"If no one is truly responsible for their actions, then there is no justification for punishing people". Discuss.

The assertion that moral responsibility is a necessary precondition for punishment strikes at the heart of legal philosophy. The justification of this as a credible theory most prominently comes from claims that 'punishment' can remain justified even without metaphysical responsibility, provided it serves consequentialist aims and the strongest societal benefits, such as through deterrence or social protection. Yet, it is my intention within this essay to defend this statement, aided by the strongest arguments from hard incompatibilists such as Pereboom, who posits that causal inevitability voids all blame. The attempts to rescue the retributive model of justice only further highlight the flaws within the concept. Despite this, the conclusion will acknowledge that this redefinition would fundamentally alter the character of modern day punishment.

The retributive model of justice rests upon a metaphysical foundation that hard incompatibilism, the view that we cannot be morally responsible for our actions, renders untenable. Retributivism in its purest form maintains that punishment is morally justified because the offender deserves to suffer in proportion to their wrongdoing (Walen, 2023). This desert claim, however, presupposes a notion of libertarian free will that is incompatible with determinism. When we recognise, as Pereboom (2001) compellingly argues, that human actions are ultimately the product of causal chains extending beyond the individual's control- encompassing genetic inheritance, environmental influences, and the inexorable laws of physics- the moral logic of retribution collapses. The offender can no more be said to truly deserve punishment than a hurricane can be said to deserve condemnation for its destructive path. Kleinig's (1998) definition of punishment as "hard treatment" that "involves not merely condemnation, rebuke, or censure, but some substantial imposition on a wrong-doer" delivers the final philosophical blow to retributivism under hard incompatibilism. For if we accept that no individual is truly the ultimate author of their actions, then the intentional infliction of substantial suffering, the very essence of punishment according to Kleinig becomes morally indefensible, reducing it to little more than institutionalised cruelty directed at individuals who ultimately had no more control over their causal determinants than they did over the circumstances of their birth. This conceptual pairing- Pereboom's demolition of desert coupled with Kleinig's unflinching definition of what punishment actually entails leaves retributive justice incoherent. The substantial imposition that Kleinig identifies as punishment's necessary condition becomes precisely what cannot be justified once Pereboom's arguments have severed the link between action and ultimate responsibility.

Yet, the human intuition that punishment is somehow appropriate or necessary in response to wrongdoing remains cognitively entrenched. Strawson's (1962) landmark work on reactive attitudes explores how our moral emotions of resentment and indignation are so deeply ingrained in our psychologies that they feel inescapable, regardless of our philosophical commitments. However, as Königs (2013) astutely observes, the mere fact that we experience these emotions does not constitute a moral justification for institutionalising them. History provides numerous examples of deep-seated emotional responses that we have overcome, such as the paternalistic impulses that once justified oppressive social structures. Society has learned to recognise that raw emotional responses must be tempered by

reasoned ethical principles, and it is necessary this is applied to a retributive model of justice.

This philosophical contention becomes even more astute when we consider contemporary neuroscientific findings. Libet's (1983) famous experiments, along with more recent work by Soon et al. (2008), demonstrate that our brains initiate actions before we become consciously aware of having decided to act. These empirical results strike at the very heart of our folk psychological conceptions of 'volition' and 'choice'. If our experience of decision-making is in fact, neuroscientifically, post hoc rationalisation of neural processes that occur outside our conscious control, then the moral basis for holding individuals accountable in the strong retributive sense evaporates entirely. The implications are profound: our entire criminal justice system may be founded upon what Smilansky (2000) aptly terms an "illusion of desert."

