

Utilitarianism and Prison Abolition: A Normative Perspective



Marina Beraha

(Independent Researcher, Milan, Italy; marina.beraha@hotmail.it)

ORCID: 0000-0002-3516-1619

Abstract: This paper develops a utilitarian case for prison abolition grounded on utilitarianism's commitment to the reduction of pain, challenging the traditional forward-looking utilitarian justifications of punishment, whose empirical foundations are increasingly contested. To construct the utilitarian abolitionist argument, traces of an abolitionist theory are identified in Bentham's penal thought, especially in his neglected analysis of indirect legislation as a means of crime prevention. His theoretical analysis is integrated with contemporary debates on punishment, incarceration and restorative justice. The paper argues that a consistently utilitarian approach should prioritize harm reduction through social restructuring, transformative approaches, and restorative responses to wrongdoing, rather than relying on incarceration and deterrence-centred punishment.

Keywords: Abolition; Utilitarianism; Bentham; prison; punishment; justice; prevention.

I. Introduction

Utilitarianism has been linked to the emergence, affirmation, and justification of prison as the sole form of institutionalised punishment.¹ In this paper, I aim to demonstrate that a deeper understanding of utilitarianism and a correct reading of Bentham's works would instead support an abolitionist stance aimed at replacing punishment-centred systems with social, economic, and relational structures that prevent harm and address it without incarceration. My normative framework will be that of hedonistic act utilitarianism in the tradition of Jeremy Bentham, which interprets the principle of "the greatest happiness of the greatest number" in terms of maximising pleasure and minimising pain, evaluating actions on a case-by-case basis by their consequences for all affected.

Bentham's *Introduction to the Principles of Morals and Legislation* is often considered the foundational text of utilitarianism, presenting the first clear-cut formulation of the principle of utility. The book was intended 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

¹ An earlier exposition of some of the arguments presented here can be found in Chapter 3.3 of my book (Beraha 2025), of which this paper constitutes a revised, updated version.

principle of utility is presented in the *IPML* as a guide for the legislator, specifically with respect to 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 agent as well, since the prospect of the disadvantage inflicted by punishment would offset the expected personal advantage.

Moreover, Bentham himself designed the prison model that inspired many of the first prisons built after the penal reform, especially in the United States: the (in)famous panopticon (Bentham 1995), which is now the symbol of omnipresent oppressive power, of a quest for efficacy that leaves no room for humanity.

However, it should not be forgotten that he was coming from a system of punishment that relied on physical violence and was part of the movement that aimed at changing it. These efforts now seem utterly insufficient (and the panopticon a dystopian *dispositif*), but they were revolutionary at a time when torture, even during interrogation to ascertain culpability, and the death penalty were the norm: in this context, the panopticon could effectively diminish the violence perpetrated within the penal system. Now reminiscent of totalitarian techniques and practices, it was part of the process of softening penalties that had started with Beccaria and the other reformers.

Given this context, the idea of a utilitarian theory of prison abolition might seem less paradoxical: although utilitarianism is rooted in the penal reform that generated the expansion of the prison system and made punishment one of the first and central subjects of its reflection, it really aimed at making it less cruel, minimising the suffering involved. Utilitarianism would, ideally, aim at overcoming punishing practices.

II. Utilitarian Justifications of Punishment and Their Crisis

Nonetheless, utilitarianism has indeed shaped the contemporary debate on punishment in many ways, most notably by introducing its forward-looking justifications. 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 importance” (Bentham 2015, 180) 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 (Bentham 2015, 181).

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 to the crime. Hence, it does not refer to the reintegration of the offender into the social web in a meaningful way, thereby preventing them from reoffending not out of fear of punishment but, like the great majority of their fellow citizens, because they recognise the legitimacy of the law (Tyler 2006). Bentham's reforming tendency corresponds instead to specific deterrence: the greater the punishment the greater the reforming effect. Again, obviously, this goes against the principle of frugality, which is a crucial constraint on punishment in the utilitarian quest for minimising suffering. 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 Beccaria believed, has been disproven).

For Bentham, the central justification for punishment is general deterrence: punishment has beneficial consequences not primarily for those on whom it falls but for "such persons as are under temptation to commit it [the criminal act]: in which case, in as far as it tends to restrain them from committing such acts, it is of a beneficial nature" (Bentham 2015, 147).

