

The Evolution of Retribution: Theological Implications

“Till on that cross, as Jesus died,
The wrath of God was satisfied...”¹

Introduction

Our understanding of hell and divine wrath has profound implications for how we interact with God in worship. A central question on which these implications hinge concerns God’s attitude toward us when we act in opposition to God or harm God’s creations. Does God *literally* take on an attitude of wrath toward us? Is God *literally* motivated to seek vengeance for our transgressions? If so, then appeasement of divine wrath is a central constituent of worship. For instance, much of Christian liturgy is structured around the narrative arc of the crucifixion. But as suggested by the epigraph, if God is motivated to avenge, then Jesus work of atonement is most aptly understood as satisfying that motive and appeasing God’s anger. It follows that much of the shape of Christian worship celebrates, participates, and perhaps even reenacts an act of appeasement. Moreover, on this picture, where our liturgies celebrate this work of salvation, they celebrate salvation from a hell that is fueled by God’s wrath.

While I find this view of worship troubling, it harmonizes with an intuitive view of punishment: that the value of punishment consists in giving wrongdoers what they deserve. When we experience wrath, the objects of our experience are an offense that we feel deserves redress and an offender who we feel deserves hard treatment. When we contemplate the worst kinds of offenses, the punishments that satisfy us are not necessarily the ones that secure a good outcome (e.g. deterrence or rehabilitation). Rather, they are the punishments that adequately

¹ Stuart Townend and Keith Getty, “In Christ Alone”, Kingsway Thankyou Music, 2001

repay the offense. My purpose here is to explore the source of these intuitions about punishment. I argue that these intuitions have an evolutionary explanation, and that this explanation has important implications for our understanding of hell and divine wrath.

As I will argue in the second section, the traditional doctrine of hell presupposes the truth of a retributive principle: that punishment has value aside from its consequences. This retributive principle seems to be supported by many of the moral intuitions evoked by particular offenses. Nevertheless, in the third section, I will suggest that our moral judgments about punishment may be a product of evolutionary forces, and I argue that if so, retributive inclinations do not actually provide evidence for the retributive principle. If my argument is correct, this calls into question whether punishment in hell could possibly be justified (as traditionally conceived).

In the fourth section, I will consider the possibility of drawing on Christian scripture to support or otherwise evaluate the retributive principle—specifically, the Bible presents God as a God of wrath who seeks punishment as an end in itself, and the plausibility of retributive justifications depend on how one understands these scriptures. One option is to understand them as *informative* (or perhaps propositional): scripture is supposed to provide information about God's attributes. While these interpretations tend to support the retributive principle, they come with hermeneutic and systematic costs. By contrast, another option is to understand these scriptures as *evocative*: they are intended to evoke certain responses in an audience, responses like worship, submission, awe, and respect. I think this view can provide a foundation on which to build an alternative to the troublesome perspective above.

Hell as (deserved) punishment

What is the traditional view of hell and why does it presuppose a retributive principle?

Jonathan Kvanvig captures the traditional view of hell with four propositions:

- 1) The Punishment Thesis: the purpose of hell is to punish those whose earthly lives and behavior warrant it;
- 2) The No Escape Thesis: it is metaphysically impossible to get out of hell once one has been consigned there;
- 3) The Anti-Universalism Thesis: some people will be consigned to hell;
- 4) The Eternal Existence Thesis: hell is a place of conscious existence.²

It is also traditionally assumed that punishment produces suffering via harsh treatment, and this assumption is close to the heart of prominent views of hell as “eternal conscious torment.”³ Thus, from these four theses it follows that some people will suffer consciously and endlessly in hell, producing an infinite amount of suffering. As many philosophers have noted, it is ordinarily wrong to impose hard treatment or suffering on another person. Thus, we need an explanation of why these impositions are justified in the case of punishment.

I can see two ways of giving such an explanation. First, one could point to some valuable consequences of punishment, which outweighs the disvalue of suffering. Here consequences are understood as the effects of punishment “as opposed to...the intrinsic nature of the act or anything that happens before the act.”⁴ Compare punitive suffering to the athletic suffering experienced as one trains for a marathon. The suffering of training increases the athlete’s perseverance and one might think that the moral value of the perseverance greatly exceeds the cost of suffering. There is a commonality between this justification of athletic suffering and some justifications of punishment: a valuable outcome is produced as a consequence of suffering that outweighs the disvalue of suffering. However, this justification is effective only insofar as suffering is necessary to bring about the valued

² J Kvanvig, “Heaven and Hell,” *Stanford Encyclopedia of Philosophy*, 2011.

