Quest Journals
Journal of Research in Humanities and Social Science
Volume 13 ~ Issue 10 (October 2025) pp: 01-06
ISSN(Online):2321-9467
www.questjournals.org

Research Paper

Retributivism vs. Utilitarianism: Rethinking Punishment in Modern Society

Qianqian Zhao
(Middlesex School)

ABSTRACT: *This essay contrasts retributivism, which seeks to punish offenders according to what they morally deserve, with utilitarianism, which justifies punishment by its social benefits. It argues that retributivism is impractical because moral responsibility cannot be measured with precision, making proportionate punishment inevitably arbitrary and prone to bias. Utilitarianism, especially in its rule-based form, avoids this flaw by focusing on empirically observable outcomes, and by adapting punishment policies to evidence of what best promotes overall well-being. While retributivism captures an important intuition about fairness, the essay concludes that a rule-utilitarian approach provides a more reliable and ethically defensible foundation for modern criminal justice.*

KEYWORDS: *Justification for Punishment, Retributivism, Utilitarianism*

*Received 25 Sep., 2025; **R**evised 03 Oct., 2025; **A**ccepted 05 Oct., 2025 © The author(s) 2025.*
P**ublished with open access at **www.questjournas.org

I.INTRODUCTION

When a society is deliberating how it should punish an offender, it confronts an inevit-able choice: is punishment a matter of giving people what they deserve, or is it about producing the best consequences for everyone? This was more than a matter of abstract philosophy, but affected every facet of life, from sentencing policy and prison regulations to public confidence in the legal system. The battle between retributivism and utilitarianism has endured for centuries due to the fact that each seems, on its face, able to capture something that is crucial about justice. The retributivist doctrine taps into our moral intuition that wrongdoers must receive punishment that commensurates the wrong they committed, a sense of fairness that we have that is deeply based on reciprocity. Utilitarianism meanwhile aims to reduce future harm to the society, which are principles particularly relevant to the practical issues of crime control.

The argument against retributivism matters because how we frame our system determines the core moral ground on which we have established our judicial system. Grounding the legal philosophy in a retributivist model would require the court to decide how responsible someone is for the crime they committed. But utilitarianism has a wholly different way of thinking about punishment, one that punishes by focusing on actual consequences. Both views have to deal with constraints, particularly the reality that human behavior is complex and forms under conditions often beyond an individual's control.

What is so powerful about this debate is that it forces us to consider not only what is effective in punishment, but also what is ethically permissible. By placing these two strategies in discussions, we can discuss the philosophy behind, the reality of our penal choices, and whether justice can be conceptualized solely under one theory or another. It is here argued that while retributivism offers a conceptually coherent vision of justice by punishing wrongdoers according to their moral desert, its reliance on accurately measuring responsibility makes it unworkable in practice; in contrast, an utilitarian approach, grounded in the promotion of societal well-being, avoids this epistemic flaw and provides a morally defensible basis for punishment in contemporary society.

II.RETRIBUTIVISM

In retributive theory, deserts mean that wrongdoers should get a punishment that is proportional to the crime. This idea of "just deserts" is a reflection of the moral instinct that punishment is not simply a means of deterrence, or rehabilitation but a return on a debt of wrongdoing. Retributivists believe that punishment rights moral scales by giving each person his or her just deserts and nothing more and nothing less. It is not a matter of looking ahead to future results, but of looking back to past acts and the offender's inherent moral culpability.

Philosopher Michael Moore expands on this by stating, "Retributivism is the view that we ought to punish offenders because and only because they deserve to be punished." According to Moore, other outcomes of punishment, such as deterrence, incapacitation, education, or social cohesion, are merely "a happy surplus" that may follow punishment but do not justify it. "For a retributivist," he explains, "deserving offenders should be punished even if the punishment produces none of these other, surplus good effects." Retributivists argue that justice should incline no more to one side than to another. Immanuel Kant famously argued that "if you insult him, you insult yourself; if you steal from him, you steal from yourself; if you kill him, you kill yourself." Therefore, according to retributivism, what one deserves is receiving the pain they imposed on others, with the specific punishment decided by the court.