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, namely the manipulation of the individual, and therefore must be coupled with the retributive justification in order to avoid it.² 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 carry forward the Enlightenment quest to reduce arbitrariness in the penal system by classifying crimes and punishments clearly, establishing proportional punishments, and instituting laws from which no *ad hoc* derogation could be made (Burns & Hart 2015, lxv).

Although he does not address the problem of the punishment of the innocent explicitly, it is clear that Bentham takes for granted 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 in the long run in terms of utility. This implies a far-reaching utilitarian view that aims to take into account indirect and more distant consequences rather than narrowly focusing on the direct and nearest

² 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 is 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 2008, 8–13).

outcome: a practice that does, say, reduce crime through deterrence or incapacitation but that substantially contributes to the evolution of society in an autocratic and repressive direction cannot be justified on utilitarian grounds.

The same answer given by Bentham is offered by many present-day utilitarians in response to the same criticism, although without much success in dispelling the objection. Others, like J. J. C. 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 in which 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 – only Kant might allow the bomb to go off in order to defend his categorical imperative. This demonstrates the hollowness of such criticism, for there is always a quantitative threshold people are not prepared to cross in order to uphold 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 claimed, that utilitarians are “happy” to punish the innocent (McCloskey 1963). 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.³ 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 (Smart 1973, 71).

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, rational reflection should override. The unease generated by the possibility of justifying the punishment of the innocent should not be overlooked, but it should not be misunderstood either. It constitutes a warning sign of the problematic nature of the infliction of suffering on others: we should not be happy, as McCloskey suggested, about punishing the innocent – but we could 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 was guilty in order to punish them, but that

³ The fact that taking the course of action that maximises utility is deemed by a fervent utilitarian like Smart as “unjust” is surprising. “Unfair” may be more accurate.

someone had been punished while, in truth, they had not been harmed – or had been harmed much less than the public was led to suppose. 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" (Foucault 1995, 94). 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 (Bentham 2015, 179–180).

Despite all this, the accusation of justifying the innocent's punishment remains a prominent topic in discussions of utilitarian justifications for punishment (see, for example: Sandel 2009, 37–40; Ryberg 2011, 89).

To this theoretical criticism, practical problems have been adding up, as decades of research on the general and specific deterrent effects of incarceration, incapacitation and rehabilitation have yielded very discouraging results (Mosconi 2025, 39–51). Moreover, to justify a practice on utilitarian grounds⁴ it is not sufficient to show that it has some positive impact: first, its positive effects need to be greater than the harms 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 in the U.S. alone, a number that grows to 5.5 million if those under probation and parole are added to the incarcerated population (Wang 2023); it must also not be forgotten that the negative consequences of incarceration affect not only the individual, but also their families. The relatives of incarcerated people, in fact, don't only suffer from the pain of affective deprivation and the disappearance of a practical support element in their family life. They often have to face harsh economic consequences: the loss of a salary, frequently the main, if not the only, source of income for the family; the legal costs; the burden of time spent on bureaucratic requirements in criminal proceedings. Also, people are not rarely incarcerated far from their homes, forcing their relatives to undertake costly journeys to visit them – travels that can be incompatible with the family's economic situation and work schedule. When mothers are incarcerated, in most Western countries their infants and toddlers are sent to prison with them. Finally, social stigma is no less significant and is often underestimated: as an example, in January 2026 the parents of a man accused of killing his wife committed suicide as a consequence of the pressure of public scorn.⁵

4 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 & Hart 2015, lxix).

5 See: https://roma.corriere.it/notizie/cronaca/26_gennaio_25/suicidi-i-genitori-di-carlomagno-in-auto-da-roma-per-morire-ad-anguillara-la-lettera-all-altro-figlio-e-il-nuovo-interrogatorio-

Additionally, 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 jurists to the common idea that “nothing works” (Duff 2011, 64; Martinson 1974; Ryberg 2011, 86) 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, this has not led to questioning the practice of punishment itself. Rather, it has brought retributive justifications back, and they have been, if not dominant, at least central in legal philosophy since the Seventies (Ryberg 2011, 86). Philosophers have started defending them as more *humane* than the instrumentalising utilitarian justifications to the point that Herbert Morris, in his 1968 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 (Morris 1968, 476). These scholarly positions fit well with the popular sentiment that demands that offenders receive “their just deserts.”