Faced with this retributive impasse, consequentialist approaches attempt to reconstruct a justification for punishment. The strongest argument comes from the deterrence models which argue that punishment can be justified by its crime preventative effects regardless of metaphysical questions about responsibility. Ostensibly, consequentialist approaches like deterrence theory appear to circumvent the challenge posed by hard incompatibilism- if punishment can be shown to effectively reduce future harm, perhaps metaphysical questions about free will become irrelevant to its justification (Johnson, 2019). Yet this superficially appealing solution disintegrates under rigorous philosophical examination. At its core, the deterrence model commits what Kant would identify as the fundamental moral error of instrumentalisation- it transforms persons into mere means for achieving social goods, violating the categorical imperative's most basic ethical constraint (Kerstein, 2024). More devastating still, the model's purely consequentialist logic contains no intrinsic limitation on punitive severity. Theoretically, one could endorse executing petty thieves if empirical evidence suggested this maximised deterrence, or punishing innocent scapegoats if doing so effectively maintained social order. This recto ad absurdum argument exposes the theory's fatal flaw: in its single-minded pursuit of utility, it jettisons the very moral boundaries that make punishment distinguishable from tyranny. The model's inability to generate principled constraints on state power reveals it as philosophically bankrupt, regardless of its practical efficacy.

A sophisticated objection might counter that we can salvage deterrence theory by supplementing it with side constraints, maintaining consequentialist benefits while adding deontological protections against the most egregious violations of individual rights. However, this attempted recovery merely papers over the fundamental contradiction at deterrence theory's heart. Once we admit the need for non-consequentialist limits, we've already conceded that pure deterrence cannot stand on its own moral merits. Moreover, as Nozick's (1974) "utility monster" thought experiment demonstrates, any hybrid model that tries to balance utility with rights inevitably collapses back into one or the other when pushed to its limits. In practice, this means that during times of social crisis or perceived emergency, the utilitarian impulse invariably overwhelms the theoretical protections, leading precisely to draconian outcomes the model claims to avoid. The supplemental constraints function as ad hoc appendages rather than organic outgrowths of the theory's core logic reveal deterrence's inherent inability to provide a robust, internally consistent moral framework that respects persons as agents rather than treating them as manipulable objects in a social equation .

Pereboom (2014) endorses a more sophisticated assertion- the quarantine analogy. By comparing dangerous individuals to carriers of contagious diseases, the model seeks to justify isolation purely on protective grounds without invoking problematic notions of moral responsibility. There is undeniable strength to this approach- it avoids metaphysical debates about free will while providing a plausible public safety rationale for confinement. However, the analogy ultimately fails to fully persuade. Quarantining disease carriers is morally permissible precisely because viruses lack moral status; human beings, by contrast, possess inherent dignity and qualia that demands more robust justification for their restraint. Moreover, the quarantine model cannot account for the distinctly punitive elements that characterise our current justice systems- the deliberate imposition of harsh conditions, the symbolic stigmata of criminal conviction, the expressive function of condemnation (Lavazza et al., 2023). These features, which go beyond mere protective sequestration, remain unjustified under the model.

A critic might object that abandoning punitive frameworks risks undermining the law's expressive function- its capacity to communicate society's moral boundaries (Duff, 2001). If justice merely mitigates harm without condemning wrongdoing, could this not erode the normative fabric that sustains social order? This challenge warrants serious consideration, yet it conflates condemnation with retribution. A reparative system can still affirm communal values through rituals of accountability such as public acknowledgments of harm, victim impact dialogues, and symbolic reparations (Braithwaite, 2002), while avoiding the metaphysical complications of desert. The objection thus reveals not a flaw in hard incompatibilism, but the need to reimagine moral communication in non-punitive terms.

The philosophical untenability of both retributive and consequentialist justifications under hard incompatibilism necessitates a fundamental reconceptualisation of justice- one that transcends the traditional punitive framework altogether. A reparative model emerges as the only theoretically coherent alternative, eschewing the metaphysical fiction of desert while addressing the practical exigencies of social harmony (Zedner, 1994). This paradigm shift reorients justice around three interlocking principles: the restitution of harm to victims (Zehr, 2002), the rehabilitation of offenders through empirically validated interventions (Andrews et al., 2010), and the rectification of systemic conditions that engender criminal behavior (Clear, 2007).

