominant in common thought as demonstrated by the diffusion of the symbol of the scale of justice and the insistence that offenders should "pay their debit to society", but are also vindicated by philosophers who apparently do not question their moral significance[701].

[700] Gottfried Wilhelm Leibniz, "De Summa Juris Regula," in *Mittheilungen Aus Leibnizens Ungerdruckten Schriften*, ed. Georg Mollat (Leipzig: H. Haessel, 1893), 85, my translation.

[701] John Sabini and Maury Silver, "Emotions, Responsibility, and Character," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987), 172.

Beccaria and the reformers triggered the process of punishment mitigation; in the last decades, we have witnessed the return of "tough on crime" policies leveraging retributive emotions and implementing harsh punishment, like supermax prisons in the US and their counterparts in other countries, despite studies having shown that more severe punishments do not have more deterrent effect than milder sentences – therefore being hardly justifiable on consequential grounds. The tension between forward- and backwards-looking justification that I have highlighted in many modern philosophers is far from being resolved.

Moreover, theological determinism has given way to physical determinism – as in Hobbes and Spinoza. Internalist explanations of free will, like that of Malebranche, have been maintained almost unaltered, as in Chisholm[702], and the difficulties of explaining God-like originator action are sometimes accounted for through reference to quantum indeterminism at the level of the brain[703], still leaving the structure of the argument unchanged. The backwards-looking justification for punishment is still very much related to the issue of free will, and this, throughout the years, has been incorporated in the criminal trial, with the growing interest in the inner life of the defendant, their so-called *mens rea*.

Foucault included this fact in the more general shift from the perpetrator's corporal dimension to their mental dimension, a de-corporalisation of punishment that is attested by its substantial transformation from torture to incarceration, the objective of which is allegedly the change of the offender's character and not the affliction of their body. In the criminal trial, the focus has also been shifted from mere actions to intentions: whereas once they were believed to only be knowable by God, while earthly justice should only be concerned with actions[704], it became gradually accepted that even the intimate sphere of intentions was accessible to judges and juries. This also determined the progressive medicalisation of trials, which included consulting psychiatrists, psychologists, and criminal anthropologists in the determination of

[702] Chisholm, "Human Freedom and the Self."

[703] Randolph Clarke, "Towar A Credible Agent-Causal Account of Free Will\," *Noûs* 27, no. 2 (1993): 193; Marco Masi, "Quantum Indeterminism, Free Will, and Self-Causation," *Journal of Consciousness Studies* 30, no. 5–6 (2023): 32–56.

[704] For example, Peter Abelard, whose ethics is centred on the notion of intention, nonetheless justifies the punishment of a mother who involuntarily but unavoidably kills her baby while trying to keep them warm in an act of love. Peter Abelard, *Ethics*, ed. and trans. D. E. Luscombe (Oxford: Oxford University Press, 1971), 39.

appropriate sentences.[705] As noted by Susan Jacoby, this has generated widespread disappointment with the penal system, for sentencing discounts based on psychiatric diagnoses generate the sense of incorrect pay, not satisfying public anger[706].

In present-day legal philosophy, the general existence of free will usually is not questioned. Philosophical analysis rather focuses on cases in which, given the general existence of free will, an action is not motivated by it and, therefore, moral responsibility could be absent, making punishment not justifiable on retributive grounds. In a way, the limits of moral responsibility are only analysed in order to reaffirm its general existence and the practices that depend on it. A prominent example of this mechanism is the focus on excuses and exemptions.

Excuses deny the intentionality of action and account for cases of physical constraint, as when one really wanted to arrive on time and did everything they could to do so but found themselves stuck in traffic due to an unforeseeable car accident and arrived late. Although, as Wallace pointed out, one might think that, in such a situation, the lack of responsibility would depend on the absence of alternatives, the real reason we excuse people in conditions of constraint is that the action is not an expression of their intention[707]. This also explains why we attribute responsibility in Frankfurt-style cases[708]. On the other hand, exemptions declare the non-accountability of the agent who lacks reflexive self-control, i.e. the ability to understand and apply moral reasons and to regulate one's behaviour according to such reasons[709]. Exemptions apply in cases of hypnosis, mental disorders, or minor age, for example.

In analysing the conditions of general moral accountability, that is, in Wallace's terms, of reflective self-control, philosophers have proposed various – albeit at heart similar – criteria. Some of the most relevant include Frankfurt's appeal to second-order desires, that is, desires to have certain desires and, thus, the agent's identification with

[705] Foucault, *Discipline and Punish*, 18–22.
[706] Jacoby, *Wild Justice*, 324.
[707] Wallace, *Responsibility and the Moral Sentiments*, 141–42.
[708] In his 1969 paper *Alternate Possibilities and Moral Responsibility*, Harry Frankfurt exposed his famous scenarios in which an evil neuroscientist can control from the outside the outcome of someone's deliberative process so that, should they not spontaneously choose to perform a certain action, the scientist would make them choose it. We would not think that the person spontaneously choosing to perform the action would be deprived of their responsibility because the scientist takes their alternate possibilities away. Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66, no. 23 (1969): 829–39.
[709] Wallace, *Responsibility and the Moral Sentiments*, 157.

the action[710]. Angela Smith advances a similar proposal, that of a "rationalist view" of responsibility, which links responsibility to one's "own evaluative judgements or appraisals"[711], emancipating it from the control over one's volitional process and making it a function of one's values and beliefs – in this way accounting not only for Frankfurt-style cases but also for involuntary actions, like forgetting a friend's birthday.

John Martin Fischer advanced another criterion for moral accountability: reason-responsiveness, i.e., the agent's susceptibility to changing their course of action when their knowledge of the information relevant to putting the action in place changes[712]. More recently, Fischer elaborated on this conception and comprised it in that of "guidance control", which couples reason-responsiveness with "mechanism ownership", that is, the recognition of the historical connection between the agent and their action, which generates a sense of ownership of the action and makes an "individual become responsible in part at least by taking responsibility"[713].

Finally, Susan Wolf's proposal of "sanity" as the condition for accountability seems particularly relevant in relation to what we have seen so far concerning the medicalisation of the criminal trial. According to Wolf, a sane individual is one who knows what they are doing and can discern whether it is right or wrong[714]. As she put it, this does not depend on one being able to control certain processes but rather on being "*controlled by* the world in certain ways"[715]. "It is not in our ordinary power to determine whether we are sane or not sane. Most of us, it would seem, are lucky, but some of us are not"[716], writes Wolf.

Even if these authors usually refer to cases related to interpersonal everyday life (should I be angry at my husband for being late again, or is he not responsible for his

[710] Frankfurt, "Freedom of the Will and the Concept of a Person"; Harry Frankfurt, "Identification and Wholeheartedness," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987), 27–45.

[711] Angela M. Smith, "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115, no. 2 (2005): 237.

[712] John Martin Fischer, "Responsiveness and Moral Responsibility," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987), 81–106.

[713] John Martin Fischer, *Deep Control. Essays on Free Will and Value* (Oxford: Oxford University Press, 2012), 10; See also: John Martin Fischer, *My Way : Essays on Moral Responsibility: Essays on Moral Responsibility* (Oxford: Oxford University Press, 2006).

[714] Susan Wolf, "Sanity and the Metaphysics of Responsibility," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987), 55.

[715] Wolf, 55.

[716] Wolf, 55.

fault this time?), excuses and exemptions, discussions on intentionality and accountability, are central in criminal courts and help determine a defendant's sentence. Although different in some respects, all these proposals share the common idea that a special relation of the agent with its action, mediated by beliefs, makes one accountable and, when intentionality is also given (thus excluding excusing conditions), morally responsible.

A theoretical problem that they share is their inability to respond to the problems posed by constitutive moral luck, as I clarified earlier: the fact of being the kind of person that has a second-order desire to commit a crime (Frankfurt), that reflectively approves of such action (Smith), that would not decide to offend given different conditions but, when they do, take ownership (Fischer) and, finally, is aware of what they are doing although recognising it to be abstractly wrong (Wolf) – all this is ultimately a matter of bad luck. The fact of being that kind of person cannot depend on a second-order desire, a judgement, etc. – it is their general condition instead.

However, I find the real problem of these theories to be more radical concerning the general conception of individual moral responsibility that underlies them. Fischer presents the issue he is addressing in this way:

> An agent is morally responsible for an action insofar as he is rationally accessible to certain kinds of attitudes and activities as a result of performing the action. The attitudes include resentment, indignation, respect, and gratitude; and the activities include moral praise and blame, and reward and punishment.[717]

The conditions proposed by the authors aim at identifying exceptions that prove the rule: in a peculiar case, responsibility is not given, but this serves the purpose of reaffirming the general framework made of free will, moral responsibility, and retributive punishment. It might be suspended in peculiar situations but is not generally questioned. What they do not tackle is the question: Why should a free action, one that is desired, judged, or reason-responsive, justify a retributive attitude? Why don't the conditions of accountability only generate reliability and possibly deception but also a moral entitlement to blame and punishment? Is it only because they are the conditions in which a wrongful action would generate person-directed anger in us? If so, as I tried to make clear in a previous section, we should acknowledge the moral inadequacy of

[717] Fischer, "Responsiveness and Moral Responsibility," 81.

such a justification. If justice is to remain a moral concept, we should try to build it on different grounds. This is what I aim to do in the next chapter.

3 The Future: Normative Perspectives

3.1 Reshaping Moral Responsibility

3.1.1 Interconnectedness

The retributive justice paradigm, based on free will, moral responsibility, and desert, has been questioned by authors who have followed the pioneering steps of Hobbes and Spinoza. Firstly, we have just cited Wallace's notion of "reflective self-control" in the previous chapter. Whereas, as mentioned there, it falls squarely within the scope of what we have called the conceptual framework of the contemporary conception of justice, and particularly of the expressive conception of responsibility based on Strawson's conception of resentment and the moral role of reactive attitudes, it must also be noted that the author does timidly attempt to put it into question at the end of his essay, where he mentions that

> […] we need the reactive emotions if we are to take seriously the moral obligations we accept as a basis for our common social life. In holding people to these obligations, however, we should bear in mind that the powers of reflective self-control that justify this stance are a fragile achievement. This is equally a fact of moral life. To do justice to it, we must aspire to respond in ways that we can continue to regard as fair even after we have a full and accurate picture of the agents whose deeds we are judging.[718]

This frailty that he mentions is related to the fact that the circumstances of one's life can – and often do – damage one's reflective self-control. Wallace refers explicitly to "common phenomena as childhood abuse, psychological trauma, and the persistence of extreme poverty and violence in the midst of general affluence."[719] All of them can actually be subsumed under the category of "trauma", an emotional shock that can impair – that is, dysfunctionally influence[720] – future reasoning and behaviour. As we can see, "trauma" refers to a very common and general phenomenon that most people face at some point in their life. Wallace's insistence on traumatic experiences is interesting but problematic.

Wallace seems to imply that there is a specific category of experiences that has a major influence on one's reflective self-control, whereas other non-traumatic

[718] Wallace, *Responsibility and the Moral Sentiments*, 234–35.
[719] Wallace, 234.
[720] By "impairment" and "dysfunctionality" I mean the bearing of negative consequences for the agent.

experiences would have no effect on it. However, traumas are only one end of the spectrum of experiences; they are experiences whose influence on a person's behaviour is more easily recognisable, but they do not constitute something radically different from all the other experiences an individual goes through. Just because, in the case of trauma, most of the times we clearly see the connection between the experience and the person's behaviour, it does not mean that other experiences do not influence our behaviour in the same way – only that, being the resulting behaviour not as dysfunctional, and the originating experience less relevant in one's perception, it is generally harder to trace one's character traits or behaviour back to specific experiences.

We are dealing with the same issue that was addressed previously when discussing the case of murderer Robert Harris: when we are told about the traumatic experiences he has been through, we let go of retributive attitudes and open up a space of compassion, despite his heinous crimes, because we easily recognise the role of bad luck in his personal history. These extreme cases do not serve as "exceptions that confirm the rule"; rather, they highlight more clearly processes that are always at work in what we have called constitutive luck. However, Wallace's stance is not unexpected given his hesitation towards a determinist position, to which he gets close but cannot eventually espouse, suggesting a difficulty in holding together the constitutive influence of experience and the desire to preserve individual moral responsibility.

However, in considering the actual development of reflective self-control, Wallace adds to the influence of trauma what he calls in a previous passage "moral education". This sort of education is constituted, according to the author, by "love and emotional support of the sort that encourage self-esteem, and freedom from hypocrisy and arbitrariness in the promulgation and enforcement of moral standards."[721] To these, I think we can safely add an analysis of the actual values that are dominant at home, at school, and that are in general transmitted by the reference figures in one's development process. There is no need to bother Montaigne's essay *On Habit* – a beautiful and always instructing read, by the way – to realise that we can come to consider good and evil, in a way that motivates our actions, the most disparate things, depending on what we are accustomed to.

If the values that are transmitted do not correspond to those that are dominant in the society one belongs to and one is part of what Durkheim calls an "*anomie*" – that is,

[721] Wallace, *Responsibility and the Moral Sentiments*, 165.

a minority group that shares different values from the majority – then the moral obligations that they hold to will not be recognised by their fellows. Does this imply an impairment of their reflective self-control? It may seem so, for one's most fundamental values can be, and often are, so ingrained in one's identity that they become the basis for one's reasoning and decision-making process rather than their object – and we have seen when discussing intuitionism in the previous chapter that many philosophers do not consider this a problem: one's moral intuition is enough as the basis for one's moral action.

So, even if we believe that the moral obligations shared by the majority are the good ones and that, ideally, we should get everyone to agree on them – of course, this is anything but a given, as a society could be a victim of moral delusion and hold on to it because moral intuitions are flawed and biased, as this has happened innumerable times across history, and is still happening everywhere, the only reason why one may not recognise it is because one is a victim of that same delusion, but let us assume it for the sake of simplicity – we should consider that those who do not share them are likely to do so because of an impairment of their reflective self-control, specifically in its *reflective* part: they cannot question the values that are ingrained in their identity.

Hence, if trauma and a certain kind of moral education, understood in the broad sense just explained, generate impaired reflective self-control, it looks like quite a significant share of the world's population might result as non-reflectively self-controlled – especially, as Wallace himself remarks, when the action performed is forbidden by common morality[722]: we would hardly say that one lacks reflective self-control, or that their self-control is unpaired, if they perform a virtuous action as a consequence of a strict education that conforms to shared values. The problem seems to lie only in the *anomie*: it is not a general reflection on the motives of action but specifically of those actions that are prohibited by ordinary moral standards. Wallace's principle of avoidability, in fact, states: "it is unfair to expose people to the harms of moral sanction if they lack the specific ability to exercise the powers of reflective self-control in the circumstances, and thereby avoid those harms."[723] However, Wallace later moderates this statement, maintaining that these kinds of social conditions make

[722] Wallace, 204.
[723] Wallace, 205.

accountability "impaired without being altogether undermined"[724], leaving open the space for individual moral responsibility. However, the threshold between impairment and disruption is never clear, and the difference in the response to be given to one whose reflective self-control is impaired and one who possesses it integrally is unspecified and perplexing.

Other authors, however, took a more radical stance. Foremost among them is Bruce Waller, who, over the course of forty years of work, developed a composite critique of moral responsibility, which builds on a rejection of free will that "would have to occur outside the natural world: if it makes sense at all, it could only be in a world of miracles"[725], and delves into the concept of desert. Notably, Waller differentiates between justice- and non-justice kinds of desert; non-justice deserts are constituted by act-, talent-, and effort-deserts, depending on whether the action is evaluated against the result obtained (e.g., the winning of a race, even if by chance), the actual abilities of the agent (e.g., the fact of being the fastest, despite losing the race because of an injury), or their commitment (e.g., the fact of training and participating to the race, despite unfavourable genetics for running). Non-justice deserts are acceptable and valuable in everyday assessment of actions, as long as they are not confused with moral desert because they bear the same name.

Justice-desert is the moral desert that serves as a basis for punishment and, to Waller, is nonsensical because, based on the rejection of free will, "all justice-deserve precisely the same. Or another way of putting it: no one justice-deserves anything at all."[726] The absence of justice-desert, then, is the basis for a rejection of retributive justice[727] and, more in general, negative reactive attitudes:

> Obviously the denial of moral responsibility undercuts such reactive attitudes as vengeance, but it does not undermine attitudes of affection, gratitude, love, and friendship. Looking carefully at our ties of affection and friendship, we may have difficulty tracing their origins and basis; but it is clear that they are not founded on just deserts. […] Nor does the universal denial of moral responsibility imply that we cannot argue with one another. It does, however, mean that we should not blame our opponents

[724] Wallace, 234.

[725] Bruce N. Waller, *Against Moral Responsibility* (Cambridge/London: The MIT Press, 2011), 27.

[726] Waller, "Just and Nonjust Deserts," 232.

[727] "[R]eflective deliberation is a good means of making a choice my own, a choice that I have made myself, a choice for which I am 'responsible' in the sense that I am responsible for making my own decisions, and no one has authority over me. But none of this establishes moral responsibility for my choices. It does not establish that I justly deserve praise or blame, reward or retribution, or any form of 'just deserts' for my reflective (or less than reflective) choices." Bruce N. Waller, "Deep Thinkers, Cognitive Misers, and Moral Responsibility," *Analysis* 59, no. 264 (1999): 227.

> for shallow reasoning, nor judge them deserving of punishment for failing to give due consideration to our compelling argumentation.[728]

According to Waller, moral responsibility is like a dark glass that hinders the clear vision of facts and people: "[p]erhaps the greatest harm caused by adherence to moral responsibility is the severe limits it places on our inquiries and understanding"[729]. This fact seems to be recognised even by retributive justice advocates, as demonstrated, for instance, by former British prime minister and tough-on-crime advocate John Major, who, in a 1993 interview, claimed: "Society needs to condemn a little more and understand a little less."[730]

The moral responsibility paradigm, in fact, limits inquiry to the individual as if they generated their actions in a vacuum – as if the individual itself were an abstract entity in the void. This is precisely what I aim to criticise in the first part of this chapter. Waller, hence, rather subscribes to what Elizabeth Beardsley has called "the perspective of ultimate moral equality"[731]. Beardsley opposes it to two other moral perspectives: the perspective of moral worth – that classifies agents according to the positive or negative moral value of their action[732] – and that of moral credit – that, to the judgement on moral worth, adds the investigation of the factors that make the action effortless or challenging for the agent[733].

While through the perspectives of moral worth and of moral credit one formulates their judgement on another agent considering just the action's outcome (in the former case) or the individual's conditions (in the latter), the perspective of ultimate moral equality is causally unlimited[734]: it takes into account all causal circumstances that generate a specific act, making the individual agent only the last link, or the proximate cause, of the causal chain that leads to the event. On the other hand, the perspectives of moral worth and credit are causally limited perspectives that only take into account a limited number of causal circumstances, stopping arbitrarily in moving up the causal chain. In a word, they are myopic perspectives, and it is the lens of moral responsibility that impairs the vision. Short-sightedness is also highlighted by

[728] Bruce N. Waller, "Denying Responsibility without Making Excuses," *American Philosophical Quarterly* 43, no. 1 (2006): 86.

[729] Waller, *Against Moral Responsibility*, 283.

[730] John Major and Jonathan Holbrow, "Interview with the Editor of the Mail," *The Mail*, 1993, 8.

[731] Elizabeth Lane Beardsley, "Determinism and Moral Perspectives," *Philosophy and Phenomenological Research* 21, no. 1 (1960): 11.

[732] Beardsley, 4.

[733] Beardsley, 7.

[734] Beardsley, 12.

Smilansky, who, as we have mentioned, is a paradoxical defender of free will and individual moral responsibility, both denying them at the fundamental level and asserting them as guides for everyday life. He considers the perspective that is based on them as "shallow",[735] strengthening the metaphor of myopia: individual moral responsibility only allows one to look at the surface of things. According to Smilansky's own words, this transforms life into a "masquerade"[736] – once again, a mask refers to something that is put on the surface and conceals a deeper truth.

At the same time, Smilansky is worried that abandoning the moral responsibility system would lead to a moral "semi-blind[ness]"[737] – keeping the sight-related metaphor going and depicting a worse scenario than myopia – that would erase the differences in the treatment of the guilty and the innocent. He bases his reasoning on the idea that the progressive concern for the individual action in judgments of blameworthiness or praiseworthiness has been a critical moral progress and that something deeply valuable would be lost if they were to be let go.

If moral responsibility is a lens that impairs vision, it looks like more archaic systems that were based on pure action, without inquiry into the intention or control of the agent, those based on the simple *lex talionis*, were fully blind. In reality, we have seen in the first chapter how the circumvention of the punishment of the guilty in stateless societies does not stem from the inability to identify or valorise causal responsibility and guilt. Rather, as Girard maintains, it is the acute awareness of the implication of direct vengeance, of hitting the guilty, that makes these societies carefully avoid punishing them.

This is not to say that moral progress should not be ascribed to State societies and to the abandonment of systems that openly accept the innocent's sacrifice and that are ready to put in place forms of collective punishment, that regulate behaviour towards others on the basis of their gender, religion, class, or skin colour[738], or that do not discriminate between harm inflicted purposefully or inadvertently[739]. However, it allows systems of moral responsibility to be inscribed in the course of moral progress as one of its steps rather than its end. It also falsifies Strawson's thesis, according to which there is something intrinsic to human nature in the beliefs and practices related to moral

[735] Smilansky, *Free Will and Illusion*, 190.
[736] Smilansky, 278.
[737] Smilansky, 99.
[738] Smilansky, 21–21.
[739] Waller, *Against Moral Responsibility*, 12.

responsibility based on our reactive attitudes, and to resentment specifically, as we have seen. If the introduction of responsibility in our practices of justice is relatively recent, as Waller maintains[740] and as I have tried to show in the first chapter analysing what I have called the "origins" of justice through the work of Girard, it means that alternative systems *are possible*: the fact that in pre-State societies practices of justice were not based on individual moral responsibility is the empirical proof that interpersonal relationships can be based on different paradigms. Of course, contrary to Verdier, Courtois and their research group, this does not mean that we are or should be advocating for a return to that paradigm.

As Waller maintains, a compelling critique of the responsibility paradigm requires reasoning outside of it: Strawson regards responsibility as the normality and lack of responsibility as the exception and then proceeds to extend the cases of exceptionality to encompass all cases, reaching his conclusion that not to consider our fellows morally responsible would necessarily mean to take on what he calls the "objective attitude". This implies to let go of any feeling of emotional participation and consider the other as an object, something "to be managed or handled or cured or trained"[741]. But an effective critique cannot be rooted in the soil of the very system it wants to criticise. A radical critique requires us to think without moral responsibility from the very beginning of the reasoning[742]: it is not, as Strawson suggests, that we should start treating everyone as we treat those whom we believe not to be responsible now. To continue the pursuit of moral progress, we need to take a different, a broader perspective.

Actions, seen from the perspective of moral responsibility, remain obscure because they are considered to originate within the individual according to mysterious processes: the vision is too narrow. It allows one to be disinterested in the systemic issues that generated the specific problem – the criminal or morally wrong action. Moral responsibility wants to convince us that the problem begins and ends with the individual. Instead, overcoming the perspective of moral responsibility leads us to see more clearly the factors that led to the occurrence of the action and, through this knowledge, to have more control over their recurrence.

---

[740] Waller, 12.

[741] Strawson, "Freedom and Resentment," 9.

[742] Bruce N. Waller, *The Stubborn System of Moral Responsibility* (London - Cambridge: The MIT Press, 2015), 83.

Whereas moral responsibility jumps from the moral classification of an action to the moral judgement on the agent and then stops, a broader perspective inserts the agent into a context, not treating them like an abstract entity but rather like one embedded in a complex network that constitutes their environment, and it is made of other people, social structures, economic status, physical setting, housing conditions, and so on. Considering all these factors allows us to pass from what Hilary Bok calls possibility in "a general sense"[743] – that is, considered abstractly: the possibility that water has to freeze or evaporate, but not to recite a poem[744] – to what she calls "possibility *tout court*"[745], the one defined on the basis of the relevant elements specific to the situation. In this sense, the frozen water in my freezer cannot evaporate.

> If we define our possibilities thus, then we can say that all of those actions that we would have performed had we chosen to perform them are possibilities for us, just as we can say that it is possible for a given amount of water to vaporize if we disregard the fact that it is sitting in my freezer.[746]

When we view an agent's possibilities as possibilities *tout court*, we start considering their responsibility more like *causal* responsibility than *moral* responsibility. We have mentioned that moral responsibility is strictly linked to moral desert and that it descends from some form of reflexive self-control – of course, a perspective like the one we are advocating, as we have seen discussing the frailty of reflexive self-control, undermines moral responsibility at its bases. The actual agent is the proximate cause of the action, but to focus on them only seems now arbitrary. In ascribing causal responsibility to them, we consider the individual's contribution, and that serves as a basis for action directed to the individual – in case of a criminous action, restraint, expulsion, or even killing could be considered. But it also reminds us that acts do not begin and end with the individual as they are the result of a large web of factors: if we want to change the outcomes, then the focus should not narrow down to individual actions but rather enlarge and work on multiple elements.

In psychology, this prospective expansion has been operated by attribution theory, theorised by gestalt psychologist Fritz Heider in his pioneering 1958 *The Psychology of Interpersonal Relations*[747], where he first formulated what would later be

[743] Hilary Bok, *Freedom and Responsibility* (Princeton: Princeton University Press, 1998), 94.
[744] Bok, 93.
[745] Bok, 94.
[746] Bok, 94.
[747] Fritz Heider, *The Psychology of Interpersonal Relations* (New York: John Wiley & Sons, 1958).

known as the "fundamental attribution error"[748]. It was observed that people tend to attribute specific behaviours to an individual's character when they are, in fact, better explained by environmental influences. Behaviours that are perceived as deviant and exceptionally blameworthy rather fall within the statistical norm given equal conditions – what is exceptional is not the behaviour given the conditions but the conditions themselves. Famously, it is the case epitomized by Milgram's experiment on compliance[749].

As is well known, in his social psychology series of experiments, Milgram told experimental subjects that they would participate in an experiment on memory: they would need to administer electric shocks of increasing intensity to another experimental subject in case they made mnemonic errors in a given exercise. The experiment was, in reality, targeting obedience to authority: the scientists wanted to see when subjects would stop administering electric shocks in the face of the other's evident suffering, request to stop the experiment, and even risk of death, if the chief scientist ordered to continue. Of course, the second subject was, in reality, an accomplice who would only pretend to receive the shocks.

Before conducting the experiment, it was expected that most people would abandon the procedure at a pretty early stage. Milgram had asked a class of Yale seniors majoring in psychology what percentage of experimental subjects would reach the final stage of the shock series: the estimates ranged from 0 to 3%, with a mean of 1.2%[750]. Milgram's colleagues also informally claimed that only a few – if any – would go beyond the shock labelled as "Very Strong" on the machine[751]. Milgram himself said in an interview that he believed that "only a very, very small portion of people going out to the end of the shock generator, and they would constitute a pathological fringe."[752] In fact, his study was at least partly motivated by the horrors of the Shoah – to the point that Milgram's experiment has also been dubbed "the Eichmann experiment"[753] – that he believed ordinary German citizens could have accomplished because of specific traits of German people that made them more prone to obedience and compliance.

[748] Lee Ross, "The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process," *Advances in Experimental Social Psychology* 10 (1977): 173–220.

[749] Stanley Milgram, *Obedience to Authority* (New York: Harper & Row, 1974).

[750] Stanley Milgram, "Behavioral Study of Obedience," *Journal of Abnormal and Social Psychology* 67 (1963): 371–78.

[751] Milgram.

[752] Philip Meyer, "If Hitler Asked You to Electrocute a Stranger Would You?," *Esquire*, 1970.

[753] Milgram, *Obedience to Authority*, 178.

Hence, he planned on conducting his experiments on U.S. citizens first and then replicating them in Germany in order to emphasise the differences in the results. He never did, for the results on Americans were astonishing in themselves.[754]

In all versions of the experiment, the majority of the experimental subjects, after some discomfort and hesitation, once encouraged by the scientist in a white coat, administered high-intensity shocks, even when the other experimental subject-accomplice stopped responding to stimuli. Approximately 65% of experimental subjects reached the highest voltage, marked with an ominous "XXX" tag (right after the "Severe Shock" label) on the machine.

The gap between the expected result – that very few people would reach the end of the scale – and the actual results can be explained by the fundamental attribution error: the influence of the environment is discounted, and behaviours are entirely attributed to an individual's personality and free will. On the other hand, findings suggest that personal character – whatever that may mean – plays a relatively minor role, and environmental factors strongly influence that behaviour.

The concept of "personal character" is in itself very vague. In ordinary talk, many use it in the context of an implicit essentialist conception of the person, considering individuals to have an immutable character already formed at birth or in the womb: an untouchable soul that gives a person their specific individuality, that they will retain for their whole life. Evidently, this conception is quite hazy: it fits well with the idea of an immortal soul to which God gives life, a spiritual substance that is infused in the body, but it makes little sense if we consider humans to be embodied entities that form their way of being in their interaction with the world. However, we can also use the notion of personal character in a weaker sense to indicate those regularities we find in a person's behaviour as a result of their experiences, sometimes crystallised as a consequence of early childhood occurrences but often susceptible to further evolution over the course of one's life.

In this second sense, the concept of personal character is situated in a continuum and not in opposition to environmental factors. It may also become so thin that one wonders whether it makes sense to use it at all, if it is not just an abstraction, a narrative fiction – especially in the light of the results presented here. If people we perceive as so

[754] Robert P. Abelson, Kurt P. Frey, and Aiden P. Gregg, *Experiments with People. Revelations from Social Psychology* (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2004), 247.

different from one another tend to behave in the same way given the same circumstance, taking predictive powers away from the notion of personal character, then what is it good for? Still, I think the term can be useful to refer to the set of expectations we have, based on our conception of others and ourselves, whether it is reasonable to have them or not.

The fact that such expectations do not seem reasonable – insofar as statistical norms are better predictors than preconceived ideas about a person's response to a certain situation – brings us back to Heider's attribution theory. Heider theorised a five-step development of attribution, starting with the attribution to the individual of any effect that may be linked to them: "[f]or example, […] a person may be congratulated upon the victory of his school's football team"[755] says Heider. An example more meaningful for the current research may be found in previously analysed vindicatory systems, illustrated by Verdier: members of the group bear the responsibility for what other members do to other groups. This means that they may be victims of adversarial retaliation or peace-making rituals and are required to participate in retributive expeditions when fellow members are victims of other groups' offences.

In the second stage, attribution regards everything an individual may cause, whether the act's effects were foreseeable to the agent or not. Then, in the third, everything that could, in principle, be foreseen – even if it is not part of the agent's goal – is attributed to them. In the fourth stage, only what was intentionally carried out is attributed to the agent. "Finally there is the stage at which even [the subject]'s own motives are not entirely ascribed to him but are seen as having their source in the environment."[756] As Ferdinand Schoeman remarks, this calls into question the notion of personality itself:

> Those who deny the predictive role of personality claim to have discovered that behavior in one context is not a good predictor of behavior in diverse contexts and that fundamentally there is no *pattern of behavior* that is context-independent. What turns out to be a better predictor, according to these researchers, is knowing the base rates of how people in general act under particular circumstances. So if you see me behave in a particular way in context A and want to know how I am likely to behave in context Z, knowing how people behave in Z is a better predictor of *my* behavior in Z than is

[755] Heider, *The Psychology of Interpersonal Relations*, 113.
[756] Heider, 114.

> knowing how I behaved in A and assuming that my personality, which gave rise to my A-relevant behavior, will also govern my Z-relevant behavior.[757]

It looks like a progressive decrease in attribution to the individual is considered by Heider as a form of moral progress: not only not everything that one does or is connected to one is one's responsibility, but, even when acts and effects are intended, once individuals are not considered abstract entities acting into the void, but rather situated beings, their responsibility is shared with the environment.

We have experienced the shift between different attribution stages in Gary Watson's analysis of Robert Harris's case[758] that we previously analysed. For the wanton and cold-blooded murder of two young boys, Harris was sentenced to death. He never showed signs of remorse; once incarcerated, he harassed other inmates, making their life on death row even harder, and he did not hesitate to say that he "chose the road to hell"[759] . The reference to choice is particularly relevant because it eliminates the possibility of referring to compulsive behaviour. The decision to kill the two boys is the result of a Frankfurtian second-level desire: Harris reflexively recognises himself in it, takes the choice upon himself and makes it a coherent part of the whole that forms his identity.

Watson initially details Harris's criminal actions, his lack of empathy, his cold-bloodedness and his cruelty. In doing so, he arouses in his reader a feeling of revulsion, but not only that: one feels a wave of moral anger and the familiar sense of justice, if not regarding his execution, at least his imprisonment. However, we have seen how this account is followed by an interview with Robert's sister, Barbara, who informs the reader that: Robert was born prematurely as a result of a kick on his mother's belly by his father, who was convinced that the child was the fruit of his wife's cheating; because of his father's hatred and anger, his mother also began to hate the infant, violently rejecting his requests for affection and care; from the age of 14 to 19, he was detained in a juvenile prison for stealing a car, and he was there repeatedly raped and attempted suicide twice; etc.

After this story, the reader's attitude is radically changed: the revulsion at the unjust murder of the two boys remains, it is a repugnant and objectively wrong action.

---

[757] Ferdinand Schoeman, "Statistical Norms and Moral Attributions," in *Responsibility, Character, and the Emotions*, ed. Ferdinand Schoeman (Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987), 292–93.

[758] Watson, "Responsibility and the Limits of Evil. Variations on a Strawsonian Theme," 267–84.

[759] Watson, 270.

The attitude towards the action, therefore, is not changed; what does change is the attitude towards the agent: moral anger is diminished, blame is replaced by compassion and the sense of justice is replaced by its opposite, a profound sense of injustice for the fact that some people are born into such unfortunate conditions: we noticed in the previous chapter that Harris is a clear example of moral bad constitutive luck. Getting to know his full story, delving deeper into the conditions that brought up the development of his behaviour, makes the reader jump from Heider's fourth stage of attribution – which comprises what is intentionally carried out – to his fifth stage: the lenses of moral responsibility are removed, and the agent is considered in their connection with their environment.

Does this mean that we need to let go of all participative feelings towards a subject such as the one just described, as Strawson suggests? If we let go of anger and blame, do we need to let go of gratitude and love? From what we know, Harris's story does not give us much grip to feel love for him, but it can generate compassion – another reactive attitude. Here is yet another example Strawson is mistaken in neglecting the differences among the various reactive attitudes and, more specifically, between positive and negative emotions. To deny individual moral responsibility entails letting go – or, at least, downsizing their weight in our lives and relationships – the latter without eliminating the first.

Although positive and negative reactive attitudes, such as love and hate, compassion and resentment, praise and blame, seem to be opposite to one another – the same kind of reaction only of opposite sign, so that it seems that to let go of one necessarily entails letting go of the other – things are not quite so at a closer look. Smart remarks, for example, the asymmetry between praise and blame, the latter being always connected to moral responsibility, while the former is utilised for both responsible actions – praising someone for having helped a person in distress, for example – and for features that have nothing to do with intentionality or control – like praising a person for their beauty. Praise has both a moral and a non-moral meaning[760]. This is also true for love, gratitude, compassion: although they are reactive attitudes in that they are generated in us by the behaviour of some other being, we do not need to ascribe individual moral responsibility to them for the feeling to be appropriate. Indeed, it would be not only startling but also very improper to say that one does not love an

[760] J. J. C. Smart, "Free Will, Praise, and Blame," *Mind* 70, no. 279 (1961): 303–5.

infant because they lack moral responsibility. Many people also appropriately feel such feelings for non-human animals (especially pets) without needing to consider them morally responsible. On the other hand, it would be very unfitting to bear resentment or anger, to blame infants and non-human animals, precisely because we do not ascribe that kind of responsibility to them.

Smilansky himself, when he holds that letting go of the concept of moral responsibility would be morally "barbaric", never mentions the need for praising the good or rewarding the upright, nor the anguish at not being legitimate in feeling gratitude or love. This need, this anguish, in fact, are not felt at all because that is not what retribution fundamentally is about: it is not about giving each their due but about satisfying vengeful desires instead. This is consistent with the evolutionary explanation we gave previously: many animals have an urge to strike back when in danger – obviously, nothing like an urge to reward, praise, or be grateful can be found.

Fundamentally, however, I think the moral asymmetry between positive and negative reactive attitudes rests on the fact the negative reactive attitudes imply, in various degrees, the infliction of suffering on the one to whom we are reacting: to blame someone implies the desire for the person that we are blaming to feel bad about what they have done, to be angry implies a desire of revenge, and, of course, the wish to punish corresponds precisely to the desire of inflicting suffering on the other. When we are thinking in moral terms, the infliction of suffering on someone is problematic and must be justified – whether we do this in consequentialist terms, with an approximative calculation of pleasures and pains, or in deontological terms, appealing to moral principles that tell us that suffering in a specific case is just. In this respect, positive reactive attitudes are radically different: they are inclined towards the pleasing of the other, and pleasure does not bear the same moral complexity of suffering. From the moral point of view, we do not need to justify ourselves for loving someone or feeling gratitude or compassion.

As Martha Nussbaum remarks, negative reactive emotions and attitudes, far from being the guardians of human dignity, are "central threat[s] to decent human interaction",[761] and she shows how the non-violent movement guided by Gandhi, Mandela and King – albeit with their differences – relies precisely on the rejection of negative, retributive attitudes, but not of their positive counterpart. It may be most

[761] Nussbaum, *Anger and Forgiveness*, 14.

evident in Mandela's dealing with the aftermath of the end of the apartheid regime in South Africa and the choice to refrain from indulging in vengeance and retributive practice, instituting the Truth and Reconciliation Commission, a court-like justice body that aimed at establishing the truth concerning the acts of violence that were perpetrated during apartheid. Victims were heard, investigations were conducted, culprits identified and then granted amnesty if, in return, they guaranteed full disclosure of their crimes.

Evidently, not everyone in South Africa was happy with this resolution, many complaining that they had been "robbed of their right to justice"[762], that is, the satisfaction of their desire for vengeance. Nevertheless, the Truth and Reconciliation Commission, as we will see at the end of the chapter in more detail, stands as a pillar of alternative forms of justice in recent history, showing that the remedy for violence need not be intrinsically violent itself. Of course, it does not consist of indifference and oversight, for rejecting the concept of individual moral responsibility does not entail that "anything goes" nor that moral distinctions should be denied *tout court*. It rather consists of the investigation of the conditions that made crime possible; as Faizel Randera, one of the commissioners, claimed: "If we cannot understand what made people think and do what they did these conflicts will arise again within our society."[763]

3.1.2 Co-responsibility

While discussing Hilary Bok's work, we opposed causal responsibility to moral responsibility: it means that recognising one causal and even crucial role in producing a certain event does not necessarily entail desert ascriptions in general and blame and punishment in particular. We now see – and I consider this last quote by Mr Randera to go in this direction – that a shift to causal responsibility also opens up the space for another kind of responsibility that is also not so intrinsically linked to blame, punishment and desert as traditionally moral responsibility is, is not so backwards-looking and is not so abstract. It is the space of shared responsibility, or "co-responsibility", as Charles Taylor called it precisely in relation to the Truth and Reconciliation Commission[764].

Causal responsibility generates shared responsibility because a causal outlook requires taking into consideration the whole causal chain – whether every single link is

[762] "Apartheid Enforcer Sticks to 'farcical' Story on Biko Killing," *The Independent*, 1997.
[763] Greg Barrow, "South Africans Reconciled?," *BBC News*, 1998, http://news.bbc.co.uk/2/hi/special_report/1998/10/98/truth_and_reconciliation/142673.stm.
[764] Charles Taylor, *A Secular Age* (Cambridge/London: Harvard University Press, 2007), 710.

known, or many, as one could confidently say after having delved into Robert Harris's story, for example, or not. In fact, even when causal connections are not so clear, or we don't have as much information concerning the circumstances, to take a causal perspective entails acknowledging that, however obscure, causal circumstances exist and influence the outcome; hence, getting to know them is of paramount concern.

It is important not to confuse these forms of responsibility with collective responsibility, which is typically invoked in order to justify collective punishment in total institutions like prisons, army barracks, and boarding schools and is usually enforced by authoritarian regimes. Collective responsibility fits entirely into a retributive logic and is positioned at the first stage of Heider's scale of attribution, as we have seen. Shared responsibility, or co-responsibility, is located at the other end of the scale, where the agent is just a node in a larger network and their responsibility is weighted against that of the other known or unknown factors that contribute to the event. It is no surprise, then, that Taylor resumes the concept of original sin[765] that we have seen in the first part of this work to have played a big part in Luther's and Calvin's critique of individual responsibility. In fact, Taylor remarks, only by not excluding oneself from evil, by the "recognition of common, flawed humanity",[766] can we escape the retributive logic.

In reality, Taylor talks of "forgiveness", a concept, on which we shall later return, that has a solid link to Christian thought, being Taylor a fervent Catholic. Forgiveness, however, does not elude the retributive logic: to forgive, to be merciful, means to ascribe individual moral responsibility, to believe blame and punishment would be appropriate responses to one's action, yet to renounce such a response as a supererogatory act. In such a view, abstaining from blame and punishment is not due.

There is a fundamental ambiguity here: blame and punishment are deserved, so assigning them to the agent would not only be *appropriate*, but it would be *right*. If we are to consider justice as a moral value, then it means that something *good* is achieved

[765] See for example: "This modern humanism prides itself on having released energy for philanthropy and reform; by getting rid of "original sin", of a lowly and demeaning picture of human nature, it encourages us to reach high. Of course, there is some truth in this. But it is also terribly partial, terribly naïve, because it has never faced the questions I have been raising here: what can power this great effort at philanthropic reform? This humanism leaves us with our own high sense of self-worth to keep us from backsliding, a high notion of human worth to inspire us forward, and a flaming indignation against wrong and oppression to energize us. It cannot appreciate how problematic all of these are, how easily they can slide into something trivial, ugly or downright dangerous and destructive." Taylor, 698.

[766] Taylor, 709.

when someone deserving blame and punishment is blamed and punished – if only the fact of having done something that is one's duty[767]. But then, in Christian ethics, to forgive – hence to abstain from something that is right, i.e., punishing, something that, as we have seen at length, also characterises God's action – is a morally supererogatory act.

As we have seen, I believe this logical short-circuit derives from the acknowledgement of the moral inadequacy of the retributive stance coupled with the propensity for vengeance that we have analysed. In fact, early Christianity in the first place, and the Reform later, tried to unhinge retribution either through indiscriminate mercy or through indiscriminate damnation, but these attempts have collided with the strong emotional impulse to retribution, and the inclination towards ethical intuitionism that led both philosophical analysis and common sense thought to try to make the two points of view coexist, giving rise to all sorts of inconsistencies.

So, contrary to Taylor, I believe that what we should morally aim for is not forgiveness, but rather the recognition that individual moral responsibility is a delusion, that desert is a myth, that punishment – once again, defined as the voluntary infliction of suffering for its own sake – is a moral aberration. Therefore, the concept of original sin, even in a secularised version, is not enough to illuminate the significance of this recognition. Secularised original sin simply contends that to separate the good from the evil is a fairytale, for, given certain conditions, anyone would find themselves on the villain's side, but this conception does not take us to the last stage of Heider's attribution development and what it implies. The space of co-responsibility, as Taylor calls it, generated here is more radical. It is not opened by Jesus' saying, "Let him who is without sin cast the first stone", for it is not the participation in the abstract category of sin that unsettles moral responsibility; it is not by being abstractly united in doing evil, but each on their own, that co-responsibility is created. The space of co-responsibility, I believe, is opened by the recognition of the interconnectedness of beings, of the common participation to the structures that generate deviance. We are not "not innocent" because we are merely able to do evil, and we all have done it in various

[767] I refer here to David Ross conception of "right", that is, "as synonymous […] with 'what is my duty'". Ross, *The Right and the Good*, 6. Although, according to Ross, a right act does not necessarily correspond to a morally good action – the difference between acts and actions being the first referring to a thing that is done, the latter to the doing of something in relation to the agent's intentions – a right act still is a morally good act: not an action, because something right could be done for bad motives, but an act that is completed and that corresponds to what ought to be done. Ross, 7.

degrees, but because we are nodes in the web that generate that very action that we condemn.

Such a stance should also respond to the proposal of the quarantine model of punishment that we have analysed in the previous chapter, when discussing forward-looking justifications for punishment. It is a justification that has been advanced by some critics of moral responsibility[768] who claim that just like we accept that people who bear no responsibility but constitute a danger to their fellows because they carry an infectious disease should be quarantined – hence deprived of their freedom of movement and their affective rights, among other things – we can accept that people who commit violent actions should be isolated from their fellows, too, because they proved to be a threat and we assume that the risk of recidivism is high enough[769]. We have already seen why the present carceral system does not correspond to the quarantine model, but let us imagine a carceral system that would respect the limitations this model imposes.

