

The Wormwood of Penal Substitution and the Miscarriage of Justice

Phil Stilwell

Independent Scholar

This paper advances a focused critique of penal substitutionary atonement by developing a set of novel arguments that highlight its incoherence. The centerpiece is the *Scapegoat Normalization Problem*, wherein penal substitution canonizes one of humanity's most destructive social mechanisms. Additional arguments explore the impossibility of transferring desert, the futility of voluntariness as a justification, the Bayesian implausibility of a justice model rejected across legal traditions, the confusion between civil and criminal frameworks, the epistemic costs of biblical contradictions, and the historical-genealogical contingencies of the doctrine, alongside cross-cultural justice contrasts. Each section includes a formal syllogism to clarify the inferential structure. Drawing from anthropology, legal philosophy, biblical studies, and comparative law, the paper contends that penal substitution is not merely unjust by human standards but conceptually self-defeating as a model of divine justice.

Introduction

Penal substitutionary atonement has been central in many streams of Western Christian theology and remains a live point of contention in contemporary apologetics and analytic philosophy of religion. It figures prominently in Reformation thought, shapes homiletic practice, and continues to frame major debates about the nature of divine justice

and forgiveness. Its coherence therefore matters both theologically and practically: if penal substitution fails as a model of justice, then central claims about God's character and redemptive action lose their explanatory footing in precisely the domain—justice—where clarity is most needed.

The thesis of this paper is straightforward: *penal substitutionary atonement is incoherent as a model of divine justice because it sanctifies scapegoating, severs punishment from agency-grounded desert, mistakes voluntariness for culpability, and contradicts stable cross-cultural constraints on criminal liability while resting on contested and historically contingent theological premises.*

The argument proceeds in seven parts. First, the *Scapegoat Normalization Problem* shows that penal substitution elevates scapegoating to a divine principle. Second, the *Transferred Desert* critique demonstrates that desert cannot be relocated without destroying its meaning. Third, the *Futility of Voluntariness* shows that willingness cannot generate guilt or justify punishing the innocent. Fourth, a *Bayesian* analysis explains why, given the strong prior against vicarious criminal punishment, the doctrine remains implausible. Fifth, the paper clarifies why appeals to civil analogies misconstrue criminal law. Sixth, it considers biblical tensions and their epistemic costs. Seventh, it sketches the historical genealogy of the doctrine and contrasts it with cross-cultural justice norms before concluding with implications for theological method and legal-moral reasoning.

The Scapegoat Normalization Problem

Scapegoating has long been recognized as a primitive but destructive mechanism for preserving group cohesion. When crises threatened a community, blame was often displaced onto an outsider, an animal, or a marginalized individual, who was then punished or expelled to relieve tension (Douglas, 2002; Girard, 2005). This ritual provided temporary catharsis but failed to address the true sources of harm, allowing cycles of violence to

persist.¹ In many societies, scapegoating was understood as provisional and, over time, was gradually exposed as unjust (Heim, 2006).

Penal substitutionary atonement, however, does not merely echo scapegoating—it enshrines it as the very foundation of divine justice. What was once a temporary and arbitrary human practice becomes, in certain forms of Christian theology, a permanent and universal principle: guilt is shifted from the guilty onto the innocent. This represents not an elevation of justice, but a regression beneath the progress achieved by many legal and cultural traditions. Whereas restorative models emphasize direct accountability, substitutionary atonement canonizes the logic of blame-shifting (Levine, 2020; Zehr, 2002).

The psychological consequences are significant. By making guilt transferable, penal substitution interrupts the development of conscience: an offender learns that responsibility can be shifted away, eroding the link between agency and accountability. When divine justice itself functions by punishing the innocent in place of the guilty, human responsibility loses its anchoring.

Culturally, legitimized scapegoating does not remain confined to theology. Witch trials, pogroms against Jewish communities, and colonial blame-shifting have been justified in contexts where substitutionary logics were valorized (Carter, 2011; Levine, 2020). Far from dismantling scapegoating, penal substitution risks sanctifying it as a normative pattern of divine–human relations, thereby reinforcing its social outworkings in persecution and violence.