III.THE PHILOSOPHICAL FLAWS IN RETRIBUTIVISM

The problem with the successful use of just deserts is that we need to know exactly how responsible someone really is for their actions, and that assumption collapses when confronted with the realities of human life. Responsibility isn't a set of scales; it can't be measured like weight or distance. Rather, it is a profoundly intricate mix of being placed in certain positions that gives one set of options, perceptions, and motivations. Think about a child growing up in abject poverty without education or social support, where crime is essentially the only means of survival. This person's agency is massively thwarted by forces of reality; and though it would be too simple to say that the person has no moral agency at all, it would be equally simple to take the moral responsibility of such a one to be anywhere near the level of responsibility as someone who was comfortable and aware of their alternates. Indeed, as criminological research demonstrates, childhood adversity, which includes neglect and abuse, is strongly associated with future criminal behavior, meaning that what we call culpability frequently reflects conditions concreted in place long before the act itself. Even issues of mental health complicate the equation, as some disorders interfere with impulse control and risk assessment in a way that effectively obfuscates the distinction between voluntary choice and compelled response.

A child matures in a rich family, and is continuously taught by his parents and fellow competitors that to succeed in personal terms is the supreme good. As a result, he is also subject to relentless psychological pressure. Everything he does is held against how well he might do in the future, and every failure a mountain to be scaled. Eventually he gets paranoid and addicted to drugs to stimulate his performance. When he is caught stealing from his employer, the superficial story line is clear: a smart, adult man who knew better did something illegal to enrich himself. But when you factor in those formative years in which his worth was inextricably bound to achievement, and the pathological anxiety fueling his control issues, the line that separates choice from compulsion becomes more blurry. Who is to take the blame? He was mentally tortured by his family and friends, and this psychological pressure compelled him to do what he did. Not that it excuses his behaviour or removes his responsibility, but it disputes our ability to allocate a very exact quantum of blame. His guilt is bound up in a mental architecture constructed over the course of decades, and that architecture was influenced by factors he never had a hand in choosing.

A proportionate penalty for misdeeds rests on an estimate of moral blame, but our estimations are inevitably clouded by incomplete information and confederate factors. Although courts will try to balance the mitigating and aggravating factors, those judgments rest on narratives assembled from limited evidence, subject to biases in perception, and often constrained by procedural rules that exclude relevant context. Even if one has the facts, it is an interpretation to make them into a scalar of moral fault, and there is no objective measure for that. Any effort to proportion punishment to it will be by nature imprecise at best and arbitrary at worst. This arbitrariness does not disappear with experienced judges or prescriptive guidelines, because the problem is epistemic rather than procedural. We cannot see into a person's formative history and mental state with the clarity that would be necessary for fine-tuned proportions. Furthermore, social injustices cause the error to fall unevenly; those from marginalized groups are both more likely to be deemed fully responsible and to also be the people whose life circumstances most plausibly diminish that responsibility. The rich defendant can have his views of his character heard more charitably, but the poor defendant is subject to bias.

The mismatch between the retributivist ideal and the epistemic reality means that the doctrine, while coherent in the abstract, fails as a practical guide in our present world. Retributivism may be conceptually sound in that wrongdoers should not be punished more than they deserve, yet in practice the desert cannot be fixed with confidence because the current technologies and rules for identification make it impossible to accurately measure how much a person is truly responsible for the crime. Any system that purports to measure out to each his or her just deserts but operates based on such uncertain estimates risks putting a stamp of authority on punishment levels that actually overstate or understate the significance of the acquitted conduct, converting what should be a principle of fairness into an instrument of injustice.

So long as we lack a significantly richer and more exact analysis of human agency capable of drawing a reliable distinction between what one elects to do and what one is forced or influenced to do as result of the circumstances of one's life, the idea of just deserts in retributivism will not function in the society. The moral

calculus that just deserts demands will be unsolvable because the inputs are not fully known, and perhaps are not even fully knowable. Until we are able to objectively assess the level of moral responsibility involved in any single act, the ideal of assigning punishment solely consistent with desert will remain impossible.

IV.UTILITARIANISM

The utilitarian view of punishment sees punishment, not as a matter of retribution, but as a method of achieving beneficial consequences for society at large. Punishment is justified if and because it is effective and produces better collective well-being. This makes the theory forward-looking. Its question is not, "What pain does this person deserve for what they did?" but rather, "How will the reaction most prevent future harm and bring stability to the community? In utilitarian thinking, punishment has only instrumental value: It matters only insofar as it sends useful signals.