In this quarantine model, we are faced again with a lesser evil justification: it is not, in itself, "right" or desirable to inflict the suffering of isolation on sick people or wrongdoers. Nonetheless, it is better than doing nothing, for refraining from quarantining someone in both cases would hurt more people. Such a proposal clearly falls within the carceral paradigm – quarantine is its perfect analog, and the analogy would not work as well if the punishment administered were different. It seems to be aiming at justifying the *status quo* given a new conceptual framework, that of causal rather than individual moral responsibility: one shall remove the cause to solve the problem. Incarceration may not be *deserved*, but it is still our best solution.

However, as I tried to show, this implies a superficial reading of the shift to causal responsibility, for it avoids considering the causal chain and remains anchored to the idea that problems begin and end with the individual. Although not entirely to be discarded, as we shall see, this idea fails when confronted with the co-responsibility framework: to isolate one, therefore breaking all their social bonds, may incapacitate one for a limited amount of time but hardly makes any progress in changing the social conditions that generate crime – quite the opposite, we have seen how prison creates

[768] Pereboom, *Living Without Free Will*, 174.

[769] The assumption that a person who has committed a crime is more likely to reoffend than someone without a criminal record is to offend for the first time would need to be assessed both in itself and for different categories of crime, but we will take it for a given for the sake of argumentation.

new social connections that encourage recidivism, often through more severe crimes, as we should not forget that most people who commit wrongdoing are not incarcerated for their whole life.

This perspective also challenges so-called "therapeutical"[770] approaches to wrongdoing, like the dystopian proposal outlined by Herbert Morris. Just like in the quarantine scheme, therapeutical approaches equate wrongdoing with disease – a moral disease that needs to be treated. The moral treatment that is then to be administered looks very much like brainwashing, closely reminiscent of the aversion therapy depicted in *A Clockwork Orange*. However, the problem with such a stance is not merely aesthetic; it runs deeper than the revulsion that we feel when Alex DeLarge has induced nausea while witnessing rape imagery. Once again, the greater context of the causal chain is forgotten, the agent receives individual treatment and the surrounding conditions are left untouched, which means that the problem is far from being resolved. Even more, we could say that through the therapeutical approach these conditions are perpetuated. If we believe wrongdoing to be the result of societal conditions that prove themselves criminogenic, to try to instil yet again the same prevailing societal norms into "deviant" individuals, although it might render harmless that specific person, at the same time reproduces the very structure that generated wrongdoing in the first place. As Ted Honderich put it,

> What our law, police, courts and prisons now do, in general, is to create, maintain, strengthen or defend an economic, political and social system and indeed a conventional morality that does more for the improvement of already good lives than for the prevention of bad lives.[771]

The West is, in fact, homogeneously framed by the neoliberal culture that is characterised by extreme inequalities, marginalised groups – especially those plagued by poverty – "and a strong belief in the ideal of individuals as self-sufficient and self-made."[772] Quite simply, according to the neoliberal perspective, everyone gets what they deserve: poverty is a symptom of laziness, wealth of industriousness, crime of inherent wickedness, and so on. It is a culture of individual moral responsibility that moral therapy wishes to inculcate in its patients. A culture that relies on a dangerous delusion but rises to the paradigm of civilization.

---

[770] Morris, "Persons and Punishment," 480.
[771] Honderich, *Punishment. The Supposed Justifications Revisited*, 217.
[772] Waller, *The Stubborn System of Moral Responsibility*, 209.

In January 1989, Bruce Waller, concluding one of his papers, considering whether punishment should be considered treatment, asked: "Which virtues should be promoted?"[773]. It is somewhat ironic that Waller, a ferocious critic of neoliberalism, should publish this question just a few months before the collapse of the Berlin wall, which would extend the neoliberal order to most of the planet, erasing its principal (and deeply flawed) competitor and sanctioning the model of social organisation based on individual action and responsibility as the only suitable one.

The "sanitary" approaches of quarantine and moral treatment, despite negating the intrinsic value of the voluntary infliction of suffering on wrongdoers on the basis of a challenge to moral responsibility, continue to operate within the neo-liberal paradigm, as if the consequences of the rejection of individual moral responsibility were not fully appreciated. This results in forms of instrumentalization of the individual for the benefit of the neo-liberal society rather than a broader rethinking of practices and theories of justice.

When this is the assumed point of view, it is not surprising that those who fall within the scheme of retributive justice may claim their retributive punishment as a form of respect, even a right, as Morris put it: "a right to be punished" that "derives from a fundamental human right to be treated as a person; […] a natural, inalienable, and absolute right"[774] rooted in the individual's right for "institutions [to] respect his choices"[775] and moral status[776], something that "[o]ur punishment system does; our therapy system does not"[777]. While one could agree with Morris that therapeutical approaches to wrongdoing do not respect the wrongdoer (whatever that may mean, as

[773] Waller, "Denying Moral Responsibility: The Difference It Makes," *Analysis* 49, no. 1 (1989): 47.
[774] Morris, "Persons and Punishment," 476.
[775] Morris, 486.
[776] Morris, 487, 490. I'd like to add briefly that the retributive system based on moral responsibility is also favoured by its victims – the wrongdoers – because it appeals to the just world bias analysed previously. Whereas paradigms based on moral luck can make one feel as a victim of the circumstances, apparently relegating them to a passivity that one would hardly desire for themselves, the idea that suffering is *deserved* as a result of free choice (of free will) makes one feel as an active forger of one's destiny: I could have avoided punishment and I will be able to avoid it in the future. It is a very powerful psychological mechanism that, as we have seen, allows people to face tragedies avoiding the feeling of helplessness. Of course, positive psychological consequences do not count as elements that make a certain worldview *true*, however, they may be reasons to choose a deceptive approach like Smilansky's. Contrary to that, I believe that an approach that lets go of individual moral responsibility but still recognises the active contribution of individuals to an outcome is possible and desirable, as I will try to show shortly.
[777] Morris, 486.

Morris does not bother to explain it), why the retributive system should grant that respect remains an utter mystery.

As Waller, referring once again to Robert Harris's case, points out:

> Locking Harris in a cage until we drag him out and strap him down in an execution chamber: this shows respect for him as a human being. But recognizing that Harris is, like all of us, shaped by forces that were ultimately beyond his control: that is a patronizing and disrespectful stance.[778]

Morris opens *Persons and Punishment* with a quote from Lester Maddox, governor of Georgia from 1971 to 1975, who states, "We have no right to treat a man like a dog."[779] Leaving all comments concerning the speciesism of this sentence aside (how are we allowed to treat a dog?), it remains unclear why the treatment reserved for Harris just described should be different or better than that reserved for a dog.

In a word, the fact that sanitary paradigms are utterly inadequate should not count as a reason to maintain a retributive stance. Co-responsibility unhinges retribution without falling into an individual's instrumentalisation, but understanding it requires a far broader gaze; it needs to shift focus away from the individual and think of systemic solutions. Co-responsibility denies individual moral responsibility as the core of the reflection of justice in a more radical way, thereby unhinging not only retributive justice but also all the solutions to crime-related problems that focus on individuals only. However, contrary to what it may seem, agents in this paradigm are encouraged to *take* responsibility. If we deny individual moral responsibility, what might this mean? Once again, Bruce Waller comes to our aid:

> The confusion is generated by two senses of responsibility: moral responsibility, and taken responsibility. Though we use the same word for both, they are in fact quite different senses of responsibility: distant relatives at best, with little in common other than, unfortunately, the same last name. Moral responsibility is the stuff of just deserts and retribution and right rewards; in contrast, taken responsibility is the responsibility we claim for our roles or offices, our acts, or even ourselves. Taken responsibility is vitally important. If we could never take and exercise responsibility, then our lives would be impoverished: we might be slaves, or infants, or incompetents, but not moral agents. However, the responsibility one takes is not the moral responsibility of just deserts.[780]

[778] Waller, *The Stubborn System of Moral Responsibility*, 139.
[779] Morris, "Persons and Punishment," 475.
[780] Bruce N. Waller, "Virtue Unrewarded: Morality without Moral Responsibility," *Philosophia* 31, no. 3–4 (2004): 435–36.

To take responsibility means to make decisions for oneself, take ownership of one's actions, and not passively follow someone else's lead[781]; it means to make one's choices their own through a deliberative process[782] – in Frankfurt's terms, it means to be a person instead of a wanton, that is, to be guided by second-order desires, i.e., the conscious choice to pursuit a certain (first-order) desire[783]. In this perspective, taking responsibility means recognising oneself in one's life. It also means allowing others to rely on oneself as an active agent who recognizes one's role in the intricated web of relationships and does one's best to be a trustworthy peer, fulfilling commitments to the best of one's ability[784]. However, should one fail to live up to the standards, should one deceive their fellows and themselves, the fact that they have taken responsibility would not be in itself a good basis for moral blame, for deserving punishment. As Waller claims, there is nothing intrinsically moral in taken responsibility: it concerns "our roles or offices, our acts, or even ourselves",[785] and it must not be confused with "the stuff of just deserts and retribution and right rewards"[786].

Among many who may agree with the analysis developed so far, the idea remains that, on a practical level, nothing should change with respect to our practices of justice: social policy considerations – deterrence, incapacitation, and rehabilitation – ordinarily blunt the radicalism of the denial of moral responsibility. Beardsley does hint at the compassion and equanimity that should arise from the assumption of a causally unlimited perspective[787] , but an institutional reflection is completely lacking, and it is not uncommon for moral philosophers to just accept the *status quo* as the least-worst option.

Not surprisingly, as Courtois remarks, at the very beginning of the philosophical criticism of anger, vengeance, and the related attitudes are the Stoics, who, living in a

[781] Waller, 438.
[782] Waller, "Deep Thinkers, Cognitive Misers, and Moral Responsibility," 227.
[783] Frankfurt, "Freedom of the Will and the Concept of a Person."
[784] "Dave can *take* responsibility for designing the departmental website. It's his responsibility, and if you take over the project, he will be resentful: you are implying that he is incompetent to carry out the responsibilities he claimed, and you are undercutting his opportunity to make choices and exercise control. If Dave's opportunities to take responsibility are consistently blocked, then he has little or no scope to function as an autonomous moral agent who exercises effective control. But this vitally important *taken* responsibility is not *moral* responsibility. For if Dave totally botches the website project, we might reasonably and consistently say: Dave has complete (taken) responsibility for the web site; but he is not (morally) responsible for failing at the task, […]." Waller, "Virtue Unrewarded: Morality without Moral Responsibility," 436.
[785] Waller, 436.
[786] Waller, 436.
[787] Beardsley, "Determinism and Moral Perspectives," 17–18.

period of centralisation of power and collapse of the *polis* in ancient Greece, shunned participation in public life and, hence, rejected institutional reform thinking[788]. The stoic aims at individual autonomy in the context of an institutional framework that does not interest them and is simply accepted as it is.

The need to accept current institutions and the intuitions that underpin them is the main point of Smilansky's theory of the "necessary illusion" of free will and moral responsibility, as we have repeatedly seen. He then maintains that the moral world needs to be permeated by what he calls "structural injustice"[789], that is, the punishment of those who do not deserve to be punished and its illusory justification, which at the same time generates an

> astounding and grotesque picture of life as a masquerade in which maintaining human dignity depends on utilizing weakness (our proneness to self-deception), and thus degrading fellow people, in which enormous dishonesty and ignorance are conditions of maintaining happiness, much of common social interaction, morality, and justice, and where the love of knowledge and rationality are so often dangerous 'enemies', […].[790]

The reference to "human dignity" makes clear Smilansky's deontological approach to ethics – one that is particularly prone to these kinds of paradoxes, for certain principles are taken as pillars of morals unwavering in their confrontation with the empirical reality of the world. It reminds Kant's stance that someone guilty of a crime should be sentenced to death and executed even in the face of a society that is collapsing and that could not gain any benefit from the criminal's punishment. Similarly, Smilansky assumes *a priori* that justice in particular and morals in general must be based on moral responsibility and desert, and when his analysis of free will brings him to the conclusion of the fundamental moral equivalence of beings, he is unable to let go of his *a priori* principles and ends up in the paradox that, in order to "save" morals, one needs to perpetuate injustice. And more than just that, it is precisely this disguising of injustice as justice that fundamentally constitutes morals.

Unfortunately, it is not as if the consequentialist side of the debate is in a better position. On the contrary, utilitarians are the champions of the lesser evil solution – one could say that, at least, compared to Smilansky, they have the virtue of accepting

[788] Courtois, "Le Sens et La Valeur de La Vengeance, Chez Aristote et Sénèque."

[789] "Structural Injustice by its very nature is part of *any* social fabric, however moral, and cannot be avoided." Smilansky, *Free Will and Illusion*, 257.

[790] Smilansky, 278.

injustice and unavoidable suffering for what they are, evils in themselves, and not misrepresent them as values. This approach is well exemplified by Smart, who is ready to accept punishment (both of the guilty and the innocent in exceptional cases[791]) on the basis of "a clear pragmatic justification which is quite consistent with a wholehearted belief in metaphysical determinism"[792] in full Benthamian style. Not only that, he also defends praise and blame as ways of dispassionately[793] evaluating actions for which responsibility is ascribed[794] – although Smart does not make that clear-cut distinction, he seems to be referring to a kind of responsibility similar to Waller's "taken responsibility" and not "moral responsibility" in the traditional sense. Hence, Smart apparently essentially misrecognises the fundamental retributive character of praise and blame: despite mentioning that "most men do not […] praise and blame people in this dispassionate and clear-headed way"[795], he does not seem to believe that using praise and blame as simple evaluative judgements fundamentally mistakes the meaning of the words. In this, I think Smart fails to recognise that praise and especially blame are intrinsically connected to our emotional reactions. As Wallace points out:

> [W]e might say that to hold someone morally responsible is to be willing to blame the person for violating moral precepts that we ourselves accept for purposes of practical reasoning, deliberation, and public normative discussion, where blame, in turn, simply expresses the belief that those precepts have been breached. But I take it that an account along these lines would not capture what is distinctive about the stance of holding people morally responsible. Blame would be rendered superficial on this account, reduced to a way of describing what an agent has done, and perhaps registering a causal connection between the agent and the action so described. True moral blame, by contrast, is a form of deep assessment, reflecting an attitude toward the agent who has acted wrongly that finds its natural expression in sanctioning behavior (avoidance, denunciation, reproach, censure, and the like). The reactive emotions are needed to explain this attitudinal aspect of true moral blame, and to account for its natural connection with sanctioning behavior.[796]

Although there is much to be said about utilitarianism and alternative approaches to justice, I will leave it to the end of this work.

If we are to find a space for reflection on the consequences of the dismissal of moral responsibility, we need to look at political philosophy, where thinkers have been bolder in the exploration of alternative paradigms.

---

[791] Smart, "An Outline of a System of Utilitarian Ethics," 70–71.
[792] Smart, "Free Will, Praise, and Blame," 302.
[793] Smart, 305.
[794] Smart, 304–5.
[795] Smart, 305.
[796] Wallace, *Responsibility and the Moral Sentiments*, 78.

3.2 A Social Justice Approach to Crime

We have seen how, in common thought, the concept of justice naturally refers to retributive justice: when something is deemed "unjust" or "unfair", it is typically to indicate something that is considered *undeserved*, whether it refers to how someone is treated by someone else, a disease, a natural disaster, an unfortunate circumstance. I have also already mentioned that when one types "justice" into Google, among the first fifty images, forty represent a two-pan balance, often held in the hand of a blindfolded goddess who is also holding a sword in the other hand. That is the shared imagination of justice: an impartial, violent balancing of wrongs. This means that, in common thought, justice is both retributive and distributive: retribution is achieved by the distribution of a specific kind of burden within society, i.e., punishments.

Punishments are special burdens because, contrary to those that are otherwise the objects of distributive justice – like the taxes that have to be paid to support the welfare state or the dwellings that will be disturbed by the construction of a new highway – they do not exist independently of an individual's action as externalities that are generated by the needs of communal living in society. Rather, they are created by the State and distributed to a class of its citizens (the offenders) because it is commonly believed that something good will be achieved through this distribution, i.e., *justice*. We have also seen how this burden creation and distribution that we call punishment has been justified on consequentialist grounds as a lesser evil. It is still widely considered as the appropriate (because necessary) response to crime: common thinking moves within the retributive paradigm, generating a naturalisation of retribution, for the inability to think of alternative responses makes retaliation akin to natural law. However, retaliation *is not* a natural law.

At the end of the 19th century, Guyau, a philosopher who has been closely associated with anarchist thinking[797], was writing:

> The classic moralists are accustomed to see in natural sanction an idea of the same order as that of expiation, nature begins, according to them, that which human

[797] Guyau's political affiliations are unclear as he never actively participated in the political life of his time and died very young, at the age of 34. However, his *Sketch* had a strong influence on Pyotr Kropotkin, who identified Guyau as one of the founders of the anarchist ethic. Peter Kropotkin, *La Morale Anarchiste*, ed. Jérôme Solal (Paris: Mille et une Nuits, 2024); Federico Testa, "Jean-Marie Guyau's Presence and Absence in the History of Philosophy," *Plí: The Warwick Journal of Philosophy* 30 (2019): 237.

> conscience and God have to continue. Whosoever violates natural laws, therefore, finds himself already punished in a way, if we may believe them, which forewarns of, and prepares, the punishment resulting from moral laws.
>
> In our eyes nothing is more inexact than this conception. Nature punishes no one […], and nature has nobody to punish, for the reason that nobody is truly guilty with respect to nature. One does not violate a natural law, or it would not be, in that case, any longer a natural law. […] [T]he indigestion of a glutton and the drunkenness of a tippler have by nature no kind of moral and penal character whatever; they simply allow the patient to calculate the force of resistance which his stomach or his brain can offer to the hurtful influence of a certain mass of food or a certain quantity of alcohol.[798]

Although it does happen that when one behaves contrary to certain accepted norms, this would generate some form of suffering – the stomach ache of the gluttonous, the STD of the promiscuous – this is far from a universal rule: considering it such is part of the just-world bias, with all the related problems that we have seen – especially the blaming of the victim, which happens so often because it is anything but rare that someone would find themselves in unfortunate circumstances without having done anything wrong. If, however, some commonsensical moral rules reflect some (limited) regularities between a certain action and ensuing suffering, like the fact that a promiscuous sexual life increases the chances of getting STDs, this, of course, does not mean that nature punishes the vice of promiscuity, but that many moral rules – like the encouragement to chastity – reflect practical insights that facilitate communal life in society. Common sense morality often reinforces rules motivated by practical reason, so the link between the violation of a rule and a subsequent form of suffering is not accidental, but it is not intrinsic to the moral nature of the interdiction either: it is not a form of natural retribution.

If one wants to find something "natural" in retribution, one needs to go back to the idea of the retributive emotions: as we have seen, such emotions arise spontaneously in human and non-human animals for evolutionary reasons, and evolution is indeed a natural mechanism; however this does not mean that they are bound to occur – that offences are connected to retributive emotions and attitudes by an inviolable law, that anger and the desire for vengeance must necessarily arise as a consequence of an injury – nor, by any means, it gives them a moral value of any sort.

3.2.1 Solidarity

We have also seen how retributive justice constitutes the primal function of the State itself according to most social contract theories: humans in a lawless state of

[798] Guyau, *A Sketch of Morality Independent of Obligation and Sanction*, 154–55.

nature live precarious lives exposed to the risk of the other's violence; hence, they form a rational agreement to delegate the power of violence to the State that, in turn, protects them. In a word, they ban violent action in civil society and delegate the right of retaliation to the State when the prohibition is violated: the State is the citizens' avenger. Retributive justice is also the fundamental justification according to Girard's theory of the State that, as we have seen, is based on opposite foundations compared to social contract theories, not falling into the seduction of rationalism and establishing the foundations of the State on his theory of mimetic desire. Still, the structure of the theory is the same: the starting point is that mimetic desire generates a mimetic crisis, i.e., a crisis of generalised violence, that corresponds to the social contract theories' state of nature. The crisis, however, is not overcome by a rational agreement but by the accidental convergence of violence on a single innocent victim, the scapegoat.

It is not the State that originates from this process but religion, and the State is introduced by Girard as an ensuing structure which fulfils the same function of religion: the managing of human violence. Whereas religion serves an entirely preventive function, inhibiting the first acts of violence that would lead to the eruption of further crises, the State – the details of whose origin are not made explicit by Girard – is preventive through repression: once violence occurs, the State takes retaliation into its hands, implementing an act of violence that cannot be avenged in its turn because the avenger is too powerful to be attacked, therefore putting an end to the cycle of the imitation of violence. Again, the fundamental function of the State lies in retribution – although, as we have seen, the victim of the State's retribution is also sacrificial, victimisers becoming scapegoats through their own victimisation. It is not surprising: internecine violence is the most significant danger for a human group, and, as we have seen, because of our reactive attitudes, the spontaneous reaction to violence takes the form of retribution. This is not to say that the State's retaliation is justified as a hypostatisation of the spontaneous animal reaction. Rather, the State's response to offences needs to take the form of retaliation to provide satisfaction for the angry victims, for if their desire for vengeance remains unsatisfied, they will take vengeance into their own hands. As Girard maintains, there is no difference in the principle underlying "justice" and vengeance: the only difference lies in their actor and, hence, their consequences.

To corroborate these theses, we can recall, as Durkheim has observed, that law in archaic societies is entirely penal in character[799]: the first laws concern punishments. Not only that, as Durkheim himself notices, penal laws do not simply "concern punishments" but consist entirely of punishments.

> [Penal law] does not first state, as does civil law: This is the duty; but states immediately: This is the punishment. Undoubtedly if an act is punished, it is because it is contrary to a mandatory rule, but this rule is not expressly spelt out.[800]

Durkheim also observes that, given the fact that across different cultures the category of acts that are punished, i.e., crimes, varies so widely, what makes them all crimes cannot lie in the specific content of the action that is punished. Instead, crimes are united by the public aversion that is addressed to the punished action: what makes an action a crime is the offence to the collective conscience, a conscience embodied by the rule that is violated[801]. This means that something can be a crime without being a violent act, without anybody being hurt by the criminal action: incest (even between consenting adults) and blasphemy, for example, have been criminalised in virtually all cultures.

Does this disprove the thesis I have been advancing so far? That criminal law serves the purpose of avenging personal offences in order to avoid the generalisation of violence through vendetta cycles? No, I do not think the thesis is disproven; rather, it is enriched. Durkheim himself claims that even though there are victimless crimes, on the other hand, it must be noted that there is almost no victimisation that does not generate crime; that is, violence is always criminalised. Although it has been maintained that archaic societies did not recognise crimes against the person, like murder or theft, these claims are based on cherry-picked misleading data.[802] The examples brought up by the authors making these claims, like Cesare Lombroso, always relate to exceptional circumstances – war, sacrifice, patriarchal authority – that are generalised for the political goal of depicting such societies as deprived of morality but, as Durkheim maintains,

> from the first moment when men form together in a society, however rudimentary it may be, there are necessarily rules that govern their relationships, and

[799] Durkheim, *The Division of Labour in Society*, 36.
[800] Durkheim, 35.
[801] Durkheim, 40.
[802] Durkheim, 174–75.

> consequently a morality which, although not resembling our own, nevertheless exists. In addition, if there is a rule common to all these moral codes, it is certainly the one that forbids attacks against the person, for men who are similar to one another cannot live together without each feeling for his fellows a sympathy that revolts against acts of any kind that will bring suffering upon them.[803]

As should be clear, however, I do not agree with Durkheim that the reason why attacks against the person are always criminalised lies in reciprocal sympathy; rather, I believe, with Girard, that violence against the person is the seed for generalised violence through vendetta cycles that destroy societies, and that penal law serves to prevent it through institutionalised retaliation.

However, it is true that not all acts that generate victims are criminalised: Durkheim speaks of economic crises, crashes of the stock markets and bankruptcies as events that are "disastrous for society without suffering the slightest repression"[804]; similarly, more recently, Victoria Law has mentioned the tycoon family that commercialised OxyContin and, with its aggressive business strategy, provoked the American opioid crisis, or the board members of corporations responsible for environmental disasters, as examples of agents that generated enormous suffering without incurring in institutionalised punishment as a consequence[805].

This, again, does not refute but rather further substantiates my thesis: it shows that the problem retributive justice systems in the State address is not harm in itself, but that kind of harm that elicits retaliation. This happens when there is someone who is recognisably victimised and a perpetrator who is clearly identifiable: if I think *you* did something wrong *to me*, then I will seek vengeance. In a State, this task is delegated and not waived. An economic, environmental, or public health crisis, on the other hand, is perceived as the result of much broader, more complex processes. Although some people are harmed as a result of some other people's choices and actions, connections are much harder to establish and, even as they are discovered, they are not perceived as personal attacks and, therefore, do not generate a desire for vengeance.

But how do we explain the criminalisation of victimless acts, then? As Durkheim remarks, in these cases the victim is the public conscience – but an abstract entity is not a victim who can take revenge; hence, the thesis I advance seems to be disproven. But the collective consciousness does not exist in and of itself: it is embodied

[803] Durkheim, 116–17.
[804] Durkheim, 33.
[805] Law, *"Prisons Make Us Safer,"* 17–18.

in the individuals that constitute the community. Just like we remarked that, in *ancien régimes,* the violation of a law corresponds to an offence towards the monarch who then seeks revenge, so, as the members of the community have internalised the collective consciousness, they are all offended by the breach of the law. However, another, more concrete element should be added to the picture: once a rule has been breached – whether the breaching creates direct harm to an individual or not – all the members of the community are put in danger: if someone does not respect our laws, just like they did not respect a victimless prohibition or one that does not victimise me personally, so they may violate another law and I may be personally affected the next time. A harmless crime bears the seed of potential harm: if one does not respect the community's authority then everyone is in danger, so everyone feels attacked, hence desires retaliation.

To this, it may be added that victimless crimes reconnect the preventive-through-retaliation State system with more ancient preventive-through-sacrifice systems: they remind us that vengeance and justice are originally dealt with through religion, the creation of transcendence. In Durkheim's theory, it is through the feeling of *effervescence*[806], the ecstatic thrill that is experienced when individuals living in small communities gather in bigger groups, generating this feeling of sacredness, both fascinating and scary – the sacred is the outcome of the feast. Girard's theory is similar but of opposite sign: whereas Durkheim finds religion in the positive excitement of the joyful gathering, Girard locates it in the negative frenzy of a much more macabre feast: the original lynching of the scapegoat, the collective founding murder. It is still its collective element that generates the feeling of transcendence that is inherent to sacredness: in both cases, what is ultimately divinised is the community itself.

This helps us understand the importance of crimes whose victim is the immaterial collective consciousness: they are the bridge connecting these older preventive systems based on the divinisation of solidarity bonds and the institutionalisation of retaliation. They are a reminder that before we can build a State with its institutions to protect individuals, we need a community, and it is the community that is symbolised by the collective conscience that is offended by victimless crimes. Retaliation against the offender of the collective conscience is

[806] Durkheim, *The Elementary Forms of Religious Life*, 218.

comparable to the ritualised sacrifice, a form of renovation of the communal bond through the reaffirmation of the values that hold the society together.

As we mentioned before, Durkheim recognises that "penal law is not only of essentially religious origin, but continues always to bear a certain stamp of religiosity. This is because the acts that it punishes always appear as attacks upon something which is transcendent"[807], and although this transcendence can be identified with personified Gods – through the divinisation process described by Girard – what it really is is the power of the community itself, a sacred power as it is both scary and marvellous since human gathering can bring both a crisis of generalised violence and the tranquillity and comfort of peaceful shared life. This explains why national flags are considered sacred or quasi-sacred almost everywhere, and most countries have laws that criminalise "flag desecration" – the term here used is in itself very telling – and punish it with several years of imprisonment[808]. National flags are the symbolisation of the national community, and their widespread sacralisation is an offshoot of the origin of the sacred itself.

As Dumouchel remarks, the rise of the nation-State based on territoriality has restructured the relationships of solidarity that were previously organised according to the three concentric circles theorised by Verdier that were analysed in the first chapter. To recall them, Verdier theorised that archaic societies were organised through three concentric circles defined by the social distance among their respective members. The first circle gathers the family, the people that are closest to one another: it is characterised by full mutual solidarity and the prohibition of vengeance as a means of dealing with interpersonal problems. Quarrels are solved through reconciliation rites or, in extreme cases, expulsion. The intermediate circle introduces a degree of social distance: different family groups recognise each other as peers and establish relationships of reciprocity; it is the reign of both gifts and vengeance, the latter being regulated by traditional norms that impose proportionality, averting the risk of escalation. Finally, the third and last circle regards the relationship with other outer

[807] Durkheim, *The Division of Labour in Society*, 56.

[808] Just to give a few examples: Italy punishes flag desecration with up to two years of imprisonment, Germany, Israel and Ukraine with up to three years. Non-Western countries have even more severe penalties, with Algeria and South Korea reaching up to ten years, Thailand up to six years and Turkey up to eighteen years. The United States and the United Kingdom, on the other hand, do not consider flag desecration a crime. Of the forty-eight countries on which we have data on flag desecration laws, thirty-four punish it, of which only two, Finland and New Zealand, limit punishment to the payment of a fine.

societies that do not share culture and traditional norms. They are not recognised as peers or equals; there is no reciprocity; they are strangers and are the object of unlimited hostility through pillage and war.

Scubla had noticed how the modern State made the first two circles collapse onto one another: as it obtained the monopoly of legitimate violence, he maintains, the inner circle grows to incorporate the intermediate one, making national identity a sort of clan identity.[809] While I agree that the two circles do collapse to form a single space, I do not share Scubla's view that this results in the expansion of the first circle at the expense of the second – to me, it looks like the two merged, preserving original elements of both instead. The function of the State as the institutionalised avenger of wrongdoings shows that vengeance is not excluded as it was in the family circle, although it is also not meted out personally as was in the intermediate circle but delegated to a third, neutral, and more powerful entity.

This expansion of retributive attitudes to the innermost circle, i.e., the family, is not without problems: if the State system requires crimes to be dealt with publicly through the criminal trial and subsequent incarceration, victims whose perpetrators are within the household tend not to report crimes for several reasons. First and foremost, such victims often have feelings of love and affection for their perpetrators, especially in cases of domestic violence and child abuse: although victims are hurt, perpetrators are still their partners, parents, and siblings, and the perspective of sending them to prison is often daunting[810]. Moreover, close relatives play a crucial role in the economic support of the family, which could encounter difficulties if someone were to be imprisoned[811]. Finally, from a more Girardian perspective, it is worth mentioning that crime report and subsequent incarceration can generate long-lasting quarrels within the family that might have harsh repercussions on other family members: the dispute may expand and escalate – for, although the actual avenger is the State, reporting the crime can be considered a form of violence and revenge to be avenged in return. As we can see, the prohibition of vengeance within the internal circle was far from unjustified.

But what concerns us mainly in this chapter is the expansion of solidarity bonds that now encompass all members of the nation-State that Scubla emphasises: fellow

[809] Lucien Scubla, "La Place de La Nation Dans Les Sociétés Individualistes," *Droit et Culture* 39, no. 1 (2000): 191–210.
[810] Law, *"Prisons Make Us Safer,"* 136, 143.
[811] Law, 17.

citizens are all "friends", meaning people whose destiny is taken into account and for whom sacrifices can be made, whereas those who are not part of the nation-State are "foes", potential enemies, still part of that outer circle where war is waged. This consideration leads us to consider another form of justice that has been theorised by political philosophy mainly as complementary to retributive justice: social justice.

Although we have traced the origins of the concept of justice back to retribution, and we have seen how retributive justice dominates the collective imagination on justice, at least since the publication of Rawls' *A Theory of Justice*, social justice has prevailed in the research on justice in political philosophy. Social justice concerns the structuring of society so that the benefits and burdens of living together are distributed fairly. This distribution is generally interpreted in economic terms: collective solidarity within the nation-State, at least after the end of World War II, takes the form of the welfare state that attends to inequalities, making sure that economic hardship does not prevent a fellow citizen from enjoying their fundamental rights. In Rawls' theory, this gets to the point of justifying redistribution if growing inequalities disfavour the worse off. It is, in fact, commonly accepted that economic misfortune is often due to bad luck, and it is considered fair that the society as a whole should care to alleviate the unfortunate's undeserved burden – if the unfortunate is a citizen of one's own nation-State.

Social justice dispenses with the whole conceptual edifice that is related to retributive justice and that we have analysed in the previous chapter, the one composed of desert, individual action, personal responsibility, and free will. The shift to social justice has been, indeed, revolutionary in changing how justice is conceived, leading to its depersonalisation. As Dumouchel has claimed, it has brought

> […] a radical transformation of the concept of justice as it had been understood until then, for it severed the notion of injustice from that of individual or collective fault. Just as there can be victimless crimes, victims of social injustice would be victims of injustice not committed by criminals. […] Justice and injustice become characteristics of social states rather than normative features specific to relations between people.[812]

So, is social justice actually complementary to retributive justice? Do they cover different and non-overlapping fields of life in society, the former concerning society's economic structure and the latter individual action and personal moral responsibility?

[812] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 145–46.

They are often presented as such[813], but we have seen, analysing Spinoza, that, like social justice, retributive justice is also a form of distributive justice: the *suum cuique tribuere* is the distribution of benefits and burdens according to one's desert. If social justice is a form of distributive justice emancipated from the concept of desert, could it not encompass the field normally covered by retributive justice?

3.2.2 Social Justice: More Than an Economic Issue

Although generally understood in economic terms, the redistribution that is carried out by social justice, in truth, is instrumental for the distribution of another kind of goods: rights[814], which are proxies for some minimal conditions that make life tolerable. It is believed – and rightly so – that without having access, even if limited, to a certain degree of education, healthcare, personal safety, or even leisure, life would not be dignified: it is a task of the State to provide these services through redistribution. Hence, the wealthiest are taxed more to allow those who are worse off to access these amenities even if they cannot afford them. We see how economic redistribution is an instrument for the more fundamental distribution, that of rights.

However, many of the rights that are considered universal and independent from desert are revoked when one commits a crime: the offender is excluded from the community of fellow citizens, and social justice no longer concerns or includes them, at least in part. Luckily, in Western countries, offenders do not lose *all* their rights – at least on paper: they retain their right to healthcare, personal safety, and education, to name a few. However, the total institution of prison, so secluded from all other spaces of public life, makes it hard to verify the actual respect of these rights and disinterest towards the fate of incarcerated people is justified on the basis of desert. It is far from uncommon to hear that carceral institutions violate even the most fundamental rights, where incarcerated people suffer from delays in healthcare that can prove fatal[815], are

[813] See, for example, David Miller's discussion of desert-judgements: "Such judgement, I believe, play at best a marginal role in our thinking about social justice. They are more important in […] discussions of punishment." Miller, "The Concept of Desert," 197.

[814] Given the utilitarian framework of this work, talk of rights might be surprising. However, public policies are often based on rights, on the idea that there are things that are due to people, morally and politically, and that are universal and inalienable. From a utilitarian perspective, rights are always liable to be revoked, but the concept can still be used as a rule of thumb for what, in normal conditions, we have reason to bestow citizens in order to grant them a dignified life. The concept, although thus transformed, can still be saved.

[815] Innumerable cases are reported by Victoria Law in her book's chapter on healthcare, including a woman who died of meningitis in 2017 because she was accused of faking symptoms and, hence, was delayed access to the hospital; a man who was denied medical treatment and died after entering a diabetic coma in 2014; and a woman who died of heart attack as she was being mocked by officers during heroin

precluded from access to education[816], and are victims of police violence[817]. When these cases hit the headlines, they are prosecuted by the law, but this happens rarely, and it is also commonly believed that the incarcerated person is, ultimately, the one to blame: if they had not broken the law, they would not find themselves in this situation.

These observations seem to corroborate the idea that social justice and retributive justice coexist complementarily, but not as it might have seemed, one having to do with the economic distributive aspect of justice and the other with individual action and its moral consequences. Rather, social justice is thoroughly applied as the main practice of solidarity within the community of fellow citizens; retributive justice, on the other hand, is the form of de-solidarisation of contemporary societies. At first sight, practices of retributive justice seem to reconstitute the second circle of Verdier's societal structure, the circle of reciprocity that Scubla believed to have disappeared from contemporary society. However, reciprocity, as we have seen, does not only imply retaliation – which in this case is the State's punishment – but is essentially defined by the consideration of the other as a peer, an equal. This is precisely what is denied to the wrongdoer who, having offended the collective conscience, is degraded: they are not our fellows anymore.

Breaking the law becomes, thus, a victimary sign and the offender, who is now situated on the border of the national community – for they are born and raised as a fellow member, most likely sharing the language and the predominant culture and religion, but have then repudiated the sacralised collective conscience through the commission of a crime – can become the sacrificial victim, repeating Girard's sacrificial scheme.

If we think of retributive justice in these terms, then we see that it is not excluded, in principle, that its complementarity with social justice may be overcome:

withdrawal in 2014: Law, *"Prisons Make Us Safer,"* chaps. 9. People in prison "jump the line" for life-saving medical care.

[816] A recent case in Italy involved incarcerated people having to choose whether to attend school or take advantage of air time: https://www.lastampa.it/torino/2024/01/26/news/istruzione_studio_carcere_lorusso_cutugno-14022485/ (accessed on 29/03/2024, 14:56). I have personally met incarcerated people who had to choose between attending school or accessing the facility's showers.

[817] Unfortunately, these cases are so common that examples are barely needed. Here are a couple that reached the headlines in Italy recently: 105 prison police officers at the Santa Maria Capua Vetere prison have been recently remanded for trial for torture, injury, violence, abuse of authority, and manslaughter for events that occurred in 2021: https://www.ildubbio.news/carcere/torture-ai-detenuti-nel-carcere-di-santa-maria-capua-vetere-105-rinviati-a-giudizio-bwybgz88 (accessed on 29/03/2024, 14:57); 10 prison police officers at the Foggia prison have been arrested on March 18th 2024 for torture: https://www.ildubbio.news/carcere/detenuti-torturati-nel-carcere-di-foggia-arrestati-10-agenti-penitenziari-cqtlw47s (accessed on 29/03/2024, 15:01).

social and retributive justice do not occupy adjacent but discrete spaces of justice, in which the term justice is used in radically different and incommensurable connotations. This perspective opens up the possibility for social justice to erode the space of retributive justice, to supplant institutionalised violence based on desert, if one believes that societies can give up the institution of sacrifice, at least in the institutionalised form of the State's punishment.

If we are to follow Girard to the end, humans cannot give up on sacrificial violence; all societies are sacrificial systems, as human communities need an escape route for their internecine violence. However plausible, whether this is true remains an open question verifiable only on the empirical level – as is the possibility that the violence of war, directed towards the inhabitants of the third circle, the circle of hostility, would be enough. In Girard's theory, the fact that the members of the third circle are external to the group – in fact, this same exteriority defines them – does not allow for proper sacrifices, for their victims need to be on the margin of the community itself in order to preserve the connection between the sacrifice and the act that originated it: the founding homicide.

The victim, then, needed to be not too integrated to avoid the risk of retaliation by their peers, but also not wholly external, for the group needed to believe that they were responsible for the crisis of generalised violence that was endangering the community's existence. Hence, the sacrificial victim now occupies a similar liminal place, being the further substitution of the original substitute victim. However, Girard recognises that sacrificial violence "is always less internal than the original violence"[818] and that "it tends toward the exterior"[819], concealing "the site of the original violence, thereby shielding from this violence, and from the very knowledge of this violence"[820]. In the same passage, this tendency is connected precisely to "'foreign' wars: aggressive tendencies that are potentially fatal to the cohesion of the group are redirected from within the community to outside it"[821].

Contemporary societies, on the one hand, have territorialised and institutionalised solidarity through the nation-State, but, on the other hand, we shall remember, through institutionalisation, they have loosened the bonds of solidarity: the

[818] Girard, *Violence and the Sacred*, 249.
[819] Girard, 249.
[820] Girard, 249.
[821] Girard, 249.

fact that the State takes responsibility for the worst off further takes responsibility away from their groups, from those individuals who once were part of that first circle that bore the burden of solidarity obligations, i.e., risky retaliatory expeditions, but also care for the fragile.

This is the culmination of a process that started way before the emergence of the welfare state and the thinking on social justice. For example, Dumouchel begins his *The Barren Sacrifice* by narrating the story of the 1870 lynching of Alain de Monéys during the Hautefaye fair, in France. De Monéys became the victim of a furious mob because his cousin was accused of being a republican[822]. What at first sight would seem like a traditional founding homicide, de Monéys being clearly a substitute victim – punished for his cousin's fault – in a time of distress (due to the Franco-Prussian War) for the community that diverts and converges its violence on a scapegoat, is revealed by Dumouchel to be an example of how the organisation of modern society changed the mechanisms of sacrificial violence.

First, although many took part in the lynching, the violence was far from unanimous: of the eight hundred people attending the fair, only two hundred participated in the lynching, while the others were simply bystanders or even went on minding their own business[823]. There was also a small group of friends who, although not physically protecting de Monéys, still tried to plead for him. This would not have been possible in the face of a founding homicide that required full unanimity: what has changed since the time of the original lynchings narrated and transfigured in myths is the structure of the society. When the first two circles merge on the basis of territoriality, reciprocal bonds loosen: "The crowd was modern because it brought together isolated individuals rather than people linked to one another by reciprocal solidarity obligations."[824]

This, in a way, makes the crowd undifferentiated from the very beginning – undifferentiation being, in Girard's theory, the characterising element of the mimetic crisis of generalised violence, in which the distinctions that inhibit violent interactions are lost[825] – for the people that go to the fair are not linked by personal bonds that assign to each a specific role within a community. In fact, there is no community, no hierarchy.

[822] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 5.
[823] Dumouchel, 6.
[824] Dumouchel, 10.
[825] Dumouchel, 10.

At the same time, however, and for the same reason, it also implies indifference: since the potential victim and executioner are not people that are relevant *to me*, I will not trouble myself to help them[826] – as we have seen in the first chapter, solidarity obligations are very demanding on the individual. It is precisely this widespread indifference that allows for those "peacetime crimes" that we mentioned in the previous chapter.

In solidarity-based societies, there was no need for a concept such as that of social justice because redistribution within the first circle in the social distance scheme happens spontaneously. Moreover, in those societies, only outcasts could be victims of sacrificial violence. In contrast, in modern societies, everyone runs the risk of victimisation; if we consider the violence of retributive justice as an example of sacrificial violence, even those who once were especially excluded from retaliation are now primarily sacrificial victims. I am talking about victimisers themselves who, in vindicatory systems, were often purposefully not targeted because direct retaliation towards them was considered too hazardous and at risk of failure since they had proved dangerous with their initial aggression. But it is especially those who, for one reason or another, struggle to keep up with what it takes to earn a living – the elderly, as we have seen, or the mentally ill, the sick in general, the disabled, single parents, etc. – that suffer from the erosion of solidarity bonds and find themselves lonely in the face of their struggles.

Two are the problems that come out of the erosion of solidarity bonds: not only is anyone now potentially a sacrificial victim, like Alain de Monéys, but also anyone can find themselves in the situation of struggling for survival – anyone can potentially end up in poverty, and most will at some point be old and unable to care for themselves, for example. To tackle these problems, institutionalised social justice needed to be introduced. After all, this is the idea that stands behind Rawls' theory of the veil of ignorance: rational agents hidden behind a veil of ignorance would create a fair structure for the society, for they do not know whether they were part of the privileged stratum or, instead, would belong to those struggling categories we just mentioned. Hence, such rational agents would accept distributive inequalities only as long as they would be proven to benefit the worst off.

[826] Dumouchel, 11.

Leaving aside the lengthy discussion that the publication of Rawls' *A Theory of Justice* generated, for it goes beyond the scope of this research, I will simply note that offenders are typically excluded from these distributive calculations: they are not part of the "worst off" category that needs to benefit from redistribution because it is *fair* for them to suffer, they *deserve* hardship, retributive justice requires their punishment, and this coexists without problems with social justice. One need not worry about being an offender when they act under the veil of ignorance because, unlike illness, poverty, or being part of a discriminated minority, committing a crime is a *free choice* one can abstain from. However, we have seen that things are not so easy, and the whole conceptual building of free will, desert, and individual moral responsibility serves to justify retributive practices and not the other way around.

This brings us back, once again, to constitutive luck. As we have mentioned, whereas luck has been integrated into the discussion on economic distributive justice, accepting that poverty is the result of factors beyond the agent's control – therefore recognising that the welfare state is necessary for achieving justice, whence the need for social justice – things have not worked the same way for criminal justice. Although some mitigating circumstances can be included in the criminal trial, and luck is taken into account in the form of excusing and exempting conditions, these, as we have seen, constitute the exceptions rather than the rule. Constitutive luck, on the other hand, wipes out desert, individual moral responsibility and retributive justice from the moral landscape altogether.