The theological tension is acute. Some interpreters influenced by Girard argue that the crucifixion unmasks scapegoating by revealing its injustice. Yet penal substitution can be read to invert this insight: instead of exposing scapegoating as a human distortion, it can

¹ For an account of how myth and ritual can obscure the innocence of victims within scapegoating systems, see Girard (1987). This footnote is descriptive.

appear to canonize it as God's means of justice. If divine justice is modeled on scapegoating, then the very system meant to redeem humanity perpetuates one of its most destructive patterns.

In contrast, restorative justice traditions—whether in Māori, Navajo, or contemporary frameworks—emphasize accountability, restitution, and reintegration (Braithwaite, 1989; Zehr, 2002). These approaches insist that culpability cannot be transferred: the offender must confront and repair harm. Penal substitution, by contrast, elevates scapegoating to the highest theological principle.

Logical Core (Syllogism):

P1: Scapegoating is unjust because it punishes innocents while leaving offenders unaccountable.

P2: Penal substitution canonizes scapegoating as divine justice.

C: Therefore, penal substitution canonizes an unjust practice as divine justice.

Note. Scapegoating punishes an innocent to relieve tension and is rejected across mature legal and moral traditions; making it central to divine justice undermines accountability.

The Incoherence of Transferred Desert

A central difficulty concerns the very notion of *desert*. In legal and philosophical accounts of justice, desert is inseparable from agency: it arises from an individual's intentional action, causal role, and moral awareness (Duff, 1986; Feinberg, 1970; Moore, 1997). To say that punishment is deserved is to say that the offender, by virtue of their conduct, merits a response in proportion to guilt. Once desert is detached from the agent, its meaning dissolves.

Penal substitution presumes that desert is transferable. On this view, humanity's guilt can be imputed to Christ, and divine punishment can be discharged upon him. Yet this

treats desert as though it were a fungible property—something detachable and reassignable. As Murphy and Hampton argue, desert is anchored in the moral relationship between the offender and the victim; it cannot be migrated without severing that relationship (Murphy & Hampton, 2003).²

Defenders sometimes appeal to Christ's willingness to bear this desert. But voluntariness cannot manufacture culpability. If an innocent person volunteers to serve a murderer's sentence, their willingness does not transform them into the murderer. Consent may affect whether a harm is wrongful, but it does not generate desert where none exists. Grounding justice in transferred desert renders the concept incoherent.

Logical Core (Syllogism):

P1: Desert arises from an offender's agency, intent, and causal role.

P2: These features cannot be transferred to another person.

P3: Penal substitution requires transferring desert to an innocent.

C: Therefore, penal substitution renders the concept of desert incoherent as a theory of justice.

Note. Desert connects a wrongdoer to their own act; treating it as transferable evacuates punishment of its meaning.

The Futility of Voluntariness

A prominent defense of penal substitution emphasizes that Christ willingly chose to suffer in humanity's place. On this account, voluntariness transforms what would otherwise be an injustice into an expression of love. Yet voluntariness does not generate culpability, create desert, or convert punishment of the innocent into justice.

In ethical and legal reasoning, voluntariness alters the meaning of suffering but not

² Feinberg emphasizes that desert is not an object that can be transferred; it is a relational assessment linking an agent to an act (Feinberg, 1970). This footnote is descriptive.

its moral target. Organ donation and military service illustrate this distinction: voluntariness makes such acts non-coercive, but it does not create wrongdoing where none occurred. If an innocent party volunteers to endure another's prison term, their willingness may make the act sacrificial, but it cannot transform them into the offender. The crucial link between agency, intent, and accountability remains missing (Moore, 1997; Rawls, 1971).³

This exposes a categorical mistake: the conflation of sacrifice with punishment. Sacrifice may be voluntary, but punishment is by definition imposed in proportion to guilt. Once voluntariness is treated as sufficient, punishment and sacrifice collapse into one another. The result is conceptual confusion: punishment without guilt is not punishment at all, even if the person accepts it (Murphy & Hampton, 2003).