One question, at this point, arises: if punishment was practised, both individually and institutionally, long before its utilitarian justifications, and it was justified in a retributive manner, and if the utilitarian justifications have been proved ineffective while the practice of punishment in itself has not been questioned, does this not mean that utilitarianism has served as a sort of cover, a satisfactory 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 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.

III. Utilitarian Abolition

III.1. Bentham's Abolitionist Stance

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 (Bentham 2015, 180, 176) – 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 (Bentham 1962, 534).

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, thus challenging 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 that underlies the “lesser evil” kind of reasoning and taking into account the bigger picture. He does so in the *Principles of Penal Law*,⁶ a text much less well known than the *IPML*, whose last edition dates back to John Bowring’s 19th century collected works of Bentham.

Bentham analyses the topic in the third and last part of the *Principles*, titled *Of Indirect Means of Preventing Crimes* (Bentham 1962, 533). Whereas “direct legislation” concerns punishment and has been addressed in the previous part of the *Principles*, as well as in the *IPML*, indirect legislation covers the actions that can be taken to prevent crimes before any punishment is administered by acting on the factors that make people want to commit crime in the first place (Bentham 1962, 533). Once again, Bentham enters 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 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,”⁷ in which he advocates both welfare-state-type aid and employment support (Bentham 1962, 543–544).⁸ Poverty aside, in the *Principles*, Bentham focuses primarily on education and knowledge sharing (Bentham 1962, 536–540, 563, 575–578), seeming to espouse a moderate ethical intellectualism: the positive consequences of ignorance, deception, and manipulation are only illusory in that they are limited – they may seem to provide the best outcome in a specific situation, but are negative contributions to the overall picture.

After Bentham, the utilitarian discussion on punishment has been chiefly

6 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 2015, 164, my emphasis).

7 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 (2001 and 2010) of *The Collected Works of Jeremy Bentham*, the outcome of the Bentham Project’s (UCL) ongoing effort to publish all of Bentham’s writings.

8 Their proposed implementation sounds 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 cannot always be transposed unmodified (and sometimes cannot be transposed at all) to the present day.

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 variant, shouldn't the intentional infliction of suffering, i.e., punishment, be avoided as much as possible? It is sometimes said that punishment should be considered a last resort, *extrema ratio*, when everything else has proven ineffective. However, its effectiveness compared to the other ineffective measures must also be assessed, whereas *extrema ratio* arguments tend to take it for granted. Utilitarians, as Bentham has shown, should be the first to consider alternatives to punishment, but the path he had traced has remained unexplored in the utilitarian field. Bentham's position is not only consistent with present-day prison abolitionists' demands; it anticipates them. Abolitionists, on the other hand, have neglected engaging with a utilitarian moral framework. It is therefore time to examine the abolitionist stance and assess its compatibility with hedonistic act utilitarianism.

III.2. Prison Abolition: Roots and Limitations

The prison abolition movement emerged in recent decades, proposing a different paradigm, aiming at abolishing the institution of prison and developing alternative responses to harm. Incarceration is not equivalent to 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: no prison abolitionist advocates for other kinds of institutionalised punishment.

Prison abolition was initially theorised by anarchist thinkers in the 19th century (for a comprehensive overview of anarchist thinking on prison abolition, see: Nocella, Seis & Shantz 2020). 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 U.S.,⁹ has lost that wide breadth. As many abolitionist scholars aim at changing practices of justice, it is vital that their research tackle specific contexts and their peculiarities (e.g., McLeod 2019), and research with a wider approach cannot substitute for the valuable work they are doing. However, the two approaches should complement each 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.

⁹ A notable exception is European abolitionist sociology, embodied by Norwegian Nils Christie (e.g., Christie 1977) and Dutch Louk Hulsman (e.g., Hulsman 1991), whereas the subject has been completely neglected by philosophers.