Punishment can work in several ways to produce these salutary effects. The first is that of deterrence, since the threat of punishment dissuades the offender as well as others from doing the same act. Another is incapacitation: taking a dangerous person off the streets or out of a home where they might hurt someone protects people in the short term. The third is rehabilitation: a sanction can be used to change an offender's behavior, habits, or perhaps give the person skills or support to live without further offending. There is also the reinforcement of societal norms, since a public response to wrongdoing reaffirms the rules that hold a community together and communicates that certain harms will not be tolerated. These functions can overlap, and different crimes might require various combinations of them.

Utilitarianism avoids the problem highlighted in the just deserts form of retributivism. In retributivist reasoning, the moral weight of the punishment might hang on the degree to which the person was responsible for his act. If this criterion is mistaken, the punishment is mistaken. And this is precisely where the retributivist model fails in application. We do not have the kind of instruments that would allow us to measure responsibility with accuracy, because human behavior is determined by a complex web of psychological and social determinants, most of which are beyond the person's control. Utilitarianism avoids this problem by grounding punishment in observable and predictable outcomes rather than in an exact calculation of moral blame. Accountability still matters in the sense that punishing people who did not act badly is wrong and undermines the rule of law. But utilitarianism does not ask us to distinguish between someone who is seventy percent to blame and someone who is eighty percent to blame for the crime they have committed. Rather, the decision is described in terms of what kind of intervention will minimise the harm in the future.

For example, in determining how much to punish a burglar, a retributivist system asks how much blame the burglar deserves given all the details of his upbringing and his state of mind, and tries to match punishment precisely to desert. In the utilitarian approach the accent is elsewhere. The only thing that matters is, which will protect the public, and prevent future burglaries the better? Would this punishment serve as enough of a deterrent, or would it create greater challenges for reintegration and lead to higher rates of crime over time? The utilitarian answer is whatever has the best chance of making things work better for everyone. As a result, it avoids the impossible question of how much the person is to blame in the first place and instead drives punishment based on its effects, rather recognizing that the causes of human actions are multiple and frequently elusive.

Utilitarianism is also able to address the root causes of crime. If some communities bear structural disadvantages making them more likely to offend, retributivism might just punish the offenders without putting anything back for dialectic rectification of those disadvantages. Utilitarianism, however, takes such factors to be directly relevant to the issue of future harm. If good education, mental health care or economic opportunity can prevent the commission of a crime, then those are not just social goods but actually part of the moral frameworks for punishment itself.

In short, utilitarianism justifies punishment as a tool for producing the best possible consequences for society. It roots punishments in their effect on the real world rather than a pristine match to moral desert, meaning it's less reliant on the impossible task of measuring how responsible someone is. In a culture in which the roots of wrongdoing are irreparably interwoven with circumstance beyond the individual's control, the moral demands robust utilitarianism makes may well be the only coherent and defensible alternative to retributivism.

V.ADDRESSING OBJECTIONS TO UTILITARIANISM

A major objection to the utilitarian approach to punishment comes from C. S. Lewis. Lewis warns that when punishment is justified solely by its usefulness, such as its ability to deter others, it opens the door to punishing even the innocent. If the goal is to make an example, then it doesn't actually matter whether the person punished committed the crime. What matters, Lewis says, is that "the public should draw the moral, 'If we do such an act we shall suffer like that man.'" So long as people believe the person is guilty, the deterrent effect remains. In fact, Lewis points out that punishing someone who is actually guilty but seen as innocent "will not have the desired effect," while punishing someone who is innocent but *thought* to be guilty could work just as well. Because modern states have the means to "fake a trial," Lewis argues that when an example is urgently

needed, the state could just as easily punish an innocent person, calling it "cure" or "treatment," as long as the illusion of guilt is maintained. This, he insists, becomes possible only when we detach punishment from the concept of desert. "The punishment of an innocent man is wicked," he says, *only if* we believe punishment should be deserved. Once that belief is abandoned, then "all punishments have to be justified…on other grounds," and in some cases, punishing the innocent could be considered just as moral as punishing the guilty. Any discomfort with this, Lewis concludes, is simply "a hang-over from the Retributive theory."