Socially, a system that allowed volunteers to assume others' penalties would erode accountability. Actual offenders could evade responsibility if willing substitutes could be found. Justice would be reduced to a lottery of volunteers rather than a system anchored in responsibility and deterrence. No functioning criminal justice system operates on such terms, and for good reason: voluntariness cannot bear the normative weight of desert.

Logical Core (Syllogism):

P1: Punishment is justified only when directed at the guilty.

P2: Voluntariness cannot create guilt or desert.

P3: Penal substitution treats voluntariness as sufficient to justify punishing the innocent.

C: Therefore, penal substitution collapses punishment into sacrifice and empties justice of its meaning.

Note. Choosing to suffer can be noble, but it does not make one guilty; punishment

³ Rawls distinguishes voluntariness from coercion but does not treat voluntariness as generating culpability; responsibility remains tethered to the agent who acted (Rawls, 1971). This footnote is descriptive.

must track guilt to remain just.

The Bayesian Implausibility of Penal Substitution

If penal substitution were the core of divine justice, we would expect either: (a) strong independent reasons to believe that justice permits punishing innocents for the guilty, or (b) overwhelming evidence that divine revelation uniquely mandates this outlier model. We have neither. Across legal traditions, criminal liability is consistently tethered to the person who *acted* with culpable intent; vicarious criminal punishment is rejected in principle (Duff, 2001; Feinberg, 1970; Hart, 1968). This furnishes a powerful prior against any theory that makes transferable culpability central to justice.

Let H_1 be the hypothesis that divine justice respects the nontransferability of culpability (convergent with robust human justice constraints). Let H_2 be the hypothesis that divine justice centrally endorses punishing the innocent for the guilty (penal substitution). Let E be the total evidence available (scriptural ambiguity, doctrinal diversity, and the persistent rejection of vicarious criminal punishment in actual legal systems). Bayes' theorem gives the posterior odds:

$$\frac{P(H_2 | E)}{P(H_1 | E)} = \frac{P(E | H_2)}{P(E | H_1)} \times \frac{P(H_2)}{P(H_1)}.$$

Given the near-universal rejection of vicarious criminal punishment, the prior odds $\frac{P(H_2)}{P(H_1)}$ are small. Moreover, E does not strongly favor H_2 over H_1 : scriptural testimony is internally contested, and historical theology displays significant plurality. Thus, unless the Bayes factor $\frac{P(E|H_2)}{P(E|H_1)}$ is extraordinarily large, the posterior for H_2 remains small.⁴

Logical Core (Syllogism):

P1: When a hypothesis contradicts stable cross-cultural justice constraints, its prior probability is low.

⁴ This footnote references a representative probabilistic defense of substitution: Swinburne (1989). This footnote is descriptive.

P2: Penal substitution contradicts the stable constraint that culpability in criminal law is nontransferable.

P3: The total evidence does not provide an overwhelming Bayes factor in favor of penal substitution.

C: Therefore, the posterior probability of penal substitution as divine justice remains low.

Note. Because criminal guilt is not transferable, any theory that depends on transferable guilt begins with a low prior; the available evidence does not overcome that starting point.

Civil vs. Criminal Liability Confusion

Defenders of penal substitution sometimes appeal to analogies in civil law, where one party may bear financial liability for another (for example, corporate liability for employees). This appeal equivocates between civil and criminal frameworks. Civil liability concerns restitution—compensation for loss—not punishment for culpable wrongdoing. Criminal liability turns on culpability itself: punishment is proportioned to guilt and directed at the offender (Duff, 2001; Hart, 1968).⁵

Invoking civil analogies to defend substitution in criminal contexts is a category mistake. Financial responsibility can, in limited cases, be shifted without conceptual confusion. But criminal punishment without personal culpability is incoherent. Penal substitution blurs this boundary, weakening the distinction between restitution and punishment that undergirds functioning legal systems.