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 in the aftermath of the American Civil War, which resulted in the presumptive abolition of slavery, as 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 who were victims of the convict-leasing system were often far worse than they had been when slavery was in force (Davis 2003, 32). Angela Davis wrote:

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 (Davis 2003, 25).¹⁰

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. 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 in racism. Morally, we would need prison abolition even if incarceration rates perfectly mirrored the ethnic, religious, and financial 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 most basic feature, that of the belief that some good is achieved through the voluntary infliction of suffering, whether the backwards-looking good of retribution or the forward-looking good of utility. Present-day abolition thinking lacks universal breadth.

III.3. Framing Abolition: Utilitarianism as a Driver of Moral Progress

It should be clear by now that I think hedonistic act utilitarianism offers here the perfect solution. Utilitarianism has been so closely linked with the justifications for punishment for accidental reasons rather than because of any intrinsic punitive character of the normative framework: the first consequentialist justifications for punishment were elaborated in a specific context, that of the reform that led 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

¹⁰ Similar positions can be found in many abolitionist thinkers (e.g., Bell 2021, 45, 61; McLeod 2019, 1617; Law 2021, 49–55).

subsequent utilitarian discussion on crime and punishment. This could be partly explained by the *Principles*' publishing vicissitudes, whereas the *IPML*, which treats direct legislation and punishment justifications at length, 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, the retributive system of punishment, that was already in place; second, because, as Bentham himself remarked, punishment is a simple, intuitive solution (Guyau even called it a "popular remed[y], like boiling oil into which (...) the limbs of the wounded were plunged"; [Guyau 1891, 172]) – 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 our feelings more than our rational analysis.

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 breadth. 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, in his famous stances against the criminalisation of homosexuality (Bentham & Crompton 1978), opposing slavery, racism, and even the abuses perpetrated against non-human animals, making him, to all intents and purposes, the father of contemporary antispeciesism (Bentham 2015, 283, n. b). Utilitarian ethics is based on sentience, a characteristic that we share among varying ethnicities and species, as well as across different life choices and pathways. The suffering of people who committed crimes has, from 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 morality in the promotion of pleasure and the prevention of suffering, represents the logical normative theory within which to frame abolition. Utilitarianism, as we have seen, has been focused on crime prevention through deterrence, conceiving punishment as a necessary lesser evil. On the other hand, abolition thinking questions the idea that this evil really is necessary at all and aims at harm prevention through a restructuring of society that reduces the conditions that lead to violence, just like Bentham himself.

This revolutionary role is one that 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. This is not to say that people do not tend to think about their moral duties in deontological terms, 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. 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.

Common sense morality may, hence, be a kind of lazy utilitarianism, consisting in following rules that are expected to have positive overall consequences, albeit on the basis of intuitions and custom rather than the meticulous analysis required by act utilitarianism, and allowing exceptions in cases where there is clear-cut evidence that waiving the rule will produce positive outcomes. 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.

Utilitarians have also been concerned with emphasising the continuity between their normative framework and common sense because they had to counter the dystopian accusation of their critics. The utilitarian ideal world, in fact, is often depicted as a world of “secrecy, deception and manipulation” (Primoratz 1991, 390), 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 (Sidgwick 1907, 489). 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 (de Lazari-Radek & Singer 2010, 47). 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 utilitarians’ tendency to conservatism – and the general population’s wariness towards utilitarianism – is the fact that, in order to promote the utility of the greatest number, one needs to believe that they know where that

utility may lie. This assumption generates some legitimate vigilance, for it quickly evokes ethical states that intrude into their citizens' lives to impose moral values and impede self-determination. 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" (Narveson 2019, 159), then to think of it as a moral framework that promotes imposition and manipulation means completely misunderstanding it.

Nevertheless, the need to know the other's utility is a problem for utilitarians and has, once again, pushed them 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 we'd better support traditional liberal stances rather than allow intrusion into people's choices on the assumption that policymakers know their preferences and utility.

Utilitarian policymaking seems to be the most problematic: 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, negative hedonistic utilitarianism – aiming at reducing pain rather than maximising pleasure – may constitute the preferable framework, being 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.