Comparative penological analysis substantiates the superiority of this approach. Norway's correctional system, which privileges rehabilitation over retribution, demonstrates markedly lower recidivism rates than punitive counterparts (Meagan, 2016), recorded to be only around 25% after five years of release (First Step Alliance, 2022) compared to the US's 70% (First Step Alliance, 2021) a disparity that reflects the deterministic insight that behavior is shaped by corrigible environmental and psychological factors rather than autonomous moral failure. Similarly, restorative justice practices, by prioritising dialogue and reparation over adversarial punishment, have proven particularly effective in contexts ranging from juvenile justice to post-conflict reconciliation (CPS, 2023). These empirical successes underscore the practical viability of a justice system that aligns with hard incompatibilism's metaphysical commitments.

The most formidable objection to this model stems not from its practical efficacy but from the visceral human demand for retributive justice- particularly in cases of egregious wrongdoing.

This objection, while psychologically compelling, constitutes what Nietzsche (1954) termed the "metaphysics of the hangman": an atavistic moral psychology that our best philosophical and scientific understandings have rendered obsolete. Just as modern jurisprudence has progressively abandoned other morally problematic instincts, such as trial by ordeal, it must recognise that the retributive impulse, however deeply felt, cannot withstand rational scrutiny when divorced from libertarian assumptions about agency.

Implementing this paradigm would require nothing less than a Copernican revolution in legal thought. The very language of justice would need reformation, replacing the morally laden vocabulary of 'guilt' and 'punishment' with the more precise terminology of 'causal responsibility' and 'harm mitigation' (Hoskins, 2001) . Such conceptual restructuring is not merely semantic pedantry; it reflects the profound philosophical realisation that the law's true function is not to enact cosmic justice through suffering, but to pragmatically address the consequences of determined actions within a deterministic universe.

The central claim- that without moral responsibility, punishment is unjustified- stands unchallenged by any coherent theory of justice. Retributivism collapses when hard incompatibilism severs the link between action and desert, rendering the intentional infliction of suffering morally indefensible. Consequentialist alternatives like deterrence fail equally, either instrumentalising persons or justifying excessive severity. While the quarantine model offers a pragmatic alternative, it cannot account for punishment's inherently condemnatory nature. The persistence of punitive institutions reflects not philosophical necessity but psychological inertia- a reluctance to abandon retributive instincts despite their metaphysical incoherence. A justice system adequate to determinism must transcend punishment altogether, prioritising reparation and prevention over retribution. Therefore, if no one is truly responsible, punishment, as traditionally conceived, has no justification.

Bibliography:
Andrews, D.A. and Bonta, J. (2010) The psychology of criminal conduct. 5th edn. New Providence, NJ: LexisNexis

Braithwaite, J. (2002) Restorative justice and responsive regulation. Oxford: Oxford University Press

Braithwaite, J. (2002). Setting Standards for Restorative Justice. British Journal of Criminology, 42(3), pp.563–577. doi:https://doi.org/10.1093/bjc/42.3.563. (Accessed: 14 April 2025).

Braun, M.N., Wessler, J. and Friese, M. (2021) 'A meta-analysis of Libet-style experiments', Neuroscience & Biobehavioral Reviews, 128, pp. 182–198. doi:10.1016/j.neubiorev.2021.06.018 (Accessed: 14 April 2025).

Chappell, R.Y. (2021) 'Negative utility monsters', Utilitas, 33(4), pp. 417–421. doi:10.1017/S0953820821000169 (Accessed: 8 March 2025).

Clear, T.R. (2007) Imprisoning communities: How mass incarceration makes disadvantaged neighborhoods worse. Oxford: Oxford University Press

CPS (2023) Restorative justice. Available at: https://www.cps.gov.uk/legal-guidance/restorative-justice (Accessed: 14 April 2025).

Duff, R.A. (2020). Punishment, Communication, and Community. [online] Philpapers.org. Available at: https://philpapers.org/rec/DUFPCA-3 (Accessed 14 April 2025).

First Step Alliance (2021) Reducing recidivism: Creating a path to successful reentry. Available at: https://www.firststepalliance.org/post/reducing-recidivism (Accessed: 12 April 2025).