The question now is: is justice achieved when those having behaved "irresponsibly" – i.e., gambling, losing a job because of laziness, committing a crime – endure suffering as a consequence, or does this retributive conception of justice make no rational sense, being based on reactive attitudes and emotions, difficult to unhinge but nonetheless morally irrelevant? Well, in the first case, just like "economic luck" makes no one deserve poverty, so constitutive luck makes no one deserve punishment. If we take what was called by Beardsley the causally unlimited perspective, even if we think that the concept of desert, *in principle*, makes sense, we see that it has no scope of application in real life.

This perspective would still subscribe to a retributive conception of justice, but since no one deserves punishment, or the struggles of poverty, then we should not administer suffering-based sanctions and engage actively in the redistribution of wealth so that everyone gets what they deserve. In truth, what one positively deserves under

these premises, given that everything is the result of luck, is unclear. However, negatively, no one deserves suffering, and, hence, relieving suffering is a major step towards the achievement of justice when it is so conceived. If one embraces this point of view, then social justice comes to be an instrument for the realisation of the more fundamental retributive justice: the structure of society must grant a certain minimum level of welfare on the basis of desert.

This is roughly the thesis of luck egalitarianism, developed by Richard Dworkin as an elaboration of Rawls' own idea that the principles of social justice must redress the profound undeserved inequalities that affect our societies[827]. While Rawls mainly focuses on starting conditions – like the wealth of one's family that determines one's opportunities in life – Dworkin acknowledges that to limit the scope of social justice-driven redistribution would still be unfair for there are undeserved inequalities, like those of talent, that have continual effects in one's life, even if everyone was granted fair initial conditions. Hence, he advocates for periodic redistribution.[828]

Contrary to what we are doing here, Dworkin distinguishes two kinds of luck: brute and option luck. Brute luck does not depend on one's choices, and it is the one whose consequences must be redressed through social justice. On the other hand, option luck is the fruit of a "deliberate and calculated gamble"[829]; hence, the gambler must bear its consequences. Of course, constitutive luck makes this distinction crumble. Dworkin himself hints at this possibility when he writes: "In one way, however, my argument might well be thought to overstate the distinction between handicaps [examples of brute luck] and at least certain sorts of what are often considered preferences" for one can have "a 'preference' […] he does not want, and it makes perfect sense to say that he would be better off without it." This may include tastes that people "have (perhaps unwittingly) themselves cultivated. […] They regret that they have these tastes, and

[827] "The basic structure is the primary subject of justice because its effects are so profound and present from the start. The intuitive notion here is that this structure contains various social positions and that men born into different positions have different expectations of life determined, in part, by the political system as well as by economic and social circumstances. In this way the institutions of society favor certain starting places over others. These are especially deep inequalities. Not only are they pervasive, but they affect men's initial chances in life; yet they cannot possibly be justified by an appeal to the notions of merit or desert. It is these inequalities, presumably inevitable in the basic structure of any society, to which the principles of social justice must in the first instance apply." Rawls, *A Theory of Justice*, 7.

[828] Ronald Dworkin, "Equality of Resources," in *Social Justice*, ed. Matthew Clayton and Andrew Williams (Malden, MA: Blackwell Publishing, 2004), 129.

[829] Dworkin, 117.

believe they would be better off without them, but nevertheless find it painful to ignore them. These tastes are handicaps […]."[830]

Dworkin does propose, in fact, treating these desires as handicaps, that is, brute (bad) luck. In his proposal, handicaps are dealt with through an insurance system that transforms brute luck into option luck: it is up to the individual's choice to decide whether to subscribe to insurance, hence having the possibility to have their bad luck compensated or not, given their risk appetite[831]. According to him, this is a form of "respect" for the individual: simply forcing everyone to subscribe to all possible insurances – that is, to allow for an unrestrained welfare state that covers all possible misfortunes one could suffer – would deprive individuals of the possibility of actually choosing their preferred life, one that allows for a certain degree of risk in exchange for greater wealth, for example[832]. However, he does not get to the point of wondering whether one's own risk propensity and choosing ability might result from brute luck – that form of brute luck that we have called constitutive luck.

Interestingly, this is one of the aspects of Dworkin's theory harshly criticised by Elizabeth Anderson. Her broader critique of luck egalitarianism consists of an accusation that Dworkin missed the point of social justice, as we will shortly see. However, one of the points that she touches on is the fact that his conception does not allow for social justice to care for everyone's suffering and misfortune; it "fails to express concern for everyone who is worst off"[833] because it puts too much weight on individual moral responsibility. This allows for justifying a considerable amount of suffering in the face of which one is not supposed to do anything because, being the result of the individual's free choice, it is deserved.

> Luck egalitarians tell the victims of very bad option luck that, having chosen to run their risks, they deserve their misfortune, so society need not secure them against destitution and exploitation. Yet a society that permits its members to sink to such depths […] hardly treats them with respect.[834]

[830] Dworkin, 124.
[831] Dworkin, 117.
[832] Dworkin, 118. Dworkin compares here this situation to that of gamblers: if all losses were to be compensated by the winners "then no one would gamble […] and the kind of life preferred by both those who in the end win and those who lose would be unavailable."
[833] Elizabeth S. Anderson, "Against Luck Egalitarianism: What Is the Point of Equality?," in *Social Justice*, ed. Matthew Clayton and Andrew Williams (Malden, MA: Blackwell Publishing, 2004), 165.
[834] Anderson, 163.

Anderson does not criticize the concepts of desert and moral responsibility *tout court*; on the contrary, another critique she moves to Dworkin is that he requires the State to pass judgements on the degree of responsibility of its citizens, which she deems "disrespectful"[835]. Desert and moral responsibility are acceptable concepts in her system; it is just not the State's task to judge them – at least within the framework of social justice.

Her broader critique of luck egalitarianism implicitly targets the idea, again implied but not explicitly stated by Dworkin, that social justice consists in retributive justice: in everyone getting what they deserve. Just like for responsibility and desert, this does not mean that she criticises the idea of retributive justice itself; in fact, she refers to Dworkin's conception of retributive-distributive justice as "cosmic justice"[836], hence understanding it as a form of justice, but again, she does not believe that is what social justice is about. To understand social justice, Anderson believes that philosophers should take the requests of egalitarian political movements more seriously. What these movements ask has nothing to do with desert – on the contrary, I suggest, recent zeal for affirmative action shows a tendency to exclude merit as a criterion of justice.

According to Anderson, egalitarian movements reclaim "the equal moral worth of persons"[837] independently of their specific identity – in terms of talents, abilities, gender, ethnicity, social status, and so on. But what does this "moral worth" consist of? Positively, of the recognition that "everyone equally has the power to develop and exercise moral responsibility, to cooperate with others according to principles of justice, to shape and fulfil a conception of their good."[838] Negatively, of freedom from oppression[839].

The moral equality she advocates for is turned into social and political equality: the problem of the inequality generated by brute luck, she maintains, does not lie in the inequality itself but in its social and political outcome – it is human relationships that transform natural inequality into oppression, and it is oppression that egalitarians should fight[840]. Inegalitarian societies are not defined as such by unequal distributions of goods: they are inegalitarian because they are organised around hierarchical

[835] Anderson, 168.
[836] Anderson, 155.
[837] Anderson, 170.
[838] Anderson, 170.
[839] Anderson, 170.
[840] Anderson, 170, 180.

relationships that classify people as superior and inferior[841]. It is interesting to notice here that there is only one case where oppression is justified, according to Anderson, i.e., the "just punishment for crime"[842]. Apparently, punishment can take the form of "social relationships by which some people dominate, exploit, marginalize, demean, and inflict violence upon others"[843] – precisely what is to be fought against for all the other citizens, those who have not committed any crime.

This is consistent with the central place that Anderson reserves for moral responsibility, despite the criticism moved to Dworkin that we saw at the beginning of this analysis. In fact, we now see that, although we started off from a critique of the excessive weight given to individual moral responsibility leading to "disrespectful" outcomes, Anderson's view on social justice is, in truth, very individualistic. Once individual responsibility is granted – and, as we have just seen, it plays a central role in Anderson's conception of social justice – social and political equality are expressed in the form of negative freedom – freedom from oppression – that is achieved through egalitarian relationships: "What citizens ultimately owe one another is the social conditions of the freedoms people need to function as equal citizens."[844] This requires a certain amount of redistribution but not in order to compensate for brute luck and get everyone to get what they deserve (whatever that may mean), not even to relieve everyone from suffering, but to achieve the minimal conditions that allow for the development of every individual, without concern for the actual achievement of such development[845]. To this, Anderson adds a safety net that protects "even the imprudent"[846] so that they are not "deprived of the social conditions of their freedom"[847].

In a way, then, Anderson's view of egalitarianism and social justice is composed of two elements that are meant to free individuals from oppression: on the one hand, equal "respect" for each individual, which means that everyone is considered

[841] Anderson, 169.
[842] Anderson, 170.
[843] Anderson, 170.
[844] Anderson, 176.
[845] "Effective access to a level of functioning means that people can achieve that functioning by deploying means already at their disposal, not that the functioning is unconditionally guaranteed without any effort on their own part. […] democratic equality guaranteed not effective access to equal levels of functioning but effective access to levels of functioning sufficient to stand as an equal in society." Anderson, 174–75.
[846] Anderson, 178.
[847] Anderson, 179.

responsible for their own fate and master of their own choices and their consequences – I will call this the responsibility aspect of Anderson's egalitarianism, and I will stress here once again that this grounds her conservative approach to criminal justice. On the other hand, the welfare state as the safety net to ensure such consequences do not result in relationships of domination – I will call this the redistributive aspect of Anderson's egalitarianism.

While much of her critique of Dworkin's luck egalitarianism is shareable, the responsibility aspect is grounded on the old concepts of individual moral responsibility and desert – although limited by the redistributive aspect – that constitutive luck completely unhinges. As we have seen, this does not mean that no place is left for any kind of responsibility, but the one Anderson has in mind is undermined. Still, her view of social justice as freedom from oppression, inspired by Iris Marion Young's analysis of oppression as "marginalization, status hierarchy, domination, exploitation, and cultural imperialism"[848], is compatible with constitutive luck. It is actually more than compatible: it brings freedom from oppression to its extreme consequences, for it makes no exception based on desert, as Anderson makes with respect to "just punishments".

This brings me to highlight another aspect of Anderson's critique of Dworkin's conception that I find relevant: that what really is at stake here is not primarily economic; hence, redistribution – although relevant – should not be central but only instrumental in order to allow for relationships of equality. This difference is relevant because it reveals a deeper divergence between the two approaches: whereas Dworkin conceives of the human being as a *homo oeconomicus*, Anderson considers it fundamentally as a *homo politicus*.

To summarise this last analysis of luck egalitarianism and its weaknesses, if one believes that the concept of retributive justice makes sense, constitutive luck still undermines the practice of punishment based on desert. While the fact that luck plays a significant role in one's opportunities and outcomes is widely accepted on the economic level, hence justifying the safety net of the welfare state – without forgetting, however, that this is a recent achievement of Western societies – things have not followed the same path in criminal matters, an area often disregarded by social justice theories, like Dworkin's and Anderson's themselves. Nonetheless, I believe that Anderson's

[848] Anderson, 169–70; Iris Marion Young, *Justice and the Politics of Difference* (Princeton: Princeton University Press, 1990).

relational approach to equality and justice can be smoothly transposed and extended to criminal justice and I will shortly explain how.

But before moving to that, let us recall the question that originated the analysis of Dworkin's and Anderson's theories: does the idea of retributive justice ultimately make sense? Compared to the idea analysed so far that luck eliminates the scope of application of retributive justice, the project I carry on is more ambitious, and the thesis defended here is more radical. I am arguing that retributive justice does not make sense from a moral point of view, as retribution is a rationalisation of an emotional response that is grounded on ancestral impulses favoured by natural selection to increase the chance of survival in case of conflict. To speak of retributive justice is actually an oxymoron from my perspective because the idea of the intrinsic goodness of the intentional infliction of deserved suffering is essential to the concept of retribution. In contrast, justice, being a value, from the utilitarian perspective that infuses this work, should aim at minimising suffering. Obviously, this claim calls for a better understanding of the faults of the utilitarian justification for punishment, which have only been hinted at for now and will have to be addressed soon. But first, I wish to tackle another question that these last points may have raised.

The critique of retributive justice presented here is not essentially based on the critique of free will, desert, and individual moral responsibility: it is not because free will is an illusion that retributive justice is made nonsensical. The origin of these concepts and, more importantly, the strenuous attachment to them, is shown to be connected to the idea of retributive justice, the crumbling of which casts a shadow on them too – hence turning the traditional argument upside down: it is not the collapse of free will to dissolve retributive justice, it is rather the falling apart of retributive justice to disrupt free will and the related concepts.

Does this mean that all the analysis devoted here to luck, desert, the absence of free will and individual moral responsibility has been pointless for the purposes of this work? Not quite so. The rational discourse on retributive justice uses them as pivotal elements; their reworking can be crucial in our rethinking of the concept of justice and its related practices. To understand how, Iris Marion Young's conception of structural injustice will come to the aid.

According to Young, social justice deals with the background conditions of people's lives. As we said at the beginning of this section, it is an innovative conception of justice because it does not address individual action but society's structures instead.

Hence, structural injustice is the result of the failures of social justice. While this might be thought to lead to a compensation theory like Dworkin's, Young takes on a novel approach in which, instead of having each wonder what they may be entitled to receive – because of their bad luck in being victims of some form of structural injustice – one has to wonder what they owe to others by reason of being part of the structures that generate the injustice[849].

However, is it not to be part of such structures a product of luck as well? Young does not address the question, and I think she is not interested in the issue because although the recognition that one *is* part of the structures generating injustice serves, in Young's work, to produce responsibility and hence take action against said injustices, the responsibility she advocates for is perfectly compatible with brute and constitutive luck. In fact, Young's conception of responsibility does not correspond to what we have here referred to as "individual moral responsibility" but to something more similar to Waller's taken responsibility and Taylor's co-responsibility.

Young, in part inspired by Hannah Arendt's distinction between guilt – which in her opinion can only be strictly personal – and responsibility – which can be both individual and collective[850] – opposes two different "models" of responsibility: the liability model, which is backwards-looking and inherently related to guilt and blame, and the social connection model, which is forward-looking and leans towards taking action and making changes[851]. Young notices that

> Conceptions of moral responsibility, both among scholars of ethics and in ordinary language, most often assume a liability model of responsibility. […] To say that an agent is responsible means that they are blameworthy for an act or its outcome.[852]

This, in turn, is looked for as a basis for "sanctioning, punishing, or exacting compensation or redress"[853].

The liability model of responsibility serves the purpose of isolating guilt: it conceives individuals as isolated beings who need to internalise the costs of their

[849] Iris Marion Young, *Responsibility for Justice* (Oxford: Oxford University Press, 2011), 26.

[850] Hannah Arendt, "Collective Responsibility," in *Amor Mundi: Explorations in the Faith and Thought of Hannah Arendt*, ed. James W: Bernauer (Boston: Martinus Nijhoff Publishers, 1987), 43–50; Young, *Responsibility for Justice*, 77–78.

[851] Young, *Responsibility for Justice*, 96–97.

[852] Young, 97.

[853] Young, 98.

actions[854]. It is a model that has been strengthened since the 1980s, with the rise of neoliberal thinking, that stopped considering poverty as the result of institutional factors and generated the erosion of the welfare state, which came to be considered as part of the problem of poverty rather than part of its solutions, for it reinforces laziness and negligence instead of encouraging individual responsibility for one's fate.[855] In fact, this kind of responsibility is strictly personal since, as Young remarks, it concerns the outcomes of one's actions, but also, we may add, it is assessed through a close investigation of one's internal states and processes, their motivations, the knowledge they had access to, their alternative possibilities.

We have mentioned in the previous chapter how this has led to the development of what Susan Jacoby has termed "psychiatric justice"[856], that is, the intrusion of the psychiatric evaluation in the criminal trial[857] – the natural development of the process that shifted the focus of retribution from the body to the "soul" of the punished person that is the object of Foucault's *Discipline and Punish*[858]. To Foucault's analysis, we may add that this is part of the broader process of individualisation of responsibility and guilt that were originally shared by the members of the family group – in Heider's first stage of attribution, as we have seen in the previous section – making it possible and reasonable, in archaic societies, to exercise retaliation against the offender's relatives.

Young brings us once again to the last stage of Heider's attribution scale that, from collective responsibility, passing through individualised responsibility (and progressively more specifically individualised, not only concerning one's actions but one's mental life, too), broadens yet again leading to the fifth and last stage of attribution, where responsibility reverts back to being non-individual, though it is not collective in the sense of the first stage. It is shared instead.

---

[854] Young, 10–11.

[855] Young, 3–10.

[856] Jacoby, *Wild Justice*, 324.

[857] Jacoby's conception of psychiatric justice must not be confused with Thomas Szasz's, the first to have used the term, although the two are of course related. Jacoby understands psychiatric justice as an instrument to achieve milder punishments for offenders because it grants exempting conditions to people on trial. This, according to her, generates a sense of unachieved retribution because sentences, taking into account the mental condition of the offender, do not match the gravity of the crime they committed (Jacoby, 324.). Szasz, on the other hand, also stresses that psychiatric justice generates unfair punishment, but in the opposite sense. Psychiatry is used in criminal justice as a tool to skip the fair trial and lock up people arbitrarily on the basis of the presumed danger generated by their mental condition. Asylums are the perfect prison substitutes, as they are not different in their restraining and afflictive procedures. Thomas Szasz, *Psychiatric Justice* (Syracuse: Syracuse University Press, 1988).

[858] Foucault, *Discipline and Punish*.

Young's shared responsibility is not about blaming and finding fault and guilt. It is about recognising one's role in the perpetration of injustice. Her choice to call this the "social connection model" of responsibility is, in my opinion, especially telling. In fact, it underlies the interdependency of social actors, and of individuals more in general, debunking the myth of the isolated person that ideally internalises the costs of their action. One's deeds, choices, opportunities, and even abilities are always related to the context in which they are and in which they have been previously. Given that constitutive luck and interdependency often generate a sense of passivity, a feeling that a person's destiny is taken out of their hands, it is interesting to notice how Young's conception of shared responsibility turns the matter upside down, using the social connection model to emphasise agency.

The liability model of responsibility, although investigating one's actions and motives, presents an intrinsic feature of passivity due to its backwards-looking position: the agent *has committed* a wrong in the past, and their responsibility must now be assessed in order for them to be blamed or punished. The responsible agent is, ultimately, the object, not the subject of responsibility; they are analysed and will *receive* according to what is considered their desert. The forward-looking posture of the social connection model, on the other hand, considers responsible agents primarily as subjects of responsibility: to *take* responsibility (to recall Waller's vocabulary) means to assume an active disposition towards the structures one is part of; it means to question them and recognise one's role in order to work actively for change.

In this model, one can be part of unjust social structures, be responsible for them, and still not be blameworthy: they may, at the same time, be taking advantage of the structure and be one of its victims. Most of the time, one is just born in a certain position within the structure and does their best not to be crushed by it or try to thrive within its limits. Where one is situated is ultimately a matter of constitutive luck, and this is compatible with the social connection model of responsibility. Within this framework, advocating for shared responsibility does not entail finding faults and requiring compensation. However, it does call for the recognition that to be part of a structure implies having some power over it; hence, if one has power over a structure that generates injustice, and if they care for justice as something valuable, then they *should* exercise that power to change the structure. As Young put it, people should take responsibility *for* justice, realising that social structures are maintained collectively, and

it is, therefore, each and every person's duty to work to realise the structures they consider just[859].

It is a kind of personal responsibility, for each must take it for themselves and take action on it, but it is shared in that it makes sense only if agents work together to produce change[860]. Simple individual action that on the liability model of responsibility can be considered virtuous if not supererogatory, hence praiseworthy, is simply pointless in the social connection model and if this feels somewhat unfair on a first reading – for why should I be deprived of the praise I deserve for my noble intentions, even if they have no relevant consequence due to the shortcomings of others? – it is because we are too immersed in the liability model.

My reading of the social connection model of responsibility implies that taking part in the collective enterprise that the creation of a just society should be does not generate any rewards – like not taking part does not generate any punishment. The motivation for taking action does not lie in the expected reward – even just in the form of social recognition and praise – or the fear of punishment – even just blame. Rather, to contribute positively to the creation of a just social structure is simply inherent to what it means to live in a society where to be a responsible person that "tries to deliberate about options before acting, makes choices that seem to be the best for all affected, and worries about how the consequences of his or her action may adversely affect others"[861] is supposed to be the norm.

Simple, responsible action does not generate any reward, just the collective positive outcome of living in a safer, more peaceful community. Likewise, failure to contribute in such a way does not make one blameworthy – which does not mean that they may not be *criticised*[862]. Again, the inability to appreciate the difference between blame and criticism is telling of a culture *immersed* in retribution, where criticism is moulded with an emotional reactive attitude and shortcomings call for punishment more than redress.

---

[859] This thesis evokes Étienne de La Boétie's *Discourse of Voluntary Servitude*, that Young, however, never mentions. It is, in fact, La Boétie's most interesting theory that tyranny is possible because most of a tyranny's subjects take advantage of the hierarchical structures to which they are subjected because, although they mean that someone is above them and has power on them, it also implies, in most cases, that they will have someone below to exercise power over. Étienne de La Boétie, *The Politics of Obedience: The Discourse of Voluntary Servitude*, ed. Murray N. Rothbard, trans. Harry Kurz (Auburn: The Mises Institute, 2010), 71–72.

[860] Young, *Responsibility for Justice*, 111.

[861] Young, 25.

[862] Young, 144.

As Young points out, the social connection model of responsibility is not different from the liability model "quantitatively" as a form of mitigated responsibility due to a lesser contribution to the outcome – as complicity to the creation and maintenance of the unjust social structure – but qualitatively: it is a different kind of moral responsibility that cannot be understood unless the liability model is set aside[863]; the two are not part of the same spectrum, they are alternative.

From the first pages of the book, it is clear that Young's perspective is mainly economic: she addresses the problem of social justice in relation to the problem of poverty and does not mention any interaction between her approach and criminal justice. However, just like we previously noticed that redistribution is instrumental to the equality of rights (whether one considers rights from a deontological perspective, as primary and irreducible concepts, or from a consequentialist perspective, as negotiable proxies that stand for the general characteristics of a decent life), so we can now remark that the problem of the socio-economic structure that she addresses is instrumental to the broader problem of granting the "best possible life" – with all the difficulties in determining what this may mean – to the members of the society, which should include wrongdoers.

Of course, offenders are a minority in society, so it is reasonable for Young not to address the issue in her seminal work on the topic, and her untimely death makes it impossible for us to know if she would have ever expanded her analysis to cover the other category of justice that is criminal retributive justice with its related concept of responsibility: would she have claimed that the liability model of responsibility should have been preserved in society's dealings with crimes, hence supporting the thesis that social justice and retributive justice complement each other in that they cover different areas of justice? Or would she have tried to extend the social connection model of responsibility to completely replace the liability model? Would she have had any thoughts on the subject at all? Unfortunately, these questions are destined to remain without an answer.

What I am trying to do here, however, is precisely to extend her model of responsibility and justice to criminal justice through the recognition that the liability model of responsibility, whose substance is constituted by the concept of desert and that is grounded in retributive impulses, is ultimately, philosophically empty.

[863] Young, 103–4.

An attempt at overcoming the separateness and complementarity of (what we now call) social and retributive justice can be found in Guyau. Writing at the end of the 19th century, before the concept of social justice was even established, Guyau introduced a conception of justice that did not consist of "normative features specific to relations between people"[864] – as Dumouchel characterised traditional conceptions of justice. Guyau, in fact, aimed at building "another larger justice, which gives kindness to all – a justice which not only ignores the hand with which it gives, but will not even know the hand which receives it."[865] Though not speaking of social structures, background conditions, welfare state or safety nets, Guyau advocates for a kind of justice that shares with contemporary conceptions of social justice the emancipation of the notion of justice from that of desert, and of individual action in general.

What is even more interesting, however, is the fact that he presented this view on justice in the context of a discussion of retribution. Differently from contemporary thinkers of social justice, Guyau did not think of this desert-free justice as something covering a different area than criminal justice; instead, it is precisely through a critique of criminal justice that he reached a conception of justice that resembles that of present-day social justice. Ultimately, he identifies in *charity* the fundamental principle of justice:

> For distributive justice, which is an entirely individual and personal justice, a justice of privilege (if such words are not a paradox), must, therefore, be substituted a justice of a more absolute character, and which at bottom is really charity. Charity for all men, whatever may be their moral, intellectual, or physical worth, should be the final aim to be pursued even by public opinion.[866]

We have already noted that the boundaries between distributive and retributive justice are blurred, as retributive justice consists of a distribution of burdens and benefits on the basis of desert. I chose to favour the use of the term "retributive" to emphasise the connection between criminal justice and vengeance. Guyau, on the other hand, spoke of distributive justice and maintained that it consists of "a justice of privilege" because his aim was to highlight the unfairness of this distribution: "Suffering shocks in itself, independently of its point of application; a distribution of suffering is, therefore, morally unintelligible."[867]

---

[864] Dumouchel, *The Barren Sacrifice. An Essay on Political Violence*, 146.
[865] Guyau, *A Sketch of Morality Independent of Obligation and Sanction*, 167.
[866] Guyau, 186.
[867] Guyau, 202.

When people treat each other justly, in the sense proposed by Guyau, they enter a relationship of brotherhood. The choice of using the term "brotherhood" is interesting in that it takes us back to Scubla's and Dumouchel's analysis of social justice in terms of Verdier's theory of the concentric circles of social distance and their claiming that the establishment of the nation-State has meant an enlargement of the inner circle to encompass the intermediate one. I have pointed out before that, in my opinion, the nation-State has generated a more balanced fusion between the two circles. In any case, Guyau's talk of brotherhood evidently points in this same direction: the dilation of the family relations – more specifically, that non-hierarchical relationship that is brotherhood (we may now say "siblinghood"), to be contrasted with hierarchical parental relations – into the extra-familiar world. In Dumouchel's and Scubla's case, however, this expansion remains circumscribed to fellow citizens, with social justice being limited on a territorial basis within the nation-State's borders. Guyau, on the contrary, aspires for a universal brotherhood[868].

In the next and last section, I will begin to show what such a conception of social justice, one that, like in Guyau's attempt, encompasses criminal justice too, is supposed to mean on a practical level. For the moment, I simply remark that, like poverty or mental illness, criminality can be considered a matter of social concern and that if data show its substantial correlation with both poverty and mental illness, it is not by chance. This correlation points in the direction of treating the problem of crime in a similar way to the other two, though not denying its specificity. If, on the one hand, relief from poverty and mental illness will reduce criminality, it will probably not eliminate it entirely, but with regard to those offenders that will remain a similar approach should be maintained: an approach based on relief and resolution rather than on blame, punishment, and personal satisfaction. An approach also based on shared responsibility rather than individualised culpability, one that considers the individual in the web of relations that shaped them and upon which society should *take* responsibility and act to modify the criminal outcome.

[868] Interestingly, following Derrida, Young rejects the connotation of political relationships in terms of "brotherhood" and contrasts it to "solidarity", which she favours. In her view, brotherhood entails a denial of "all personal and social difference that might produce asymmetrical experiences" and "appeals to a natural origin in the unspoken mother, already there to be counted on." (Young, *Responsibility for Justice*, 120.). Solidarity, on the other hand, "is a relationship among separate and dissimilar actors who decide to stand together, for one another". It is also a relationship which "must always be *forged* and *reforged*. Solidarity is firm but fragile. It looks to the future because it must constantly be renewed." (Young, 120.).

It should now be clear that even if my main criticism of retributive justice does not primarily stem from determinism and the challenging of free will, the analysis here dedicated to luck and the absence of free will is decisive in the *pars construens* of my attempt at rethinking criminal justice. It is based on the idea that actions and motives are not born, developed and executed under the complete independent control of the isolated individual but are rather the result of the complex web of relationships that envelop the agent. This grounds the social connection approach just analysed, allowing all members of the society to take responsibility for generating the conditions that brought some people to commit crimes and act on them.

I here welcome Young's criticism of Dworkin's – and my own – choice to speak of luck instead of structures[869]. Speaking of luck tends to elide responsibility, which is precisely what we have been meaning to do. However, Young seems to imply, it does not do so only for responsibility based on the liability model but also for that based on the social connection model. Her criticism recalls that of Anderson when she pointed out against Dworkin that social justice does not correspond to cosmic justice because it does not aim at correcting and compensating the unfair distribution of natural advantages and drawbacks assigned by the natural lottery. Rather, it seeks redress for socially constructed inequalities that take the form of oppression. According to Young, similarly, it is essential to distinguish injustice from destiny[870], social justice aiming to rectify the former and not the latter, a distinction that the concepts of brute luck or constitutive luck fail to make.

Whereas it is clear to me that social inequalities that take the form of oppression, being manifestly created by human-generated structures, are *easier*[871] to address and therefore might well be the first to be targeted for correction, I do not understand why, if brute bad luck generates suffering which cannot be relieved through the eradication of oppressive relationships, society should not aim at assuaging its consequences. If, for example, a baby is born with a disease that causes them a tremendous amount of pain – actually, whatever amount of pain – and their family is not in the financial position to pay for treatment, even if only palliative treatment, it seems right that society should bear the cost of the treatment. The baby's suffering is not caused by oppression but by the bad luck of being born with such a condition.

[869] Young, *Responsibility for Justice*, 28.
[870] Young, 32.
[871] By this, I do not mean *easy* by any means.

An element of oppression might be identified in their family's condition of poverty, and hence its inability to pay for treatment, thus turning the problem into one of economic redistribution. However, treatments can be so expensive that a family may be well off and still not be able to sustain the price or only be able to sustain it at the cost of considerable deprivation. The family is not in a situation of oppression due to financial hardship; it just faces a case of acute bad luck in having to face extreme costs for the infant's wellbeing. In another version, the baby might even be an orphan, their parents deceased in, say, un unfortunate car accident or a natural disaster, or any situation that we can imagine where the problem is not originated by economic or social oppression. If one wishes to classify these cases under that category, then what they are really doing is making the distinction between luck and injustice that Anderson and Young tried to introduce crumble, for I cannot think of a case of bad luck – a situation of undeserved suffering – that cannot be ultimately traced back to an injustice of this kind: it is the fundamental injustice – if one wants to call it like that – entailed by the fact of being living sentient beings, a condition which inherently carries with it the risk of suffering and premature death.

Talking of both bad luck and injustice and not focusing exclusively on socially created oppression also has another advantage, in my opinion. Although speaking of injustice encourages the members of society to take responsibility and be agents of change, speaking of bad luck highlights the fact that not all undeserved suffering can be relieved and that this must be accepted as a fact of life, helping to avoid falling into the just-world fallacy which makes us blame and assign responsibility to victims in order to cope with our inability to soothe undeserved suffering. Structural injustice and bad luck are, hence, both essential concepts in that the former promotes the assumption of responsibility (of the social connection kind) in order to make progress, and the latter serves as a reminder that the fact that suffering cannot be eradicated from society should not result in the attribution of responsibility, on the liability model, as a coping strategy.

How to extend the concept of social justice and the social connection model of responsibility to address the problem traditionally tackled by criminal retributive justice will now be the object of the next and last section.

3.3 Utilitarianism and Prison Abolition: Rethinking Justice

Opening her book *Are Prisons Obsolete?*, Angela Davis remarks that the imaginary related to incarceration is so ingrained in our view of society that we are unable even to imagine a society without prisons[872]. However, we have seen that prisons are a relatively new tool to tackle the problem of crime and that human societies have gone prison-free for most of history. The answer to crime has been mostly punitive, but through the torture system that preceded the 18th-century criminal reform, and the justification for punishment has been purely retributive: in monarchies, violation of the law amounted to lese majesty, and the punishment that followed was the manifestation of the monarch's anger that required satisfaction. Consequential justifications – incapacitation, deterrence, and rehabilitation – were only introduced with the criminal reform.

It is my aim, in this section, to show that taking utilitarianism seriously leads to questioning the prison system and that, eventually, the position most consistent with utilitarianism is the abolition of punishment. The prison abolition movement, however, has ignored – and has been ignored by – utilitarianism, and it is my aim to fill this gap. Considering that it was the first utilitarian thinkers who both popularised thinking of punishment in terms of its positive consequences and shaped the criminal reform that gave birth to the prison, my position will sound surprising. The fact that modern utilitarians have followed this line of thought will make it even more startling. However, I will show that prison abolition itself also has roots in the works of the founding fathers of utilitarianism, and, to reach this end, I will start by analysing that same reform that disseminated the prison model and the principles that inspired it.

3.3.1 The Missing Match

Beccaria was one of the reformers who led to the abandonment of the torture system in the 18th century. Whether he is considered utilitarian or not – as this classification is an object of debate[873] – he undoubtedly advanced consequentialist arguments against maintaining the torture system, whose punishments are accused of

[872] Davis, *Are Prisons Obsolete?*, 9–10.

[873] Though Bentham himself identified Beccaria as an advocate of the principle of utility: "Before it [the principle of utility] was mine it was M. Beccaria's. Before it was his, it was Helvetius's: before it was Helvetius's, it was in some sort everybody's." University College London Library, Bentham Papers, Box xxvii, fo. 100, as quoted in: Philip Schofield, "«The First Steps Rightly Directed in the Track of Legislation»: Jeremy Bentham on Cesare Beccaria's Essay on Crimes and Punishments," *Diciottesimo Secolo* 4 (2019): 68.

being a "*useless* cruelty that is the instrument of fury and fanaticism or of weak tyrants"[874] while claiming that

> The purpose of punishment […] is none other than to prevent the criminal from doing fresh harm to fellow citizens and to deter others from doing the same. Therefore, punishments and the method of inflicting them must be chosen such that, in keeping with proportionality, they will make the most efficacious and lasting impression on the minds of men with the least torment to the body of the condemned.[875]

Beccaria and the reformers, as Foucault has analysed at length, did not advocate imprisonment as the universal punishment. Rather, they imagined what Foucault called "the theatre of punishments"[876], composed of "a whole new arsenal of picturesque punishments"[877], whose variety was necessary for the "analogical punishment"[878], a penalty that matches the specificity of the crime, creating an intuitive connection with the crime, so that the idea of committing a crime would give rise to the idea of the specific, associated punishment, and deter potential offenders from wrongdoing[879].

However cruel these punishments proposed by the reformers may now seem to us, comprising not only isolation, deprivation, and humiliation but the infliction of intense and prolongated physical pain[880], compared to the torture system that was based on disproportion, reformers brought about a softening of penalties. Disproportion was intrinsic to the conception of crime as lese-majesty because the offence to the monarch by violating their law is immeasurably greater than the actual harm produced (like stealing food). Disproportion was also necessary in older times to grant deterrence in the face of the already mentioned greater difficulty in the process of investigation, identification of the guilty, and prosecution due to a lack of infrastructures, a lack that has been remedied in recent times, allowing for more moderate but more widespread punishments. As is evident in Beccaria's work, the criminal reform aimed at making

[874] Beccaria, *On Crimes and Punishments and Other Writings*, 26, my emphasis.
[875] Beccaria, 26.
[876] Foucault, *Discipline and Punish*, 106.
[877] Foucault, 114.
[878] Foucault, 105.
[879] Foucault, 104–10.
[880] "Take Vermeil [a reformer], for example: those who abuse public liberty will be deprived of their own; those who abuse the benefits of law and the privileges of public office will be deprived of their civil rights; speculation and usury will be punished by fines; theft will be punished by confiscation; 'vainglory' by humiliation; murder by death; fire-raising by the stake. In the case of the poisoner, 'the executioner will present him with a goblet the contents of which will be thrown into his face; thus he will be made to feel the horror of his crime by being offered an image of it; he will then be thrown into a cauldron of boiling water' (Vermeil, 68-145; cf. also Dufriche de Valaze, 349)." Foucault, 105.

punishments milder on the basis of the consideration that harsher punishments, and especially torture to extort confession, were useless[881].

Bentham's *Introduction to the Principles of Punishment and Legislation*[882] is often considered the founding book for utilitarianism, presenting the first clear-cut formulation of the principle of utility. However, it is sometimes forgotten that the book was meant to be an introduction to a penal code that Bentham eventually never wrote. Utilitarianism is so rooted in the 18th-century penal reform that the principle of utility was presented in the *IPML* as a guide for the legislator precisely in their consideration of punishment. Punishment, in fact, is conceived as the key instrument for the creation of an artificial harmony between the individual and the general interest, making socially harmful but personally advantageous actions unfavourable for the individual, too, since the prospect of the disadvantage of the inflicted punishment would negatively compensate the expected personal advantage.

Moreover, Bentham himself designed the prison model that inspired many of the first prisons that were built after the criminal reform, especially in the United States: the (in)famous panopticon[883]. The ideal structure of the panopticon is composed of a central tower surrounded by cells on several levels. Thanks to a play of light and shades, those in the cells, having a window behind and therefore visible in backlighting, are always perfectly observable by those in the central tower, which, conversely, are invisible to those in the cells. The central tower might even be empty, but the prisoners would never know: they would always feel that they may be watched at any time, therefore preventing any escape attempt or breaching of the rules.

The panopticon is now the symbol of an omnipresent oppressive power, of a quest for efficacy that leaves no room for humanity. It is, in fact, anonymous: the power is not even exercised by people; it is the structure itself that dominates the individuals, who are prevented from acting freely not by armed officers but by the building's architecture and what it suggests. It is a *dispositif* of power, a power structure that crashes individuals through dehumanisation. It is often pointed at as an example of the dystopic consequences of the adoption of the utilitarian principle, the concrete manifestation of utilitarianism itself as a worldview that paved the way for

[881] Beccaria, *On Crimes and Punishments and Other Writings*, 32.
[882] From now on: *IPML*.
[883] Jeremy Bentham, *The Panopticon Writings*, ed. Miran Bozovic (London and New York: Verso, 1995).

totalitarianism[884]. Not surprisingly, Orwell's "telescreens" (the two-way televisions that operate as cameras and microphones depicted in *Nineteen Eighty-Four*) are considered technologically advanced panopticons, which are now a reality in most prisons, where surveillance is put in place through recording devices, somehow reaching Bentham's ideal.

However, I believe that, when judging these aspects of Bentham's theories, it is necessary to consider the context in which he was writing. He was coming from a system of punishment that relied on physical violence and was part of the movement that aimed at changing it. While, nowadays, the project of the panopticon seems indeed dystopian, it must be understood as a continuation of this softening of punishments initiated by Beccaria and the other reformers. These efforts seem now utterly insufficient, but were revolutionary at a time when torture (even under interrogation to ascertain culpability) and the death penalty were the norm: in this context, the panopticon could effectively diminish the violence that was perpetrated in the criminal system.

To understand the conditions of prisons and jails right after the reform and to look more leniently at Bentham's proposal, Charles Dickens's *American Notes* can be helpful. During his 1842 trip to North America, Dickens visited the prisons of Boston, New York, and Philadelphia and described the different experiences in his travel diary. New York's prison is nicknamed "The Tombs"[885], where conditions are so atrocious that the administration had to take the hooks for hanging clothes off the walls for prisoners too commonly used them to hang themselves[886]. Philadelphia's famous Eastern Penitentiary – the same that Foucault analyses in *Discipline and Punish* as an example of the monastic model of prison, where incarcerated people are made to live in complete isolation in order to generate repentance[887] – "is rigid, strict, and hopeless", "cruel and wrong"[888].

On the other hand, Dickens views Boston's prison so favourably that he hoped "that the next new prison we erect in England may be built on this plan."[889]. It is no

[884] Janet Semple, *Bentham's Prison: A Study of the Panopticon Penitentiary* (Oxford: Clarendon Press, 1993), 3–4.

[885] Charles Dickens, *American Notes for General Circulation. Vol. 1* (New York: Cambridge University Press, 2009), 199.

[886] Dickens, 203.

[887] Foucault, *Discipline and Punish*, 124–26.

[888] Dickens, *American Notes*, 238.

[889] Dickens, 124.

coincidence that he speaks of "plan", for the particularity of Boston's prison is architectural; although not entirely a panopticon, Dickens' lengthy description[890] recalls the principles of the Benthamian *dispositif*. The structure, in fact, allows for a single officer to guard half of the cells in a single glance: the certainty of being seen deters any attempt to escape, so much so that officers do not carry any weapon, for they are rendered useless[891]. Contrary to the other prisons, Dickens does not perceive Boston's penitentiary as a place of violence, hopelessness and suffering.

We can now see how Bentham's panopticon, although being *now* reminiscent of totalitarian techniques and practices, could be part of the process of softening penalties that had started with the reformers. This consideration should hopefully begin to make the idea of a utilitarian theory of prison abolition seem less paradoxical: utilitarianism, it is true, is rooted in the penal reform that generated the expansion of prison, and it also made punishment one of the first and central subjects of its reflection, but aimed at making it less cruel, at minimising the suffering that was involved.

Nonetheless, utilitarianism has indeed shaped the contemporary debate on punishment in many ways, most notably by introducing its forward-looking justifications. We have just seen how Beccaria believed that the central aim of punishment should be to prevent criminals from offending again, which now takes the name of specific deterrence: punishment should reduce recidivism. However, we have already observed that recidivism rates are so high that if this is the primary aim of punishment, results are so disappointing that our practices should be at least questioned, if not entirely discarded already.

In the *IPML*, Bentham mentioned the three contemporary consequentialist justifications for punishment: deterrence, incapacitation, and rehabilitation. Incapacitation and rehabilitation, however, are addressed as "properties of inferior

[890] "In the centre of a lofty area, lighted by windows in the four walls, are five tiers of cells, one above the other; each tier having before it a light iron gallery, attainable by stairs of the same construction and material: excepting the lower one, which is on the ground. Behind these, back to back with them and facing the opposite wall, are five corresponding rows of cells, accessible by similar means: so that supposing the prisoners locked up in their cells, an officer stationed on the ground, with his back to the wall, has half their number under his eye at once; the remaining half being equally under the observation of another officer on the opposite side; and all in one great apartment. Unless this watch be corrupted or sleeping on his post, it is impossible for a man to escape; for even in the event of his forcing the iron door of his cell without noise (which is exceedingly improbable), the moment he appears outside, and steps into that one of the five galleries on which it is situated, he must be plainly and fully visible to the officer below. […] [The cell] is small, of course; and the door being not solid, but grated, and without blind or curtain, the prisoner within is at all times exposed to the observation and inspection of any guard who may pass along that tier at any hour or minute of the night." Dickens, 123–24.

[891] Dickens, 125.

importance"[892] compared to the main aim of punishment, that is, general deterrence. Concerning incapacitation, Bentham speaks of the possible "disabling efficacy" of penalties, of which he claims:

> The inconvenience is, that this property is apt, in general, to run counter to that of frugality: there being, in most cases, no certain way of disabling a man from doing mischief, without, at the same time, disabling him, in a great measure, from doing good, either to himself or others.[893]

It is, therefore, discarded on utilitarian grounds.

Rehabilitation is addressed only in the form of the "reforming tendency" of punishment, which consists of the generation of aversion towards the crime. I hesitate to encompass this "property" under the category of rehabilitation because the latter seems to me relative to the reintegration of the offender in the social web in a meaningful way, subsequently preventing them from reoffending not for fear of punishment but, like the great majority of their fellow citizens, as we have already mentioned because they recognise the legitimacy of the law. Bentham's reforming tendency corresponds instead to what we have previously called "specific deterrence".

Reformation so conceived, according to Bentham, can derive from either the *quantity* or the *quality* of punishment. The former means that the greater the punishment, the greater the reforming effect – again, obviously, this goes against the principle of frugality, which is a crucial constraint to punishment according to Bentham in the utilitarian quest for minimising suffering, as we shall shortly see. Being reformation, just like incapacitation, a "minor property", it cannot outweigh frugality: it is, therefore, not legitimate to inflict disproportionate punishments on the basis of their supposed reformative effect (which, by the way, as we have seen, and as Beccaria believed, has been disproven). The latter refers to the kind of punishment meted out, which must be specifically related to the crime committed: we have already seen that the 18th-century penal reform advocated for analogical punishment, a conception that has been left behind.[894]

For Bentham, however, the central justification for punishment is clearly that of general deterrence: punishments have beneficial consequences not primarily on those on which they fall but on "such persons as are under temptation to commit it [the criminal

[892] Bentham, *An Introd. to Princ. Morals Legis.*, 180.
[893] Bentham, 181.
[894] Bentham, 180–81.

act]: in which case, in as far as it tends to restrain them from committing such acts, it is of a beneficial nature."[895] This is, of course, coherent with the utilitarian principle: a morally good action is one that maximises utility, i.e., pleasure in Bentham's hedonistic utilitarianism, and hence justifies the infliction of suffering if it serves the purpose of pleasure maximisation.