Lewis' imaginary scenario, in which utilitarianism would justify punishing an innocent person for the furtherance of society in general, is generally targeted at act utilitarianism, but does not successfully apply as a legitimate criticism from the perspective of rule utilitarianism. Rule utilitarianism contrasts with act utilitarianism by also considering the consequences of following general rules of justice, promise-keeping, truth-telling, and the like. Whereas act utilitarianism would continue to allow this if the increase in utility were sufficiently higher than it would have been not to punish an innocent, a rule utilitarian test could prevent this by considering the consequences of adopting such a rule, i.e. if the net utility was negative. A human society that accepted a rule that would allow innocent people to be punished would be doomed to lawlessness, unreliability and personal insecurity. These harms could, over time, greatly exceed in value whatever short-term benefit may be derived from deterrence. So rule utilitarians would say that we should adhere to rules against punishing the blameless or punishing people who are not guilty because, generally speaking, those rules produce the most overall good when consistently obeyed over time. This model circumvents the moral pitfall that Lewis cites, and it demonstrates Lewis's thought experiment generalizes utilitarianism by neglecting its more nuanced expressions. From this perspective, punishing the innocent is not only morally objectionable, but it is also self-defeating under rule utilitarianism, because it has systemic consequences that make everyone in a position lower their overall level of happiness, rather than raise it. So, at most, Lewis's case is a good criticism of act utilitarianism, but it does not challenge utilitarianism as a whole.

In addition, while punishing an innocent person might, in theory, produce short-term utilitarian benefits, such as deterring future crimes or reassuring the public that justice is being done, it would almost certainly lead to greater long-term harm, which fundamentally violates utilitarian principles. The core of utilitarianism focuses on making the greatest happiness for the most people, and preventing as much pain as possible. Therefore, any action that produces more harm than good, especially in the broader and sustained view, would be morally impermissible from a utilitarian perspective. If the punishment of an innocent individual was discovered, the public's trust in its justice system would be badly shaken. People trust in the theory that courts and other legal institutions are just, and will protect individual rights. They would stop believing that legal decisions are real, leading to less support for law enforcement, fewer people willing to sit on a jury or testify, and more cynicism in the community about the rule of law.

Moreover, deliberately punishing the innocent might lead to greater social instability and terror. Anxiety and paranoia may grow if people can imagine that they might be accused falsely and suffer punishment for crimes they never committed. The victims of this are more than just those falsely accused; it leads to a chilling effect across society. Such emotional and psychological torment would diminish total happiness: directly opposing the utilitarian focus.

Lewis' second rejection is that of the consequences based on punishment for the sake of society rather than retributivism. Under "the traditional or classical theory of the punishment of the offender was justified in that the individual deserved punishment; his guilt and his desert were determined before he was made an example," he declared. So the punishment was doubly useful: it served as a signal to others and it gave the would-be reprobate their just deserts. But without deserts, there is no moral basis for punishment at all. Without the desert, Lewis wonders, why should anyone be sacrificed or harmed for the greater good of society? Why should an individual agree to be a means of social utility unless they really are deserving of their fate?

To answer Lewis' question, we need to understand that, from a utilitarian perspective, the formation of society has the sole purpose of creating the conditions for the greatest happiness by creating the environment which allows for the safe living of people under predictable conditions, cooperating effectively. Thus, punishment clearly serves as an imperative for maintaining societal norms and preventing the disruption of its stability. If punishing an individual is necessary to uphold the social order and prevent greater harm, it aligns with the original purpose of the social contract. In other words, individuals accept that in certain circumstances, their interests may be subordinated to the collective good because this collective good is precisely what society was formed to achieve. Punishment serves as a deterrent and a means of protecting the rights and welfare of the majority. Punishment helps ensure that society remains a place where individuals' rights and well-being can be safeguarded in the long run. Ultimately, the utilitarian calculation suggests that the punishment of an individual is morally right if it benefits the interests of social stability and the happiness of society.

Finally, people may point out that utilitarianism is subject to the same challenge as retributivism. Utilitarianism's reliance on predicting consequences can lead to serious errors. One might never know if a

presumed punishment can have the good outcome as intended, as it might exacerbate crime by worsening social inequalities.

Nevertheless, this essay maintains that one of the greatest attractions of utilitarianism, and rule utilitarianism in particular, lies in its possibility to predict outcomes based on empirical evidence. In the area of punishment, rule utilitarianism is susceptible to empirical assessment. We can ask whether a rule of punishment, for example, rehabilitative vs. retributive sentencing, tends towards maximization of social welfare. This focus in turn enables the analytical study of punishment policies with robust social-scientific instruments. For example, criminologists can aggregate studies in a meta-analysis and test if particular punishment rules prevent crime and recidivism or foster rehabilitation. These empirical results can serve as strong arguments for or against the continuation of specific punishment rules. Rule utilitarianism does not focus on one-off predictions, but concentrates on long-term evidence of stable rules. For example, meta-analyses show that purely punitive incarceration often fails to reduce reoffending, while rehabilitative programs lower recidivism rates. Similarly, studies on deterrence suggest that the *certainty* of punishment is far more effective than the *severity* of punishment. These insights help policymakers adopt evidence-based punishment rules rather than relying on fallible short-term predictions.