First Step Alliance (2022). Rehabilitation Lessons from Norway's Prison System. Available at: https://www.firststepalliance.org/post/norway-prison-system-lessons (Accessed: 12 April 2025).

Hoskins, Z. and Duff, A. (2001) Legal punishment. Available at: https://plato.stanford.edu/entries/legal-punishment/ (Accessed: 3 March 2025).

Johnson, B. (2019) Do criminal laws deter crime? Deterrence theory in criminal justice policy: A primer. Available at: https://www.house.mn.gov/hrd/pubs/deterrence.pdf (Accessed: 9 March 2025).

Kerstein, S. (2024) Treating persons as means. Available at: https://plato.stanford.edu/archives/sum2024/entries/persons-means/ (Accessed: 1 March 2025).

Kleinig, J. (1998) 'The hardness of hard treatment', in Ashworth, A. and Wasik, M. (eds.) Fundamentals of sentencing theory. Oxford: Clarendon Press

Königs, P. (2013) 'The expressivist account of punishment, retribution, and the emotions', JSTOR. doi:10.2307/24478744 (Accessed: 4 March 2025).

Laird, J. (1931) 'The justification of punishment', The Monist, 41(3). Available at: https://www.jstor.org/stable/27901306 (Accessed: 14 April 2025).

Lavazza, A., Levin, S. and Farina, M. (2023) 'The quarantine model and its limits', Philosophia, 51(5), pp. 2417–2438. doi:10.1007/s11406-023-00683-4 (Accessed: 7 March 2025).

Libet, B., Gleason, C.A., Wright, E.W. and Pearl, D.K. (1983) 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential)', Brain, 106(3), pp. 623–642. doi:10.1093/brain/106.3.623 (Accessed: 6 March 2025).

Markowitz, M. (2016) 'Rehabilitation in the punitive era: The conflicting logics of punishment and treatment in Norwegian prisons', Law & Social Inquiry, 41(1)

Meagan, D. (2016) 'Norway's prison system: Investigating recidivism and reintegration', Bridges: A Journal of Student Research, 10(10). Available at: https://digitalcommons.coastal.edu/bridges/vol10/iss10/2 (Accessed: 11 March 2025).

Murtagh, K. (2013) 'Free will denial and punishment', Social Theory and Practice. Available at: https://www.jstor.org/stable/23562596 (Accessed: 27 February 2025).

Nietzsche, F. (1954) Twilight of the idols, or, how to philosophize with a hammer. Translated by W. Kaufmann, in The portable Nietzsche. New York: Viking Press (Accessed: 14 April 2025).

Nozick, R. (1974) Anarchy, state, and utopia. New York: Basic Books
Pereboom, D. (2001) Living without free will. Available at: https://www.stafforini.com/docs/pereboom_-_living_without_free_will.pdf (Accessed: 1 March 2025).

Smilansky, S. (2000) Free will and illusion. Available at: https://philpapers.org/rec/SMIFWA (Accessed: 14 April 2025).

Snowdon, P. and Gomes, A. (2023) Peter Frederick Strawson. Available at: https://plato.stanford.edu/archives/sum2023/entries/strawson/ (Accessed: 5 March 2025).

Soon, C.S., Brass, M., Heinze, H.-J. and Haynes, J.-D. (2008) 'Unconscious determinants of free decisions in the human brain', Nature Neuroscience, 11(5), pp. 543–545. doi:10.1038/nn.2112 (Accessed: 2 March 2025).

Walen, A. (2023) Retributive justice. Available at: https://plato.stanford.edu/archives/win2023/entries/justice-retributive/ (Accessed: 3 March 2025).

Zedner, L. (1994) 'Reparation and retribution: Are they reconcilable?', The Modern Law Review, 57(2), pp. 228–250. Available at: https://www.jstor.org/stable/1096809 (Accessed: 9 March 2025).

Zehr, H. (2002) The little book of restorative justice. Intercourse, PA: Good Books