The justification of general deterrence still plays a central role in the contemporary conception of punishment; however, it is commonly believed that utilitarian justifications, based primarily on deterrence, bear an intrinsic moral risk, that of the manipulation of the individual, and therefore must be coupled with the retributive justification in order to avoid it[896]. It is the age-old criticism of utilitarianism, according to which the principle of utility, if not complemented with other moral principles, would lead to the punishment of the innocent if the overall consequences were believed to be positive. This accusation is ironic because Bentham's aim in the *IPML* was precisely to follow up on the Enlightenment quest to reduce arbitrariness in the penal system by classifying crimes and punishments clearly, establishing proportional punishments, and instituting laws that could not be derogated *ad hoc*[897].

Although he does not address the problem of the punishment of the innocent explicitly, it is clear that Bentham takes for a given that only the guilty should be punished and from his insistence on the importance of transparency, education, and shared knowledge, as well as from his anti-tyrannical stance, it is evident that he thinks that passing innocent people off as guilty to have someone to punish in virtue of the supposed deterrent effects of punishment would not be a wise strategy on the long run in terms of utility. This implies having a far-reaching utilitarian view that aims at taking into account indirect and more distant consequences rather than just narrowly focusing on the direct and nearest outcome: a practice that does, say, reduce crime through deterrence or incapacitation but that substantially contributes to the evolution of society in a autocratic and repressive direction cannot be justified on utilitarian grounds.

---

[895] Bentham, 147.

[896] This, for example, is done by Hart when he distinguishes the General Justifying Aim of punishment from its distributive principle: we may have a utilitarian General Justifying Aim, but it must be coupled with a retributive distributive principle that only allows for the punishment of the guilty. Even when this principle of distribution is breached and the innocent in punished, the fact that the punishment of the innocent is perceived as worse than the punishment of the guilty cannot be explained in utilitarian terms, which means that we intuitively adopt a retributive distributive principle even when generally justifying punishment in utilitarian terms. Hart, *Punishment and Responsibility*, 8–13.

[897] Burns, James H. and Herbert L. A. Hart, "Introduction," in *An Introduction to the Principles of Morals and Legislation*, ed. H. L. A. Hart and J. H. Burns (Oxford: Oxford University Press, 2015), lxv.

The same answer provided by Bentham is given by many present-day utilitarians to the same criticism, although without much success in ending the objection. Others, like Smart, decided to accept, albeit reluctantly, that the punishment of the innocent is an inescapable consequence of the application of the utilitarian principle. Of course, extreme examples where the punishment of the innocent would eventually be justified can always be formulated, and I challenge any non-utilitarian thinker to say that the punishment of an innocent would not be acceptable if it could prevent, say, the explosion of an atomic bomb. Maybe only Kant would allow the bomb blast to defend his categorical imperative. I think this demonstrates the hollowness of such criticism, for there is always a quantitative threshold people are not ready to trespass to respect other principles: the fact that people who identify as utilitarians may present lower thresholds might just be the result of greater awareness of this fact.

This does not mean, as McCloskey tried claiming, that utilitarians are "happy" to punish the innocent[898]. Smart neatly replied to him:

> Even in my most utilitarian moods I am not *happy* about this consequence of utilitarianism. Nevertheless, however unhappy about it he may be, the utilitarian must admit that he draws the consequence that he might find himself in circumstances where he ought to be unjust[899]. Let us hope that this is a logical possibility and not a factual one. In hoping thus I am not being inconsistent with utilitarianism, since any injustice causes misery and so can be justified only as the lesser of two evils. The fewer the situations in which the utilitarian is forced to choose the lesser of two evils, the better he will be pleased.[900]

However, it is true that an ethical theory based only on the principle of utility, one that does not admit retribution as a value, does not recognise an intrinsic difference between the punishment of the guilty and that of the innocent, except on an emotional level that moral reflection should overpower. It is the same consequence of the discredit of desert and individual moral responsibility that I have outlined so far, as well as of determinist worldviews. It derives, for example, from Beardsley's causally unlimited perspective or the determinist approach sketched by Kevin Murtagh, who ends his paper *Free Will Denial and Punishment* accepting that his view implies that it is "at least theoretically justifiable to punish a nonoffender"[901]. Among the determinists, a different

[898] McCloskey, "A Note on Utilitarian Punishment."

[899] The fact that taking the course of action that maximises utility is deemed by a fervent utilitarian like Smart as "unjust" in surprising. "Unfair" may be more accurate.

[900] Smart, "An Outline of a System of Utilitarian Ethics," 71.

[901] Kevin J. Murtagh, "Free Will Denial and Punishment," *Social Theory and Practice* 39, no. 2 (2013): 239.

approach is taken by Spinoza, who, instead of focusing on deterrence, focuses on incapacitation – a forward-looking aspect of punishment that, as mentioned, is neglected by Bentham. In his 1676 letter to Oldenburg, Spinoza writes: "Someone who is crazy because of a dog's bite is indeed to be excused; nevertheless, he is rightly suffocated."[902]

The disgust that these stances generate in us tells us that to inflict suffering on others is *always* appalling: not that we should be happy, as McCloskey suggested, of punishing the innocent, but that we should be equally scandalised by the punishment of the guilty. Instead of accepting the innocent's punishment with revulsion, as Smart does, this aversion should push us to look for alternative approaches to the problem of crime management.

In addition, it must be noted that if Bentham is to be accused of any deceptive strategy in his approach to punishment, it would rather be in the opposite direction: not to make the public believe that an innocent were guilty in order to punish them, but that someone had been punished while, in truth, they were not harmed – or were harmed much less then it is supposed. As the principal aim of punishment is general and not specific deterrence, punishment only needs to be *apparent*, a principle established by Bentham and reflected in Foucault's rule of "sufficient ideality" mentioned previously. In Bentham's words:

> It is the idea only of the punishment (or, in other words, the *apparent* punishment) that really acts upon the mind; the punishment itself (the real punishment) acts not any farther than as giving rise to that idea. It is the apparent punishment, therefore, that does all the service, I mean in the way of example, which is the principal object.[903]

Although increasing the real punishment generally leads to the increase of the apparent punishment, therefore augmenting the positive effect of the penal system, sometimes the apparent punishment can be magnified through means of particular qualities and specific rituals ("s*olemnities*") that accompany the penalty. When this can be done, generating the impression of a greater imposed suffering than the one truly endured by the condemned, utilitarianism requires this route to be pursued: utilitarians must be *frugal* in the distribution of pain, the supplement of suffering otherwise

[902] Spinoza, "Letter 78 (OP). To the Most Noble and Learned Mr. Henry Oldenburg From B. d. S.," 480.

[903] Bentham, *An Introd. to Princ. Morals Legis.*, 179–80.

imposed through the actual punishment is superfluous and should be cut.[904] Despite all this, the accusation of justifying the innocent's punishment is still a hot topic in the discussion of utilitarian justifications for punishment[905].

To this theoretic criticism, practical problems have been adding up, as decades of research on the general and specific deterrent effects of incarceration, incapacitation and rehabilitation have shown very discouraging data, as we have seen at length in the previous chapter. We have seen there, moreover, that to justify a practice on utilitarian grounds[906] it is not sufficient to show that it has some positive impact: first, its positive effects need to be greater than the harm the practice itself may cause – harms that are not negligible when we consider the carceral system, especially in this era of mass incarceration: suffice it to say that, as of 2023, almost 2 million people were incarcerated just in the U.S., a number that grows to 5.5 million if, to incarcerated individuals, those under probation and parole are added[907]; it must also not be forgotten that the negative consequences of incarceration do not just affect the individual only, but their whole family too. Second, a utilitarian justification must be shown to be the most economical in terms of suffering: again, it is not enough that more suffering is avoided than is created; the suffering that is generated must be as little as possible: if similar results in terms of crime prevention can be achieved with the infliction of less suffering, this alternative path must be taken. That this is the case for the carceral system is far from being demonstrated.

The disheartening data on deterrence, incapacitation and rehabilitation have led, as previously seen, to the common idea that "nothing works"[908] and that punishment cannot be justified on a utilitarian basis, for it does not have the positive consequences it was hoped it would have. Surprisingly, however – or not-so-surprisingly, now that we have analysed the reasons for our deep-rooted retributive attitude – this has not also led to a questioning of the practice of punishment itself. Rather, it has brought retributive

[904] Bentham, 180.

[905] See, for example: Sandel, *Justice. What's the Right Thing to Do?*, 37–40; Ryberg, "Punishment and Desert-Adjusted Utilitarianism," 89.

[906] As Herbert Hart and James Burns remark in their introduction to the *IPML*, Bentham, like Hume, may have believed that "there might not be a general justification of practices apart from existing conventions", which means that he was not aiming for a utilitarian general justification for punishment, a practice that was not put into question. More limitedly, in the *IPML* he is justifying the reform. Burns, James H. and Hart, "Introduction," lxix.

[907] Leah Wang, "Punishment Beyond Prisons: Incarceration and Supervision by State," 2023.

[908] Duff, "Responsibility, Restoration, and Retribution," 64; Robert Martinson, "What Works? Questions and Answers about Prison Rform," *Public Interest* 35 (1974): 22–54; Ryberg, "Punishment and Desert-Adjusted Utilitarianism," 86.

justifications back, and they have been, if not dominant, central in legal philosophy since the Seventies[909]. Philosophers have started defending them as more *humane* than the instrumentalising utilitarian justifications to the point that, as we have seen, Herbert Morris in his 1986 paper *Persons and Punishment* advanced the Kantian argument that to punish an offender for retributive reasons is a form of respect towards them, that they even have a "right" to being punished[910]. These scholarly positions fit well with the popular sentiment that demands that offenders receive "their just deserts".

One question, at this point, arises for me: if punishment was practised, both individually and institutionally, way before its utilitarian justifications, and it was justified in a retributive manner, and if the utilitarian justifications have been proved ineffective, but the practice of punishment in itself has not been questioned, doesn't it mean that utilitarianism has been a sort of cover, a satisfying rationalisation used to hide the real not-so-moral and not-so-rational reasons why we punish at all? Of the vengeful impulses that could not be accepted in contemporary bureaucratic States, that aimed at the rationalisation of public life? Utilitarians did provide some rational justifications for a pre-existing practice based on irrational impulses, but did not question the practice. At least, to my knowledge, not contemporary utilitarians in their reflections on punishment.

It would be unfair to say the same about Bentham himself. Like all utilitarians, he is unequivocal in underlining that the infliction of suffering is, in itself, always evil; he defines punishment as "mischief" – the same word that he uses when referring to offences[911] – and even refers to penal justice as "a train of evils":

> […] evils in the threats and constraint of the law – evils in the pursuit of the accused, before the innocent can be distinguished from the guilty – evils in the infliction of judicial sentences – evils in the inevitable consequences which reverberate upon the innocent.[912]

Contrary to most utilitarians, however, Bentham dedicated lengthy analysis to the topic of crime prevention, not simply passively acknowledging that punishment, as the lesser of two evils, must be accepted and executed, but actively working to make it unnecessary, unhinging the idea that the only alternative to it is the greater evil of leaving crime unpunished. He implicitly calls for shifting away from the binary logic

[909] Ryberg, "Punishment and Desert-Adjusted Utilitarianism," 86.
[910] Morris, "Persons and Punishment," 476.
[911] Bentham, *An Introd. to Princ. Morals Legis.*, 180, 176.
[912] Jeremy Bentham, "Principles of Penal Law," in *Works of Jeremy Bentham*, ed. John Bowring (New York: Russell & Russell, 1962), 534.

that underlies the "lesser evil" kind of reasoning and taking into consideration the bigger picture. He does so in the *Principles of Penal Law*[913], a text much less well-known than the *IPML*, whose last edition dates back to John Bowring's collected works of Bentham published in the 19th century.

Bentham analyses the topic in the third and last part of the *Principles*, titled *Of Indirect Means of Preventing Crimes*[914]. He begins with a question:

> Some actions are hurtful: what ought to be done to prevent them? The first reply which presents itself to all the world is – *Prohibit such actions; punish them*. This method of combating offences is the most simple, and the first adopted […].[915]

This method is covered by what he calls "direct legislation", which consists of the art of punishing and has been addressed in the previous part of the *Principles*, as well as in the *IPML*. Direct legislation is contrasted with indirect legislation, which covers the actions that can be taken to prevent crimes before any punishment is administered, acting on the factors that make people want to commit crime in the first place "by acting principally upon the inclinations of individuals, for the purpose of diverting them from evil and impressing on them the direction most useful to themselves and others"[916]. Once again, Bentham enters a slippery ground, for sly critics might interpret his aim as one of manipulation, of totalitarian thought control, but before jumping to such conclusions let us first analyse his proposal.

First of all, his recognition of the criminogenic influence of poverty, which we have already remarked and which is well substantiated by data on the correlation between poverty and offence rates, motivates the many pages Bentham dedicates to the so-called "Poor Laws"[917] in which he advocates both for welfare state-type aid[918] and employment support[919] – the implementation of which, it must be noted, sounds

---

[913] From now on: *Principles*. A hint to the importance of crime prevention as opposed to repression, however, is present in the *IPML*, too, when Bentham discusses "needless" punishments, which take place "where the purpose of putting an end to the practice may be attained as effectually at a cheaper rate: *by instruction*, for instance, as well as by terror: by informing the understanding, as well as by exercising an immediate influence on the will." Bentham, *An Introd. to Princ. Morals Legis.*, 164, my emphasis.

[914] Bentham, "Principles of Penal Law," 533.

[915] Bentham, 533.

[916] Bentham, 533.

[917] Other than the few pages dedicated to the topic in the *Principles*, Bentham's lengthy *Writings on the Poor Laws* are gathered in two volumes (Oxford: Oxford University Press, 2001 and 2010) of *The Collected Works of Jeremy Bentham*, the outcome of the Bentham's Project's (UCL) ongoing effort to publish all of Bentham's writings.

[918] "The effects of indigence can therefore only be guarded against, by providing necessaries for those who have them not." Bentham, "Principles of Penal Law," 543.

[919] Bentham, 544.

nowadays anachronistic and paternalistic at best, Bentham being a supporter of the workhouses. This reminds us that Bentham, despite being very progressive on many subjects, was still a man of his time and that *his* utilitarianism bears the marks of the Georgian era. This fact, however, is not a reason to reject utilitarianism, but to distinguish the *principles* of utilitarianism, which Bentham set out in his work in an exemplary manner, and the practical and political proposals that, being grounded in Georgian Britain, as we have noticed commenting the panopticon, cannot always be transposed unmodified (and sometimes cannot be transposed at all) to the present days.

Poverty aside, in the *Principles*, Bentham focuses primarily on education and knowledge sharing[920]. Passages like "I intend to show […] [t]hat the diffusion of knowledge is not hurtful upon the whole; crimes of refinement being less hurtful than those of ignorance"[921] or "The more we become enlightened, the more benevolent shall we become; because we shall see that the interests of men coincide upon more points than they oppose each other"[922], just to quote a couple, are barely ever cited[923]. Bentham seems to purport a moderate ethical intellectualism, where the positive consequences of ignorance, deception, and manipulation are only illusory in that they are limited: they may provide the best outcome, apparently, in a specific situation, but they are negative contributions to the overall picture. This is true with respect to human action in general and to criminal reduction in particular.

We can see now how Bentham's position is not only consistent with the requests of present-day prison abolitionists, it anticipates them. After Bentham, the utilitarian discussion on punishment has been chiefly focused on deterrence, following solely Bentham's *IPML*. It looks like utilitarians have jumped on the bandwagon of punishment upon shared assumptions about efficacy, but if utilitarianism aims at maximising social utility, and specifically pleasure in its hedonistic current, shouldn't the intentional infliction of suffering, i.e., punishment, be avoided as much as possible? So, shouldn't recourse to punishment only be considered a last resort when everything else has proven ineffective? Utilitarians, as Bentham has shown, should be the first to

[920] Bentham, 536–40, 563, 575–78.
[921] Bentham, 536.
[922] Bentham, 563.
[923] I do not wish to hide that other passages do support more malicious interpretations, like Bentham's mention of "a secret art of governing opinion, so that it shall not perceive, so to speak, the manner in which it is led." Bentham, 563. However, they must be contextualised in the greater picture of Bentham's proposal, which is clearly and firmly anti-authoritarian.

consider alternatives to punishment, but the path he had traced has remained unexplored in the utilitarian field.

The prison abolition movement emerged in the last decades proposing a different paradigm, alternative not to retributivism but to punishment itself. We have remarked on several occasions that incarceration is not equivalent to punishment, prisons being only recently introduced as forms of punishment. However, the carceral institution has occupied the entire space of institutionalised punishment in the Western world, so much so that we do not even consider that State-administered punishment might take other forms on the grounds that physical punishments (except for death sentences, notably and shamefully in some Western countries) are considered retrograde and barbaric. From this, it follows that the prison abolition movement corresponds to the punishment abolition movement: there is no prison abolitionist, to my knowledge, who advocates for other kinds of institutionalised punishment[924].

Prison abolition was initially theorised by the anarchists in the 19th century[925]. Anarchists, of course, based their abolitionist stances on their rejection of the State and hierarchical systems of power in general: the scope of their project was, as is clear, very broad. Present-day abolition thinking, which developed mainly in the US, has lost that wide breadth. First of all, abolitionists are often activists who are linked to specific situations and experiences, cities or even neighbourhoods and their peculiarities, as is reflected in some of the academic literature on the subject[926]. As many abolitionist scholars aim at changing practices of justice, it is vital that their research tackle specific contexts and their peculiarities, and research with a wider approach cannot substitute the valuable work they are doing. However, the two approaches should complement each

[924] Verdier and Courtois may, here, come to mind, as they advocate for the abolition of the State's criminal system in favour of a return to the vindicatory system, letting citizens settle discords without intermediation, through vengeance regulated by customary law. They are, in a way, prison abolitionists, although they do not take part in the prison abolition movement for sure. However, they still do not advocate for another form of State-administered punishment because, first, what they challenge is the institution of the State itself as an entity that has the power to inflict punishments because it holds the monopoly of violence. Second, because, although it is my thesis that punishment derives from vengeance, the two do not coincide: punishment needs to be administered by an authority, even a temporary one, a peer who is recognised as entitled to harm another. On the contrary, vengeance can be pursued by anyone, without any special kind of hierarchical relationship between the avenger and the one on whom vengeance is perpetrated. Hence, Verdier and Courtois are not prison abolitionists advocating for a different kind of punishment.

[925] For a comprehensive overview of anarchist thinking on prison abolition see: Anthony J. II Nocella, Mark Seis, and Jeff Shantz, *Classic Writings in Anarchist Criminology: A Historical Dismantling of Punishment and Domination* (Chico: AK Press, 2020).

[926] See, for example: Allegra M. McLeod, "Envisioning Abolition Democracy," *Harvard Law Review* 132, no. 6 (2019): 1613–49.

other, and more abstract and theoretical work may ground hands-on research and proposals on a firmer foundation, also helping context-sensitive results expand to other settings.

More importantly, contemporary abolition thinking is largely intertwined with the history of discrimination in the U.S., specifically with respect to slavery and racism. The American prison system has, in fact, developed dramatically after the end of the American Civil War, which resulted in the presumptive abolition of slavery. Angela Davis remarks that the abolition was merely putative[927], as the 13th amendment itself reads: "Neither slavery nor involuntary servitude, *except as a punishment for crime* whereof the party shall have been duly convicted, shall exist within the United States, or any place subject to their jurisdiction."[928] Before even stating the outlawing of slavery, an exception is specified: punishment is the means for preserving slavery.

Not only that: after the end of the civil war in the United States, so-called Black Codes were put into force, that is, bodies of law that were only applied to the African-American population. Such codes criminalised actions like leaving or neglecting a job, being drunk, or carelessness with one's money, which may be punished through incarceration and subsequent forced labour[929]. It is not surprising, then, that the triad of criminalisation, incarceration, and forced labour became effectively a surrogate for slavery in many of the Southern States after its formal abolition, and conditions of black incarcerated people victims of the convict-leasing system were often far worse than they had been when slavery was in force[930].

The American carceral system bears the marks of these origins, as has been perfectly illustrated by Michelle Alexander in her book *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*[931]. This recognition also underpins much of the work of abolitionists, and it is clearly expressed by Angela Davis herself when she writes:

[927] Davis, *Are Prisons Obsolete?*, 28.
[928] 13th Amendment to the U.S. Constitution, my emphasis.
[929] Davis, *Are Prisons Obsolete?*, 29.
[930] Davis, 32.
[931] Michelle Alexander, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness* (New York: The New Press, 2010).

> If we are already persuaded that racism should not be allowed to define the planet's future and if we can successfully argue that prisons are racist institutions, this may lead us to take seriously the prospect of declaring prisons obsolete.[932]

The work on the connection between the American carceral system and racism is essential, and the de-racialisation of crime is even more important. However, I believe that the broader prison abolition movement should primarily stand on different grounds. Of course, it is true that the U.S. have arisen as a paradigm, a model for prisons all over the world, where they have exported their punitive practices, making mass incarceration a worldwide problem by now. However, we have seen how the carceral institution as the one and only form of punishment was not originally born in the U.S., nor were punitive practices in general. Other countries, especially European countries – as in most cases in Africa, Asia, and South America, the diffusion of incarceration as institutionalised punishment is a consequence of colonisation – have their own "history of the prison", as was well illustrated by Foucault. Those histories are not related to slavery.

It is true that, almost everywhere, minorities are targets of police violence and face higher incarceration rates; however, the basic problem with incarceration does not lie here. Morally, we would need prison abolition even if incarceration rates perfectly mirrored the ethnic, religious, and financial[933] distribution of the general population: the fundamental problem with the punitive system is not discrimination; it is punishment itself. To limit abolitionist thinking to the critique of the racist aspects of the American system leaves intact the punitive structure in its more basic feature, that of the belief that some good is achieved through the voluntary imposition of suffering, whether the backwards-looking good of retribution or the forward-looking good of utility. In a word, present-day abolition thinking lacks universal breath.

It should be clear by now that I think utilitarianism offers here the perfect solution. Utilitarianism has been so closely linked with the justifications for punishment

---

[932] Davis, *Are Prisons Obsolete?*, 25. Similar positions can be found is many abolitionist thinkers, see for example: Bell, "Abolition: A New Paradigm for Reform," 45, 61; McLeod, "Envisioning Abolition Democracy," 1617; Law, *"Prisons Make Us Safer,"* 49–55.

[933] The addition of this particular element is meant to be provocative, as whereas ethnicity and religion are not themselves criminogenic and higher rates of incarceration in minorities are often the result of discrimination, things are different when we talk about economic inequalities since, as we have seen, poverty *is* criminogenic (of course, this is not to say that wealthier people do not commit crimes, unfortunately they do and, being in positions of power, their wrongdoings tend to affect many more people). Things get more difficult when we consider that minorities are often more affected by poverty than the general population, which may, in turn, result from discrimination.

for accidental reasons rather than an intrinsic punitive character of the normative framework: the first consequentialist justifications for punishment were elaborated in a specific context, that of the reform which brought to the overcoming of the torture system and the softening of penalties. Moreover, Bentham's justifications for punishment were accompanied by a fervent critique of the practice of punishment itself, the recognition that punishment is the mark of a system's failure and that action should primarily be taken through indirect legislation so that people would not commit crimes in the first place.

This latter aspect of Bentham's thought has been remarkably neglected in the subsequent utilitarian discussion on crime and punishment. This could be partly explained by the *Principles*' publishing vicissitudes, while the *IPML*, which treats direct legislation and punishment justifications at length, being one of the founding books for utilitarianism, has established the traditional bond between utilitarian thought and punishment. However, Bentham's justifications have probably been so successful for three other less accidental reasons: first, because they could fit into a system that was already in place, as we have seen in the previous chapter; second, because, as Bentham himself remarked, that of punishment is a simple solution (Guyau even called it a "popular remed[y], like boiling oil into which […] the limbs of the wounded were plunged"[934]) – though the problem it addresses is most complex, hence the utter failure of simple solutions; and, finally, because it complies with deeply-rooted emotional impulses, validating what we intuitively feel.

So, utilitarianism has been a driver of sentence softening, and many of Bentham's arguments are plainly compatible with – if not entirely in support of – prison abolition. Most importantly, however, utilitarianism offers precisely what abolition thinking currently lacks: universal breath. The utility principle is inherently universal in scope and has already served as a powerful catalyst for moral inclusivity and the expansion of the moral circle. This was palpable in Bentham's writings, with his famous

[934] Guyau, *A Sketch of Morality Independent of Obligation and Sanction*, 172. Interestingly, Guyau connects this desire for simple solutions to what we have here called the "just-world fallacy" in the compelling passage that immediately follows, which reads: "Really the desire to see the guilty one punished comes from good nature. It is chiefly explained by the impossibility for man to remain inactive, indifferent in the presence of any kind of evil. He wants to try and do something, to touch the wound, either to heal it or to apply a revulsive; and his intelligence is enticed by the apparent symmetry offered by the proportionality of moral and physical evil." Guyau, 172.

stances against the criminalisation of homosexuality[935], opposing slavery, racism[936], and even the abuses perpetrated towards non-human animals[937] , making him, to all intents and purposes, the father of contemporary antispeciesism. The passage in which these progressive stances are advanced in the *IPML* ends with the famous articulation of the fundamental utilitarian criterion for discerning moral from non-moral objects, which is still at the basis of utilitarian thought: "The question is not, Can they reason? nor, Can they talk? but, Can they suffer?"[938]. Utilitarian ethics is based on sentience, a characteristic that we share among varying ethnicities and species, but also different life choices and pathways. The suffering of people who committed crimes has, under this perspective, the same value as that of the innocent.

This universalism allows for a reformulation of abolition in broader terms. Moreover, hedonistic utilitarianism, which grounds morals on promoting pleasure and preventing suffering, represents the logical normative theory to frame abolition. Utilitarianism, as we have seen, has been focused on crime prevention through deterrence, conceiving punishment as a necessary evil. On the other hand, abolition thinking questions the idea that this evil really is necessary and aims at crime prevention through a restructuring of society that reduces the conditions that lead to violence, just like Bentham himself.

This revolutionary role is one utilitarians have been afraid to take on. Mostly, utilitarians have emphasised the continuity between utilitarianism and common-sense moral rules, showing that adopting the utilitarian perspective would not lead to an upheaval of ordinary morality. This is in part due to the fact that an implicit consequentialist element that weighs the benefits and burdens of a choice is already present, in most cases, in common-sense ethics – which is why Bentham could claim that the utility principles, before being Helvetius's, Beccaria's, and his own, "was in

[935] Jeremy Bentham and Louis Crompton, "Offences Against One's Self," *Journal of Homosexuality* 3, no. 4 (1978): 389–406.

[936] "The French have already discovered that the blackness of the skin is no reason why a human being should be abandoned without redress to the caprice of a tormentor." Bentham, *An Introd. to Princ. Morals Legis.*, 283, note b.

[937] "It may come one day to be recognized, that the number of the legs, the villosity of the skin, or the termination of the os *sacrum*, are reasons equally insufficient for abandoning a sensitive being to the same fate? What else is it that should trace the insuperable line? Is it the faculty of reason, or, perhaps, the faculty of discourse? But a full-grown horse or dog, is beyond comparison a more rational, as well as a more conversible animal, than an infant of a day, or a week, or even a month, old." Bentham, 283, note b.

[938] Bentham, 283, note b.

some sort everybody's"[939]. This is not to say that people do not tend to think about their moral duties in deontological terms, indeed in part because it is easier to commit to definite and supposedly unquestionable rules rather than engage in difficult and always uncertain calculations that need to adapt to specific situations. However, when faced with examples where the positive consequence is evident, as in the aforementioned case of killing an innocent person to prevent the explosion of a nuclear bomb, common sense morality becomes consequentialist.

The utilitarian *felicific calculus* is undoubtedly a critical element for utilitarianism, for calculating precise utilities in concrete situations whose outcomes are ultimately unpredictable is impossible, and even having confidence in expected results is extremely hard. Real-life situations are composed of too many unknown factors to allow for calculation. This has led some authors to endorse rule utilitarianism, which involves the formulation of moral rules according to the utilitarian principle and then following them even when a bad outcome is foreseen for a specific action. However, not allowing exceptions – as in the nuclear bomb example – would entail a form of rule worship that is fundamentally incompatible with utilitarianism; on the other hand, if exceptions are allowed, then rule utilitarianism is reduced to act utilitarianism, for the rule is only followed as long as following it is expected to generate the best consequences. Given the uncertainty that permeates real life, it makes sense, from a utilitarian perspective, to actually follow well-designed rules instead of needing to analyse thoroughly every single case, as that would be too time- and energy-consuming, hence inefficient, and, also, would not lead to precise predictions anyhow, though allowing for exceptions for particular cases.

Common-sense morality may, hence, be a kind of lazy utilitarianism, consisting of the following of rules that are expected to have positive overall consequences, albeit this expectation relying on intuitions and custom rather than on the meticulous analysis required by utilitarianism, and allowing exceptions in cases of clear-cut evidence of positive outcomes for waiving the rule. This reading of common-sense morality explains why, in most cases, utilitarians can claim that utilitarianism would mostly respect the moral stances of common sense.

[939] UCL, Bentham Papers, xxvii, 100, as quoted in: Schofield, "«The First Steps Rightly Directed in the Track of Legislation»: Jeremy Bentham on Cesare Beccaria's Essay on Crimes and Punishments," 68.

Utilitarians have also been concerned with emphasising the continuity between their normative framework and common sense because they had to counter the dystopic accusation of their critics. The utilitarian ideal world, in fact, is often depicted as a world of "secrecy, deception and manipulation"[940], as Primoratz put it.

These accusations are not groundless. Famously, indeed, one of the founding fathers of utilitarianism, Henry Sidgwick, did defend what is now called "esoteric morality", a morality that allows and even encourages duplicity in some cases. In his *The Methods of Ethics*, he argues:

> [O]n Utilitarian principles, it may be right to do and privately recommend, under certain circumstances, what it would not be right to advocate openly; it may be right to teach openly to one set of persons what it would be wrong to teach to others; it maybe conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world; and even, if perfect secrecy can be reasonably expected, what it would be wrong to recommend by private advice and example.[941]

This view has also been defended by contemporary utilitarians such as Peter Singer and Katarzyna De Lazari-Radek, who maintained that, when shown to have better consequences than honesty, duplicity is morally required[942]. However, again, let us think of the nuclear bomb case: if dishonesty was needed in order to avert the explosion, wouldn't it be widely considered morally acceptable to lie and deceive? We face another case where a utilitarian approach might well support a general rule – that of avoiding duplicity and secrecy and, as Bentham recommended, promoting education and knowledge sharing – that might nonetheless be broken in (rare and extreme) clear-cut cases.

A final critical point that helps explain the utilitarians' tendency to conservativism – and the general population's wariness towards utilitarianism – is the fact that, in order to promote the utility (or wellbeing, or interests, or pleasure) of the greatest number, one needs to believe that they know where that utility (however it may be conceived) may lie. This assumption generates some legitimate vigilance, for it quickly evokes moral States that intrude into their citizens' lives to impose moral values and impede self-determination through negative freedom – the freedom to pursue one's

[940] Igor Primoratz, "Utilitarianism, Justice and Punishment: Comments on Smart and Flew," *Israel Law Review* 25, no. 3–4 (1991): 390.

[941] Henry Sidgwick, *The Methods of Ethics* (London: MacMillan, 1907), 489.

[942] Katarzyna de Lazari-Radek and Peter Singer, "Secrecy in Consequentialism: A Defence of Esoteric Morality," *Ratio* 23, no. 1 (2010): 47.

objectives when it does not interfere with the other's same pursuit. Again, this accusation is paradoxical since, contrary to deontological morality, which establishes moral rules on intuitive grounds and raises them to the status of objectivity, therefore justifying their imposition on those who have different intuitions, utilitarianism is based on respect for individual preferences, on Bentham's idea that every person is the best judge of their own interests. If the "central point of utilitarianism", as Jan Narveson put it, is "not to project our values indiscriminately on other people"[943], then to think of it as a moral framework based on manipulation means completely misunderstanding it.

Nevertheless, the need to know the other's utility is a problem for utilitarians, which, once again, were pushed to defend a conservative approach: just like it is better, in general, to follow a commonsensical rule instead of calculating the precise utility of a specific action, for the risks of error are too high so that we better not trade something that may not be perfect but is still decently efficient for something that might be better but at significant risk of leading to a way worse off outcome, so we'd better support traditional liberal stances instead of allowing intrusion into people's choices on the assumption that policymakers know their preferences and utility.

Utilitarian policymaking seems to be the most problematic in this respect because, whereas in our private relationships we can simply passively accept other people's choices and focus on our own, in policymaking, if one is guided by the aim of maximising the utility of the greatest number, some understanding of the others' utility must be presumed. In this case, I believe that assuming negative utilitarianism as a rule of thumb may provide a solution to the problem and constitute the preferable framework, one that is less intrusive into people's lives, hence less dangerous from the liberal point of view, without losing too much of utilitarianism's aspiration of rationality and universality. Negative utilitarianism is the view that utilitarianism should aim at (in its strong version) or prioritise (in its weak version) reducing suffering (disutility) rather than maximising pleasure (utility).

The famous critique moved to negative utilitarianism by Roderick Ninian Smart, the author who coined the very term "negative utilitarianism" in 1958, is that if one does not give weight to positive experience, then the preferred utilitarian outcome would be human extinction[944] – from a less speciesist point of view, it would actually be

[943] Jan Narveson, *Morality and Utility* (Baltimore: Johns Hopkins University Press, 2019), 159.
[944] Roderick Ninian Smart, "Negative Utilitarianism," *Mind* 67, no. 268 (1958): 542–43.

the extinction of all sentient beings as they are capable of suffering. Suffering is inherent to life; it is one of its foundations, as it constitutes the elementary mechanism for survival. Even a weak negative utilitarianism is exposed to this risk: it is not enough to give any weight to positive experience to end up with a positive balance and escape the extinction outcome. *Enough* weight must be given, as suffering is a basic, recurrent, and often very intense experience for every sentient individual. How much precisely is unsure, just like the result of the calculation is hard to determine, but, for present purposes, it is sufficient to say that weak negative utilitarianism in itself is not enough, and it is possible that to allow for outcomes to be positive – which means to morally justify the perpetuation – of life something very close to traditional utilitarianism would be required.

I am always wary of philosophical arguments that begin from the end, that is, from the desired outcome: just like I find arguments supporting free will based on the "intolerable" moral consequences of determinism deeply flawed, so the idea that negative utilitarianism should be rejected because we are offended by the idea that accepting it would mean that, if one had the opportunity, they morally ought to painlessly kill all sentient life seems to me unsound. Not that *I* believe that we should aim at killing all sentient life, but a better argument against such a conclusion, if one is persuaded by the stance posited by negative utilitarianism, would be, in my opinion, to allow for a certain degree of moral uncertainty.

Although negative utilitarianism might seem convincing, as humans, there is always the possibility that we may be mistaken. But killing all sentient life is an action with no chance of a return, so that if we are, in fact, wrong, we may have generated immeasurable (literally, for we are not only affecting the current generation but all possible future generations, too) harm. In the face of an action with such possible negative consequences, we should always lean on the safe side and avoid ending up in undesirable scenarios from which there is no turning back. As low as we may consider the probability of error to be, the immensity of its consequences gives it massive weight in the utilitarian calculation, so we should always refrain from these kinds of actions.

However, luckily, I am not here defending negative utilitarianism *in itself*, but only as the safest form of normative framework for policy making in utilitarian terms. It is another rule of thumb that may be waived but serves as the most efficient guiding principle in a specific context. We have seen that the context of policymaking, with the possibility of the State interfering with a citizen's private life, is a slippery slope.

However, if the State's objective is not to positively maximise utility but to negatively minimise disutility, the risk of intrusion into one's own private life and choices should be reduced, the State's action tending rather to alleviate hardships citizens are going through – which are more easily identifiable compared to utilities – basically supporting the welfare state and social assistance. This perspective also has manifest consequences on the punitive system since, if the State's aim is to minimise suffering, it is evidently paradoxical that it should voluntarily inflict suffering on its citizens. Of course, the lesser evil justification always looms, but this framework constitutes a reminder that recurring to the lesser evil should be the very last resort.

Caution in all these matters must not lead to underestimating the revolutionary power of a normative theory based on a principle that can be tested on empirical data. A moral framework of this kind calls to question institutions and practices on the basis of an analysis of reality and has an intrinsic inclination to change, for the desirable configuration does not depend on the respect of fixed rules but on the concrete experience of the actors involved. The fact that a weak form of rule utilitarianism that allows for exceptions and does not end up in rule worship is compatible with the workings of common sense morality and the fact that the content of these rules, under the utilitarian perspective, may, in many cases, remain unchanged, does not mean that a rigorous analysis of real-life cases would not lead to the formulation of different rules, which is precisely what I have aimed at doing in this work with the critique of punishment. Utilitarians have been wisely focussed on rejecting the dystopic accusations that we have seen but have thus depowered utilitarianism itself, transforming it into a justification, rather than a challenge, to the status quo.

With respect to incarceration, the basic intuition underlying it is that justice is achieved through the imposition of suffering on those who harm us. Utilitarians have elaborated alternative justifications to support this intuition, but utilitarian ethics requires the exploration of alternative systems – an exploration that can be conducted on the basis of Bentham's own suggestions. Here, the alliance with the prison abolition movement comes into play. Hence, it is time to address the question that looms over abolition: what do we do with offenders?

3.3.2 Transforming Conceptions of Justice

> Where do you put the prisoners? The 'criminals'? What's the alternative? First, having no alternative at all would create less crime than the present criminal training

> centers do. Second, the only full alternative is building the society that does not need prison.[945]

As Fred Moten and Stefano Harney put it, abolition does not aim at abolishing prison; rather, it aims at abolishing societies that can have prisons[946]: the criminalisation of the individual must be overcome to address the social order's own criminality[947]. This is the central aim of what has come to be called "transformative justice", initially theorised by the legal reformer Ruth Morris[948]. What this practically required has been well synthesised by Angela Davis: a restructuring of the social configuration that subtracts space from prison in favour of physical and mental healthcare[949], school, and education on all levels[950] – the reasons behind the need for intervention in these areas should be evident now, and I will not delve into the practicalities of the proposed reforms.

As Marina Bell remarked, this allows for a reformulation of the concept of accountability, too: not accountability of the individual in the face of their crime, but of the institution that generated the conditions for crime[951]. This conception of justice, hence, is directly anchored to the notion of responsibility that has been analysed in the first two sections of this chapter: the *responsibility for justice* that is based on causality rather than blame, is shared rather than individual, and most importantly, forward- rather than backwards-looking. Transformative justice enhances the social connection model proposed by Iris Marion Young and frees justice from the murkiness of the concepts of free will, desert, and individual moral responsibility.

However, to believe that these interventions would completely eradicate interpersonal violence would be naïve wishful thinking, so the question remains: what do we do with the "criminals"? Angela Davis endorses "a justice system based on reparation and reconciliation rather than retribution and vengeance"[952], supporting the shift from the criminal punishment-based traditional justice system to restorative approaches. Restorative justice is a concept introduced in the last decades of the 20th

---

[945] Arthur Waskow, Institute for Policy Studies, "Saturday Review", 8 Janurary 1972, quoted in Davis, *Are Prisons Obsolete?*, 105.

[946] Fred Moten and Stefano Harney, "The University and the Undercommons: Seven Theses," *Social Texts* 22, no. 2 (2004): 114.

[947] McLeod, "Envisioning Abolition Democracy," 1615.

[948] Ruth Morris, *Stories of Transformative Justice* (Toronto: Canadian Scholars' Press, 2000).

[949] Davis, *Are Prisons Obsolete?*, 107.

[950] "Schools can therefore be seen as the most powerful alternative to jails and prisons." Davis, 108.

[951] Bell, "Abolition: A New Paradigm for Reform," 52.

[952] Davis, *Are Prisons Obsolete?*, 107.

century and has gained traction since the beginning of the 21st. Criminologist John Braithwaite has defined it as:

> [A] process where all stakeholders affected by an injustice have an opportunity to discuss how they have been affected by the injustice and to decide what should be done to repair the harm. With crime, restorative justice is about the idea that because crime hurts, justice should heal. It follows that conversations with those who have been hurt and with those who have inflicted the harm must be central to the process.[953]

Sometimes put in opposition with transformative justice, as it has practically coexisted and, actually, prospered within the same punitive system that transformative justice aims at supplanting, it seems to me to be, rather, an imperative component of the societal structure that transformative justice aims at building. The reasonable expectation of substantial crime reduction must be complemented with a non-retributive and non-punitive crime management system. The fact that restorative justice has partly developed within the traditional criminal system must be understood as an accidental occurrence and not as an inherent feature of restorative justice itself, the relationship between the two being extrinsic and, if not paradoxical, at least dissonant.

The traditional system, in fact, aims at severing – for presumed security reasons – the bond between the offender and the community, from which they are temporarily expelled[954]. Restorative justice, on the other hand, is based on the Durkheimian assumption that crime has broken a social bond of solidarity[955] and that justice should aim at rebuilding the broken tie. Of course, if one party is severed from society, this reconstruction is bound to fail or at least remain very partial. Restorative justice processes have taken place within prisons under the (valid) assumption that it would be better to put them into practice in these contexts than not to implement them at all. However, not all restorative processes of justice have followed this scheme, nor are they bound to.

It has been noted that many indigenous societies have dealt with crime through proto-restorative justice systems. This should not be surprising given the concentrical circles structure of archaic society previously exposed: within the inner circle, as we have seen, breach of rules is handled in a ritualistic manner with the goal of

[953] John Braithwaite, "Restorative Justice and De-Professionalization," *The Good Society* 13, no. 1 (2004): 28.

[954] It is no accident that prisons are typically located far away from city centres, in sparsely populated areas, and that the access of volunteers – the few who are willing to reach these places – is strictly limited.

[955] Durkheim, *The Division of Labour in Society*, 31.

reintegrating the transgressor within the group, healing the broken bond. Some of the studied examples of these restorative processes can be found among the Māori in New Zealand, who were a driver for the implementation of restorative procedures in the juvenile system of the country[956].

Lévi-Strauss contrasted our contemporary severance-based, "anthropopoemious"[957] judicial practices with anthropophagous – in both concrete and figurative senses – practices of some indigenous populations of the North American Planes. In *Tristes Tropiques*, he argued that our own practices would seem as absurd and barbaric to these indigenous populations as their cannibalism appears to us. In fact, this quest for "incorporation" of the other, which takes the form of cannibalistic rituals, is equated by Lévi-Strauss to the judicial practices of these North American natives; in these societies, an individual who would break the law would indeed be punished, but the police force in charge of meting out the punishment would then need to make amends and compensate the offender for the harm they had caused them. These compensations needed, then, to be recompensated from the offender, who was helped by the whole society – police included – to make a counter-gift. This exchange would go on in a back and forth of reciprocal gifts until the social bond was re-established, the former criminal re-incorporated as a legitimate member of the society.

Lévi-Strauss interprets these practices as a form of infantilisation of the offender who, like a child, is punished but then, again like a child, needs to be consoled. One can object that we should not treat offenders like children and also that children should not be punished when they do something wrong, but I think Lévi-Strauss got it right when he claimed that the idea that the State should educate adult offenders (a task that it performs through punishment) is extremely infantilising, and that, with this respect, our punitive system is the most paradoxical of all. In fact, he noticed, it both infantilises the offender by punishing or "educating" them, and then treats them like adults by not consoling them after they have been punished.[958]

---

[956] Howard Zehr, *Changing Lenses: A New Focus for Crime and Justice* (Scottdale: Herald Press, 1990), 257.

[957] "From the Greek *emein*, to vomit". Lévi-Strauss, *Tristes Tropiques*, 386.

[958] "It is simply grotesque to believe that we have made a 'great spiritual advance' simply because, instead of eating our fellow human beings, we prefer to mutilate them, both physically and morally", wrote Lévi-Strauss (Lévi-Strauss, 387.) almost quoting Montaigne's essay *On Cannibals*, where he contrasts cannibalism with the judicial torture system of his times and argues: "I consider it more barbarous to eat a man alive than to eat him dead; to tear by rack and torture a body still full of feeling, to roast it by degrees, and then give it to be trampled and eaten by dogs and swine – a practice which we

Although Lévi-Strauss's description of the North American Planes natives' system indeed seems better than our present system, and I think it can be appropriately included among restorative justice systems because it aims at restoring the broken bond between offender and society rather than severing it definitely, I will not subscribe to a view of this kind and will rather argue for non-punitive approaches to crime management, in which offenders are not infantilised, as they are not recipients of an education imposed from above, but rather enter into a discursive process among peers, as we will shortly see.