Logical Core (Syllogism):

⁵ This footnote clarifies the category difference: civil liability can allocate costs without attributing criminal guilt, whereas criminal punishment in most traditions is tied to the offender's culpable act. This footnote is descriptive.

P1: Civil liability (restitution) may be transferred, but criminal punishment (culpability-based) may not.

P2: Penal substitution defends criminal punishment by appeal to civil analogies.

C: Therefore, penal substitution equivocates and misapplies legal categories.

Note. One can pay another's bill, but one cannot justly serve another's sentence; mixing these categories makes justice incoherent.

Biblical Contradictions and Epistemic Costs

Scripture displays tensions regarding culpability. Ezekiel 18:20 insists: "*The soul who sins is the one who will die. The son will not share the guilt of the father.*"⁶ Yet Exodus 20:5 states that God "punishe[s] the children for the sin of the fathers to the third and fourth generation."⁷ These stances destabilize appeals to biblical authority in support of substitution (Barr, 1999). If divine justice were communicated consistently, we would not expect such variance.

From an epistemic perspective, this contradiction undermines the reliability of scripture as a source of consistent justice principles for substitutionary punishment. If justice is revealed yet oscillates between prohibiting and endorsing substitution, appeals to biblical authority carry diminished evidential weight for penal substitution.

Logical Core (Syllogism):

P1: Ezekiel affirms that each person bears responsibility for their own sin.

P2: Exodus affirms intergenerational punishment in certain covenantal frames.

P3: Contradictory principles cannot jointly define a consistent model of justice.

⁶ Ezekiel 18 is often read as centering responsibility in the individual agent rather than in a kin-group. This footnote summarizes a common scholarly reading and is descriptive.

⁷ Many scholars see Exodus 20:5 as reflecting an ancient Near Eastern covenantal context in which group identity bears legal and moral significance. This footnote situates the verse historically and is descriptive.

C: Therefore, scriptural foundations for penal substitution are epistemically unstable.

Note. Divergent biblical principles about who bears guilt weaken appeals to scripture as a stable basis for a substitutionary model of justice.

Historical Genealogy of Penal Substitution

Historically, penal substitution is not an original or universal doctrine. Early Christian writers often emphasized *Christus Victor* (Christ's triumph over evil) or moral exemplar accounts. Penal substitution, as classically formulated, rose to prominence in Reformation contexts, especially in Calvin's work, drawing on contemporary legal metaphors.⁸ This historical contingency suggests the doctrine is not a timeless feature of Christian justice but a theological development responding to particular cultural and legal concerns.

If penal substitution were the unique and divinely necessary mechanism of justice, we would expect early, consistent, and universal recognition. Instead, we find a patchwork: multiple atonement theories co-exist, with substitution only later taking center stage. This genealogical fragility weakens its claim to divine necessity. Moreover, even if a strong genealogical pedigree for penal substitution could be demonstrated, it would not undercut the logical arguments already advanced in this paper. Scapegoat normalization, transferred desert, voluntariness, Bayesian priors, civil–criminal confusion, biblical contradictions, and cross-cultural justice constraints would all continue to weigh decisively against it. Historical continuity cannot rescue a model that fails conceptually.

Logical Core (Syllogism):

P1: If penal substitution were divinely necessary, it should appear early, consistently, and universally.

⁸ For a historical sketch of substitutionary atonement's alternatives, see Aulén (1969). For justification doctrines across Christian history, see McGrath (2005). These footnotes are descriptive.

P2: Historically, penal substitution emerged late and competed with alternatives.

P3: Even if a consistent genealogy could be shown, it would not repair the conceptual incoherences.

C: Therefore, penal substitution lacks the marks of divine necessity and remains conceptually deficient.

Note. The historical record displays contingency, not necessity; even a stronger pedigree would not cure the doctrine's logical defects.

Cross-Cultural Justice Contrasts

Across legal and moral traditions—Confucian, Islamic, Hindu, African customary law, and Indigenous justice practices—punishment is anchored in the accountability of the agent. While degrees of collectivism vary, a widely shared principle is that one person does not bear criminal liability for another's intentional act. This cross-cultural convergence raises the burden of proof on any view that centrally requires transferable culpability.