The famous critique levelled against negative utilitarianism by Roderick N. 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 extinction (Smart 1958). However, I only defend negative utilitarianism as the safest normative framework for policymaking, as another rule of thumb that may be waived but serves as the most efficient guiding principle in a specific context, avoiding the slippery slope of ethical states. If the State's objective is not to positively maximise utility but to negatively minimise disutility, the risk of intrusion into private life and personal choices is reduced, as the State's action tends rather to alleviate hardships – which are more easily identifiable than utilities – thus supporting the welfare state and social assistance. And if the State's aim is to minimise suffering, it is evidently paradoxical that it should voluntarily inflict suffering on its citizens with a punitive system.

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 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 that the content of these rules may frequently remain unchanged under a utilitarian perspective, 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. Utilitarians have been focused on rejecting the dystopian 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 and punishment, the basic intuition underlying them 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. The question that looms over abolition, once indirect legislation to reduce crimes is put in place, of course is: what do we do with offenders? However successful indirect legislation might be in preventing offences, the prospect of eradicating all forms and occasions of violence is clearly unrealistic. The issue is broad and complex; I have explored it in another publication (Beraha 2025) and will therefore confine my concluding remarks to outlining the main direction for addressing it.

IV. Conclusions and Future Directions: The Restorative Paradigm

Whereas the carceral solution to the problem of violence, which is a major form of social bond breakdown, further accentuates the disconnection of the offender from society (in which they are supposed to be, eventually, reintegrated), an abolitionist proposal should aim at social connection. Restorative justice has emerged, since the Truth and Reconciliation Commission in post-apartheid South Africa, as the principal tool for responding to harm, respecting the survivor’s need for truth, understanding, and reparation (which, by the way, are not respected at all in traditional criminal prosecution) while also not persecuting the offender.

Restorative processes are based on the acceptance that the people involved in the harmful situation hold different and legitimate narrative truths about the event, which forms the basis for investigating the conditions that gave rise to it, all within a dialogical process that, by its very nature, creates new forms of social bonds (see, e.g., Mazzucato 2015; Ceretti 2015; Mosconi 2025, 181–210). Restorative justice maintains the symbolic and expressive aspect of punishment, constituting a form of institutionalised response to harm that signals censure of the act and attention to the offended persons’ needs, while rejecting the idea that these aims can only be achieved by intentionally inflicting pain on the offender, therefore radicalising the utilitarian principle of frugality: no punishment is needed for general deterrence. Its details, its supposed deterrent effect, its transformative potential and its

overall positive consequences, however, must be left to further investigation.

References

- Bell M. 2021. "Abolition: A New Paradigm for Reform," *Law & Social Inquiry* 46(1):32–68.
<https://doi.org/10.1017/lsi.2020.21>
- Bentham J. 1962. "Principles of Penal Law," in J. Bowring (Ed.), *The Works of Jeremy Bentham* (pp. 365–358). New York: Russell & Russell.
- Bentham J. 1995. *The Panopticon Writings*. Edited by Miran Bozovic. London and New York: Verso.
- Bentham J. 2015. *An Introduction to the Principles of Morals and Legislation*. Edited by J. H. Burns and H. L. A. Hart. Oxford: Oxford University Press.
- Bentham J. & Crompton L. 1978. "Offences Against One's Self," *Journal of Homosexuality* 3(4):389–406. https://doi.org/10.1300/J082v03n04_07
- Beraha M. 2025. *Ripensare la giustizia. Oltre il modello retributivo*. Roma: Carocci.
- Burns James H. & Hart H. L. A. 2015. "Introduction," in H. L. A. Hart & J. H. Burns (Eds.), *An Introduction to the Principles of Morals and Legislation* (pp. xxxi–lxix). Oxford: Oxford University Press.
- Ceretti A. 2015. "Per una convergenza di sguardi. I nostri tragitti e quelli della commissione per la verità e la riconciliazione sudafricana", in G. Bertagna, A. Ceretti, & C. Mazzucato (Eds.), *Il libro dell'incontro. Vittime e responsabili della lotta armata a confronto* (pp. 219–250). Milano: Il Saggiatore.
- Christie N. 1977. "Conflicts as Property," *The British Journal of Criminology* 17(1):1–15.
<https://doi.org/10.1093/oxfordjournals.bjc.a046783>
- Davis A. Y. 2003. *Are Prisons Obsolete?* New York: Seven Stories Press.
- Duff R. A. 2011. "Responsibility, Restoration, and Retribution," in M. Tonry (Ed.), *Retributivism Has a Past: Has it a Future?* (pp. 64–85). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199798278.003.0004>
- Foucault M. 1995. *Discipline and Punish. The Birth of the Prison*. New York: Vintage Books.
- Guyau J.-M. 1891. *A Sketch of Morality Independent of Obligation and Sanction*. London: Watts & Co.
- Hart H. L. A. 2008. *Punishment and Responsibility. Essays in the Philosophy of Law*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199534777.001.0001>
- Hulsman L. 1991. "The Abolitionist Case: Alternative Crime Policies," *Israel Law Review* 25(3–4):681–709. <https://doi.org/10.1017/S0021223700010694>
- Law V. 2021. *"Prisons Make Us Safer" And 20 Other Myths About Mass Incarceration*. Boston: Beacon Press.