Finally, the Rwandan "*gacaca*", a traditional judicial system that was in place before the European colonisation of the region, is also depicted as a restorative system. In Rwanda, German and, subsequently, Belgian colonisers exported the carceral system, effectively supplanting the *gacaca*. Interestingly, however, Rwanda exhumed the pre-colonial system after the 1994 genocide, the aftermath of which presented a considerable practical problem: the perpetrators, the *genocidaires*, were so numerous that Rwanda did not possess the infrastructures to incarcerate them all[959]. To face the crisis, Rwanda's president, Paul Kagame, instead of accelerating the building of new prisons and jails, decided to progressively decarcerate all *genocidaires*, except for the ten thousand that were considered to bear the greatest responsibility – roughly a million cases would be judged in *gacacas*[960].

In the *gacaca*, the accused is evaluated by a judge popularly elected by the community they have hurt in an open trial taking place in a field. The aim of the process is to achieve reconciliation, which traditionally could be produced through material compensation, beatings or even death[961]; in 21st-century *gacacas*, however, those found guilty mainly were sentenced to "works of public interest", i.e., construction of public infrastructures that were needed to rebuild the country after what had happened. The process did not aim at deterrence, incapacitation, or rehabilitation, to recall the classical consequential justifications for the traditional criminal system. To discard these

---

have not only read about but seen within recent memory, not between ancient enemies, but between neighbours and fellow-citizens and, what is worse, under the cloak of piety and religion – than to roast and eat a man after he is dead." Montaigne, *The Complete Essays*, 113.

[959] Baz Dreisinger, *Incarceration Nations: A Journey to Justice in Prisons Around the World* (New York: Other Press, 2016), 28.

[960] Dreisinger, 28–29.

[961] Dreisinger, 29. I interpret the idea that reconciliation is achieved by death, which is itself a paradox, that death was meted out when reconciliation was considered impossible, hence the killing of the offender ended up being the only way to bring peace back into the community.

objectives implies having a different view of the offenders and the rest of the citizens in the society: the other citizens are not potential offenders that the authority needs to scare in order for them not to commit crimes, and offenders are not dangerous creatures to be severed from the community, to be incapacitated and rehabilitated. Offenders are citizens who did something wrong, sometimes something horribly wrong, but remain people to listen to, to talk to, with whom new relationships can be built.

With regard to the retributive justification for punishment, however, the comparison is not so flattering. *Gacacas*, like other restorative justice systems, give voice to the victims that traditional Western criminal systems silence instead; in the case of Rwanda, given the scale of the crimes, the victim is the entire society, especially the Tutsi population. Restorative justice reclaims victims' right to be heard, to be part of the process that follows the offence, allowing the shift from victim – a passive receiver of the offence – to survivor – one who takes agency with respect to what has happened. As criminologist Nils Christie, one of the pioneers of restorative justice, has said, the traditional criminal system *steals* the conflict from victims and offenders – for whom the offence might be the most important and determinant event of their life – and hands it to professionals who have no personal interest in the matter. Instead, conflicts should be considered *property* of victims and offenders, and they should be granted some level of autonomy in dealing with them[962].

However, as has been noted by Barbara Hudson, in practice, in restorative justice processes, victims tend to monopolise the conflict. Whereas they had been completely excluded from the traditional criminal trial, restorative justice has focused on their needs, even at the expense of offenders' rights[963]. This is why a form of justice that could lead to the beating or even the killing of the wrongdoer, like the old *gacacas*, could count as a precursor to restorative justice. I said that the comparison with traditional retributive justifications is not so flattering because, if for restorative justice we mean that the meeting of the victim and the offender has the sole aim of letting the victim express their needs and have them fulfilled through the encounter with the offender, we are going back to a purely Girardian framework, the same that explained and justified the traditional criminal system. We have seen how, according to Girard,

[962] Nils Christie, "Conflicts as Property," *The British Journal of Criminology* 17, no. 1 (1977): 1–15.

[963] Barbara Hudson, "Victims and Offenders," in *Restorative Justice and Criminal Justice: Competing of Reconciliable Paradigms?*, ed. Andrew Von Hirsch et al. (Oxford and Portland: Hart Publishing, 2003), 187.

the point of criminal justice is to provide satisfaction to the victim, who fundamentally wants revenge, averting the risk of entering into destructive, never-ending blood feuds. The State enters into play to avenge the victim and put an end to violence. The idea that restorative justice aims at providing satisfaction to the victim is just a repetition of this same pattern, only recognising that the criminal system has failed the victims and thus proved incapable of fulfilling its original task. But I believe that this understanding of restorative justice misconceives it: restoration is not achieved once the victim has been satisfied, not even once they have forgiven. Restoration is a process that needs to be bidirectional; the offender must be taken into account, too.

As Hudson understands it, restorative justice is a form of *discursive justice* as opposed to *process justice*. The latter consists of the "application of right rules to standard cases"[964], entailing the problematic categorisation of complex human experiences, especially in the case of conflict, with the need to examine and classify the involved parties' motives and mental states into standardised classes. Discursive justice, on the other hand, is expressed through the "empathetic reflection on events"[965]: the gathering between the victim and the offenders is bidirectional, and so empathy must be for the process to work and reconciliation to be achieved. Instead of boxing up experiences into predefined cells, discursive justice requires considering the individuality of each of the conflicting parties, taking into account the needs and specific situations of both.

Traditional courts focus on the offender as the *object* of the trial, with the court aiming to decide their punishment. It may look like restorative justice could maintain the same structure, just adding the victim as an additional *subject* to the process, expressing needs and making decisions for the offender's destiny. But restorative justice really disrupts the traditional structure: because of its discursive nature, it requires considering both the victim and the offender as subjects. Both of their lives are at stake; both have viewpoints on what happened; and both need to be taken seriously and listened to.

Christie remarked that excluding the victim from the process of justice would leave them angry and humiliated, making them

[964] Hudson, 192.
[965] Hudson, 192.

> need all the classical stereotypes around 'the criminal' to get a grasp on the whole thing. He has a need for understanding, but is instead a non-person in a Kafka play. Of course, he will go away more frightened than ever, more in need than ever of an explanation of criminals as non-human.[966]

However, understanding cannot be achieved if the victimiser is not listened to either: contrary to monsters, who hurt others just because they're evil, human action is moved by reasons – reasons that we may consider improper, mistaken, or selfish but reasons that we, as fellow human beings, can try to understand. Offenders can explain themselves, and they can offer apologies; they may or may not be forgiven, but regardless of this type of outcome, the dialogue can be crucial for the healing process of both parties.

This kind of discursive justice has been put in place in the aforementioned South African Truth and Reconciliation Commissions after the end of the apartheid regime. The choice to forego the traditional criminal trial was due to the belief that, apart from being a massive waste of resources for an already struggling country, such trials would permeate society with vengeance and further deepen the fracture between the black and the white population. However, general amnesty would "revictimize the victims, by failing to acknowledge their suffering."[967] TRCs aimed at establishing truth, that is, at having culprits acknowledge and explain what they had done, taking responsibility. The traditional criminal system incentivises the denial of responsibility in order to avoid punishment, consisting of a second confrontation between victim and offender – this time vicariously through the lawyers – where each aims at the victory as if the trial were a zero-sum game: the victim is winning if the offender is condemned, the offender if they are absolved. TRCs took punishment away from the picture: the result of acknowledgement and explanation would be amnesty. This opened up the possibility (which obviously was not always realised) for the transformation of confrontation into dialogue, encounter, reconciliation.

Of course, the processes put in place to restructure a society after decades of institutional discrimination and violence cannot be directly transposed to address interpersonal harm and crime. In particular, the relative positions of victim and offender change in that offenders, for the most part, are not in this case sustained by a wider structure that endorses their action: they are often not in an actual position of power,

[966] Christie, "Conflicts as Property," 8.
[967] Nussbaum, *Anger and Forgiveness*, 240.

although they impose themselves on their victims through their crimes. As we have mentioned, interpersonal crime frequently arises from contexts of deprivation, from poverty, mental illness, and abuse. This allows for the creation of a deeper dimension for the discursive process of restorative justice, in which the parties may strip off the rigid role of victim and perpetrator and meet each other as human beings, each with their own personal history and sorrows.

To quote Guyau again, that kind of justice must be overcome "which refuses kindness to him who is already unhappy enough to be guilty"[968] – their sorrow, their struggle, their humanity must be recognised too. And if they are recognised as struggling too, it is a powerful moral stance to take those struggles into consideration: to act on them, to overcome the intuitive and emotional response to harm, not forgiving – which would imply accepting anger as the appropriate reaction that, in a supererogatory act, is renounced – but simply addressing it and working to relieve it.

This is something that the community must commit to. Restorative justice, as part of the broader framework of transformative justice, does not leave the victim, offender, and facilitator to deal with reparation alone, as would happen in restorative programmes within the traditional retributive system. This is partly because it understands that crime is not a private affair between two individuals: it is not a simple quarrel, a betrayal of mutual agreement and, hence, of mutual trust. A crime is the infringement of a social bond, and although the direct victim is obviously primarily affected, it touches the whole community. This is the idea behind the incapacitation-based justification for incarceration or death penalty: crime creates a fracture in the social fabric, the offender has betrayed all society for they have not abided by public rules, and the fact that they have behaved so once is an indicator that they may do it again. Once a victim has been hurt, everyone is in danger: the offender cannot be trusted anymore as a reliable member of society.

However, under the perspective I have been developing here, crime is a community issue under another critical respect: that of the generation of crime resulting from structural injustice or constitutive bad luck. The communal restorative justice process also conforms to the social-connection model of responsibility: the community's involvement in the healing of the broken social tie is a form of taking *responsibility for justice*, as Young advocated for. Responsibility for justice does not

[968] Guyau, *A Sketch of Morality Independent of Obligation and Sanction*, 167.

stop at prevention; it must address restoration, too. While, as we have seen, the retributive justice system is based on the individualisation of guilt through the concept of a moral responsibility based on an incorruptible free will, restorative justice rests on shared responsibility, shifting the focus from backwards-looking guilt to forward-looking reparation.

Obviously, this does not mean that, the offender can be a passive agent in the process; that the community is taking responsibility for them. Responsibility is taken *together* by all actors involved, each from their specific position – and, of course, offenders need to occupy a special place if the process is to be successful. It also does not mean that, to achieve restoration, the offender would not (virtually) inevitably need to go through some forms of suffering. However, it is not the voluntarily-administered-because-deserved suffering of retributive punishment; it is the suffering that is generated by public censure, by being accused of having been socially inadequate – which typically generates shame – and, most importantly, that of the repentance or regret that is inherent to the recognition that one has done something wrong, that one, retrospectively, has not lived up to one's own standards, even that one may not be able to make up.

Nonetheless, I disagree with Anthony Duff, who claims that the fact that these kinds of suffering need to be involved in the restorative process means that retributive and restorative justice can not only coexist, but that the former is a constitutive element of the latter because punishment is necessary for restoration[969]. Retributive punishment, we now know, is the voluntary infliction of suffering based on desert, not the suffering that is an inevitable component of a certain process: the fact that a woman giving birth necessarily undergoes suffering in the process does not mean that she is punished for giving birth. The fact that Duff, after having talked about regret as an appropriate kind of suffering for a restorative process, maintains that public censure needs to involve the voluntary infliction of suffering as a way of showing that the victim's own pain has been taken seriously by both the offender and the community demonstrates his misunderstanding of restoration, mistaken for the Girardian satisfaction of the victim's desires that we have addressed earlier.

[969] R. Antony Duff, "Restoration and Retribution," in Restorative Justice and Criminal Justice: Competing of Reconciliable Paradigms?, ed. Andrew Von Hirsch et al. (Oxford and Portland: Hart Publishing, 2003), 43–59.

Of course, the discursive restorative processes that involve the suffering of regret might fail: the offender may not repent, reconciliation may not be achieved. It looks like Duff is trying to find a solution to this problem: since we cannot know whether the offender's repentance is sincere – as we cannot know whether to their words an inner state of regret corresponds – we'd better inflict on them an objective amount of suffering to complement or entirely substitute that of contrition. But this is not how restoration works: the offender's suffering may be a necessary step for achieving reconciliation (which I assume for the present discussion but am far from certain), but, in contrast to retributive justice, this suffering is not the point. At most, it is an indispensable yet accidental phase; punishment, although also involving suffering, does not lead to the same end. Facilitators and victims can make efforts to clarify to the offender the harm that they have caused but have no control over their feelings of regret, just like the victim can fail to overcome anger and trauma and reject reconciliation. The restoration of broken bonds is not the inevitable result; restorative justice processes are not always successful, unfortunately.

In the words of Brown and Schept:

> Abolitionists do not assume that we will ever exist in harm-free spaces; rather, they emphasize that vulnerability is bound up with the human condition, that the pursuit of total security is a life-damaging pursuit, and that there are ways to better address, collectively, the harms we face and produce.[970]

Through the idea of incapacitation and deterrence, the carceral system has sold the public a fake perception of total safety. However obvious it may be that the current system does not provide this kind of safety, it is perceived as the system that gets the closest to it: the deployment of police forces, the thorough investigations, the multi-stage criminal trial, together with black chronicles (and, in recent years, true crime online content, which has proven enormously successful) both create a sense of insecurity – that justifies the deployment of these forces – and one of security: the problem is being taken seriously, justice will be made and this will positively affect the future through crime reduction. We just need *more* of whatever the current system is doing.

Foucault analysed a similar twofold psychological mechanism that serves the same purpose of justifying the carceral system through the black chronicle and its

[970] Brown and Schept, "New Abolition, Criminology and a Critical Carceral Studies," 449.

precursors, as the crime novel. In these accounts, "delinquency appears both as very close and quite alien, a perpetual threat to everyday life, but extremely distant in its origin and motives, both everyday and exotic".[971] To him, this serves the purpose of getting citizens to accept police surveillance, for they need its protection since they may be the next delinquents' victims. At the same time, they do not feel at risk of becoming delinquents themselves, and hence the police and penal justice victims, because crime, although being something close "to us", is also something different "from us". It is a theory that presents somewhat conspiratorial overtones: the State puts in place this system so as to justify surveillance through delinquency, hence the protection of the population from harm, while, in truth, its real objective is not delinquency but political crime instead – not the protection of private citizens, but of structures of power. "So successful has the prison been that, after a century and a half of 'failures', the prison still exists, producing the same results, and there is the greatest reluctance to dispense with it."[972]

My own idea of the duplicity I emphasized is simpler; it does not entail a malicious mastermind that aims at taking advantage of the carceral system to protect their power. Instead, the system creates a sense of safety because it can show the implementation of concrete actions that correspond to our intuitions concerning the correct response to harm. On the other hand, it creates a sense of insecurity that sustains it – for if security was really achieved, then we would not need it anymore. If we aim at perfect security, and, of course, we haven't reached it, then we need more repression, more retribution, more of our intuitive response to crime. Failure is attributed to the insufficiency of the forces utilised. This view allows us to feed our retributive intuitions. But leaving wishful thinking aside, it is clear that the carceral system, which is not providing us with perfect security now, will never be able to do so. Total prevention would look more like a dystopic scenario where illicit behaviour is repressed *before* it has been performed, thanks to a very accurate human behaviour prediction system. Evidently, to live in such a world is far from desirable.

Moreover, it must be recognised that we already accept that the current system does not criminalise and punish *all actions* that victimise and generate harm. We have seen that Durkheim had noted that those responsible for market crashes and economic

[971] Foucault, Discipline and Punish, 286.
[972] Foucault, 277.

crises often do not undergo any consequences for their actions. More recently, Law has similarly analysed public health crises, like the opioid epidemic that has been afflicting the U.S. since the Nineties, or environmental disasters. John Braithwaite analysed the absence of redress for white-collar crimes, especially pharmaceutical malpractice[973]. These are all events that generate enormous amounts of suffering and harm, incomparable to most of those that are judged through the criminal justice system and they show that we, in fact, do accept living in a system where we may be harmed, and those (causally) responsible for this harm may not be punished as a consequence. The reason they are not punished, moreover, does not lie in a failure of the system – i.e., we cannot identify them – but is part of the system's design. This is not to say that it is desirable that these people, as in Durkheim's and Law's examples, should not undergo *any* consequences for their actions. However, it shows that criminalisation is not, at present, our only response to harm. Sometimes, we give no response at all, and these examples show that even when this is the case, we do not live in constant and extreme danger and fear for lack of deterrence as a consequence. But the absence of punishment does not necessarily mean the absence of consequences or responses at all.

As Braithwaite noticed, once we realise that some harmful actions are not criminalised, instead of calling for their penalisation as it often happens, we can learn that "mainstream criminal justice [may] become more like the restorative, compliance-inducing processes"[974] that are put in place in other cases, like in addressing corporate crime: we can learn that we do not *need* punishment. Given the evident failure of the current system in providing prevention and rehabilitation, the least one should do is explore alternative systems. Alternative systems may sometimes fail, too, of course. But these failures must not be compared to a perfect system that is on the path to eliminating all crime but to a system that has never provided, nor will it ever provide perfect safety. On the contrary, as we have seen, it has a negligible deterrent effect and criminogenic influences. From this perspective, the failures of restorative processes may seem less concerning.

The alternative I have outlined here is based on the rejection of reactive emotions as valid moral guides. It is rooted in a broad-scope hedonistic utilitarian theory, one that aims at taking into account the consequences of actions on a scale that

[973] Hudson, "Victims and Offenders," 177.
[974] Hudson, 178.

is wider than the one that utilitarians have used in their justifications for punishment. Many utilitarians, betraying Betham's pioneering approach, focused on crime as an isolated event to be addressed on the individual level. Following Bentham, I instead proposed inscribing crime into the bigger picture of social organisation and primarily acting on the structural causes of crime. Once they are addressed, it is again utilitarianism to require that the restorative justice approach, which is based on a utilitarian aspiration – that victims, offenders and communities can be reconciled without the infliction of more suffering than that which has already been created – should be explored.

Whether the results will be satisfactory, whether societies can overcome the practices dictated by reactive emotions and base their moral and political stances on rational grounds, whether individuals can do without substitutive sacrificial victims, letting go both of actual sacrifices and of the victimisation of victimisers; these are all empirical questions that will need to be assessed as a transformative justice approach takes over policy reforms.

Conclusion

In this thesis my aim has been to address two main questions: firstly, where does the idea of retributive justice come from? In other words, how did we come to believe that inflicting suffering on another may count, in itself, as a moral value? Secondly, could justice take other forms? What does non-retributive justice look like? The first question has been tackled in the first two chapters, the second in the third and last.

In the first chapter, I have introduced my theory on the archaic origins of retributive justice. I have started from the problem that is chiefly addressed by means of retributive justice: that of interpersonal violence, the fundamental problem of human coexistence. I have embedded this subject within a Girardian framework, in which human violence has existential implications because of human mimetic behaviour: once the spark of interpersonal violence has been ignited – and, according to the mimetic theory, this spark will necessarily be ignited because of the rivalries that mimetic desire inherently generates – then it will spread like wildfire, endangering the very existence of the entire community. Violence will become the object of imitation itself and lead to a crisis of generalised violence.

As I have shown, ancient communities did not solve this fundamental problem through retribution. Due to the lack of a strong, centralised authority, retaliation – which is precisely the imitation of the previous violence – cannot solve the problem of interpersonal violence; rather, vengeance is the first step of the escalation that leads to the generalised crisis. Hence, archaic societies needed interpersonal violence to be prevented rather than cured: this preventive role is taken on by religion through sacrifice, the ritual repetition of the original solution that communities randomly found for the mimetic crisis, i.e., the founding homicide. The founding homicide is the killing of a scapegoat who is blamed for having caused the crisis: their collective murder unifies the community, allowing the temporary overcoming of rivalries, hence bringing peace back to the community. The scapegoat is then perceived as supremely powerful in that they have generated both the crisis and its resolution and is therefore divinised: religion in born.

Sacrifice, which lies at the core of most known religions, repeats this violent act, similarly canalising the community's internal tensions, periodically giving violence an outlet and creating a sense of commonality that allows for the prevention of

interpersonal violence. The founding murder before and the sacrifice after establish both religion and morality, the first through the divinisation of the original scapegoat, the latter through the distinction of bad violence (individual and warmongering) from good violence (unanimous and peacekeeper). Hence, the need for dealing with the fundamental problem of human coexistence, i.e., interpersonal violence, generates religion and morality. Not only that, the management of interpersonal violence is also, in fact, the reason behind the institution of the State, as has been emphasised by social contract theories (from Lucretius to Rawls, passing through Hobbes, Locke, and Rousseau, among others). These theories' fictional state of nature corresponds to the Girardian crisis of generalised violence, and the abdication of one's rights of retaliation in favour of an authority that monopolises violence is the tool to break the cycles of reciprocal violence and put an end to reprisals: the State is the avenger against which one cannot take their revenge, hence being able to put a final word to violence.

While social contract theories are meant to be fictional stories and not historical accounts of the birth of the centralised modern state, they do highlight its primary function and reason for being, and it is the same as religion's. They, in fact, accomplish the same task – the management of human violence in order to allow for communal life and avoid destructive internecine conflict – religion originally doing it preventively, the State through both preventive and repressing means. I have claimed that this explains the relevance of the topic of justice and retribution in all religions, both in their intertwining with the State's preventive-repressive procedures and in their theology. Concerning the latter, I have chiefly analysed Christian conceptions of justice and have shown how the tendency to unhinge retribution that characterised both early Christianity and early Protestantism was abandoned in favour of the strictest forms of retribution. Jesus's predication and early Christianity professed God's unlimited grace and forgiveness, whereas later Christianity was so focused on retribution as to introduce Purgatory in order to achieve perfect retribution in the afterlife. On the other hand, the Protestant rejection of retribution passed through extensive damnation and justification by grace alone, but in ordinary preaching, the importance of success as a *sign* of salvation was quickly transformed into its cause. Both early Christianity and early Protestantism tried to reject a retributive divine world design and historically failed: Christianity became the most strictly retributive religion. I have claimed that this fact points in the direction of a deeply rooted retributive inclination and that such religious conceptions of justice as retribution helped establish this inclination as an ideal for

society, a moral value, granting it intrinsic worth rather than accepting it as a human impulsive response to harm.

In the second chapter, I have discussed the psychological origins of this inclination, but only after having analysed the role that retribution plays in contemporary practices of justice within the framework of the modern State after the transition from a religion-centred to a State-centred management of justice was completed; I have completed the chapter investigating contemporary thinking on the subject and its origins in modern philosophy.

The modern State is presented as a rational entity: it is not a case that its most successful philosophical justification took the form of the social contract theory, according to which isolated individuals rationally decide to subscribe to a contract in order to live peacefully in a society. I have shown how retributive justice is integrated within this framework, presented as the rational, regulated counterpart to private, disproportionate, savage vengeance. This is why the State's retributive practices focus much attention on proportionality and a transparent process for ascertaining guilt; however, why transparency, guilt, and proportionality should make retribution rational is not ever explained, nor is it shown to be fundamentally different, in its principle, from the same vengeance that it opposes.

However, retribution has risen as the paradigm of a just society to the point that the Western ideal of a society's structure, that is, the meritocratic society, is based on the retributive mechanism of punishment and reward. Meritocracy does not simply entail that people who are better at a certain job should get it for efficiency reasons. It is a system – as was perfectly shown in Michael Young's dystopia *The Rise of the Meritocracy* – that categorises individuals as better or worse in more general, political, and even moral terms so that those who are better at certain tasks that are (arbitrarily) valorised by the society are considered as overall deserving more. On the other hand, the losers of the system, as noted by Dumouchel, can be used as sacrificial victims, as the meritocratic mentality blames them for their struggles.

However, I have shown how this violence that is perpetrated towards the losers of the meritocratic system – the violence that Dumouchel has dubbed "violence by omission" – has nothing to do with the one perpetrated towards those who have committed crimes, whom I believe to be the real sacrificial victims of the contemporary world. Through Nagel's concept of constitutive luck, I developed a comparison between offenders and people in financial difficulty in meritocratic societies, the elderly

abandoned in retirement homes, and, chiefly, people suffering from mental illness. While the latter – the poor, the elderly, the mentally ill – are all victims of forms of violence by omission, offenders are the only ones who are the object of violence by execution: the ones towards whom the perpetration of violence does not need to be hidden because it is plainly recognised, just like the violence of the sacrifice, as *good* violence.

I then showed that the origin of the moral justification for retribution lies in the intuitionist approach to ethics. The inclination towards retribution is based on reactive attitudes, which, in turn, rest on retributive emotions: anger and resentment above all. These emotions are rooted in our animal biology – they are, in fact, shared with other mammals. Animals in the wild, the environment in which natural selection took place, struggle to survive in the face of different kinds of threats, among which is the aggressive behaviour of other animals. To impulsively react with retaliation to aggression is a vital defence strategy as it teaches the adversary that they cannot take advantage of us. Through ethical intuitionism, the meta-ethical theory that grounds moral knowledge on intuitions, retaliatory impulses are raised to the status of moral truths. However, I have maintained that, whereas such emotions have been crucial to grant survival, they have nothing to do with morals. Ethical intuitionism is based on amoral impulses whose function is to grant the individual's survival: they, therefore, constitute self-centred, biased ethical theories. It is such theories that transform vengeful impulses into moral duties.

In opposition to ethical intuitionism, I have argued in favour of a another realist moral theory, one based on the negative and positive quality of experience (i.e., pleasure and pain) and have identified hedonistic utilitarianism as such a theory. Before elaborating a theory of justice within this normative framework in the last chapter, I have analysed the conventional justifications for punishment: utilitarianism is, in fact, the traditional framework for forward-looking punishment justifications. If one wishes to justify punishment through the classical consequentialist triad of deterrence, incapacitation, and rehabilitation, however, one needs to substantiate it with empirical evidence, and I have shown that such evidence is missing.

As for backwards-looking justifications, I have claimed that they either resolve into the satisfaction of the victim's urge for vengeance or the re-establishment of the victim's offended honour through the abasement of the offender. In the first case, the justification turns out to be consequentialist, and therefore, to justify punishment, we

need to weigh the positive consequence of the victim's satisfaction with the negative consequence of the offender's suffering. Clearly, the latter outweighs the first, and punishment is thus unjustified. As for the honour-based justification, I have followed Martha Nussbaum in her rejection of the valorisation of honour, rank, or status in contemporary societies, hence denying that the harm generated by crime that is to be amended by punishment should be interpreted in such terms: I agree with her claim that the problem with offences must not be understood as harm to one's status, but should be identified in the concrete suffering of the victim.

Despite all this, I have argued that the retributive system stays firmly in place because it perfectly fits many aspects of the contemporary society structure and worldview. I recalled Foucault's and Angela Davis's theories of the economic and political fitness of the carceral system, which aims at producing submissive citizens and consumers. I have then evoked my own theories of the "emotional fitness" of the system, which panders to the spontaneous reactions to aggression, and of the religious fitness, especially within the Christian framework and its theories of free will and afterlife retribution. Finally, I have introduced the "conceptual fitness", or the fitness of the carceral-punitive system with the conceptual framework, constituted by the interconnections of free will, desert, moral responsibility, and guilt, that is dominant in the Western contemporary world.

In order to substantiate this idea, I have chosen to analyse the major philosophical theories that have been advanced on the subject in the 17th century in the works of Hobbes, Descartes, Spinoza, Malebranche, and Leibniz. I have chosen this century and these authors because I believe that, a few decades after Luther's reform (that was grounded on the rejection of free will and desert), these modern philosophers forged, with the unprecedented attention they devoted to these issues, the conceptual horizon in which the discussion over these subjects still takes place in both philosophical reflection and common sense, on the materialistic and determinist side – through the work of Hobbes, Spinoza, and Leibniz – as well as on the libertarian, idealistic side – through the work of Descartes and Malebranche.

In connecting their theories to contemporary positions, I have shown how Descartes' and Malebranche's justifications for free will and moral responsibility based on religious dogmas have been preserved in Peter Strawson's and Smilansky's works, who only exchanged the religious authority for that of an intuitive morality. I have emphasised the continuity between Hobbes', Spinoza's, and Leibniz's consequentialist

justifications for punishment and those advanced by Beccaria and Bentham during the 18th-century penal reform, to then inform the contemporary common justifications for punishment. I have also argued that Leibniz's aesthetic considerations, his conception of retributive justice as the restoration of an infringed harmony, are also central in common ideas of justice, as in the widespread imagery of justice as a scale and punishment as payment of a debt. I have pointed out the similarity between theological and physical determinism, as well as the uniformity, over the centuries, of internalist explanations of free will like that of Malebranche and, at present, Chisholm. The problems related to the possibility of God breaking mechanistic laws of nature are now transposed onto quantum mechanics. Finally, I have shown how the importance of free will that was firmly established by these authors has been integrated into the criminal trial through the analysis of the *mens rea* and the discussion, in contemporary legal philosophy, on excuses and exemptions of the guilty party. I have ended the second chapter noticing that all these reflections on accountability, however, do not tackle the fundamental question of why accountability should generate desert and hence a moral justification for punishment, leaving retributive justice morally unjustified.

I have explored the possibility of alternative paradigms of justice that could include harm and crime management in the third and last chapter of this thesis. I have started presenting R. Jay Wallace's, Bruce Waller's, Elizabeth Beardsley's and Hilary Bok's challenges to the retributive paradigm, but have noticed how they all based their critique on the rejection of free will, and of individual moral responsibility as a consequence, rather than criticising the very idea that responsible action constitutes the foundation for desert and punishment. On the basis of Gestalt psychologist Fritz Heider's attribution theory, I have investigated the possibility of grounding on causal responsibility a type of responsibility that is shared, which I have called, following Charles Taylor, co-responsibility, unhinging the idea of individual moral responsibility. In dealing with harm and crime, this shift moves the focus from the treatment of the individual offender to a systemic reconsideration of the conditions that generate crime.

I have argued that this systemic approach is consistent with recent theorisations of social justice – the idea, developed after World War II, that justice is a feature of a society's structure and is achieved when the benefits and burdens of living together are distributed fairly among its members. Whereas social justice is typically understood as complementary to retributive criminal justice, for social justice is rooted in national solidarity and offenders are excluded from solidarity bonds, I have maintained that such

complementarity may be overcome precisely through the systemic approach to crime that I have sketched. Once constitutive luck is integrated into the picture, just like it is crucial, from a social justice perspective, to aid the poor because they are not considered responsible for their misfortune, so it is to assist offenders rather than exclude them from the circle of solidarity as incarceration does. This approach to the issue, I have argued, would constitute a broadening of Dworkin's luck egalitarianism. However, I have shown that although it eliminates the scope of application of retributive justice, it does not undermine the logic behind it. My aim has not been to show, through constitutive luck, that there is no free will and hence that retribution cannot be fairly applied, but that retribution itself is a moral oxymoron that is hidden behind a conceptual veil made of the notions of free will, desert, blame, guilt, and individual moral responsibility. In this web of concepts, I have identified moral responsibility as the key element that holds the edifice together and founds retributive justice so that a reworking of that concept can be the base for rethinking justice.

For this purpose, I have introduced Iris Marion Young's idea of structural injustice and her social-connection model of responsibility, establishing a connection with the conception of shared responsibility previously discussed. The social connection model calls for agents to actively take responsibility for structural injustices, presenting a forward-looking posture opposite to traditional liability-based responsibility that looks backwards at what an actor has already done in order to decide the appropriate response to their past action. Within this new framework, agents take responsibility for justice, that is, for creating a just social structure, without thus generating praise or blame for succeeding or failing in the task. The social connection model severs the concept of responsibility from that of desert, differing qualitatively rather than quantitively from the liability model, requiring to let go of this desert-based posture, only to focus on achieving the desired outcome of creating a just societal structure.

Young's perspective is purely economic; my aim has been to widen it in order for this conception of justice to encompass the area traditionally covered by penal justice. I have found her framework particularly suitable for this purpose because of the connection she establishes between social justice and her forward-looking conception of responsibility. Social justice is traditionally considered unrelated to responsibility, contrary to criminal justice, which is grounded in responsibility. Young's framework, already linking social justice with responsibility, allows for a very smooth transition from social justice – understood in economic terms – to criminal justice and its renewed

conceptualisation. The social-connection model of responsibility proves fruitful in dealing with harm and crime as well as with poverty.

Although I have said that my critique of retributive justice is not based on the rejection of free will, for even if offenders were perfectly free and morally responsible, with total control over their volitions and actions, retribution would still be moral nonsense, I have argued in favour of a situationist approach as an underlying basis for the social-connection model of responsibility. From this perspective, to take responsibility for justice does not only mean not to create the conditions for poverty and to help those afflicted by it to improve their financial situation, but also not to create the conditions for crime and to help those who do commit it restore the broken bonds with their victims and the whole of the society. If we consider individuals to be entangled even more than connected, then we can – at least partially – take this kind of responsibility and rethink our dealings with crime through this lens instead of individualising guilt and addressing the problem of crime acting on individual persons.

I inserted my proposal for the overcoming of traditional criminal justice based on punishment within a utilitarian framework. The coupling of utilitarianism and punishment abolition – especially in the form of prison abolition – might be surprising because utilitarianism was born in the cradle of the criminal reform of the 18th century that brought to the establishment of the prison as the universal punishment. Moreover, utilitarianism is typically used as a framework for punishment justification rather than for its abolition. However, I have argued that, once this bond between utilitarianism and the prison is contextualised, this idea will seem less bizarre: the connection between the two, I have claimed, is, in fact, of extrinsic nature, as there is nothing inherently punitive in utilitarianism. On the contrary, the role of utilitarianism within the penal reform was to push in the direction of the softening of punishments, as it must remembered that, before the universalisation of incarceration, punishment was administered through torture.

Concerning the traditional utilitarian justifications for punishment – deterrence, incapacitation, and rehabilitation – I have claimed that for them to be *real* utilitarian justifications, their effectiveness must be assessed on empirical grounds. In the previous chapter, I have already shown that empirical proof is lacking. I have argued that the lack of data in support of consequential justifications points in the direction of a misuse of utilitarianism in the discussion over punishment: not a clear-headed assessment of the most efficient ways of dealing with crime, but an ex-post rationalisation of a practice

that was already in place – that of punishment – and that was not questioned in its foundations.

However, I have shown that the seed of punishment abolition was already present in Bentham's work and that it has been neglected by subsequent utilitarians. He, in fact, believed that although punishment is the easiest solution to the problem of crime, in that it requires the least effort and is the most intuitive, it is already the mark of the failure of the system, for society should primarily aim at crime prevention through education, knowledge sharing, and poverty prevention and alleviation. I have argued that he anticipated many of the arguments purported by present-day prison abolition thinkers, but utilitarianism – probably because of its historical connection with punishment justification – was ignored by the prison abolition movement, which has mostly based its criticism of the prison system on its initial ties with slavery and racism in the U.S. However, although these ties are real and problematic, I believe that prison abolition should rather be grounded on a more general understanding of the moral inadequacy of the punitive framework: prison is wrong in itself and not only insofar as it is a racist institution.

I have shown how hedonistic utilitarianism could grant prison abolition this broader scope, as it is based on the broadly inclusive feature of sentience and, although having historically framed punishment as a necessary evil, requires us to question whether this evil is really necessary. This questioning, however, requires a revolutionary attitude that utilitarians have been afraid to assume, as they have been focussing on the continuity between utilitarianism and common sense. Doing so has been valuable to avoid the accusations of deception and manipulation derived from the fact that some utilitarians (from Sidgwick to Singer) have supported esoteric morality. However, I have argued that utilitarianism *is* intrinsically revolutionary in nature and caution when advocating for change on utilitarian grounds must not result in complete abdication if one does not want to betray utilitarianism fundamentally.

I have then argued for a utilitarianism-framed prison abolition through societal transformation with the aim of creating a society that does not need prison, as was partly advocated for by Bentham himself. Hence, social justice, taking over the space traditionally occupied by punitive justice, has taken the form of transformative justice. As for dealing with the remaining offenders that are not prevented by such societal transformation, I have claimed that the preferred approach would be that of restorative

justice, which should be integrated with, rather than seen as an alternative to, transformative justice.

Restorative justice consists of a discursive process in which victims and offenders seek reconciliation, overturning the traditional system that excludes the victim from the trial and focusses on the offender as its object. Here, both victims and offenders are subjects in a dialogue founded on egalitarian grounds: their experiences, motivations, and struggles are addressed. And just like in the social-connection model of responsibility, the community takes responsibility for the victim's healing, so it must do for the offender's rather than lock them up behind prison walls.

Of course, the restorative process is fragile, and it may fail. I have argued that, contrary to retributive punishment, where the infliction of suffering is in itself the successful response to crime – hence just requiring the correct individuation and prosecution of the offender – a discursive process involving regret and reconciliation is at a higher risk of failure. However, this does not count as a reason in support of retributive punishment, for, although it may be easier to achieve, we have seen that it really brings nothing good. As for supporting punishment for consequential reasons, in that it should grant a higher degree of safety from offenders compared to restorative processes, we have already seen that this promise of safety is an illusion and that the current system is not granting it, nor will it ever do. The promise of perfect safety is always bound to fail as fragility is an intrinsic part of life: we should aim to reduce risk, not eradicate it. Restorative justice, within the broader context of transformative justice, is reasonably expected to give better results, from a utilitarian perspective, than the criminogenic and violent system that is, in and of itself, generating an immense amount of suffering. However, its actual effectiveness may only be assessed when such a system will be put in place.

In essence, the critique of retributive justice and individual moral responsibility developed in this thesis is framed within two conceptual frameworks: on the descriptive level, concerning the mechanisms leading to reciprocal violence and the means that have been used to tackle the problem of violence, that of the Girardian theory of mimetic desire, victimisation, and sacrifice; on the normative level, concerning what, given this descriptive picture, ought to be done, that of hedonistic utilitarianism, firmly rooted in Bentham's works. The first has been introduced in the first pages of this work and has been especially relevant in my historical analysis; however, it has been carried out, silently at times, for the whole length of the work. The latter has been introduced as

an alternative normative framework to ethical intuitionism, compared to which it retained the realism but founded it on the shared experiences of sentient beings rather than inscrutable and unquestionable intuitions. Within these two frameworks, many topics revolving around the concept of justice have been analysed: from otherworldly retribution to the State's fundamental functions, from meritocracy to moral sentiments, from the long-established justifications for punishment to transformative justice and prison abolition.

My aim has been to show the fundamental interconnectedness of these elements that revolve around the concept of justice and to elaborate an original proposal that would be able to take them all into account. Obviously, a work that touches on all these subjects cannot aspire to exhaustiveness in each of them. I hope I have shown that the drawing of a larger, more comprehensive picture is essential for a better understanding of a subject as broad and pervasive as that of justice; however, I recognise that an enormous amount of work could (and should) be dedicated to each of them. The section on meritocracy is especially lacking, as only an entire work of comparable length could aim to explore it comprehensively. Moreover, conceptions of justice in non-Western religions should really be thoroughly investigated, as should ancient and contemporary restorative justice practices in non-Western societies. Finally, much could be written on theories of justice and responsibility from the 18th century to the present day, especially among socialist and anarchist thinkers. I had to limit my historical enquiry due to time and space constraints, as I did not aim to produce a text of an essentially historical character; however, I believe such an enquiry would invaluably enrich a research of this kind. I hope this work can provide a solid basis for filling these gaps.

Bibliography

Abelard, Peter. *Ethics*. Edited and translated by D. E. Luscombe. Oxford: Oxford University Press, 1971.

Abelson, Robert P., Kurt P. Frey, and Aiden P. Gregg. *Experiments with People. Revelations from Social Psychology*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2004.

Alexander, Michelle. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. New York: The New Press, 2010.

Allen, Mark W. "Hunter-Gatherer Conflict: The Last Bastion of the Pacified Past?" In *Violence and Warfare Among Hunter-Gartherers*, edited by Mark W. Allen and Terry L. Jones, 15–25. Walnut Creek: Left Coast Press, 2014.

Anderson, Elizabeth S. "Against Luck Egalitarianism: What Is the Point of Equality?" In *Social Justice*, edited by Matthew Clayton and Andrew Williams, 155–85. Malden, MA: Blackwell Publishing, 2004.

Anspach, Mark Rogin. *À Charge de Revanche. Figures Élémentaires de La Réciprocité*. Paris: Seuil, 2002.

———. "Apprivoiser La Violence. Le Sacrifice, La Vengeance, Le Don." *Revue Du MAUSS* 55, no. 1 (2020): 129–41.

———. "Penser La Vengeance." *Esprit* 128, no. 7 (1987): 103–11.

———. "Vengeance and the Gift." In *Palgrave Handbook of Mimetic Theory and Religion*, edited by James Alison and Wolfgang Palaver, 53–59. New York: Palgrave Macmillan, 2017.

Antigone, Associazione. "Il Carcere Visto Da Dentro. XVIII Rapporto Di Antigone Sulle Condizioni Di Detenzione. Numeri Del Carcere," 2022. https://www.rapportoantigone.it/diciottesimo-rapporto-sulle-condizioni-di-detenzione/numeri-del-carcere/.

"Apartheid Enforcer Sticks to 'farcical' Story on Biko Killing." *The Independent*, 1997.

Arendt, Hannah. "Collective Responsibility." In *Amor Mundi: Explorations in the Faith and Thought of Hannah Arendt*, edited by James W: Bernauer, 43–50. Boston: Martinus Nijhoff Publishers, 1987.

Aristotle. *Poetics*. Edited and translated by Malcolm Heath. London: Penguin Books, 1996.

Association, American Psychiatric. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC, 2013.

Augustine. *Confessions*. Edited and translated by Thomas Williams. Indianapolis & Cambridge: Hackett Publishing Company, 2019.

———. "On the Free Choice of the Will." In *On the Free Choice of the Will, On Grace and Free Choice, and Other Writings*, edited and translated by Peter King, 3–126. Cambridge: Cambridge University Press, 2010.

———. "Quaestiones VII Ad Simplicianum." In *Mélanges Doctrinaux; Quaestiones 83; Quaestiones VII Ad Simplicianum; Quaestiones VIII Dulcitii; De Divinatione Daemonum*, edited by Gustave Bardy, Jérôme André Beckaert, and Joseph Boutet. Paris: Desclée De Brouwer, 1952.

Ayer, Alfred. *The Problem of Knowledge*. London: MacMillan, 1956.

Baron-Cohen, Simon. *The Science of Evil. On Empathy and the Origins of Cruelty*. New York: Basic Books, 2011.

Barrett, Lee C. "Kierkegaard's Appropriation and Critique of Luther and Lutheranism." In *A Companion to Kierkegaard*, edited by Jon Stewart, 181–92. Hoboken, New Jersey: Wiley-Blackwell, 2015.

Barrow, Greg. "South Africans Reconciled?" *BBC News*, 1998. http://news.bbc.co.uk/2/hi/special_report/1998/10/98/truth_and_reconciliation/142673.stm.

Barth, Karl. *A Shorter Commentary on Romans*. Edited by Maico M. Michielin. Translated by D.H. van Daalen. Aldershot: Ashgate, 2007.

Basaglia, Franco, and Franca Basaglia Ongaro. "Premessa." In *Crimini Di Pace. Ricerche Sugli Intellettuali e Sui Tecnici Come Addetti All'oppressione*, edited by Franco Basaglia and Franca Basaglia Ongaro. Torino: Einaudi, 1975.

Beardsley, Elizabeth Lane. "Determinism and Moral Perspectives." *Philosophy and Phenomenological Research* 21, no. 1 (1960): 1–20.

Beccaria, Cesare. *On Crimes and Punishments and Other Writings*. Edited by Aaron Thomas. Translated by Aaron Thomas and Jeremy Parzen. Toronto, Buffalo, London: University of Toronto Press, 2008.

Bell, Marina. "Abolition: A New Paradigm for Reform." *Law & Social Inquiry* 46, no. 1 (February 26, 2021): 32–68.

Benedetti, Marina. "Forme Inquisitoriali Di Polizia Nel Medioevo." In *Le Polizie Informali*, edited by Livio Antonielli, 21–44. Messina: Rubbettino, 2003.

Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. Edited by J. H. Burns and H. L. A. Hart. Oxford: Oxford University Press, 2015.

———. "Principles of Penal Law." In *Works of Jeremy Bentham*, edited by John Bowring, 365–358. New York: Russell & Russell, 1962.

———. *The Panopticon Writings*. Edited by Miran Bozovic. London and New York: Verso, 1995.

Bentham, Jeremy, and Louis Crompton. "Offences Against One's Self." *Journal of Homosexuality* 3, no. 4 (1978): 389–406.