Such convergence functions as a *moral fixed point*: when cultures across time and geography converge on a constraint, the evidential weight for that constraint strengthens. Penal substitution stands as an outlier and must explain why divine justice would uniquely reverse this persistent consensus. No satisfactory explanation is forthcoming.

Logical Core (Syllogism):

P1: Cross-cultural legal traditions converge on nontransferable culpability in criminal law.

P2: Penal substitution contradicts this convergence.

C: Therefore, penal substitution is improbable as a model of divine justice.

Note. The global norm that criminal guilt is personal renders a transferable-guilt theory like penal substitution highly implausible.

Conclusion

Taken together, these critiques converge on a single conclusion: penal substitution is not merely implausible but incoherent as a model of divine justice. The *Scapegoat Normalization Problem* shows that it sanctifies a destructive social mechanism. The *Transferred Desert* critique shows that it severs punishment from culpability. The *Futility of Voluntariness* shows that willingness cannot transmute innocence into guilt. The *Bayesian* argument shows that strong priors against vicarious criminal punishment are not overcome by the available evidence. The *Civil vs. Criminal* analysis exposes a category mistake; *Biblical Contradictions* highlight epistemic instability; the *Historical Genealogy* reveals contingency rather than necessity and underscores that even a strong genealogy would not answer the conceptual objections; and the *Cross-Cultural* review underscores a broad, stable constraint against transferable guilt.

If justice is to retain coherence, culpability must remain with the agent whose action and intent generated the wrong. A theologically adequate account of reconciliation should therefore align with restorative practices that center accountability, repair, and reintegration rather than sanctifying scapegoating. On legal, moral, historical, and epistemic grounds, penal substitution fails to provide a coherent model of divine justice. Looking forward, constructive accounts that integrate restorative justice with non-substitutionary atonement motifs—such as moral transformation and *Christus Victor*—offer a more coherent and practically fruitful path for theology and moral practice.

References

- Aulén, G. (1969). *Christus Victor: An historical study of the three main types of the idea of atonement* [(Original work published 1931)]. SPCK.
- Barr, J. (1999). *The concept of biblical theology: An Old Testament perspective*. Fortress Press.
- Braithwaite, J. (1989). *Crime, shame and reintegration*. Cambridge University Press.
- Carter, J. (2011). *Scapegoats and scapegoating in early modern England*. Routledge.
- Douglas, M. (2002). *Purity and danger: An analysis of concepts of pollution and taboo* [(Original work published 1966)]. Routledge.
- Duff, R. A. (1986). *Trials and punishments*. Cambridge University Press.
- Duff, R. A. (2001). *Punishment, communication, and community*. Oxford University Press.
- Feinberg, J. (1970). *Doing and deserving: Essays in the theory of responsibility*. Princeton University Press.
- Girard, R. (1987). *Things hidden since the foundation of the world*. Stanford University Press.
- Girard, R. (2005). *Violence and the sacred* [(Original work published 1977)]. Continuum.
- Hart, H. L. A. (1968). *Punishment and responsibility*. Oxford University Press.
- Heim, S. M. (2006). *Saved from sacrifice: A theology of the cross*. Eerdmans.
- Levine, M. (2020). *Scapegoating: The psychology of blame*. Routledge.
- McGrath, A. E. (2005). *Iustitia Dei: A history of the Christian doctrine of justification* (3rd). Cambridge University Press.
- Moore, M. S. (1997). *Placing blame: A general theory of the criminal law*. Oxford University Press.
- Murphy, J. G., & Hampton, J. (2003). *Forgiveness and mercy*. Cambridge University Press.
- Rawls, J. (1971). *A theory of justice*. Harvard University Press.

Swinburne, R. (1989). *Responsibility and atonement*. Oxford University Press.

Zehr, H. (2002). *The little book of restorative justice*. Good Books.

[illegible]