- Lazari-Radek K. de & Singer P. 2010. "Secrecy in Consequentialism: A Defence of Esoteric Morality," *Ratio* 23(1):34–58. <https://doi.org/10.1111/j.1467-9329.2009.00449.x>
- Martinson R. 1974. "What Works? Questions and Answers about Prison Reform," *The Public Interest* 35:22–54. URL: https://www.nationalaffairs.com/public_interest/detail/what-works-questions-and-answers-about-prison-reform (accessed on February 27, 2026).
- Mazzucato C. 2015. "La giustizia dell'incontro. Il contributo della giustizia riparativa al dialogo tra responsabili e vittime della lotta armata", in G. Bertagna, A. Ceretti & C. Mazzucato (Eds.), *Il libro dell'incontro. Vittime e responsabili della lotta armata a confronto* (pp. 251–304). Milano: Il Saggiatore.
- McCloskey H. J. 1963. "A Note on Utilitarian Punishment," *Mind* 72(288):599. <https://doi.org/10.1093/mind/LXXII.288.599>
- McLeod A. M. 2019. "Envisioning Abolition Democracy," *Harvard Law Review* 132(6):1613–1649.
- Morris H. 1968. "Persons and Punishment," *The Monist* 52(4):475–501. <https://doi.org/10.5840/monist196852436>
- Mosconi G. 2025. *Decostruire la pena. Per una proposta abolizionista*. Milano: Meltemi.
- Narveson J. 2019. *Morality and Utility*. Baltimore: Johns Hopkins University Press.
- Nocella A. J. II, Seis M., & Shantz J. 2020. *Classic Writings in Anarchist Criminology: A Historical Dismantling of Punishment and Domination*. Chico: AK Press.
- Primoratz I. 1991. "Utilitarianism, Justice and Punishment: Comments on Smart and Flew," *Israel Law Review* 25(3–4):388–397. <https://doi.org/10.1017/S0021223700010499>
- Ryberg J. 2011. "Punishment and Desert-Adjusted Utilitarianism," in M. Tonry (Ed.), *Retributivism Has a Past: Has It a Future?* (pp. 86–100). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199798278.003.0005>
- Sandel M. 2009. *Justice. What's the Right Thing to Do?* London: Penguin Books.
- Schofield P. 2019. "«The First Steps Rightly Directed in the Track of Legislation»: Jeremy Bentham on Cesare Beccaria's Essay on Crimes and Punishments," *Diciottesimo Secolo* 4:65–74.
- Sidgwick H. 1907. *The Methods of Ethics*. London: MacMillan.
- Smart J. J. C. 1973. "An Outline of a System of Utilitarian Ethics," in J. J. C. Smart & B. Williams, *Utilitarianism. For and Against* (pp. 1–74). Cambridge – New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511840852.001>
- Smart R. N. 1958. "Negative Utilitarianism," *Mind* 67(268):542–543. <https://doi.org/10.1093/mind/LXVII.268.542>
- Tyler T. 2006. *Why People Obey the Law*. Princeton: Princeton University Press.
- Wang L. 2023. "Punishment Beyond Prisons: Incarceration and Supervision by State," *Prison Policy Initiative*, May 2023. URL: <https://www.prisonpolicy.org/reports/correctionalcontrol2023.html> (accessed on February 27, 2026).