Benveniste, Émile. *Dictionary of Indo-European Concepts and Society*. Translated by Elizabeth Palmer. Chicago: HAU Books, 2016.

Boétie, Étienne de La. *The Politics of Obedience: The Discourse of Voluntary Servitude*. Edited by Murray N. Rothbard. Translated by Harry Kurz. Auburn: The Mises Institute, 2010.

Bok, Hilary. *Freedom and Responsibility*. Princeton: Princeton University Press, 1998.

Braithwaite, John. "Restorative Justice and De-Professionalization." *The Good Society* 13, no. 1 (2004): 28–31.

Brody, David C., James R. Acker, and Wayne A. Logan, eds. "The Queen v. Dudlet and Stephens, 14 Q.B.D. 273 (1884)." In *Criminal Law*, 19–23. Gaithersbourg, Maryland: Aspen Publishers Inc., 2001.

Bronson, Jennifer, and Marcus Berzofsky. "Indicators of Mental Health Problems Reported by Prisoners and Jail Inmates, 2011-12." *Us Department of Justice, Office of Justice Programs, Bureau of Justice Statistics*, 2017.

Brown, Gregory. "Happiness and Justice." In *The Oxford Handbook of Leibniz*, edited by Maria Rosa Antognazza, 622–40. Oxford: Oxford University Press, 2018.

Brown, Michelle, and Judah Schept. "New Abolition, Criminology and a Critical Carceral Studies." *Punishment and Society* 19, no. 4 (2017): 440–62.

Burkert, Walter. *Homo Necans. The Anthropology of Ancient Greek Sacrificial Ritual and Myth*. Translated by Peter Bing. Berkeley and Los Angeles: University of California Press, 1983.

Burns, James H., and Herbert L. A. Hart. "Introduction." In *An Introduction to the Principles of Morals and Legislation*, edited by H. L. A. Hart and J. H. Burns, xxxi–lxix. Oxford: Oxford University Press, 2015.

Buzon, Frédéric de, and Denis Kambouchner. "Descartes." In *Le Vocabulaire Des Philosophes. Tome 2: La Philosophie Classique et Moderne (XVIIe-XVIIIe Siècle)*,

edited by Jean-Pierre Zarader, 13–76. Paris: Ellipses, 2016.

Calvin, John. *Institutes of the Christian Religion. Volume 1*. Edited by John T. McNeill. Translated by Ford Lewis Battles. Louisville: Westminster John Knox Press, 2006.

———. *Institutes of the Christian Religion. Volume 2*. Edited by John T. McNeill. Translated by Ford Lewis Battles. Louisville: Westminster John Knox Press, 2006.

Camus, Albert. "Reflections on the Guillotine." In *Resistance, Rebellion, and Death*, translated by Justin O'Brien, 173–234. New York: Knopf, 1961.

Chisholm, Roderick. "Human Freedom and the Self." In *Free Will*, edited by Gary Watson, 24–35. Oxford: Oxford Readings in Philosophy, 1983.

Christie, Nils. "Conflicts as Property." *The British Journal of Criminology* 17, no. 1 (1977): 1–15.

Clark, John. "Against the Corporal Punishment of Children." *Cambridge Journal of Education* 34, no. 3 (2004): 363–71.

Clarke, Randolph. "Towar A Credible Agent-Causal Account of Free Will\." *Noûs* 27, no. 2 (1993): 191–203.

Clarke, Samuel. "A Demonstration of the Being and Attributes of God." In *A Demonstration of the Being and Attributes of God and Other Writings*, edited by Ezio Vailati, 1–92. Cambridge: Cambridge University Press, 2010.

Cosmides, Leda, and John Tooby. "Evolutionary Psychology and the Generation of Culture, Part II. Case Study: A Computational Theory of Social Exchange." *Ethology and Sociobiology* 10 (1989): 51–98.

Courtois, Gérard. "Introduction : La Vengeance, Du Désir Aux Institutions." In *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume IV. La Vengeance Dans La Pensée Occidentale*, edited by Gérard Courois, 7–45. Paris: Éditions Cujas, 1984.

———, ed. *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume IV. La Vengeance Dans La Pensée Occidentale*. Paris: Éditions Cujas, 1984.

———. "Le Sens et La Valeur de La Vengeance, Chez Aristote et Sénèque." In *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume IV. La Vengeance Dans La Pensée Occidentale*, edited by Gérard Courtois, 91–124. Paris: Éditions Cujas, 1984.

Curley, Edwin M. "Spinoza's Moral Philosophy." In *Spinoza. A Collection of Critical Essays*, edited by Marjorie Grene, 354–76. Garden City: Anchor Press/Doubleday, 1973.

Darwin, Charles. *The Expression of the Emotions in Man and Animals*. Edited by Francis Darwin. Cambridge: Cambridge University Press, 2009.

Davis, Angela Y. *Are Prisons Obsolete?* New York: Seven Stories Press, 2003.

Dawkins, Richard. *The Selfish Gene*. 30th anniv. Oxford: Oxford University Press, 2006.

Deffenbach, Jerry L., Eugene R. Oetting, Gregory A. Thwaites, Rebekahh S. Lynch, Deborah A. Baker, Robert S. Stark, Stacy Thacker, and Lora Eisewerth-Cox. "State-Trait Anger Theory and the Utility of the Trait Anger Scale." *Journal of Counseling Psychology* 43, no. 2 (1996): 131–48.

Descartes, René. "Descartes à Elisabeth. [Egmond, Janvier 1646]." In *Œuvre de Descartes. Correspondance IV. Juillet 1643 - Avril 1647*, edited by Charles Adam and Paul Tannery, 351–57. Paris: Librairie Philosophique J. Vrin, 1996.

———. *Meditations on First Philosophy. With Selections from the Objections and Replies*. Edited and translated by John Cottingham. Cambridge: Cambridge University Press, 2013.

———. *Principles of Philosophy*. *The Cambridge Descartes Lexicon*. Dordrecht/Boston/London: Kluwer Academic Publishers, 1982.

———. *The Passions of the Soul and Other Late Philosophical Writings*. Edited and translated by Michael Moriarty. Oxford: Oxford University Press, 2015.

———. "To Chanut, 15 June 1646." In *The Philosophical Writings of Descartes. Volume III. The Correspondence*, edited by John Cottingham, translated by John Cottingham, Robert Stoothoff, Dugald Murdoch, and Anthony Kenny, 289. Cambridge: Cambridge University Press, 1991.

———. "To Chanut, 20 November 1647." In *The Philosophical Writings of Descartes. Volume III. The Correspondence*, edited by John Cottingham, translated by John Cottingham, Robert Stoothoff, Dugald Murdoch, and Anthony Kenny, 326–27. Cambridge: Cambridge University Press, 1991.

———. "To Mersenne, 27 February 1637." In *The Philosophical Writings of Descartes. Volume III. The Correspondence*, edited by John Cottingham, translated by John Cottingham, Robert Stoothoff, Dugald Murdoch, and Anthony Kenny, 52–54. Cambridge: Cambridge University Press, 1991.

———. "To Mesland On Freedom (February 9, 1645)." In *Philosophical Essays and Correspondence*, edited by Roger Ariew, 273–74. Cambridge/Indianapolis: Hackett Publishing Company, 2000.

Dickens, Charles. *American Notes for General Circulation. Vol. 1*. New York: Cambridge University Press, 2009.

Dölling, Dieter, Horst Entorf, Dieter Hermann, and Rupp. Thomas. "Is Deterrence Effective? Results of a Meta-Analysis of Punishment." *European Journal on Criminal Policy and Research* 15, no. 1 (2009): 201–24.

Dreisinger, Baz. *Incarceration Nations: A Journey to Justice in Prisons Around the World*. New York: Other Press, 2016.

Duff, R. Antony. "Responsibility, Restoration, and Retribution." In *Retributivism Has a Past: Has It a Future?*, edited by Michael Tonry, 64–85. Oxford: Oxford University Press, 2011.

———. "Restoration and Retribution." In *Restorative Justice and Criminal Justice: Competing of Reconciliable Paradigms?*, edited by Andrew Von Hirsch, Julian V. Roberts, Anthony Bottoms, Kent Roach, and Mara Schiff, 43–59. Oxford and Portland: Hart Publishing, 2003.

Duff, R. Antony, and David Garland. "Introduction." In *A Reader on Punishment*, 1–43. Oxford: Oxford University Press, 1994.

Dumouchel, Paul. *The Barren Sacrifice. An Essay on Political Violence*. East Lansing: Michigan State University Press, 2015.

———. "The Barren Sacrifice." In *Palgrave Handbook of Mimetic Theory and Religion*, edited by James Alison and Wolfgang Palaver, 279–85. New York: Palgrave Macmillan, 2017.

Dupuy, Jean-Pierre. "Secularization." In *Palgrave Handbook of Mimetic Theory and Religion*, edited by James Alison and Wolfgang Palaver, 271–77. New York: Palgrave Macmillan, 2017.

———. "Self-Transcendence and Tangled Hierarchies in Çatalhöyük." In *How We Became Human. Mimetic Theory and the Science of Evolutionary Origins*, edited by Pierpaolo Antonello and Paul Gifford, 233–59. East Lansing: Michigan State University Press, 2015.

Durkheim, Emile. *The Division of Labour in Society*. Translated by W. D. Halls. London: The Macmillan Press, 1984.

———. *The Elementary Forms of Religious Life*. Translated by Karen E. Fields. New York: The Free Press, 1995.

Dworkin, Ronald. "Equality, Luck and Hierarchy." *Philosophy and Public Affairs* 31, no. 2 (2003): 190–98.

———. “Equality of Resources.” In *Social Justice*, edited by Matthew Clayton and Andrew Williams, 110–34. Malden, MA: Blackwell Publishing, 2004.

Eamon, Mary Keegan. “Antecedents and Socioemotional Consequences of Physical Punishment on Children in Two-Parent Families.” *Child Abuse and Neglect* 6 (2001): 787–802.

Ehrman, Bart D. *Heaven and Hell. A History of the Afterlife*. New York: Simon & Schuster, 2020.

Ekman, Paul. “Darwin’s Contributions to Our Understanding of Emotional Expressions.” *Philosophical Transactions of the Royal Society* 364 (2009): 3449–51.

Embry, Brian. “Descartes on Free Will and Moral Possibility.” *Philosophy and Phenomenological Research* 96, no. 2 (2018): 380–98.

Erasmus, Desiderius. “A Diatribe or Sermon Concerning Free Will.” In *Discourse on Free Will*, edited by Ernst F. Winter, 3–94. New York: Continuum, 2002.

Fazel, Seena, and John Danesh. “Serious Mental Disorder in 23 000 Prisoners: A Systematic Review of 62 Surveys.” *The Lancet* 359 (2002): 345–50.

Fischer, John Martin. *Deep Control. Essays on Free Will and Value*. Oxford: Oxford University Press, 2012.

———. *My Way : Essays on Moral Responsibility: Essays on Moral Responsibility*. Oxford: Oxford University Press, 2006.

———. “Responsiveness and Moral Responsibility.” In *Responsibility, Character, and the Emotions*, edited by Ferdinand Schoeman, 81–106. Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987.

Foucault, Michel. *Discipline and Punish. The Birth of the Prison*. Translated by Alan Sheridan. New York: Vintage Books, 1995.

Fovet, Thomas, Laurent Plancke, Alina Amariei, Imane Benradia, Fanny Carton, Aminata Sy, Maeva Kyheng, et al. “Mental Disorders on Admission to Jail : A Study of Prevalence and a Comparison with a Community Sample in the North of France.” *European Psychiatry* 63, no. 1 (2020): 1–7.

Frankfurt, Harry. “Alternate Possibilities and Moral Responsibility.” *The Journal of Philosophy* 66, no. 23 (1969): 829–39.

———. “Freedom of the Will and the Concept of a Person.” *The Journal of Philosophy* 68, no. 1 (1971): 5–20.

———. “Identification and Wholeheartedness.” In *Responsibility, Character, and the*

Emotions, edited by Ferdinand Schoeman, 27–45. Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987.

Freud, Sigmund. *Totem and Taboo. Some Points of Agreement between the Mental Lives of Savages and Neurotics*. Translated by James Strachey. London: Routledge, 2011.

Fuller Torrey, Edwin. *Nowhere to Go: The Tragic Odyssey of the Homeless Mentally Ill*. New York: Harper & Row, 1988.

García-Sánchez, Efraín, Isabel Correia, Cícero R. Pereira, Guillermo B. Willis, Rosa Rodríguez-Bailón, and Jorge Vala. "How Fair Is Economic Inequality? Belief in a Just World and the Legitimation of Economic Disparities in 27 European Countries." *Personality and Social Psychology Bulletin* 48, no. 3 (March 1, 2022): 382–95.

Gassendi, Pierre. "Syntagma Philosophicum." In *Petri Gassendi Diniensis Ecclesiae Praepositi, et in Academia Parisiensi Matheseos Regii Professoris Opera Omnia*, edited by Friedrich Frommann. Stuttgart-Bad Canstatt, 1964.

Gatens, Moira. "The Condition of Human Nature. Spinoza's Account of the Ground of Human Action in the Tractatus Politicus." In *Spinoza's Political Treatise. A Critical Guide*, edited by Yitzhak Y. Melamed and Hasana Sharp, 47–60. Cambridge: Cambridge University Press, 2018.

Gershoff, Elizabeth T. "More Harm than Good: A Summary of Scientific Research on the Intended and Unintended Effects of Corporal Punishment on Children." *Law & Contemporary Problems* 73, no. 2 (2021): 31–56.

Giancotti, Emilia. "Introduzione." In *Trattato Teologico-Politico*, XI–XXXVII. Torino: Einaudi, 2007.

Gifford, Paul, and Pierpaolo Antonello. "Rethinking the Neolithic Revolution: Symbolism and Sacrifice at Göbekli Tepe." In *How We Became Human. Mimetic Theory and the Science of Evolutionary Origins*, edited by Paul Gifford and Pierpaolo Antonello, 261–88. East Lansing: Michigan State University Press, 2015.

Girard, René. "Animal Scapegoating at Çatalhöyük." In *How We Became Human. Mimetic Theory and the Science of Evolutionary Origins*, edited by Pierpaolo Antonello and Paul Gifford, 217–31. East Lansing: Michigan State University Press, 2015.

———. *Battling to the End. Conversations with Benoît Chantre*. Translated by Mary

Baker. East Lansing: Michigan State University Press, 2010.

———. *Deceit, Desire & the Novel. Self and Other in Literary Structure*. Translated by Yvonne Freccero. Baltimore: The Johns Hopkins University Press, 1965.

———. *I See Satan Fall Like Lightning*. Translated by James G. Williams. Maryknoll, New York: Orbis Books, 2001.

———. *Je Vois Satan Tomber Comme l'éclair*. Paris: Grasset & Fasquelle, 1999.

———. *Job: The Victim of His People*. Translated by Yvonne Freccero. Stanford: Stanford University Press, 1987.

———. *La Violence et Le Sacré*. Paris: Bernard Grasset, 1972.

———. "Peter's Denial and the Question of Mimesis." *Notre Dame English Journal* 14, no. 3 (1982): 177–89.

———. *The Scapegoat*. Translated by Yvonne Freccero. Baltimore: The Johns Hopkins University Press, 1986.

———. *Things Hidden Since the Foudation of the World*. Translated by Stephen Bann and Michael Metteer. Stanford: Stanford University Press, 1987.

———. *Violence and the Sacred*. Translated by Patrick Gregory. Baltimore and London: The Johns Hopkins University Press, 1977.

Girard, René, Pierpaolo Antonello, and João Cezar de Castro Rocha. *Evolution and Conversion. Dialogues on the Origins of Culture*. London and New York: Continuum, 2008.

Goff, Jaques Le. *The Birth of Purgatory*. Aldershot: Scolar Press, 1990.

Goffman, Erving. *Asylums. Essays on the Social Situation of Mental Patients and Other Inmates*. Garden City: Anchor Books, 1961.

Guyau, Jean-Marie. *A Sketch of Morality Independent of Obligation and Sanction*. London: Watts & Co., 1891.

Hammond, Michael. "Evolutionary Theory and Emotions." In *Handbook of the Sociology of Emotions*, edited by Jan E. Stets and Jonathan H. Turner, 368–85. New York: Springer, 2006.

Harper, David J., Gaham F. Wagstaff, J. Tim Newton, and Kevin R. Harrison. "Lay Causal Perceptions of Third World Poverty and the Just World Theory." *Social Behavior and Personality* 18, no. 2 (1990): 235–38.

Hart, Herbert L. A. *Punishment and Responsibility. Essays in the Philosophy of Law*. New York: Oxford University Press, 2008.

Heider, Fritz. *The Psychology of Interpersonal Relations*. New York: John Wiley &

Sons, 1958.

Hirsch, Andrew Von. *Past or Future Crimes. Deservedness and Dangerousness in the Sentencing of Criminals*. Manchester: Manchester University Press, 1985.

Hitchens, Christopher. "Mommie Dearest. The Pope Beatifies Mother Teresa, a Fanatic, a Fundamentalist, and a Fraud." Slate, 2003. https://slate.com/news-and-politics/2003/10/the-fanatic-fraudulent-mother-teresa.html.

Hobbes, Thomas. "De Cive." In *Three-Text Edition of Thomas Hobbes's Political Theory. The Elements of Law, De Cive and Leviathan*, edited by Deborah Baumgold, 3–548. Cambridge: Cambridge University Press, 2017.

———. "Leviathan." In *Three-Text Edition of Thomas Hobbes's Political Theory. The Elements of Law, De Cive and Leviathan*, edited by Deborah Baumgold, 3–548. Cambridge: Cambridge University Press, 2017.

———. "The Elements of Law." In *Three-Text Edition of Thomas Hobbes's Political Theory. The Elements of Law, De Cive and Leviathan*, edited by Deborah Baumgold, 3–548. Cambridge: Cambridge University Press, 2017.

Hobbes, Thomas, and John Bramhall. *Hobbes and Bramhall on Liberty and Necessity*. Edited by Vere Chappell. Cambridge: Cambridge University Press, 1999.

Honderich, Ted. *Punishment. The Supposed Justifications Revisited*. London: Pluto Press, 2006.

Hudson, Barbara. "Victims and Offenders." In *Restorative Justice and Criminal Justice: Competing of Reconciliable Paradigms?*, edited by Andrew Von Hirsch, Julian V. Roberts, Anthony Bottoms, Kent Roach, and Mara Schiff, 177–94. Oxford and Portland: Hart Publishing, 2003.

Hudson, John, and Mark D. West. "Stringer's Saga: Njal and The Wire." In *Emotion, Violence, Vengeance and Law in the Middle Ages. Essays in Honour of William Ian Miller*, edited by Kate Gilbert and Stephen D. White, 271–95. Leiden and Boston: Brill, 2018.

Hudson, W. D. *Ethical Intuitionism*. London, Melbourne, Toronto: MacMillan, 1967.

Huemer, Michael. "Revisionary Intuitionism." *Social Philosophy and Policy* 25, no. 1 (January 2008): 368–92.

Hume, David. "Of the Original Contract." In *Political Essays*, edited by Knud Haakonssen, 186–201. Cambridge: Cambridge University Press, 1994.

Hutcheson, Francis. *An Essay on the Nature and Conduct of the Passions and Affections, with Illustrations on the Moral Sense*. Edited by Aaron Garrett.

Indianapolis: Liberty Fund, 2002.

Izard, Carroll E., and Sandra Buechler. “Aspect of Consciousness and Personality in Terms of Differential Emotions Theory.” In *Emotion. Theory, Research, and Experience. Volume 1: Theories of Emotion1*, edited by Robert Plutchik and Henry Kellerman, 165–87. New York: Academic Press, 1980.

Jacoby, Susan. *Wild Justice. The Evolution of Revenge*. New York: Harper & Row, 1983.

James, Susan. “Politically Mediated Affects. Envy in Spinoza’s Tractatus Politicus.” In *Spinoza’s Political Treatise. A Critical Guide*, edited by Yitzhak Y. Melamed and Hasana Sharp, 61–77. Cambridge: Cambridge University Press, 2021.

Janoff-Bulman, Ronnie, and Camille B. Wortman. “Attributions of Blame and Coping in the ‘Real World’: Severe Accident Victims React to Their Lot.” *Journal of Personality and Social Psychology* 35, no. 5 (1977): 351–63.

Jaquet, Chantal. “Longing (Desiderium) for Vengeance as the Foundation of the Commonwealth.” In *Spinoza’s Political Treatise. A Critical Guide*, edited by Yitzhak Y. Melamed and Hasana Sharp, 78–92. Cambridge: Cambridge University Press, 2018.

Jefferson, Thomas. “Jefferson to Adams. Monticello, October 28 1813.” In *The Adams-Jefferson Letters. The Complete Correspondence between Thomas Jefferson and Abigail and John Adams*, edited by Lester J. Cappon, 387–92. Chapel Hill and London: The University of North Carolina Press, 1987.

Jones, Terry L., and Mark W. Allen. “The Prehistory of Violence and Warfare among Hunter-Gatherers.” In *Violence and Warfare Among Hunter-Gartherers*, edited by Mark W. Allen and Terry L. Jones, 353–71. Walnut Creek: Left Coast Press, 2014.

Kant, Immanuel. *The Philosophy of Law. An Exposition of the Fundamental Priciples of Jurisprudence Ad the Science of Right*. Translated by W. Hastie. Edimburgh: T. & T. Clark, 1887.

Keeley, Lawrence H. *War before Civilization: The Myth of the Peaceful Savage*. Oxford: Oxford University Press, 1996.

Kierkegaard, Søren. “Fear and Trembling. Dialectical Lyric by Johannes de Silentio.” In *Fear and Trembling. Repetition.*, edited and translated by Howard V. Hong and Edna H. Hong, 1–123. Princeton: Princeton University Press, 1983.

Knight, Carl. *Luck Egalitatianism. Equality, Responsibility, and Justice*. Edimburgh: Edimburgh University Press, 2009.

Kremer, Elmer J. “Malebranche on Human Freedom.” In *The Cambridge Companion to Malebranche*, edited by Steven Nadler, 190–219. Cambridge: Cambridge University Press, 2000.

Kropotkin, Peter. *La Morale Anarchiste*. Edited by Jérôme Solal. Paris: Mille et une Nuits, 2024.

Law, Victoria. *“Prisons Make Us Safer” And 20 Other Myths About Mass Incarceration*. Boston: Beacon Press, 2021.

Lazari-Radek, Katarzyna de, and Peter Singer. “Secrecy in Consequentialism: A Defence of Esoteric Morality.” *Ratio* 23, no. 1 (2010): 34–58.

LeBlanc, Steven. “Forager Warfare and Our Evolutionary Past.” In *Violence and Warfare Among Hunter-Gartherers*, edited by Mark W. Allen and Terry L. Jones, 26–46. Walnut Creek: Left Coast Press, 2014.

———. “The Origins of Warfare and Violence.” In *The Cambridge World History of Violence. Volume 1: The Prehistoric and Ancient Worlds*, edited by Garrett G. Fagan, Linda Fibiger, Mark Hudson, and Matthew Trundle, 39–57. Cambridge: Cambridge University Press, 2020.

Leibniz, Gottfried Wilhelm. “De Justitia.” In *Mittheilungen Aus Leibnizens Ungerdruckten Schriften*, edited by Georg Mollat, 35–40. Leipzig: H. Haessel, 1893.

———. “De Summa Juris Regula.” In *Mittheilungen Aus Leibnizens Ungerdruckten Schriften*, edited by Georg Mollat, 85–87. Leipzig: H. Haessel, 1893.

———. “De Tribus Jusris Naturae et Gentium Gradibus.” In *Mittheilungen Aus Leibnizens Ungerdruckten Schriften*, edited by Georg Mollat, 8–18. Leipzig: H. Haessel, 1893.

———. “Definitio Iustitae Universalis.” In *Historische Beiträge Zur Philosophie. Vol. II*, edited by Adolf Trendelenburg, 265–75. Berlin: G. Bethge, 1855.

———. “Dialogue on Human Freedom and the Origin of Evil (1695).” In *Philosophical Essays*, edited by Roger Ariew and Daniel Garber, 111–16. Indianapolis & Cambridge: Hackett Publishing Company, 1989.

———. *Essais De Théodicée, Sur La Bonté De Dieu, La Liberté De L’Homme, Et L’Origine Du Mal*. Amsterdam: Mortier, 1714.

———. “Juris et Aequi Elementa.” In *Mittheilungen Aus Leibnizens Ungerdruckten Schriften*, edited by Georg Mollat, 19–34. Leipzig: H. Haessel, 1893.

———. “Letter to Magnus Wedderkopf, 1671.” In *Confessio Philosophi: Papers*

Concerning the Problem of Evil, 1671-1678, edited by Robert C. Sleigh, 2–5. New Haven and London: Yale University Press, 2006.

———. "Meditation on the Common Concept of Justice (c. 1702-3)." In *Political Writings*, edited by Patrick Riley, 45–63. Cambridge: Cambridge University Press, 1988.

———. "On Contingency (1686?)." In *Philosophical Essays*, edited by Roger Ariew and Daniel Garber, 28–29. Indianapolis & Cambridge: Hackett Publishing Company, 1989.

———. "On Free Choice (Between Summer 1678 and Winter 1680–81?)." In *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, edited by Robert C. Sleigh, 132–37. New Haven and London: Yale University Press, 2006.

———. "On Freedom, ca. 1679." In *Philosophical Papers and Letters*, edited by Leroy E. Loemker, 263–66. Dordrecht/Boston/London: Kluwer Academic Publishers, 1989.

———. *On Freedom, Fate, and the Grace of God (1686)*. Edited by Shane Duarte, n.d.

———. "On Freedom and Possibility (1680-82?)." In *Philosophical Essays*, edited by Roger Ariew and Daniel Garber, 19–22. Indianapolis & Cambridge: Hackett Publishing Company, 1989.

———. "On the Necessity of Choosing the Best (1677?)." In *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, edited by Robert C. Sleigh, 138–39. New Haven and London: Yale University Press, 2006.

———. "On the Omnipotence and Omniscience of God and the Freedom of Man (1670–71?)." In *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, edited by Robert C. Sleigh, 4–25. New Haven and London: Yale University Press, 2006.

———. "Principles of Nature and Grace, Based on Reason (1714)." In *Philosophical Essays*, edited by Roger Ariew and Daniel Garber, 206–12. Indianapolis & Cambridge: Hackett Publishing Company, 1989.

———. "The Confession of a Philosopher (Fall 1672–Winter 1672–73?)." In *Confessio Philosophi: Papers Concerning the Problem of Evil, 1671-1678*, edited by Robert C. Sleigh, 26–109. New Haven and London: Yale University Press, 2006.

———. *Theodicy. Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*. Edited by Austin Farrer. Translated by E. M. Huggard. Charleston:

BiblioBazaar, 2007.

Lerner, Melvin J. *The Belief in a Just World: A Fundamental Delusion*. New York: Springer, 1980.

Leung, Rebecca. “The Debate Over Sainthood.” CBS News, October 19, 2003. https://www.cbsnews.com/news/the-debate-over-sainthood/.

Lévi-Strauss, Claude. *Tristes Tropiques*. Translated by John Russell. New York: Criterion Books, 1961.

Lewis, Michael. “The Development of Emotional Life.” In *After Darwin. Animals, Emotions, and the Mind*, edited by Angelique Richardson, 281–304. Amsterdam - New York: Rodopi, 2013.

Lindberg, Carter. “Luther’s Struggle with Social-Ethical Issues.” In *The Cambridge Companion to Martin Luther*, edited by Donald K. McKim, 165–78. Cambridge: Cambridge University Press, 2003.

Locke, John. *Second Treatise of Government; and a Letter Concerning Toleration*. Edited by Mark Goldie. Oxford: Oxford University Press, 2016.

Lorenz, Konrad. *On Aggression*. Edited by Julian Huxley. Translated by Marjorie Kerr Wilson. London and New York: Routledge, 2002.

Lucioni, Alfredo. “Monachesimo Occidentale e Monaci Missionari.” In *Storia Del Cristianesimo II. L’età Medievale (Secoli VIII-XV)*, edited by Marina Benedetti, 61–84. Roma: Carocci, 2015.

Lucretius. *De Rerum Natura*. Edited by Monica R. Gale. Oxford: Oxford University Press, 2009.

Luther, Martin. “The Bondage of the Will.” In *Discourse on Free Will*, edited by Ernst F. Winter, 95–138. New York: Continuum, 2002.

Major, John, and Jonathan Holbrow. “Interview with the Editor of the Mail.” *The Mail*. 1993.

Malebranche, Nicolas. “Traité de l’amour de Dieu.” In *Œuvres T.2*, edited by Geneviève Rodis-Lewis. Paris: Gallimard, 1992.

———. *Treatise on Ethics*. Edited and translated by Craig Walton. Dordrecht: Springer Science+Business Media, 1993.

———. *Treatise on Nature and Grace*. Edited by Patrick Riley. Oxford: Oxford University Press, 1992.

Martinson, Robert. “What Works? Questions and Answers about Prison Rform.” *Public Interest* 35 (1974): 22–54.

Masi, Marco. “Quantum Indeterminism, Free Will, and Self-Causation.” *Journal of Consciousness Studies* 30, no. 5–6 (2023): 32–56.

Mastrobuoni, Giovanni, and Daniele Terlizzese. “Leave the Door Open? Prison Conditions and Recidivism.” *American Economic Journal: Applied Economics* 14, no. 4 (2022): 200–233.

McCloskey, Henry John. “A Note on Utilitarian Punishment.” *Mind* 72, no. 288 (1963): 599.

McCord, Joan. “Unintended Consequences of Punishment.” *Pediatrics* 98, no. 4 (1996): 832–34.

McLeod, Allegra M. “Envisioning Abolition Democracy.” *Harvard Law Review* 132, no. 6 (2019): 1613–49.

Meyer, Christian. “Early Massacres: Mass Violence in Neolithic Europe.” In *The Cambridge World History of Violence. Volume 1: The Prehistoric and Ancient Worlds*, edited by Garrett G. Fagan, Linda Fibiger, Mark Hudson, and Matthew Trundle, 299–319. Cambridge: Cambridge University Press, 2020.

Meyer, Philip. “If Hitler Asked You to Electrocute a Stranger Would You?” *Esquire*, 1970.

Mihali, Andreea. “The Role of Freedom in Descartes’ Ethics of Belief.” *Pacific Philosophical Quarterly* 95, no. 2 (2014): 218–45.

Milgram, Stanley. “Behavioral Study of Obedience.” *Journal of Abnormal and Social Psychology* 67 (1963): 371–78.

———. *Obedience to Authority*. New York: Harper & Row, 1974.

Mill, John Stuart. “Utilitarianism.” In *Utilitarianism and On Liberty: Including “Essay on Bentham” and Selections from the Writings of Jeremy Bentham and John Austin*, edited by Mary Warnock, 181–235. Hoboken, New Jersey: Blackwell Publishing, 2003.

Miller, David. “The Concept of Desert.” In *Social Justice*, edited by Matthew Clayton and Andrew Williams, 186–200. Malden, MA: Blackwell Publishing, 2004.

Ministers, Committee of. “Recommendation Rec(2006)2-Rev of the Committee of Ministers to Member States on the European Prison Rules.” Strasbourg: Council of Europe, 2020.

Montada, Leo, and Melvin J. Lerner. “Preface.” In *Responses to Victimizations and Belief in a Just World*, edited by Leo Montada and Melvin J. Lerner, vii–viii. New York: Springer Science+Business Media, 1998.

Montaigne, Michel de. *The Complete Essays*. Edited and translated by M. A. Screech. London: Penguin Books, 1991.

Moore, Michael S. "The Moral Worth of Retribution." In *Responsibility, Character, and the Emotions,* edited by Ferdinand Schoeman, 179–219. Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987.

Mormino, Gianfranco. *Determinismo e Utilitarismo Nella Teodicea Di Leibniz*. Milano: FrancoAngeli, 2005.

———. "Il Triplo Contratto Sociale in Lucrezio e in Gassendi." In *Lucrezio e La Modernità: I Secoli XV-XVII: Atti Del Convegno Internazionale, Università Di Milano-Bicocca, 13-14 Dicembre 2007*, 141–62, 2011.

———. *Per Una Teoria Dell'imitazione*. Milano: Raffaello Cortina Editore, 2016.

———. *René Girard. Il Confronto Con l'altro*. Roma: Carocci, 2012.

Morris, Herbert. "Nonmoral Guilt." In *Responsibility, Character, and the Emotions,* edited by Ferdinand Schoeman, 220–40. Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987.

———. "Persons and Punishment." *The Monist* 52, no. 4 (1986): 475–501.

Morris, Ruth. *Stories of Transformative Justice*. Toronto: Canadian Scholars' Press, 2000.

Moten, Fred, and Stefano Harney. "The University and the Undercommons: Seven Theses." *Social Texts* 22, no. 2 (2004): 101–15.

Mulder, Jack Jr. "Grace and Rigor in Kierkegaard's Reception of the Church Fathers." In *A Companion to Kierkegaard*, edited by Jon Stewart, 156–66. Hoboken, New Jersey: Wiley-Blackwell, 2015.

Murtagh, Kevin J. "Free Will Denial and Punishment." *Social Theory and Practice* 39, no. 2 (2013): 223–40.

Nagel, Thomas. "Moral Luck." In *Free Will*, edited by Gary Watson, 174–86. Oxford: Oxford Readings in Philosophy, 1983.

Narveson, Jan. *Morality and Utility*. Baltimore: Johns Hopkins University Press, 2019.

Nesse, Randolph M. "Evolutionary Explanations of Emotions." *Human Nature* 1, no. 3 (1990): 261–89.

Nocella, Anthony J. II, Mark Seis, and Jeff Shantz. *Classic Writings in Anarchist Criminology: A Historical Dismantling of Punishment and Domination*. Chico: AK Press, 2020.

Nozick, Robert. *Anarchy, State, and Utopia*. Cambridge: Blackwell Publishers, 2001.

Nussbaum, Martha C. *Anger and Forgiveness. Resentment, Generosity, Justice*. Oxford: Oxford University Press, 2016.

Olsthoorn, Johan. "Hobbes on Justice, Propery Rights and Self-Ownership." *History of Political Thought* 36, no. 3 (2015): 471–98.

Orschiedt, Jorg. "Violence in Palaeolithic and Mesolithic Hunter-Gatherer Communities." In *The Cambridge World History of Violence. Volume 1: The Prehistoric and Ancient Worlds*, edited by Garrett G. Fagan, Linda Fibiger, Mark Hudson, and Matthew Trundle, 58–78. Cambridge: Cambridge University Press, 2020.

Parson, Anne E. *From Asylum to Prison. Deinstitutionalization and the Rise of Mass Incarceration after 1945*. Chapel Hill: The University of North Carolina Press, 2018.

Pepper-Bates, Susan. *Nicolas Malebranche: Freedom in an Occasionalist World*. London: Continuum, 2009.

Pereboom, Derk. *Living Without Free Will*. Cambridge: Camb, 2001.

Perret, Bernard. "The Economy as the Opium of the People." In *Palgrave Handbook of Mimetic Theory and Religion*, edited by James Alison and Wolfgang Palaver, 287–94. New York: Palgrave Macmillan, 2017.

Pessin, Andrew. "Malebranche's Doctrine of Freedom/Consent and the Incompleteness of God's Volitions." *British Journal for the History of Philosophy* 8, no. 1 (2000): 21–53.

Plato. *Phaedrus*. Translated by Robin Waterfield. Oxford: Oxford University Press, 2009.

———. *The Republic*. Translated by Desmond Lee. London: Penguin Classics, 2007.

Plutchik, Robert. "A General Psychoevolutionary Theory of Emotion." In *Emotion. Theory, Research, and Experience. Volume 1: Theories of Emotion*, edited by Robert Plutchik and Henry Kellerman, 3–33. New York: Academic Press, 1980.

Price, Richard. *A Review of the Principal Questions in Morals*. Edited by D. D. Raphael. Oxford: Oxford University Press, 1974.

Primoratz, Igor. "Utilitarianism, Justice and Punishment: Comments on Smart and Flew." *Israel Law Review* 25, no. 3–4 (1991): 388–97.

Ragland, C. P. "Descartes on the Principle of Alternative Possibilities." *Journal of the History of Philosophy* 44, no. 3 (2006): 377–94.

Rawls, John. *A Theory of Justice*. Cambridge: Harvard University Press, 1999.

Riley, Patrick. “Malebranche’s Moral Philosophy: Divine and Human Justice.” In *The Cambridge Companion to Malebranche*, edited by Steven Nadler, 220–61. Cambridge: Cambridge University Press, 2000.

Ross, David. *The Right and the Good*. New York: Oxford University Press, 1930.

Ross, Lee. “The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process.” *Advances in Experimental Social Psychology* 10 (1977): 173–220.

Rutherford, Donald. “Malebranche’s Theodicy.” In *The Cambridge Companion to Malebranche*, edited by Steven Nadler, 165–89. Cambridge: Cambridge University Press, 2000.

Ryberg, Jesper. “Punishment and Desert-Adjusted Utilitarianism.” In *Retributivism Has a Past: Has It a Future?*, edited by Michael Tonry, 86–100. Oxford: Oxford University Press, 2011.

Sabini, John, and Maury Silver. “Emotions, Responsibility, and Character.” In *Responsibility, Character, and the Emotions*, edited by Ferdinand Schoeman, 164–75. Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987.

Saltel, Philippe. *Hume : Oberver, Philosopher*. Paris: Ellipses, 2023.

———. “Présentation.” In *Esquisse d’une Morale sans Obligation Ni Sanction*, 15–29. Paris: Les Belles Lettres, 2008.

Sandel, Michael. *Justice. What’s the Right Thing to Do?* London: Penguin Books, 2009.

———. *The Tyranny of Merit*. New York: Farrar, Straus and Giroux, 2020.

Santos Campos, Andre. “Spinoza on Justice: Understanding the Suum Cuique.” *Epoché* 21, no. 1 (2016): 127–43.

Scarantino, Andrea. “The Philosophy of Emotions and Its Impact on Affective Science.” In *Handbook of Emotions*, edited by Lisa Feldman Barrett, Michael Lewis, and Jeannette M. Haviland-Jones, Fourth., 3–44. New York: The Guildford Press, 2016.

Scheper-Hughes, Nancy. “Coming to Our Senses: Anthropology and Genocide.” In *Annihilating Difference. The Anthropology of Genocide*, edited by Alexander Laban Hinton, 248–381. Berkeley, Los Angeles, and London: University of California Press, 2002.

Schieman, Scott. “Anger.” In *Handbook of the Sociology of Emotions*, edited by Jan E. Stets and Jonathan H. Turner, 493–515. New York: Springer, 2006.

Schoeman, Ferdinand. “On Incapacitating the Dangerous.” *American Philosophical Quarterly* 16 (1979): 27–35.

———. “Statistical Norms and Moral Attributions.” In *Responsibility, Character, and the Emotions*, edited by Ferdinand Schoeman, 287–315. Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987.

Schofield, Philip. “«The First Steps Rightly Directed in the Track of Legislation»: Jeremy Bentham on Cesare Beccaria’s Essay on Crimes and Punishments.” *Diciottesimo Secolo* 4 (2019): 65–75.

Schuurmans-Stekhoven, James Benjamin. “Just World Beliefs Mediate the Well-Being Effects of Spiritual/Afterlife Beliefs among Older Australians.” *Journal of Religion, Spirituality & Aging* 33, no. 3 (2020): 332–49.

Scubla, Lucien. “Fuir Vers Une Terre sans Mal et sans Violence ?” *Revue Du MAUSS* 55, no. 1 (2020): 171–97.

———. “L’impératif Sacrificiel et Ses Masques.” *L’Homme*, no. 156 (2000): 241–46.

———. “La Place de La Nation Dans Les Sociétés Individualistes.” *Droit et Culture* 39, no. 1 (2000): 191–210.

———. “Le Sacrifice A-t-Il Une Fonction Sociale ?” *Pardès* 39, no. 2 (2005): 143–59.

———. “Les Hommes Peuvent-Ils Se Passer de Toute Religion ? Coup d’oeil Sur Les Tribulations Du Religieux En Occident Depuis Trois Siècles.” *Revue Du MAUSS* 22, no. 2 (2003): 90–117.

———. “Sur Une Lacune de La Théorie Mimétique : L’ Absence Du Politique Dans Le Système Girardien.” *Cités*, no. 53 (2013): 107–37.

Semple, Janet. *Bentham’s Prison: A Study of the Panopticon Penitentiary*. Oxford: Clarendon Press, 1993.

Shaftesbury, Anthony Third Earl of. *Characteristicks of Men, Manners, Opinions, Times. Vol. II. An Inquiry Concerning Virtue and Merit. The Moralists; a Philosohical Rhapsody*. Edited by Douglas Den Uyl. Indianapolis: Liberty Fund, 2001.

Shapiro, Lisa. “Descartes’s Ethics.” In *A Companion to Descartes*, edited by Janet Broughton and John Carriero, 445–63. Hoboken, New Jersey: Wiley-Blackwell, 2010.

Sidgwick, Henry. *The Methods of Ethics*. London: MacMillan, 1907.

Smail, Daniel Lord, and Kelly Gibson, eds. *Vengeance in Medieval Europe: A Reader*. Toronto: University of Toronto Press, 2009.

Smart, J. J. C. “An Outline of a System of Utilitarian Ethics.” In *Utilitarianism. For and Against*, 1–74. Cambridge/New York: Cambridge University Press, 1973.

———. “Free Will, Praise, and Blame.” *Mind* 70, no. 279 (1961): 291–306.

Smart, Roderick Ninian. “Negative Utilitarianism.” *Mind* 67, no. 268 (1958): 542–43.

Smilansky, Saul. *Free Will and Illusion*. Oxford: Oxford University Press, 2000.

———. “On Free Will and Ultimate Injustice.” *Iyyun: The Jerusalem Philosophical Quarterly /* 52 *עיון: רבעון פילוסופי* (2003): 55–41.

Smith, Angela M. “Responsibility for Attitudes: Activity and Passivity in Mental Life.” *Ethics* 115, no. 2 (2005): 236–71.

Smith, Martin J., Rick J. Schulting, and Linda Fibiger. “Settled Lives, Unsettled Times: Neolithic Violence in Europe.” In *The Cambridge World History of Violence. Volume 1: The Prehistoric and Ancient Worlds*, edited by Garrett G. Fagan, Linda Fibiger, Mark Hudson, and Matthew Trundle, 79–98. Cambridge: Cambridge University Press, 2020.

Spinoza, Baruch. *Ethics*. Edited by Clare Carlisle. Translated by George Eliot. Princeton: Princeton University Press, 2020.

———. “Letter 78 (OP). To the Most Noble and Learned Mr. Henry Oldenburg From B. d. S.” In *The Collected Works of Spinoza. Volume II.*, edited and translated by Edwin M. Curley, 480–82. Princeton and Oxford: Princeton University Press, 2016.

———. “The Political Treatise.” In *The Collected Works of Spinoza. Volume II.*, edited and translated by Edwin M. Curley, 503–604. Princeton and Oxford: Princeton University Press, 2016.

———. “Theological-Political Treatise.” In *The Collected Works of Spinoza. Volume II.*, edited and translated by Edwin M. Curley, 65–354. Princeton and Oxford: Princeton University Press, 2016.

Strawson, Galen. *Freedom and Belief*. New York: Oxford University Press, 2010.

Strawson, Peter. “Ethical Intuitionism.” *Philosophy* 24, no. 88 (1949): 23–33.

———. “Freedom and Resentment.” In *Freedom and Resentment and Other Essays*, 1–28. London: Routledge, 2008.

Szasz, Thomas. *Psychiatric Justice*. Syracuse: Syracuse University Press, 1988.

Taylor, Charles. *A Secular Age*. Cambridge/London: Harvard University Press, 2007.

Testa, Federico. “Jean-Marie Guyau’s Presence and Absence in the History of Philosophy.” *Plí: The Warwick Journal of Philosophy* 30 (2019): 234–53.

Thornton, David E. “Communities and Kinship.” In *A Companion to the Early Middle Ages: Britain and Ireland, c.500-c.110*, 91–106. Hoboken, New Jersey: Wiley-Blackwell, 2009.

Tomkins, Silvan S. “Effect as Amplification: Some Modifications in Theory.” In *Emotion. Theory, Research, and Experience. Volume 1: Theories of Emotion*, edited by Robert Plutchik and Henry Kellerman, 141–64. New York: Academic Press, 1980.

Tonry, Michael. “Doing Justice in Sentencing.” *Crime and Justice* 50, no. 1 (2021): 1–12.

Tooby, John, and Leda Cosmides. “The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables.” In *Handbook of Emotions*, edited by Michael Lewis, Jeannette M. Habiland-Jones, and Lisa Feldman Barrett, Third., 114–37. New York: The Guildford Press, 2008.