The empiricist basis of utilitarianism makes it flexible and dynamic. If future research shows that a particular punishment rule is not contributing to a decrease in crime, or is actually harming society, then the utilitarian perspective would recommend eliminating such rules. In this way, instead of having to make one-time predictions, it bases moral rules on the best-known evidence and constantly updates rules as evidence becomes more refined.

VI.CONCLUSION

What the essay has shown on the retributivist strategy's reliance on misericord is that the extent to which such a blame must issue in a punishment of precisely the degree that will succeed in censuring wrongdoing is deeply flawed. In theory, giving due punishment for a criminal's deeds may lend an appearance of fairness, but in practice it requires an assessment of moral worth that is beyond our knowledge and beyond our legal institutions. Utilitarianism avoids this problem by grounding punishment in its consequences, focusing on the broader well-being of society, making it more adaptable to the complex realities of human behavior, especially when responsibility is intertwined with social and psychological factors in the environment.

Even if the essay doesn't persuade the readers that the legal system should fully embrace utilitarianism as its only framework, this discussion certainly demonstrates that the desert-based model cannot be our only guide. Real world decisionmaking requires flexibility, and the extreme harshness of pure retributivism risks what real-world intuition would find unjust.

That being said, there are some good reasons not to reject Lewis's thought experiment in this essay. His reasoning is that in a state where an innocent man is knowingly condemned to death means can be found to be just regardless of the circumstances if the guilt is assumed. Although this essay rejects Lewis on the basis that punishing an innocent person will violate an established rule that assures good for the society, one might argue that what if we set up another rule: "not to punish the innocent unless in extreme circumstances like a social upheaval?" However, this is a relatively minor problem, and thus will not undermine the argument for utilitarianism in the essay.

Indeed, for protections for the innocent to work, and for penalties to be pulled into the spotlight to make sure they are justified in relation to the individual's actions, both the retributivist and the utilitarian may need to be given voice in a just system of penalties. These are only a few of the hard cases that a retributive theory cannot satisfactorily come to terms with when it can be used as the only criterion to be followed. Even if we keep our guard up against any potential excesses of utilitarianism, a recognition of its insights permits us to progress towards a model of punishment that is principled but also responsive to the world as it is, not as theory might like it to be.

REFERENCES

[1]. Bentham, Jeremy, and Laurence J. Lafleur. "XIII Cases Unmeet for Punishment." In *An Introduction to the Principles of Morals and Legislation*. Hafner Pub., 1948.

[2]. Glover, Jonathan. *Responsibility*. Routledge & K. Paul; Humanities P., 1970.

[3]. Kant, Immanuel, and Mary J. Gregor. "On the Right to Punish and to Grant Clemency." In *The Metaphysics of Morals*, p.473. Cambridge University Press, 1996. pp. 473.

[4]. Lewis, Clive Staples, *Against the Humanitarian Theory of Rehabilitation.*

[5]. Lipsey, M. W., and Cullen, F. T. (2007). *The effectiveness of correctional rehabilitation: A review of systematic reviews*. Annual Review of Law and Social Science, 3, 297–320.

[6]. Moore MS. Justifying Retributivism. *Israel Law Review*. 1993;27(1-2):15-49.

[7]. Nagin, D. S. (2013). *Deterrence in the twenty-first century*. Crime and Justice, 42(1), 199–263.

[8]. Nakao M, Shirotsuki K, Sugaya N. "Cognitive-behavioral Therapy for Management of Mental Health and Stress-Related Disorders: Recent Advances in Techniques and Technologies." Biopsychosoc Med. 2021 Oct 3;15(1):16. doi: 10.1186/s13030-021-00219-w. PMID: 34602086; PMCID: PMC8489050.

[9]. Tadros, V., 2011, *The Ends of Harm: The Moral Foundations of Criminal Law*, Oxford: Oxford University Press.
[10]. Wilson, James, and Christine Zozula. "Reconsidering the Project Greenlight Intervention: Why Thinking About Risk Matters," November 2, 2011, nij.ojp.gov: https://nij.ojp.gov/topics/articles/reconsidering-project-greenlight-intervention-why-thinking-about-risk-matters.