Tracy, Jessica L. “An Evolutionary Approach to Understanding Distinct Emotions.” *Emotion Review* 6, no. 4 (2014): 308–12.

Trestman, Robert L., Julian Ford, Wanli Zhang, and Valeire Wiesbrock. “Current and Lifetime Psychiatric Illness Among Inmates Not Identified as Acutely Mentally Ill at Intake in Connecticut’s Jails.” *The Journal of the American Academy of Psychiatry and the Law* 35, no. 4 (2007): 490–500.

Turner, Jonathan H., and Jan E. Stets. “Moral Emotions.” In *Handbook of the Sociology of Emotions*, edited by Jan E. Stets and Jonathan H. Turner, 544–66. New York: Springer, 2006.

Tyler, Tom. *Why People Obey the Law*. Princeton: Princeton University Press, 2006.

Vaquero, Stéphan. “L’unité de La Philosophie Chez Descartes: Métaphysique et Topologie Morale.” *Revue Philosophique de La France et de l’étranger* 134, no. 4 (2009): 471–84.

Veblen, Thorstein. *The Theory of the Leisure Class*. Edited by Martha Banta. Oxford: Oxford University Press, 2007.

Verdier, Raymond. “Droits Pré-Étatiques et Théorie Du Droit.” *Revue Interdisciplinaire d’études Juridiques* 21, no. 2 (1988): 191–205.

———. “Introduction : Le Système Vindicatoire.” In *La Vengeance. Études d’ethnologie, d’histoire et de Philosophie. Volume I. Vengeance et Pouvoir Dans Quelques Sociétés Extra-Occidentales*, edited by Raymond Verdier. Paris: Éditions Cujas, 1980.

———, ed. *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume I. Vengeance et Pouvoir Dans Quelques Sociétés Extra-Occidentales*. Paris: Éditions Cujas, 1980.

———, ed. *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume II. Vengeance et Pouvoir Dans Quelques Sociétés Extra-Occidentales*. Paris: Éditions Cujas, 1980.

———. "Le Désir, Le Devoir et l'interdit : Masques et Visages de La Vengeance." *Déviance et Société* 8, no. 2 (1984): 181–93. https://doi.org/10.3406/ds.1984.1406.

———. "Violence et Intégration Socio-Culturelle : Vengeance et Serment." *Revue Européenne Des Sciences Sociales* 30, no. 94, Les formes de la violence : IXe colloque annuel du Groupe d'Etude "Pratiques Sociales et Théories" (1992): 69–72.

Verdier, Raymond, and Jean-Pierre Poly, eds. *La Vengeance. Études d'ethnologie, d'histoire et de Philosophie. Volume III. Vengeance, Pouvoirs et Idéologies Dans Quelques Civilisations de l'Antiquité*. Paris: Éditions Cujas, 1981.

Waal, Frans De. *Mama's Last Hug. Animal Emotions and What They Teach Us about Ourselves*. New York: W. W. Norton & Company, 2019.

Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge/London: Harvard University Press, 1996.

Waller. "Denying Moral Responsibility: The Difference It Makes." *Analysis* 49, no. 1 (1989): 44–47.

Waller, Bruce N. *Against Moral Responsibility*. Cambridge/London: The MIT Press, 2011.

———. "Deep Thinkers, Cognitive Misers, and Moral Responsibility." *Analysis* 59, no. 264 (1999): 223–29.

———. "Denying Responsibility without Making Excuses." *American Philosophical Quarterly* 43, no. 1 (2006): 81–90.

———. "Just and Nonjust Deserts." *The Southern Journal of Philosophy* 25, no. 2 (1987): 229–38.

———. *The Stubborn System of Moral Responsibility*. London - Cambridge: The MIT Press, 2015.

———. "Virtue Unrewarded: Morality without Moral Responsibility." *Philosophia* 31, no. 3–4 (2004): 427–47.

Wang, Leah. "Punishment Beyond Prisons: Incarceration and Supervision by State," 2023.

Watson, Gary. “Responsibility and the Limits of Evil. Variations on a Strawsonian Theme.” In *Responsibility, Character, and the Emotions. New Essays in Moral Psychology*, edited by Ferdinand Schoeman, 256–86. Cambridge: Cambridge University Press, 1987.

Weber, Max. *Economy and Society. An Outline of Interpretive Sociology*. Edited by Guenther Roth and Claus Wittich. Translated by Ephraim Fischoff, Hans Gerth, A. M. Henderson, Ferdinand Kolegar, C. Wright Mills, Talcott Parsons, Max Rheinstein, Guenther Roth, Edward Shils, and Claud Wittich. Berkeley, Los Angeles, and London: University of California Press, 1978.

———. “Politics as a Vocation.” In *The Vocation Lectures*, edited by David Owen and Tracy B. Strong, translated by Rodney Livingstone, 32–94. Indianapolis & Cambridge: Hackett Publishing Company, 2004.

———. *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Persons. London and New York: Routledge, 1992.

Weinrich, James D. “Toward a Sociobiological Theory of the Emotions.” In *Emotion. Theory, Research, and Experience. Volume 1: Theories of Emotion*, edited by Robert Plutchik and Henry Kellerman, 113–38. New York: Academic Press, 1980.

Williams, Bernard. “Moral Luck.” In *Moral Luck. Philosophical Papers 1973-1980*, 20–39. Cambridge: Cambridge University Press, 1981.

Wolf, Susan. “Sanity and the Metaphysics of Responsibility.” In *Responsibility, Character, and the Emotions*, edited by Ferdinand Schoeman, 46–62. Cambridge/New York/Melbourne: Press Syndicate of the University of Cambridge, 1987.

Young, Iris Marion. *Justice and the Politics of Difference*. Princeton: Princeton University Press, 1990.

———. *Responsibility for Justice*. Oxford: Oxford University Press, 2011.

Young, Michael. *The Rise of the Meritocracy 1870-2033. An Essay on Education and Equality*. London: Penguin Books, 1961.

Yukhnenko, Denis, Shivpriya Sridhar, and Seena Fazel. “A Systematic Review of Criminal Recidivism Rates Worldwide: 3-Year Update.” *Wellcome Open Research* 4, no. 28 (2019).

Zehr, Howard. *Changing Lenses: A New Focus for Crime and Justice*. Scottdale: Herald Press, 1990.

Résumé en français

Au-delà du mérite et de la responsabilité morale : Repenser la justice

Le passé : Les origines de la justice

Le problème de la violence humaine

Le défi fondamental de la gestion de la violence découle de la compétition humaine pour les ressources et la survie, comme l'a souligné Hobbes. Malgré le caractère inévitable des conflits liés à la proximité, les humains dépendent également des liens sociaux pour survivre, ce qui nécessite la coopération et l'organisation de groupes. Jusqu'aux années 1990, les anthropologues ont souvent adopté une vision romantique des sociétés préhistoriques comme étant pacifiques, mais des chercheurs comme Lawrence Keeley ont remis en question cette notion, suggérant que la violence était omniprésente tout au long de l'histoire : les preuves archéologiques confirment la présence de la violence entre les groupes et à l'intérieur des groupes dans les communautés préhistoriques.

Il est essentiel de comprendre comment les sociétés anciennes régulaient la violence pour comprendre les systèmes judiciaires modernes. Les théories du contrat social, illustrées par des penseurs tels que Hobbes, Locke et Rousseau, offrent des perspectives sur la manière dont les sociétés passent d'un état de nature à une gouvernance organisée et font remonter l'origine de l'État à la fonction judiciaire, c'est-à-dire à la gestion des conflits. Même si elles ne fournissent pas des récits historiques exacts, elles éclairent le rôle de l'État dans la résolution des conflits et le maintien de l'ordre social. D'autre part, la théorie victimaire de René Girard offre des perspectives sur la dynamique de la violence dans l'histoire de l'humanité, complétant les découvertes archéologiques et ethnographiques : je commencerai par analyser ces dernières, en identifiant la religion comme le moyen originel de gestion de la violence, pour ensuite expliquer la transition vers les systèmes étatiques et le rôle des théories du contrat social pour les comprendre.

Justice terrestre

Le système sacrificiel

La théorie victimaire de René Girard présente une analyse du comportement humain et de la dynamique sociétale, établissant des parallèles entre l'analyse littéraire, l'anthropologie et les études religieuses. Au cœur de cette théorie se trouve le concept de désir mimétique, que Girard identifie comme le mécanisme fondamental des désirs humains : il implique que les individus désirent des objets non pas pour leur valeur intrinsèque mais parce que d'autres les désirent, ce qui conduit à un réseau complexe d'imitation et de rivalité conséquente.

Si Girard a "découvert" sa théorie mimétique par le biais d'une analyse littéraire, il l'étend au-delà de la littérature à l'anthropologie, en soulignant comment le désir mimétique contribue aux conflits sociaux et à la violence. Il affirme que les crises mimétiques surviennent lorsque la compétition pour les objets désirés dégénère en hostilité généralisée et en guerre : dans ces crises, la violence elle-même devient un objet d'imitation, déchirant les communautés. Girard suggère que ces crises sont résolues par la désignation d'un bouc émissaire, c'est-à-dire qu'une seule victime est collectivement accusée d'être responsable des troubles et tuée pour rétablir la cohésion sociale. L'homicide du bouc émissaire est l'acte de fondation de la société humaine qui, grâce à lui, parvient à surmonter les conflits internes et à vivre en paix.

En examinant d'anciens mythes et textes religieux, Girard montre comment la notion de bouc émissaire a été ancrée dans les sociétés humaines tout au long de l'histoire. Il identifie ensuite un modèle de sacralisation qui se produit après la mort du bouc émissaire, dans lequel la victime est vénérée à la fois comme la source du mal et comme l'agent de la réconciliation ; la sacralisation du bouc émissaire est suivie par la naissance du sacrifice, la répétition rituelle du meurtre fondateur, afin de rénover périodiquement le sens de la communauté de la société et de fournir un exutoire à la violence envers une victime qui n'engendrera pas une crise de violence généralisée.

La théorie de Girard englobe un large éventail de pratiques sacrificielles observées dans diverses cultures à travers le monde, dans lesquelles il identifie des éléments communs tels que la sélection de victimes en marge de la société et l'unanimité de la violence à leur égard, et en analyse l'évolution historique, parfois masquée. Son analyse remet en cause les croyances conventionnelles sur l'objectif et la

signification des pratiques sacrificielles, qui sont généralement décrites comme des pratiques irrationnelles dues à la stupidité des anciennes civilisations : au contraire, il affirme que les rituels sacrificiels ont perduré à travers le temps et se sont développés pratiquement dans toutes les cultures parce qu'ils répondent au besoin de solidarité communautaire et qu'ils fournissent un exutoire à la violence au sein de la communauté. S'appuyant sur des sources littéraires, anthropologiques et religieuses, Girard révèle les mécanismes sous-jacents à ce phénomène et son importance durable à travers les cultures et les époques. Dans l'optique de Girard, le sacrifice apparaît non pas comme un moyen d'apaiser les forces divines, mais comme une libération cathartique de la violence communautaire, un mécanisme d'extériorisation et de réification des tensions collectives.

À l'origine, le sacrifice est décrit comme un acte transformateur, un acte ritualisé qui transcende le simple abattage pour conférer à la victime sacrificielle une signification sacrée. L'étymologie du mot "sacrifice" lui-même, enracinée dans le latin "*sacer facere*", évoque la notion de création du sacré, soulignant le lien intrinsèque entre le sacrifice et l'établissement du divin. L'exploration par Benveniste des parallèles linguistiques entre la terminologie sacrificielle et l'acte de tuer permet de mieux comprendre cette relation symbiotique.

Le schéma sacrificiel peut être étendu au-delà de l'analyse textuelle au domaine de l'enquête archéologique, où les fouilles de sites tels que Çatalhöyük et Göbekli Tepe offrent des aperçus intéressants des vestiges matériels des anciennes pratiques sacrificielles. Des peintures murales représentant des scènes d'abattage rituel et des éléments architecturaux évoquant des arènes sacrificielles sont les témoins silencieux de la place centrale qu'occupait le sacrifice dans les sociétés préhistoriques. En interprétant ces découvertes archéologiques comme des manifestations de violence sacrificielle, Girard remet en question les récits conventionnels sur l'émergence de la religion et de l'organisation sociale, affirmant que les rites sacrificiels sont antérieurs à la sédentarisation et ont joué un rôle fondamental dans la formation de la culture humaine.

L'analyse de Girard conduit à l'examen des dimensions morales de la violence, en juxtaposant les actes collectifs de sacrifice aux actes individuels de vengeance. Dans le cadre de la pensée de Girard, la violence sacrificielle apparaît comme une forme moralement sanctionnée de catharsis communautaire – un moyen d'étouffer préventivement les germes de la vendetta avant qu'ils ne prennent racine. Cette dichotomie morale, dans laquelle la violence collective est valorisée tandis que la

violence privée est vilipendée, souligne le rôle fondamental du sacrifice dans l'établissement de l'ordre et de la cohésion sociale, et dans la fondation de la moralité par la distinction entre la bonne (c'est-à-dire collective) et la mauvaise (c'est-à-dire individuelle) violence.

En naviguant sur le terrain complexe de l'anthropologie religieuse et de la théorie juridique, la théorie de Girard nous permet d'établir des parallèles entre les anciennes pratiques sacrificielles et les systèmes de justice pénale modernes. L'évolution de la punition, des spectacles publics de châtiment à l'enfermement carcéral, reflète les changements d'attitude de la société à l'égard de la vengeance et de la justice. Cependant, cette évolution est sous-tendue par une quête persistante de *catharsis* communautaire – un désir ardent de réconcilier les tensions entre les griefs individuels et l'harmonie collective par l'expiation rituelle de la violence sacrificielle – qui est aujourd'hui mise en œuvre par le système judiciaire, avec son caractère sacrificiel fondamental, comme je le montrerai.

Le système vindicatoire

Avant d'en venir au système actuel, il convient toutefois d'analyser l'une des principales critiques de la théorie de Girard, à savoir l'idée de Verdier d'un système vindicatoire développé par opposition au système victimaire de Girard. Au début des années 1980, Verdier a mené un groupe de recherche qui a travaillé sur la vengeance à travers les cultures et les époques. Dans son essai introductif, Verdier souligne que le système vindicatoire est à la base des civilisations anciennes et met l'accent sur la régulation de la vengeance plutôt que sur sa suppression. Alors que les conceptions modernes la considèrent souvent découlant d'émotions individuelles telles que la colère, Verdier affirme qu'elle peut avoir une valeur éthique lorsqu'elle est contrôlée, une notion soutenue par des penseurs tels qu'Aristote et Thomas d'Aquin.

Verdier identifie trois principes au fondement des systèmes vindicatoires : la réciprocité, qui limite la vengeance à l'offense, la solidarité, qui transforme les griefs individuels en actions collectives, et la distance sociale, qui régule la vengeance en fonction des relations entre les groupes. La théorie de Verdier divise l'espace social en zones périphériques, centrales et intermédiaires, chacune ayant des règles distinctes en matière de vengeance : elle est interdite au sein des communautés familiales pour éviter l'autodestruction, mais prescrite contre les menaces extérieures pour maintenir la cohésion du groupe. Cette approche, à mon avis, réconcilie la théorie de Girard sur la

vengeance comme violence contagieuse avec la notion de formes réglementées et ritualisées de violence : c'est cet aspect rituel qui permet le déploiement sûr de la vengeance, qui, non réglementée, serait destructrice. Verdier aussi reconnaît les aspects rituels et religieux de la vengeance dans son système. En outre, les rituels religieux collectifs facilitent souvent la réconciliation après les actes de vengeance, ce qui souligne l'imbrication du droit et de la religion dans de nombreuses cultures.

La position de Verdier diverge de celle de Girard parce qu'elle cherche à réhabiliter la vengeance comme forme de justice plutôt que de la saper moralement. En tant que penseur affilié à l'anarchisme de droite, Verdier critique le système juridique moderne pour son détachement des individus qui constituent la société et plaide pour un retour à une approche de la justice plus centrée sur la communauté, mettant l'accent sur les besoins des victimes et la solidarité plutôt que sur l'intervention bureaucratique de l'État. Il déplore la marginalisation de la vengeance dans le discours moral et juridique, et plaide pour son rétablissement comme moyen de restaurer la justice au sein de la société civile en contraste avec la justice impersonnelle de l'État.

Les travaux de Verdier remettent en question les conceptions occidentales conventionnelles de la vengeance, en la présentant comme une pratique réglementée et potentiellement juste au sein du tissu social. Bien qu'elle diffère de la perspective de Girard, la recherche de Verdier enrichit la discussion sur le rôle de la vengeance dans la formation des dynamiques sociales et des systèmes de justice à travers les cultures et les époques. Si sa proposition normative pour l'époque actuelle n'est finalement pas convaincante, son système vindicatoire peut être compris de manière fructueuse comme une étape dans le processus qui a conduit du système victimaire aux systèmes étatiques actuels, dans le processus qui, par le biais de la ritualisation, a permis aux communautés de manipuler la vengeance en toute sécurité.

La transition vers le système d'État

L'exploration des systèmes sacrificiels et vindicatoires révèle un lien profond entre la gestion de la violence et l'évolution de la religion. Girard souligne que le christianisme est une anomalie dans son exposition des mécanismes sacrificiels et de l'innocence des victimes, particulièrement illustrée par le rôle de Jésus en tant que bouc émissaire. Contrairement aux mythes anciens, les Évangiles présentent un récit véridique de la violence collective, remettant en question la valeur religieuse positive qui lui est généralement attribuée.

Paul Dumouchel développe la notion de rôle du christianisme dans l'abandon des systèmes victimaires et vindicatoires en affirmant, en particulier, que le pardon et la charité chrétiens perturbent ces systèmes en sapant le principe de réciprocité et en libérant les individus de lourdes obligations de solidarité. Cette adhésion pragmatique aux principes chrétiens s'inscrit dans un processus d'érosion progressive des liens de solidarité qui légitimaient la vengeance collective, la reléguant à une passion individuelle.

L'interprétation de Dumouchel, fondée sur une analyse matérialiste, contraste avec l'accent mis par Girard sur la religion chrétienne en tant que moteur de la découverte de la vérité, cette dernière étant un élément crucial pour le changement systémique de la société. Selon Dumouchel, la capacité des apôtres à exposer la vérité sacrificielle reflète plutôt la complexité de leur société, qui tolérait la dissidence, contrairement aux communautés plus petites et plus homogènes où les mythes fondateurs ont vu le jour. Cette transformation se produit dans un cadre politique où l'État monopolise la violence, supprimant progressivement le besoin de systèmes vindicatoires. Toutefois, ce processus s'entrelace avec les institutions et les pratiques religieuses, façonnant les premiers systèmes de justice. Comprendre la justice dans un contexte religieux permet de mieux comprendre l'évolution des cadres politiques et des normes sociétales. C'est pourquoi il est temps d'analyser plus en détail les conceptions religieuses de la justice.

Justice de l'au-delà

Reconnaître la fonction première de la religion dans la gestion des conflits par la différenciation de la bonne et de la mauvaise violence dévoile son profond lien avec la justice. Bien que la religion ne soit pas essentiellement théorique, les réflexions théologiques portent l'empreinte de ce rôle pragmatique fondamental, reflétant l'évolution des approches en matière de résolution des conflits. Au fur et à mesure que les sociétés humaines se sont développées, l'unité sacrificielle traditionnelle est devenue impraticable, nécessitant d'autres stratégies de gestion de la violence : la rétribution est devenue une pratique omniprésente dans toutes les cultures, reflétée dans le discours théologique et philosophique sur la vie après la mort. Les concepts de justice cosmique partout et toujours (de l'Antiquité grecque et romaine aux conceptions hindoues,

égyptiennes et juives, par exemple) illustrent l'ethos rétributif qui imprègne les conceptions humaines de la justice.

Enracinée dans la pulsion primitive de représailles, la justice rétributive préserve l'harmonie communautaire en endiguant le cycle de la vengeance. La propension aux représailles souligne la nécessité de mécanismes sociaux pour atténuer les dangers de la vengeance privée ; l'imagerie fictive religieuse de représailles divines joue un rôle dissuasif, en complément des pratiques sacrificielles. La dichotomie morale entre la violence pacificatrice et la violence belliciste s'étend à la vengeance, la première n'étant plus seulement sacrificielle, mais consistant en une forme institutionnalisée de vengeance, qu'elle soit exercée par l'État ou par Dieu lui-même dans l'au-delà.

Les institutions religieuses ont historiquement assumé des fonctions de justice (c'est-à-dire de vengeance) tandis que les États bureaucratiques ont lentement mûri – l'alliance entre la religion et l'État a facilité l'administration de la justice, les doctrines religieuses guidant la conduite morale et les clercs étant chargés des enquêtes et des punitions. Malgré les justifications théologiques, cependant, la réalité matérielle de la justice impliquait souvent la violence et l'oppression, révélant l'interaction complexe entre la dynamique du pouvoir et les besoins de la société. La transformation des dynamiques de domination en impératifs moraux, étayés par la justice divine, continue de façonner les notions contemporaines de justice. Les idéaux rétributifs imprègnent les cadres sociaux, favorisant une éthique méritocratique qui rationalise la souffrance et l'oppression.

L'évolution de la justice chrétienne : de la miséricorde à la rétribution parfaite

L'évolution médiévale des croyances chrétiennes concernant la vie après la mort témoigne d'une transformation significative : initialement caractérisée par des attentes millénaristes d'un jugement imminent, la conception du paradis et de l'enfer s'est étendue à un modèle tripartite avec l'introduction du purgatoire. Cette évolution est particulièrement intéressante compte tenu des enseignements de Jésus contre la violence en représailles, qui ont fait du christianisme une religion non rétributive par essence. Les passages de l'Évangile prônant l'amour des ennemis soulignent cette position anti-rétributive.

Cependant, la prédication ultérieure a été principalement axée sur le paradis et l'enfer plutôt que sur la prédication de Jésus sur la miséricorde. Le malaise que suscita

très tôt le concept d'enfer conduisit finalement à la création du purgatoire, une idée qui prit de l'ampleur sur le plan théologique au XIIe siècle, bien qu'elle ait fait partie du discours populaire sur la vie après la mort pendant des siècles déjà. La notion de châtiment dans l'au-delà a été instrumentalisée par les autorités ecclésiastiques et séculières, influençant le droit canonique et conduisant également à la vente d'indulgences.

L'association de la souffrance et du mérite, enracinée dans la théologie chrétienne, imprègne les attitudes de la société. Elle fait partie du préjugé du monde juste, un mécanisme d'adaptation psychologique qui permet de trouver des justifications aux souffrances imméritées, mais qui entrave le progrès de la société en justifiant les injustices sociales. Ce cadre de justice, du châtiment terrestre à la récompense céleste, reflète la lutte de l'humanité pour réconcilier les malheurs quotidiens avec la justice divine. Malgré les enseignements de Jésus sur la non-rétribution, la théologie chrétienne a développé des systèmes complexes de justice rétributive, devenant, par le biais du purgatoire, la religion connue la plus rétributive.

L'évolution de la justice protestante : de la damnation à la méritocratie

Le protestantisme, issu de la révolte de Luther contre les indulgences, s'écarte radicalement des notions catholiques de mérite et de justice. La formation augustinienne de Luther a fortement influencé son rejet de la capacité humaine de salut ou de mérite : le passage de l'accent mis par Augustin sur le libre arbitre à la grâce de Dieu a façonné la croyance de Luther en l'état de péché inhérent à l'humanité et en sa dépendance totale de l'intervention divine. Cette idée a été radicalisée par Jean Calvin, qui a affirmé l'incapacité de l'homme à comprendre la justice de Dieu ; en conséquence, les protestants ont aboli la notion de purgatoire et la nécessité d'une rétribution proportionnelle.

Le conflit entre Luther et Érasme reflète la confrontation d'Augustin avec Pélage, Luther défendant la souveraineté divine sur la volonté humaine. Érasme, tout en reconnaissant le déterminisme divin, plaidait pour une croyance pratique dans le libre arbitre afin de maintenir la responsabilité morale. De même, les compatibilistes modernes comme Saul Smilansky soutiennent le libre arbitre comme une nécessité sociale tout en reconnaissant sa nature illusoire.

Le rejet par le protestantisme de la justice fondée sur le mérite remet en question les conceptions chrétiennes traditionnelles aussi bien que conceptions du sens commun :

la justification par la grâce seule sape le principe de rétribution. En dépit des débats théologiques, les protestants sont aux prises avec l'éthique normative, faisant écho aux préoccupations d'Érasme concernant les implications du déterminisme divin pour la culpabilité morale. En fait, cela soulève des questions sur l'efficacité de la prière et de l'effort humain dans un cadre déterministe, illustrant le défi de réconcilier la souveraineté divine avec la responsabilité humaine.

Pour comprendre la solution de Luther à ce problème, nous devons introduire sa notion de "vocation", qu'il a tirée de sa traduction de la Bible. La vocation confère à l'activité mondaine une signification religieuse, mais elle soulève le dilemme existentiel de savoir si les luttes mondaines contribuent au salut. Le calvinisme et le puritanisme anglais ont abordé cette question en invoquant le concept de prédestination, en posant que le succès dans les activités mondaines pouvait indiquer l'élection d'une personne, et non la provoquer. Toutefois, cela a permis de recadrer les activités mondaines comme des signes de la faveur divine, rétablissant ainsi un lien entre le mérite mondain et le destin dans l'autre monde.

Alors que Luther mettait l'accent sur l'humilité et la grâce imméritée, les mouvements protestants ultérieurs, en particulier le puritanisme, ont promu une éthique de la maîtrise et de l'autonomie. Cette évolution a conduit à des valeurs individualistes qui sous-tendent les pays protestants modernes, favorisant un "providentialisme sans Dieu" caractérisé par la méritocratie. Cet *ethos* rend les citoyens aveugles aux injustices de la société et contribue à démanteler les institutions de protection sociale établies par le protestantisme original. Cette rupture avec l'éthique de Luther ne se reflète toutefois pas nécessairement dans la théologie protestante : des théologiens comme Karl Barth, par exemple, soulignent la passivité de l'humanité dans le salut, rejetant l'idée que le salut dépend des efforts de l'homme. Parallèlement, l'analyse sociologique de Weber met en évidence l'impact sociétal de l'éthique protestante en tant que moteur principal du capitalisme et de la méritocratie.

Cette évolution de la pensée protestante reflète le changement historique du christianisme primitif, qui est passé du message de grâce de Jésus au concept de justice rétributive de l'Église précédemment analysé. La théologie de Luther s'est d'abord fait l'écho de la grâce aveugle du christianisme primitif, mais en la faisant basculer vers la damnation indiscriminée : apparemment non rétributive, elle se révèle rétributive dans les textes de Luther, où elle se fonde sur la culpabilité collective découlant du péché originel. Cependant, cette position théologique est devenue psychologiquement

intenable, ce qui a conduit les pasteurs protestants à souligner l'importance des œuvres dans le salut, s'alignant finalement sur l'éthique rétributive de la méritocratie.

En résumé, l'évolution du protestantisme, depuis la théologie de Luther jusqu'aux conceptions populaires ultérieures, reflète un glissement vers la justice rétributive et la méritocratie, s'écartant de l'accent mis à l'origine sur la damnation indiscriminée et l'humilité.

Transformer les représailles en justice

Cette discussion a mis en évidence le lien persistant entre la religion et le châtiment, ce qui s'étende même dans les sociétés laïques. Alors que la religion a historiquement géré la violence par des pratiques sacrificielles, les États laïques modernes ont adopté le châtiment comme moyen de faire face à la violence réciproque ; malgré la sécularisation de la justice, les influences religieuses restent cruciales, comme en témoignent l'association des pratiques répressives à l'action divine de Dieu et les conceptions religieuses d'une justice d'un autre monde. Le châtiment est ainsi transformé en valeur, sa nature réelle, enracinée dans la vengeance, étant obscurcie par sa transcendance, qui est transposée dans les systèmes judiciaires laïques aussi, où elle conserve, à ce propos, une imagerie religieuse, symbolisée par la déesse aux yeux bandés tenant une balance et une épée. La vengeance se transforme ainsi en justice, le discours autour de laquelle occulte ses éléments fondamentaux, en particulier ses aspects sacrificiels et vengeurs.

Le présent : Théories et pratiques de la justice

L'État, la justice et leurs victimes

Nous avons vu que la religion et l'État sont liés dans la gestion du problème de la violence humaine. Si, historiquement, la religion a aidé les États à lutter contre la violence sociale par le biais de pratiques sacrificielles et de mesures de dissuasion dans l'au-delà, les États laïques modernes portent encore la marque des influences religieuses dans leurs systèmes judiciaires, en ce sens que la justice se distingue de la vengeance par sa relation perçue avec les valeurs transcendantes.

La monopolisation progressive de la violence par l'État a conduit à l'érosion des obligations de solidarité et à la montée du pardon et de la charité en tant que valeurs, lesquels ont contribué à la transformation de la dynamique sociale, rendant obsolète le système vindicatoire et ouvrant la voie à l'émergence de l'État-nation moderne. La solidarité a cédé la place à l'indifférence mutuelle et l'État est devenu le seul arbitre de la violence légitime. Le monopole de l'État sur la violence souligne son rôle primordial dans la gestion des conflits sociaux, comme les théories du contrat social le montrent.

Rétribution et rationalité

La justice de l'État était à l'origine administrée par le système de la torture, infligeant une douleur physique aux transgresseurs et aux suspects, ce qui rendait plus évidente sa relation avec la vengeance. L'institution de la prison est née de la réforme pénale du XVIIIe siècle. Bien que l'on puisse dire que les prisons servent des objectifs tels que la dissuasion et la neutralisation, elles ont principalement hérité du caractère punitif des systèmes antérieurs, où la gestion de la violence était partagée entre les institutions religieuses, l'État et les citoyens privés, en changeant la forme de la punition, mais non pas la substance.

L'État et ses systèmes de justice visaient à canaliser, et non à supprimer, les désirs de vengeance de la société, illustrant ainsi la continuité de la nature vengeresse du châtiment. La littérature et la culture populaire dépeignent souvent la justice et la vengeance comme étant intimement liées, reflétant ainsi les attitudes de la société : un cas exemplaire connu de tous est la fameuse saga "*The Avengers*". La façade rationnelle de l'État moderne masque la nature vengeresse du châtiment : elle interdit la victimisation des innocents mais n'efface pas par ce biais les éléments sacrificiels intégrés dans le châtiment. L'imposition de la proportionnalité vise à distinguer la vengeance de l'État de la vengeance privée sauvage, mais ne modifie pas le principe sous-jacent qui consiste à infliger des souffrances en réponse à un préjudice : l'accent mis sur la proportionnalité et la rationalité ne modifie pas fondamentalement la nature vengeresse du châtiment, qui n'est pas justifiée.

En outre, le mythe de l'individu autodéterminé (*self-made*) génère des reproches à l'égard de ceux qui souffrent du malheur, car ils sont considérés comme responsables de leur condition. Cela crée la catégorie de la violence par omission : l'indifférence de la société à l'égard des groupes marginalisés et l'érosion des liens de solidarité contribuent à la normalisation de la violence, tant dans les systèmes de justice pénale

que dans d'autres contextes sociaux, en particulier celui de la pauvreté. Cela nécessite une brève analyse de la méritocratie en tant que structure sociale qui justifie moralement les inégalités économiques et la violence par omission à l'égard des personnes en difficulté financière.

La méritocratie et ses perdants

Le concept de méritocratie est étroitement liée avec le rétributivisme car il est centré sur la notion d'individus recevant ce qu'ils méritent. Inventé par Michael Young en 1958, le terme a d'abord eu des connotations négatives, critiquant les conséquences potentielles d'une société gouvernée uniquement par le mérite basé sur l'intelligence. Le roman dystopique de Young, *The Rise of the Meritocracy*, dépeint un monde où l'intelligence devient le seul déterminant de la classe sociale et des avantages économiques, ce qui entraîne des troubles parmi les personnes marginalisées.

Le récit retrace l'évolution imaginaire de la méritocratie, qui est passée d'une réaction contre l'ancien système aristocratique des castes à un système fondé sur l'égalité des chances et l'accès à l'éducation. Cependant, une fois l'égalité des chances atteinte, les disparités dans les capacités individuelles deviennent plus prononcées, remettant en cause la notion d'équité de la méritocratie. Malgré les efforts déployés pour garantir l'équité par l'éducation, les divisions entre "supérieurs" et "inférieurs" persistent, entraînant le mécontentement de la société, du aux avantages économiques et éducatifs qui continuent d'influencer l'estime sociale, perpétuant ainsi les inégalités et les distinctions morales.

Contrairement à la représentation de Young, les conceptions contemporaines de la méritocratie omettent souvent les efforts visant à masquer le caractère moral de l'évaluation individuelle. Au contraire, la méritocratie est souvent perçue comme le système le plus juste, et non pas seulement le plus efficiente économiquement, dans lequel les individus reçoivent ce qu'ils méritent. Toutefois, cette moralisation du système économique perpétue des distinctions morales injustifiées au sein de la société.

La méritocratie révèle la nature rétributive inhérente aux perceptions communes de la justice distributive, où les individus sont censés mériter non seulement des emplois correspondant à leurs talents, mais aussi des avantages supplémentaires, à la fois économiques et en termes de prestige. Toutefois, la justification de ces avantages supplémentaires au-delà de l'attribution efficace des emplois est remise en question, notamment en ce qui concerne la raison pour laquelle certains individus devraient

mériter davantage pour avoir travaillé plus dur ou mieux, bien que leur travail ne conduise pas nécessairement à des résultats socialement souhaitables.

Dans la société contemporaine, la réussite économique est souvent liée à l'appréciation sociale, tandis que l'échec économique est considéré comme un échec global dans la vie. Toutefois, les critères de réussite ou d'échec sont arbitraires, récompensant des traits tels que l'intelligence et l'agressivité, qui ne sont pas nécessairement en corrélation avec les vertus morales. La valorisation de certains talents par la société est moralement arbitraire, mais une société imprégnée de valeurs méritocratiques perçoit que ceux qui excellent dans des tâches valorisées sont moralement supérieurs et méritent donc plus de richesse et de pouvoir, ainsi que l'estime sociale, devenant ainsi des modèles sociaux.

Les processus de sélection méritocratiques, fondés sur la concurrence, créent inévitablement des perdants, ce qui contribue à un sentiment de frustration et de victimisation. La pensée rétributive commune tend à rendre ces perdants responsables de leur situation, les considérant comme moralement défectueux ou incompétents, justifiant ainsi leur infortune comme une punition méritée pour leurs faiblesses perçues. La méritocratie, qu'elle soit parfaite ou imparfaite, perpétue la victimisation de ceux qui ne réussissent pas dans son cadre : cette mentalité de blâme des victimes reflète une tendance sociale à attribuer aux individus la responsabilité de leur situation, en négligeant les facteurs systémiques qui contribuent à l'inégalité et à la malchance.

Violence par omission, violence par exécution

Les reproches adressés aux pauvres reflètent la violence par omission, où leur souffrance est à la fois acceptée et justifiée par le discours sur la responsabilité personnelle et le destin mérité. Ce phénomène s'enracine dans l'idée illusoire du monde juste, un biais cognitif qui pousse à blâmer les personnes défavorisées. Cependant, le traitement des personnes punies par l'État diffère sensiblement. Alors qu'elles partagent avec les pauvres et d'autres victimes par omission le fait d'être laissés par la dissolution des liens sociaux, elles sont soumises à une souffrance infligée intentionnellement, c'est-à-dire que, contrairement à la "violence par omission", elles sont soumises à la "violence par exécution". Contrairement au traitement réservé aux pauvres, les transgresseurs sont directement visés non seulement par le blâme, mais aussi par la colère et surtout la violence directe, justifiés comme une fonction nécessaire de la

justice rétributive. Par la punition, en effet, la souffrance leur est volontairement infligée.

Si la chance joue un rôle important dans les discussions sur la justice distributive, notamment en ce qui concerne l'inégalité économique et la détresse financière, en mitigeant la culpabilisation de la pauvreté, c'est un concept qui n'a pas imprégné la réflexion sur la justice rétributive. Les transgresseurs sont considérés comme moralement responsables de leurs actes, même si des facteurs indépendants de leur volonté influencent leur comportement. La comparaison entre les transgresseurs et d'autres membres vulnérables de la société, tels que les malades mentaux, révèle des similitudes dans les processus mentaux perturbés qui conduisent aux problèmes sociaux. Cette perspective remet en question les notions conventionnelles de mérite et de responsabilité morale individuelle, en suggérant que les transgresseurs, comme les malades mentaux, peuvent également être victimes de circonstances indépendantes de leur volonté.

En fin de compte, l'institution du châtiment public sert d'exutoire sacrificiel à la violence sociétale, entremêlant justifications rationnelles, émotions rétributives et besoin de victimes substitutives. Alors que le discours sur la justice distributive a intégré des considérations de chance, la justice rétributive reste ancrée dans les notions traditionnelles de responsabilité morale, perpétuant le cycle du blâme et de la punition.

Les émotions rétributives : une illusion morale

Intuitionnisme moral

Les émotions servent souvent de guides pour nos réponses à diverses situations. Par exemple, la sympathie ressentie à l'égard d'un bébé sans défense suggère un devoir normatif de le protéger et de le soigner, tandis que le malheur face à la souffrance d'un innocent génère un sentiment d'obligation de ne pas lui infliger de mal. De même, la gratitude envers ceux qui nous apportent des bienfaits implique un sentiment de mérite, le bienfaiteur devant être traité positivement ou au moins remercié.

Ces réactions émotionnelles sont généralement acceptées par les différentes théories normatives, notamment la déontologie, le conséquentialisme et l'éthique de la vertu, car elles s'alignent sur leurs principes moraux fondamentaux. Cependant, les émotions et leur interprétation morale deviennent plus complexes lorsqu'il s'agit de situations de conflit et préjudice, générant les sentiments d'indignation et de rage. Si ces

réactions émotionnelles peuvent sembler justifiées intuitivement, leur rationalisation peut s'avérer difficile.

Cette difficulté est due aux différents rôles que l'intuition morale et les émotions jouent dans les diverses théories éthiques. L'intuitionnisme postule que l'homme a une conscience immédiate des valeurs morales, qu'il perçoit grâce à une faculté d'intuition. Ces intuitions morales s'accompagnent d'émotions : dégoût, colère, amour, etc. Cependant, la valeur épistémique des émotions, notamment dans le jugement moral, varie selon qu'elles sont considérées comme vertueuses ou non vertueuses. Les émotions vertueuses, comme la culpabilité, sont considérées comme des guides fiables pour le jugement moral, tandis que les émotions non vertueuses, comme le ressentiment ou l'indifférence, sont jugées peu fiables.

Cependant, même les émotions vertueuses peuvent s'avérer problématiques dans certaines situations. Par exemple, des sentiments de culpabilité peuvent apparaître même dans les cas où le dommage n'était pas intentionnel ou il peut être disproportionné par rapport à l'infraction. Cela met en évidence des limites de l'utilisation des émotions comme guides moraux, car elles ne s'alignent pas toujours sur les principes moraux rationnels.

Le mirage des émotions morales : Le cas de la justice

L'ouvrage de Darwin publié en 1872, *L'expression des émotions chez l'homme et les animaux*, a marqué le début de l'étude fonctionnelle des émotions, un domaine qui a connu une résurgence dans les années 1960 grâce aux travaux d'Ekman et d'Izard, qui visaient à valider la théorie évolutionniste des émotions de Darwin. Depuis, les recherches se sont étendues à d'autres aspects des émotions : non seulement l'expression physique, mais aussi les changements physiologiques et les effets psychologiques, formant ainsi la base de la théorie fonctionnelle moderne des émotions. Cette théorie postule que les émotions remplissent des fonctions adaptatives, façonnées par la sélection naturelle pour faciliter le comportement en réponse à des situations spécifiques. Les psychologues ont identifié un nombre limité d'émotions universelles fondamentales, dont toutes les autres émotions sont composées, chacune étant associée à la résolution de problèmes adaptatifs particuliers.

Cependant, la fonction adaptative d'émotions n'est pas intrinsèquement morale ; la colère, par exemple, sert plutôt des objectifs liés au pouvoir, à la hiérarchie et à la destruction. Si la colère peut naître de la perception de violations morales, elle stimule

principalement l'agression en réponse à des menaces. Des études ont également montré que la tendance à évacuer la colère par l'agression de boucs émissaires constitue une forme de défense psychologique, en évitant la confrontation directe avec les dangers perçus.

Malgré le rôle de la colère en tant que démonstration de l'engagement envers certaines normes morales, son utilisation morale a des implications problématiques, car elle engendre des représailles, justifiées ensuite par l'idée de rétribution qui repose sur une notion irrationnelle d'équilibre cosmique. Si la colère peut renforcer les jugements moraux, sa transformation en principe de justice par le biais de la rétribution est problématique. Ce processus, comme l'a observé Camus, ne fait que ratifier les pulsions naturelles.

Réalisme moral et droit d'être sûr

Huemer affirme que toute justification des croyances évaluatives découle en fin de compte de l'intuition, reconnaissant que les affirmations morales peuvent être retracées jusqu'aux apparences intellectuelles initiales. Toutefois, pour que les théories morales soient solides, ces intuitions doivent être développées et justifiées rationnellement ; Huemer présente une méthodologie critique pour discerner les bonnes intuitions des mauvaises en identifiant et en écartant les biais qui affectent la pensée morale, notamment l'influence culturelle, la programmation biologique et les intérêts personnels. En outre, la cohérence entre les croyances est soulignée comme un moyen d'atteindre la vérité morale, accordant aux individus le "droit d'être sûr" en fournissant des explications et des preuves convaincantes au-delà de la simple intuition.

En examinant les approches des déontologues, des conséquentialistes et des éthiciens de la vertu à l'égard des intuitions morales, Huemer trouve des limites à chacune d'entre elles. Les déontologues font appel à l'intuition mais manquent de justification rationnelle au-delà de l'affirmation que certaines actions sont des devoirs ou des vertus évidents. De même, les éthiciens de la vertu énumèrent des vertus sans fournir de méthode pour les discerner ou les justifier, laissant leur cadre moral dépendre de l'intuition. Si le conséquentialisme offre une solution potentielle en évaluant les actions sur la base de leurs conséquences, le défi consiste à déterminer comment discerner les conséquences positives des conséquences négatives. Cependant, l'utilitarisme hédoniste, une forme de conséquentialisme, offre une solution en fondant les jugements moraux sur le critère universel du plaisir et de la douleur.

L'utilitarisme hédoniste surmonte les défis de Huemer en adhérant à des principes cohérents étayés par l'expérience du plaisir et de la douleur. Ce cadre accorde aux individus le droit de revendiquer une connaissance morale dans un cadre réaliste, réfutant l'idée que les jugements moraux sont purement subjectifs ou arbitraires. En outre, l'utilitarisme hédoniste fournit une méthode pour résoudre les dilemmes moraux et offre une approche alternative de la prise de décision morale qui ne repose pas sur des règles ou des vertus arbitraires. L'utilitarisme hédoniste ébranle les justifications rétributives de la punition, en exigeant une réévaluation totale de la pratique.

La peine : une analyse philosophique

Entre avenir et passé : les justifications traditionnelles de la peine

La pratique de la punition peut être justifiée par des justifications prospectives ou rétrospectives. Les justifications prospectives consistent traditionnellement en la dissuasion, la neutralisation ou la réadaptation, souvent en une combinaison des trois. Les justifications utilitaires se concentrent sur la dissuasion, en jouant sur la peur pour dissuader les transgresseurs potentiels. Historiquement, la punition reposait sur des mesures exemplaires en raison de la faible probabilité d'arrestation ; au contraire, les systèmes pénaux modernes privilégient les taux de détection élevés, c'est-à-dire l'incarcération de masse, plutôt que les sanctions sévères, dans le but de dissuader par la certitude de la sanction plutôt que par sa sévérité.

Les critiques des justifications utilitaires mettent en avant des préoccupations morales, telles que l'instrumentalisation des individus et le risque de sanctions disproportionnées. Bentham, cependant, plaidait pour des peines proportionnelles afin d'éviter les souffrances inutiles, reconnaissant que des peines trop sévères pouvaient nuire à la dissuasion en imposant un coût trop élevé à la société, ce qui n'est plus justifiable en termes utilitaires. En outre, les peines excessivement sévères, telles que la mort pour vol, peuvent engendrer la peur et l'agitation sociale, ce qui, en fin de compte, diminue l'utilité globale.

L'efficacité des justifications prospectives, en particulier la dissuasion, doit être évaluée empiriquement et reste incertaine. Des études suggèrent que si la punition peut avoir des effets dissuasifs, même si modestes, pour les délits contre les biens et les infractions administratives, son efficacité contre les crimes violents est limitée. En outre, la certitude de la sanction semble avoir plus d'influence que sa sévérité, ce qui

soulève des questions quant à la nécessité des mesures punitives ou si d'autres types de réponses pourraient jouer le même rôle. En effet, la recherche comportementale suggère que le respect de la loi est principalement motivé par la perception de la légitimité et de l'autorité plutôt que par la crainte d'une sanction. Ces résultats remettent en question la justification utilitaire de la punition basée sur la dissuasion, car ses avantages doivent l'emporter sur la souffrance infligée pour être jugés justifiables.

Du point de vue utilitaire, une pratique n'est pas justifiée uniquement si elle génère plus d'utilité que de désutilité, elle doit également être le moyen le plus efficace d'obtenir des avantages pour la société : si des systèmes alternatifs permettent d'obtenir des résultats similaires, voire supérieurs, avec moins de souffrance, la punition est jugée inutile et injustifiée. Compte tenu des piètres résultats des études sur la dissuasion et de l'énorme quantité de souffrance générée par l'incarcération, le système punitif actuel ne semble pas pouvoir être justifié par la dissuasion générale sur la base de critères utilitaires.

La dissuasion spécifique, quant à elle, vise à dissuader les individus qui ont déjà commis une infraction de récidiver à travers la punition. Toutefois, les taux de récidive démontrent l'incapacité de l'incarcération à dissuader efficacement les futurs comportements criminels. Les prisons exacerbent souvent les tendances criminelles, transformant des transgresseurs mineurs en criminels endurcis, ce qui remet également en question la notion de neutralisation comme justification valable de la punition. La neutralisation est parfois comparée à la quarantaine, mais contrairement à cette dernière, elle ne remplit pas efficacement son objectif : au lieu de protéger la société d'un risque, l'incarcération remet en circulation des personnes rendues plus dangereuses. De plus, contrairement à la quarantaine, nous estimons raisonnable que les personnes incarcérées ne bénéficient pas de tout le confort possible qui viendrait compenser la privation de liberté de mouvement : l'incarcération entraîne un supplément de souffrance imposée. En outre, la stigmatisation sociale et la perte des liens sociaux subie pendant l'incarcération contribuent aux taux élevés de récidive. La nature cyclique de l'emprisonnement perpétue la criminalité au lieu de la prévenir efficacement.

La réinsertion, troisième justification prospective de la peine, vise à réintégrer les transgresseurs dans la société. Cependant, des preuves empiriques suggèrent que les approches punitives traditionnelles ne parviennent pas à réhabiliter efficacement : ils font plutôt le contraire. Les taux élevés de récidive et les mauvaises conditions de vie après la libération indiquent l'inadéquation de la peine en tant que mesure de

réadaptation. Les prisons ouvertes scandinaves offrent une approche alternative, axée sur l'éducation, la formation professionnelle et la responsabilité personnelle. Ces prisons donnent la priorité à la réinsertion plutôt qu'à la punition, ce qui se traduit par des taux de récidive nettement inférieurs. Le succès des efforts de réhabilitation remet en question l'idée que la punition, en particulier sous la forme de l'incarcération, est nécessaire pour réduire la récidive ; au contraire, l'éducation, la formation professionnelle et les opportunités d'emploi apparaissent comme des stratégies plus efficaces pour réintégrer les transgresseurs dans la société et réduire les comportements criminels futurs.

La justification rétrospective de la punition s'aligne sur la philosophie kantienne, selon laquelle la violation d'une règle justifie une punition quelles qu'en soient les conséquences. Cette position s'appuie toutefois sur des émotions vengeresses telles que la colère et le ressentiment, qui génèrent un sentiment de justice par le biais du châtiment. Si cette pratique génère quelque chose de bon, en maintenant cette perspective, c'est la satisfaction de la victime, mais cela signifie que la justification est en fin de compte conséquentialiste, puisqu'elle repose sur la gratification émotionnelle des personnes lésées plutôt que sur la valeur intrinsèque de la punition elle-même. D'un point de vue conséquentialiste, cependant, la satisfaction tirée du châtiment ne peut justifier les graves conséquences infligées à la partie punie, telles que les condamnations à mort ou les incarcérations prolongées. En outre, nourrir des émotions vengeresses conduit au malheur sociale, sapant ainsi les avantages supposés de la justice rétributive – en fait, les survivants de la violence optent souvent pour des processus de justice réparatrice plutôt que pour des procédures pénales traditionnelles quand on leur donne l'opportunité, ce qui témoigne de leur insatisfaction à l'égard des approches rétributives.

Un autre aspect apparemment positif de la rétribution réside dans son lien avec l'honneur, tel qu'articulé par Courtois et Aristote. Dans les sociétés fondées sur l'honneur, ne pas chercher à se venger peut être un signe de faiblesse et inviter à d'autres abus : la vengeance devient donc un moyen de restaurer son statut au sein de la société, en signalant sa force et sa respectabilité par la peur qu'elle inspire aux autres. Toutefois, cette théorie perd de sa pertinence dans les systèmes étatiques modernes où l'État administre la justice, rendant obsolètes les représailles individuelles et émancipant les relations de l'honneur. Martha Nussbaum critique la justification du châtiment fondée sur l'honneur, en soulignant sa nature narcissique et l'accent mis sur le statut plutôt que sur le préjudice intrinsèque causé par les actes répréhensibles : considérer le

préjudice uniquement comme une atteinte au statut d'une personne empêche d'aborder la douleur et le traumatisme des victimes de manière constructive. Nussbaum souligne l'importance de reconnaître la valeur morale intrinsèque des actions et de donner la priorité aux besoins des victimes plutôt qu'aux impulsions de représailles.

Malgré les critiques, la justification rétributive rétrospective a historiquement dominé les discussions sur la punition. Si les justifications conséquentialistes tournées vers l'avenir ont gagné du terrain lors des réformes pénales du 18e siècle, elles ont été confrontées au scepticisme au milieu du 20e siècle en raison des preuves de l'inefficacité des châtiments administrés par l'État ; cela a entraîné un déclin de l'intérêt pour les justifications utilitaristes. Bien que les critiques du système carcéral remontent à sa création, les appels à la rétribution ont persisté, soulignant l'influence durable de la pensée rétributive sur la punition. Malgré ses défauts, le principe de rétribution reste largement inchangé, reflétant la fascination persistante de la société pour la vengeance et le châtiment.

L' alignement rétributif

Foucault et Angela Davis ont examiné l' alignement du système carcéral avec les intérêts économiques et politiques : Foucault a noté que les prisons ont évolué parallèlement à d'autres institutions disciplinaires au cours de la révolution industrielle, s'intégrant de manière transparente dans le cadre socio-économique ; Davis a introduit le concept de complexe carcéro-industriel, soulignant les motivations économiques qui sous-tendent la construction et le fonctionnement des prisons, à travers l'utilisation des personnes incarcérées comme main-d'œuvre bon marché et comme consommateurs captifs. L'apparent dysfonctionnement du système est aussi fonctionnel : selon Foucault, les prisons sont comme des générateurs de délinquance, rendant l'illégalité gérable tout en instillant un sentiment d'insécurité au sein de la population, justifiant ainsi une surveillance policière accrue. Davis a souligné le rôle des prisons dans la perpétuation de l'exploitation raciale, en particulier aux États-Unis, où les lois post-esclavagistes et le système de *leasing* de prisonniers ont facilité l'incarcération des Afro-Américains.

Outre les facteurs économiques et politiques, le système carcéral s'aligne également, comme nous l'avons vu, sur les réponses émotionnelles à l'agression, ainsi que sur les croyances religieuses. Une autre raison de la persistance du système carcéral réside dans son alignement sur le cadre conceptuel de la philosophie e du sens commun

occidental qui lie libre arbitre, responsabilité morale et culpabilité. Ces idées sous-tendent la conception rétributive de la justice, qui met l'accent sur la responsabilité individuelle et le bien-fondé de la punition. Globalement, l'enracinement du système carcéral dans la société résulte d'une interaction complexe de facteurs économiques, politiques, émotionnels, religieux et conceptuels. Ces derniers trouvent leur origine dans la pensée philosophique moderne, qui sera analysée dans la section suivante.

Construire le cadre conceptuel : Justice et responsabilité dans la philosophie moderne

Les discussions sur le libre arbitre, la responsabilité morale et la justice entre les philosophes les plus influents du XVIIe au début du XVIIIe siècle, notamment Hobbes, Descartes, Spinoza, Malebranche et Leibniz, ont façonné le paysage conceptuel qui est encore d'actualité aujourd'hui : leurs débats, qui se sont déroulés dans un climat d'agitation religieuse, ont eu une influence déterminante sur les conceptions sociétales ultérieures en matière de châtiment et de récompense.

La querelle entre Hobbes et Bramhall porte sur des arguments théoriques concernant le déterminisme par opposition au libre arbitre, en particulier en ce qui concerne les implications pratiques en matière de punition et de blâme. Hobbes plaide pour une approche prospective de la punition, tandis que Bramhall insiste sur la nature rétributive de la faute et de la punition, enracinée dans la loi divine. Cette perspective religieuse, qui s'apparente au discours d'Érasme contre Luther, souligne l'enchevêtrement profond des notions théologiques et philosophiques de la justice. De plus, des penseurs contemporains comme Smilansky empruntent des voies similaires, s'appuyant sur des vérités morales apparemment incontestables pour justifier la justice rétributive malgré les défis posés par le déterminisme.

Descartes considère l'éthique comme le sommet de la sagesse et l'enracine dans la métaphysique et la physique ; cependant, il n'a jamais élaboré de théorie morale complète : il s'est plutôt appuyé sur une « morale par provision » fondée sur les coutumes, la religion et les lois, s'en remettant à la moralité du sens commun.

L'exploration du libre arbitre par Descartes joue un rôle central dans son discours éthique. Il distingue deux degrés de liberté : l'un fondé sur des perceptions claires et distinctes et l'autre sur l'indifférence. Le concept de libre arbitre de Descartes

recoupe les notions de responsabilité morale, de louange et de blâme : en fait, il suggère que les individus sont louables lorsqu'ils exercent correctement leur libre arbitre, en choisissant ce qui est perçu comme vrai et bon ; au contraire, un mauvais usage du libre arbitre conduit à des erreurs pour lesquels on est coupable.

Bien que Descartes mette l'accent sur le libre arbitre, sa position philosophique sur le déterminisme reste ambiguë. S'il reconnaît l'influence déterminante de perceptions claires et distinctes sur la volonté, il suggère également que les individus ont la liberté de choisir des actions contraires. Les opinions de Descartes sur le libre arbitre et la responsabilité morale reflètent les tensions au sein de son cadre philosophique, soulignant les défis que représente la réconciliation des perspectives déterministes et libertaires.

Spinoza rejette le libre arbitre en faveur du déterminisme, affirmant que la volonté est une cause nécessaire, et non une cause libre. Cette position remet en cause les notions traditionnelles de responsabilité morale, ce qui conduit Spinoza à critiquer des concepts tels que le mérite et le péché, qu'il considère comme des produits illusoires de la croyance dans le libre arbitre. Le bien et le mal, en outre, sont des concepts relatifs et dépendent de leurs effets sur les individus, plutôt que de leurs qualités intrinsèques.

Spinoza établit une distinction entre l'état naturel, caractérisé par la peur et la concurrence, et l'état civil, où règnent la coopération et la paix. Il présente deux théories sur la transition vers l'état civil : une vision contractuelle rationnelle et une théorie émotionnelle ancrée dans la nature humaine ; dans les deux cas, le problème que l'établissement de l'état resoude est celui des conflits interpersonnels. La justice qui en dérive est définie comme la distribution des biens et des maux en fonction de ce que les individus méritent, bien qu'il ne s'étende pas sur la manière dont le mérite devrait être déterminé. En tout état de cause, les bonnes lois sont jugées telles en vertu de leur efficacité dans le maintien de la stabilité de l'État civil, suggérant une approche conséquentialiste de la justice basée sur la réduction des conflits.

Le rejet de la miséricorde par Spinoza repose sur des considérations conséquentialistes aussi, privilégiant la sécurité publique à la compassion individuelle : la punition est, en fait, essentielle au maintien de l'ordre social et à la prévention de la violence, s'alignant ainsi sur un désir de vengeance partagé qu'il reconnait au sein de la société.

Malebranche, philosophe, théologien et prêtre, aborde les thèmes de Dieu, de la grâce, de la justice divine et du libre arbitre dans sa philosophie morale, sans se préoccuper de la politique. Mécontent de la solution corps-esprit de Descartes, il plaide pour le dualisme strict de l'occasionnalisme, qui fait de Dieu le seul être efficace ; cela soulève des questions sur la responsabilité de Dieu dans le mal, ce qui conduit à la "proto-Théodicée" de Malebranche : il postule que des volitions générales régissent les lois du monde, minimisant l'intervention directe de Dieu, ce qui justifie le mal comme une conséquence inévitable de lois fécondes.

Cependant, la réconciliation de la liberté humaine avec la providence divine s'avère complexe : la grâce restaure la liberté perdue dans le péché originel, permettant des actes immanents de volonté et de mérite, mais n'arrive pas à résoudre complètement la tension entre la responsabilité humaine et la souveraineté divine.

Leibniz étudie la justice, le libre arbitre et la responsabilité en adoptant une notion de justice comme une vérité éternelle proche des mathématiques et rejetant la vision de Descartes selon laquelle l'omnipotence de Dieu est un arbitraire illimité, soulignant que ce sont des valeurs extérieures à Dieu qui déterminent ses actions. Il défend une conception compatibiliste du libre arbitre, où la liberté réside dans la capacité d'agir selon sa volonté et de percevoir ce qui est bon : la liberté implique l'intelligence, la spontanéité et la contingence, et non pas l'indifférence.

Leibniz fait, toute fois, une distinction entre la nécessité hypothétique et la nécessité absolue, affirmant que la première ne nie pas le libre arbitre, ainsi défendent la responsabilité humaine et affirmant que la louange et le blâme sont fondés sur des actions librement choisies, même si elles remontent en fin de compte jusqu'à Dieu. Selon Leibniz, les actions humaines soient déterminées par des raisons de bonté, mais à cause de la faillibilité humaine cela ne permet pas de rester sans faute : le péché découle d'une mauvaise perception plutôt que d'une méchanceté délibérée. La faillibilité n'est pas toutefois une condamnation au péché, ce qui permet une amélioration morale.

Les mécanismes de la volonté humaine éclairent également la volonté de Dieu, dont la nature parfaite l'oblige à choisir ce qui est le meilleur, ce qui aboutit à la création du meilleur monde possible. Malgré l'existence du péché et du mal, Leibniz affirme que Dieu les permet plutôt qu'il ne les veut, les utilisant comme des moyens pour un plus grand bien, la responsabilité des actions spécifiques reposant sur le libre arbitre de l'homme enclin au mal. Leibniz justifie l'existence du mal comme faisant

partie du plan de Dieu pour un plus grand bien, comme les ombres qui rehaussent un tableau.

Dans la justice distributive, Leibniz préconise de récompenser les actions positives et punir actions négatifs, pour générer le plus grand bien commun possible. Bien qu'il reconnaisse la prédominance des principes utilitaires, le cadre de la justice de Leibniz est multiforme, incorporant des éléments d'harmonie, de responsabilité morale, et même une vision très innovante de la justice en tant que charité, c'est-à-dire un amour sans discrimination.

Le cadre conceptuel dans la philosophie contemporaine

Le discours sur la punition, la responsabilité et le libre arbitre perdure, les arguments contemporains faisant largement écho à ceux du passé. Les principes religieux incontestables ont été supplantés par des principes moraux bénéficiant du même niveau de certitude ; les justifications conséquentialistes du châtiment, défendues par des penseurs tels que Hobbes, Spinoza et Leibniz, ont gagné en importance dans le sillage des réformes pénales du XVIIIe siècle, bien que les arguments rétributifs aient perduré. Beccaria et Bentham ont notamment été les précurseurs des efforts visant à atténuer les peines, mais les dernières années ont été marquées par une résurgence des politiques de répression dure, malgré les preuves que les peines sévères n'ont pas d'effet dissuasif efficace. La tension entre les justifications prospectives et rétrospectives de la punition persiste, reflétant les défis philosophiques actuels concernant le libre arbitre.

Le passage du déterminisme théologique au déterminisme physique n'a pas modifié de manière substantielle les questions fondamentales concernant le libre arbitre et la responsabilité. Les analyses philosophiques continuent de s'attaquer aux questions de l'intentionnalité et de la responsabilité morale, en particulier dans le contexte des procès pénaux et des condamnations. La philosophie juridique contemporaine accepte largement l'existence du libre arbitre, s'attachant plutôt à définir les conditions dans lesquelles les actions peuvent être dépourvues de responsabilité morale : les excuses et les exemptions servent à atténuer la culpabilité en cas de contrainte physique ou de manque de maîtrise de soi.

Différents critères, tels que les désirs de second ordre, la réactivité à la raison et la maîtrise rationnelle de soi, sont proposés pour évaluer la responsabilité morale. Toutefois, ces cadres peinent à prendre en compte la chance morale constitutive et ne parviennent pas à fournir une justification solide des attitudes rétributives et des

punitions. Bien que ces cadres identifient des exceptions à la responsabilité morale, ils soutiennent largement le paradigme dominant sans examiner de manière critique ses fondements. Pour aller de l'avant, il est nécessaire de reconsidérer la base morale de la justice et de développer des cadres alternatifs qui s'alignent plus étroitement sur les principes éthiques.

L'avenir : Perspectives normatives

Remodeler la responsabilité morale

Interconnexion

Le discours sur la justice rétributive, enraciné dans les notions de libre arbitre, de responsabilité morale et de mérite, a fait l'objet d'une critique substantielle de la part d'auteurs s'appuyant sur les fondements posés par des penseurs tels que Hobbes et Spinoza. Wallace, par exemple, à travers son con concept d' "autocontrôle réflexif" met en évidence la fragilité de la responsabilité morale face au traumatisme, considéré comme trop impactant pour qu'on garde ce type d'autocontrôle. Cependant, les traumatismes ne représentent qu'une extrémité d'un spectre d'expériences de vie qui façonnent le comportement, d'autres expériences exerçant également une influence, bien que de manière moins évidente ; ces cas extrêmes illustrent le rôle de la chance morale constitutive, remettant en question les notions traditionnelles de responsabilité morale. Wallace introduit aussi le concept d' "éducation morale", qui englobe le soutien affectif et la transmission des valeurs sociétales, comme un facteur supplémentaire influençant autocontrôle réflexif : les valeurs inculquées au cours de l'éducation façonnent le raisonnement moral et la prise de décision, influençant la perception du bien et du mal et cela détermine différemment les choix de l'individu.

Bruce Waller adopte une position encore plus radicale, rejetant complètement le concept de libre arbitre et de responsabilité morale : il soutient que le mérite fondé sur la justice n'a pas de sens dans un monde déterministe, où tous les individus méritent la même chose. Il plaide pour une perspective d' "égalité morale ultime", qui prend en compte toutes les circonstances causales menant à une action, remettant en cause la focalisation myope sur la responsabilité morale individuelle qui empêche une

compréhension plus profonde du comportement humain et des questions sociétales. La responsabilité morale perpétue une compréhension superficielle des actions et des relations, limitant le progrès moral, en particulier en négligeant les facteurs systémiques contribuant aux actions individuelles.

Une perspective plus large, détachée de la responsabilité morale, permet une analyse plus complète de la causalité et des influences de la société sur le comportement : en considérant l'ensemble des facteurs contribuant aux actions, y compris les structures sociales, le statut économique et les conditions environnementales, une compréhension plus nuancée émerge. Cette perspective met l'accent non plus sur la culpabilité individuelle, mais sur la responsabilité partagée, en soulignant l'interconnexion entre le comportement humain et le contexte social.

En psychologie, la théorie de l'attribution, formulée par Fritz Heider en 1958, a introduit le concept de l'erreur fondamentale d'attribution, soulignant la tendance injustifiée à attribuer les comportements au caractère individuel plutôt qu'aux influences environnementales. Cette erreur est illustrée par les expériences d'obéissance de Milgram, où des sujets ont administré des chocs électriques à d'autres personnes sous l'influence de l'autorité : contrairement aux attentes, la majorité des sujets ont obéi, révélant ainsi le pouvoir des facteurs situationnels sur le caractère individuel. La théorie de Heider comporte quatre étapes d'attribution, dans une échelle qui va de l'attribution de tous les résultats à l'agent individuel, à la prise en compte de son intentionnalité, pour finalement reconnaître le pouvoir écrasant de l'influence de l'environnement, en remettent en question la notion de caractère personnel et en fournissant un outil qui permet d'expliquer les comportements de manière plus précise.

Alors que certains, Harry Frankfurt surtout, s'opposent à l'abandon des attitudes réactives telles que le blâme et le ressentiment – qui tendent à responsabiliser l'individu – parce qu'ils pensent que cela entraînerait l'abandon de toutes les attitudes participatives, d'autres, comme Martha Nussbaum, soutiennent que leur rejet favorise en fait des attitudes positives telles que l'amour et la compassion. L'approche de Mandela en matière de justice postapartheid en est un exemple, puisqu'elle donne la priorité à la compréhension plutôt qu'à la vengeance, soulignant ainsi le potentiel des alternatives non violentes dans les systèmes judiciaires. En fin de compte, le rejet de la responsabilité morale individuelle n'implique pas le relativisme moral ou l'absence de distinctions morales ; il s'agit plutôt de comprendre les conditions qui conduisent au

comportement, de promouvoir l'empathie et de s'attaquer aux causes profondes plutôt que de se contenter de mesures punitives.

La coresponsabilité

La responsabilité causale peut être distinguée de la responsabilité morale : reconnaître son rôle causal dans les événements n'implique ni blâme ni punition ; au contraire, elle ouvre la voie à une responsabilité partagée. La responsabilité causale reconnaît l'influence de l'ensemble de la chaîne causale, même lorsque ses éléments individuels ne sont pas clairs. Cette perspective met l'accent sur la compréhension des circonstances afin de s'attaquer efficacement aux causes profondes en fondent une responsabilité partagée – qui diffère de la responsabilité collective, qui justifie la punition collective et s'inscrit dans une logique rétributive. La responsabilité partagée reconnaît les individus comme faisant partie d'un réseau plus large, mettant en balance leur responsabilité causale avec d'autres facteurs contributifs.

Il s'agit d'un paradigme qui ébranle les modèles thérapeutiques (quarantaine, thérapie morale) pour la punition, qui concentre leur action sur l'individu et ne tiennent pas compte de la chaîne de causalité plus large, n'abordant pas les conditions sociétales qui contribuent à la criminalité ; au contraire, ils perpétuent les normes sociales existantes et ne permettent pas de prévenir efficacement les crimes à l'avenir. Même si elles sont des approches qui rejettent l'idée de punir simplement sur la base de la responsabilité morale, elles fonctionnent toujours dans le cadre du paradigme néolibéral, en l'absence d'une réévaluation plus large de la justice.

Le néolibéralisme, paradigme occidental dominant caractérisé par des inégalités extrêmes, des groupes marginalisés et une croyance en l'autosuffisance individuelle, encourage une culture de la responsabilité morale individuelle. L'inadéquation de ces paradigmes appelle à une évolution vers la coresponsabilité, qui remet en cause la rétribution. La coresponsabilité nie le caractère central de la responsabilité morale individuelle et met l'accent sur des solutions systémiques à la criminalité.

Bruce Waller établit une distinction entre la responsabilité morale et la responsabilité assumée, soulignant l'importance de cette dernière pour l'autonomie et l'action de l'individu : la prise de responsabilité implique de prendre des décisions, de s'approprier des actions et de rendre des comptes dans le cadre des relations sociales, sans que cela n'entraîne de blâme ou d'autres formes de punition face à l'échec. L'exploration de paradigmes alternatifs en philosophie politique peut permettre

d’aborder les conséquences du rejet de la responsabilité morale individuelle et de repenser les approches de la justice.

Une approche de la criminalité fondée sur la justice sociale

Dans la pensée commune, la justice est symbolisée par l’image d’une déesse aux yeux bandés tenant une balance et une épée. Les châtiments sont considérés comme des fardeaux distribués par l’État comme un moindre mal : cette vision tend à naturaliser le châtiment, en l’assimilant à une loi de la nature. Guyau rejette cette idée, arguant que la nature ne punit personne ; la souffrance résultant d’actions n’est pas intrinsèquement morale ou pénale. Les émotions rétributives peuvent survenir naturellement, mais elles n’ont pas de valeur morale inhérente. La morale du sens commun reflète souvent des idées pratiques pour la vie en communauté : elles peuvent être utiles, mais elles ne sont pas nécessairement moralement valables. D’autres conceptions de justice sont possibles et moralement désirables.

Solidarité

La justice rétributive, profondément ancrée dans les structures sociétales et les théories de l’État, est souvent considérée comme la fonction première de l’État : dans les théories du contrat social, les individus délèguent le pouvoir de violence à l’État pour les protéger des risques de violence ; l’État, à son tour, devient le vengeur des citoyens en administrant le châtiment lorsque les lois sont violées. La théorie de l’État de Girard met également l’accent sur le rôle de la justice rétributive : l’État émerge pour gérer la violence humaine, en prévenant d’autres crises à travers une vengeance qui ne peut pas donner lieu à une contre-vengeance.

La montée en puissance de l’État-nation provoque l’effondrement l'un sur l'autre des deux premiers cercles de solidarité du système vindicatoire de Verdier : la solidarité s’étend du cercle intérieur au cercle intermédiaire et les pratiques de rétorsion des justices, exercées par l’intermédiaire de l’État, s’étendent du cercle intermédiaire au cercle intérieur. L’élargissement des attitudes rétributives à la famille au sein de l’État-nation pose des problèmes : les victimes peuvent hésiter à dénoncer des crimes commis par des membres de leur famille en raison des liens affectifs et de la dépendance économique ; telles actions peuvent aussi conduire à des conflits familiaux internes, remettant en cause l’idée que la vengeance est désormais monopolisée par l’État.

En ce qui concerne les liens de solidarité, dans l'État-nation, tous les citoyens deviennent des "amis", tandis que les étrangers sont considérés comme des "ennemis". Cette expansion incite à concevoir la justice sociale, qui traite de la répartition équitable des avantages et des charges au sein de la société. Contrairement à la justice rétributive, la justice sociale se concentre sur les structures sociales et l'atténuation des inégalités, reflétant une évolution vers des concepts de justice dépersonnalisés. Malgré l'apparente complémentarité entre la justice rétributive et la justice sociale, je soutiens que la justice sociale pourrait être étendue pour couvrir le domaine traditionnellement couvert par la justice rétributive, comme je le montrerai dans la section suivante.

La justice sociale : Plus qu'une question économique

La justice sociale, souvent abordée en termes économiques, est le système par lequel l'État assure l'accès à des services fondamentaux tels que l'éducation, les soins de santé et la sécurité personnelle à travers la redistribution économique, qui est donc un moyen de parvenir à une répartition plus large des droits. Cependant, les personnes qui commettent des crimes sont souvent confrontées à la révocation de certains droits et à l'exclusion de la communauté de leurs concitoyens. Cette observation suggère que la justice sociale et la justice rétributive entretiennent une relation complémentaire mais contrastée : la justice sociale incarne la solidarité au sein de la communauté en garantissant une répartition équitable des ressources et des opportunités entre les citoyens. En revanche, la justice rétributive représente une forme de désolidarisation, les transgresseurs étant exclus de la communauté.

La justice rétributive retire essentiellement le transgresseur de la communauté des égaux et le positionne en tant que victime sacrificielle, répétant un schéma observé dans la théorie de la violence sacrificielle de Girard. L'érosion des liens de solidarité dans les sociétés modernes signifie que n'importe qui peut potentiellement devenir une victime sacrificielle, en particulier ceux qui ont du mal à répondre aux attentes de la société, comme les personnes âgées ou les handicapés. Pour résoudre ces problèmes, une justice sociale institutionnalisée devient nécessaire, comme l'illustre l'idée de Rawls du voile d'ignorance pour accorder des conditions équitables dans sa théorie de la justice. Cependant, les transgresseurs sont généralement exclus des considérations de justice sociale, car leurs actes sont considérés comme méritant d'être punis : cette exclusion souligne la tension entre les concepts de mérite, de responsabilité individuelle et de justice distributive dans les cadres rétributifs.

En outre, la discussion sur la chance constitutive complique les notions de mérite et de responsabilité au sein de la justice pénale. Alors que la justice distributive économique reconnaît des facteurs indépendants de la volonté d'un individu dans les résultats personnels, la justice ne prend pas en compte des facteurs similaires. Cela est enraciné dans une réticence plus générale à remettre en question les récits traditionnels du libre arbitre et de la responsabilité morale au sein des systèmes de justice rétributive. L'égalitarisme de la chance, défendu par Richard Dworkin, postule que la chance économique ne détermine pas le mérite de pauvreté d'une personne ; de la même manière, je soutiens que la chance constitutive annule le mérite de la punition. Dans cette perspective, la justice sociale devient un instrument permettant de soulager la souffrance au lieu d'administrer des mesures punitives, car personne ne mérite intrinsèquement la souffrance.

La distinction établie par Dworkin entre la chance brute, indépendante de la volonté d'un individu, et la chance optionnelle, résultant de choix délibérés, devient problématique lorsque l'on considère la chance constitutive, qui remet en question la notion même de choix, brouillant les frontières entre ce qui est mérité et ce qui ne l'est pas. Elisabeth Anderson critique l'égalitarisme de la chance parce qu'il met l'accent sur la responsabilité morale individuelle dans la chance optionnelle, arguant qu'il n'aborde pas l'impact sociétal plus large de l'inégalité. Elle plaide pour un cadre de justice sociale qui donne la priorité à l'égalité de la valeur morale et à l'absence d'oppression pour tous les individus, indépendamment de leurs circonstances ou de leurs choix spécifiques. L'approche d'Anderson combine un engagement en faveur de la responsabilité individuelle avec un État-providence redistributif afin de garantir les libertés fondamentales et de protéger contre l'oppression tous les individus indépendamment de leurs responsabilités. L'accent qu'elle met sur l'absence d'oppression s'aligne sur l'idée de chance constitutive, qui remet en cause la notion de mérite et propose une justice plus inclusive. Cependant, son argumentation va également dans la direction opposée, en mettant l'accent sur la responsabilité morale individuelle et la nécessité pour les individus de faire face aux conséquences de leurs propres choix sans l'intervention de l'État.

En critiquant l'égalitarisme de la chance de Dworkin, Anderson met l'accent non plus sur la redistribution économique, mais sur le concept plus large d'égalité dans les relations. Alors que Dworkin considère les humains comme des acteurs économiques, Anderson les voit avant tout comme des êtres politiques : la redistribution devrait donc

être un instrument plutôt qu'un élément central pour atteindre l'égalité, en soulignant l'importance des relations fondées sur l'égalité par rapport aux facteurs purement économiques. Alors que les inégalités économiques sont souvent abordées par le biais de mesures sociales, la justice pénale reste largement épargnée par les théories de la justice sociale. L'approche relationnelle d'Anderson en matière d'égalité et de justice peut être étendue à la justice pénale en se concentrant sur les structures sous-jacentes qui perpétuent l'inégalité et l'oppression.

Le concept d'injustice structurelle d'Iris Marion Young déplace l'attention des actions individuelles vers les structures sociétales. L'injustice structurelle découle de l'échec de la justice sociale, ce qui incite à reconsidérer la responsabilité individuelle au sein de ces structures. La notion de responsabilité partagée de Young, qu'elle propose à travers son modèle de responsabilité par connexion sociale, met l'accent sur l'interdépendance des acteurs sociaux et sur la nécessité d'une action collective pour lutter contre l'injustice. Ce type de responsabilité s'éloigne du blâme et de la culpabilité : la reconnaissance du rôle de chacun dans la perpétuation de l'injustice n'est qu'un instrument pour inciter à œuvrer activement en faveur du changement. Toutefois, le modèle de responsabilité fondé sur le lien social implique également que la contribution à une société juste ne donne pas lieu à des récompenses ou à des punitions, mais qu'elle soit intrinsèque à une citoyenneté responsable : il met l'accent sur des actions responsables pour le bien collectif, sans attendre de gain personnel ni craindre d'être blâmé. Si l'absence de contribution ou la faillite peuvent être critiquée, elles n'entraînent pas de blâme.

Le point de vue de Young porte essentiellement sur les aspects économiques de la justice sociale, en se concentrant sur la pauvreté, sans aborder la question de la justice pénale. Toutefois, la redistribution sert à atteindre des objectifs sociétaux plus larges, notamment à offrir la "meilleure vie possible" à tous les membres, y compris aux malfaiteurs. La criminalité est alors présentée comme une question d'intérêt social influencée par des facteurs tels que la pauvreté et la maladie mentale. On propose, donc, une approche basée sur le soulagement et la résolution, plutôt que sur le blâme et la punition, mettant l'accent sur le partage des responsabilités au sein des structures sociétales.

Si la justice sociale vise avant tout à rectifier les inégalités socialement construites, elle devrait également reconnaître et atténuer les conséquences de la chance brute. En étendant le concept de justice sociale et le modèle de responsabilité par

connexion sociale à la justice pénale, j'esquisse une approche nuancée qui tient compte à la fois de l'injustice structurelle et de la malchance, en mettant l'accent sur la responsabilité partagée et en évitant d'attribuer la faute aux victimes.

Utilitarisme et abolition des prisons : Repenser la justice

Dans *Are Prisons Obsolete?,* Angela Davis remet en question la croyance en la nécessité des prisons, en notant leur émergence relativement récente dans l'histoire de l'humanité. L'incarcération est une pratique enracinée dans la rétribution, la fonction principale de l'ancien système punitif basé sur la torture qui prévalait dans les monarchies où la punition servait à satisfaire la colère royale, et que le système carcéral a remplacé sans en remettre en cause les principes.

L'utilitarisme, souvent négligé dans les discussions sur l'abolition des prisons, soutient en fait la remise en question et, enfin, l'abolition des prisons. Mon objectif est de combler le fossé entre l'utilitarisme et le mouvement d'abolition des prisons, malgré l'influence historique des penseurs utilitaristes dans l'élaboration des systèmes pénaux modernes.

Le chaînon manquant

Beccaria a joué un rôle essentiel dans l'abandon du système de torture au XVIIIe siècle, en défendant des arguments conséquentialistes contre son utilisation. Bien que sa classification comme utilitariste soit discutée, il a soutenu que l'objectif de la punition est la dissuasion et la prévention de nouveaux dommages à la société, en mettant l'accent sur la proportionnalité et la minimisation de la douleur physique. Les réformateurs ont imaginé une variété de châtiments adaptés à des crimes spécifiques, dans le but de créer un lien intuitif entre le crime et le châtiment afin de dissuader les transgresseurs potentiels. Malgré la cruauté perçue de ces châtiments, ils représentaient un adoucissement des peines par rapport à l'approche disproportionnée du système de torture.

Les travaux de Bentham ont enraciné les principes utilitaires dans la réforme pénale. Il considérait la punition comme un moyen de créer l'harmonie sociale en décourageant les actions socialement nuisibles. Toutefois, la punition doit être la moins sévère possible, ce qui fonde le concept du panoptique – bien qu'aujourd'hui associée à une surveillance oppressive, en fait, le panoptique benthamien visait à diminuer la

violence dans le système de justice pénale en renforçant la dissuasion par une observation constante au lieu que par la violence physique, comme le mettent en évidence les observations de Dickens sur les prisons américaines. Malgré son association historique avec l'expansion des prisons, l'utilitarisme cherchait en fin de compte à minimiser la souffrance avec la reforme pénale.

La conception contemporaine de la punition repose encore largement sur la justification utilitaire de la dissuasion générale, malgré le manque d'efficacité prouvée. Cependant, les critiques soutiennent que l'utilitarisme, s'il n'est pas équilibré avec d'autres principes moraux tels que la rétribution, conduirait à la punition d'innocents pour le plus grand bien, ainsi qu'à une punition disproportionnée tant qu'elle est plus efficace. En ce qui concerne ce dernier point, il est intéressant de noter que les partisans de l'utilitarisme, comme Bentham, visaient à réduire l'arbitraire dans le système pénal par des classifications claires et des punitions proportionnelles, rejetant ainsi implicitement la punition des innocents. L'utilitarisme, lorsqu'il est considéré dans sa globalité, prend en compte les conséquences indirectes et à long terme, en évitant les stratégies qui pourraient conduire à la répression de la société.

Si certains utilitaristes acceptent à contrecœur la possibilité de punir les innocents, il est important de noter que, selon l'utilitarisme, toute douleur infligée est négative. Encore une fois, une vision plus globale de l'utilitarisme ne justifie pas la punition de l'innocent en raison de ses impacts sociétaux négatifs évidents à long terme, sauf dans des cas exceptionnels où les déontologues seraient également d'accord avec cette pratique. Il est vrai, cependant, que d'un point de vue moral, l'utilitarisme ne reconnaît pas les différences intrinsèques entre la punition du coupable et celle de l'innocent, ce qui remet en question les notions de responsabilité morale individuelle.

Ceci, combiné aux résultats décevants de la recherche empirique sur la dissuasion et la réhabilitation a conduit à la réapparition des justifications rétributives, certains affirmant que la punition est une forme de respect des droits des transgresseurs. Cette évolution reflète une demande sociale plus large pour que les transgresseurs reçoivent le châtiment qu'ils méritent et montre que la pratique du châtiment, antérieure aux justifications utilitaires, et leur survivant, les a utilisées seulement comme rationalisation superficielle pour dissimuler des motifs de châtiment plus profonds et moins moraux. Si l'utilitarisme a fourni des justifications rationnelles, il n'a pas remis en cause la pratique elle-même. Cependant, contrairement à de nombreux utilitaristes subséquents, Bentham a cherché activement des alternatives à la punition, préconisant

des méthodes indirectes pour prévenir le crime en s'attaquant à ses causes sous-jacentes, en mettant l'accent sur l'éducation, la protection sociale et l'aide à l'emploi, et reconnaissant la pauvreté en tant que facteur criminogène. L'approche de Bentham, en bref, s'aligne sur l'abolitionnisme carcéral contemporain, qui cherche des alternatives à la punition.

Le mouvement abolitionniste est né d'idéologies anarchistes mais a évolué récemment en réponse à la discrimination et au racisme au sein du système carcéral américain. Bien que l'injustice raciale soit un aspect crucial, l'abolition des prisons devrait plutôt, je trouve, s'attaquer à la question fondamentale de la punition elle-même : les abolitionnistes devraient donc élargir leur champ d'action au-delà du racisme pour remettre en question la structure punitive de manière structurelle, en reconnaissant que le problème central réside dans la croyance que la souffrance sert un bien sociétal, que ce soit par le biais de la rétribution ou de l'utilité. Ainsi, la réflexion contemporaine sur l'abolition devrait s'inscrire dans une perspective universelle, visant à démanteler le système punitif plutôt qu'à s'attaquer uniquement à ses aspects discriminatoires.

L'utilitarisme offre une solution à ce problème, bien qu'il ait souvent été associé à la justification de la punition plutôt qu'à sa remise en question fondamentale. Les idées de Bentham s'alignent sur l'abolitionnisme carcéral contemporain et l'utilitarisme fournit un cadre universel qui fait défaut à la pensée abolitionniste actuelle. Les propositions de Bentham en matière de prévention de la criminalité ont été négligées dans les discussions sur l'utilitarisme, éclipsées par l'accent traditionnellement mis sur la punition. Pourtant, l'importance qu'il accorde à la législation indirecte et sa critique de la punition comme signe d'échec systémique sont des aspects cruciaux de sa philosophie.

L'accent mis par l'utilitarisme sur la maximisation de l'utilité fournit une base rationnelle pour remettre en question le système punitif. Dans l'élaboration des politiques, je pense que l'utilitarisme négatif offre un cadre particulièrement prometteur en tant que règle générale, car il permet à la politique de ne pas rentrer dans les questions difficiles de l'identification et de la promotion de l'utilité, en se limitant à la prévention de la souffrance.

Les utilitaristes ont toujours eu besoin de se défendre contre les accusations dystopiques, mais de cette manière, ils ont affaibli l'utilitarisme en se concentrant sur le maintien du statu quo plutôt que sur sa remise en question. Cependant, l'engagement de

l'utilitarisme en faveur de l'analyse empirique et de la flexibilité dans la formulation des règles morales offre une opportunité de changement transformateur unique qu'il faut bien valoriser.

Transformer les conceptions de la justice

La question de savoir ce qu'il faut faire des transgresseurs si les prisons sont abolies amène à reconsidérer l'ensemble du cadre sociétal entourant le crime et la punition. L'abolitionnisme ne cherche pas à abolir que les prisons elles-mêmes, mais plutôt les conditions qui nécessitent leur existence. Cette approche, alignée sur la justice transformative, se concentre sur la restructuration de la société afin de minimiser le besoin de mesures punitives. Angela Davis, à cette fin, souligne l'importance de réorienter les ressources des prisons vers des besoins sociétaux essentiels tels que les soins de santé, l'éducation et le soutien à la santé mentale ; une réorientation qui permet de s'attaquer aux causes profondes de la criminalité plutôt que de se contenter de réagir à ses manifestations. En outre, il demande de redéfinir la notion de responsabilité, en mettant l'accent sur la responsabilité systémique plutôt que sur la responsabilité individuelle.

Cependant, lorsque des crimes sont encore commis, la justice réparatrice offre une alternative aux mesures punitives traditionnelles. John Braithwaite définit la justice réparatrice comme un processus dans lequel toutes les parties concernées ont la possibilité de discuter du préjudice causé et de décider d'actions réparatrices. Contrairement à la justice punitive, qui met l'accent sur la sanction, la justice réparatrice donne la priorité à la guérison et à la réconciliation.

Bien que la justice réparatrice ait coexisté avec les systèmes punitifs, elle s'écarte du paradigme punitif en visant à reconstruire les liens sociaux rompus par la criminalité, et mettant l'accent sur la participation de la communauté et le dialogue entre la victime et l'auteur du délit. Les sociétés autochtones, telles que les Māori en Nouvelle-Zélande, pratiquent depuis longtemps des formes de justice réparatrice, et ont même contribué à l'intégration des pratiques réparatrice dans le système néozélandais. Le système rwandais des *gacacas*, réactivé après le génocide de 1994, est un autre exemple de pratiques réparatrices, qui s'est concentré sur la satisfaction des victimes.

La justice réparatrice remet en question les notions traditionnelles de justice en donnant la priorité au dialogue victime-transgresseur et à la compréhension mutuelle. Barbara Hudson souligne l'importance de prendre en compte à la fois les besoins des

victimes et ceux des transgresseurs, en allant au-delà d'une approche purement centrée sur la victime, illustrée par les *gacacas* traditionnels. En favorisant l'empathie et le dialogue, la justice réparatrice offre une voie vers la guérison et la réconciliation pour toutes les parties concernées.

Les Commissions de Vérité et de Réconciliation (CVR) illustrent un processus similaire dans l'Afrique du Sud de l'après-apartheid. Les CVR ont cherché à obtenir la vérité et la reconnaissance des coupables sans recourir à des mesures punitives, reconnaissant que la punition perpétue la vengeance et la division de la société. Cette approche met l'accent sur le dialogue, la rencontre et la réconciliation plutôt que sur la punition, dans le but de transformer la confrontation en réconciliation.

La justice réparatrice, dans le cadre de la justice transformatrice, offre une alternative aux systèmes punitifs traditionnels. Elle reconnaît l'interaction complexe des facteurs conduisant à la criminalité, tels que la privation, la pauvreté, la maladie mentale et les abus ; en plus, elle encourage l'empathie et le dialogue, en remettant en question les rôles rigides de victime et d'agresseur et en reconnaissant l'humanité et les luttes des deux parties. L'implication de la communauté est cruciale dans les processus de justice réparatrice, car elle reconnaît que le crime affecte la communauté tout entière et met l'accent sur la responsabilité partagée pour la guérison et la restauration.

Cependant, l'efficacité des processus de justice réparatrice dépend de divers facteurs, notamment des remords de l'auteur de l'infraction et de la volonté de la victime d'engager le dialogue – éléments qui rendent le processus fragile et risquent d'échouer. Cependant, un certain degré de risque d'échec doit être accepté comme il l'est actuellement dans notre système punitif qui, comme nous l'avons vu, échoue dans la plupart des tâches qu'il vise à accomplir.

En fin de compte, le passage à une justice transformative exige une réévaluation des normes et des valeurs sociétales, une remise en question des émotions réactives et l'adoption d'approches rationnelles et empathiques de la justice. Le succès de ces réformes dépendra de l'évaluation empirique et de l'engagement continu en faveur des principes transformateurs.