³ Some philosophers argue that suffering is not the primary aim of punishment. Even so, almost all agree that punishment will produce some amount of suffering in almost every case, whether or not that is its primary aim.

⁴ Walter Sinnott-Armstrong, “Consequentialism,” *Stanford Encyclopedia of Philosophy*, 2003.

outcome. For instance, there are many valuable goals for running a marathon that do not necessarily require suffering. One such goal is simply the achievement of finishing out a marathon. While this goal might outweigh all the suffering that went into one's training, it does not necessarily justify the suffering. We can imagine that analgesic medications could make it possible to avoid the suffering of training entirely, such that one could accomplish the same goal without suffering. In that case, suffering in training would not actually be justified by the consequence of finishing the marathon.⁵

In ordinary cases of punishment, this strategy is promising. Punishment can produce many good outcomes:

1. General deterrence – i.e. punishing in order to deter other would-be offenders from committing similar offences.
2. Incapacitation [or specific deterrence] – i.e. punishing in order to prevent the offender from committing similar crimes while he is being detained and/or treated [or thereafter].
3. Rehabilitation and moral education – i.e. punishing in order to rehabilitate or re-educate the offender...
5. Catharsis – i.e. punishing in order to give victims and society more generally a healthy emotional release.
6. Norm reinforcement – i.e. punishing in order to highlight and reassert the importance of social values and norms.
7. Quelling revenge – i.e. punishing in order to keep the original or third parties from starting a blood feud.⁶

⁵ Here, one must notice the difference between the goal of finishing a marathon and the goal of finishing a marathon by overcoming adversity. The former is the goal I claim could be accomplished without suffering. The latter could not, and actually resembles the justification for suffering that I consider below.

⁶ Thomas Nadelhoffer et al., "Folk Retributivism and the Communication Confound," *Economics and Philosophy* 29, no. 02 (July 11, 2013): 237, doi:10.1017/S0266267113000217. I leave off the list the following: "Communication – i.e. punishing in order to communicate or express disapproval of an action." Ibid. Communication is not a consequence of punishment as I have defined "consequence". This is because punishment as communication is incoherent without some reference to the transgression, which came before the act of punishment. That is, when punishment expresses disapproval, it *constitutes* a message to the offender or an act of communication. If so, communication is *not* a consequence of punishment because it is inseparable from the act of punishment and the transgression that preceded it.

We can also add “reconciliation” to this list, whereby punishment allows a restoration of relationship between offender and victim. It is not entirely implausible that in some cases punishment is necessary to bring about some of these ends.⁷

However, none of these ordinary goods can be achieved through punishment in hell, at least not according to orthodox Christian theology. Those condemned to hell are eternally separated from God and from the saints, and are thus without hope of reconciliation. Presumably, there is no need to quell revenge in the eschaton, or at least, it seems unnecessary to punish eternally in order to quell revenge. It is unclear why norm reinforcement or catharsis requires eternal punishment, as opposed to a large but finite amount of suffering (e.g. 2 million years of suffering). Likewise, it is implausible to suppose that hell will rehabilitate its denizens. On almost any theological framework, there are severe limits to the moral improvement that can occur there. Finally, eternal suffering in hell is clearly not necessary to deter transgressions (either for general deterrence prior to the eschaton or for incapacitation after its initiation). A large but finite period of suffering would presumably be enough to deter almost any crime. Insofar as this list exhausts the valued consequences of punishment, the first strategy for justifying punishment in hell fails.

Here is a second strategy: one could point to some feature of punishment that infuses the suffering with value (rather than disvalue). By analogy, one might claim that suffering while training for a marathon is its own reward. On this view there is something about the suffering *itself*, or in the act that leads to suffering, that changes suffering from bad to good. For instance, each moment of suffering that attends training might be valuable as a necessary

⁷ Though some have argued that contemporary institutions of punishment do not ordinarily achieve these ends and thus are unjustified. See e.g. Hugo Adam Bedau and Erin Kelly, “Punishment,” *Stanford Encyclopedia of Philosophy*, 2003.

constituent of certain actions: willfully overcoming adversity, exercising one's self-control, or asserting one's agency against countervailing hardships. Each of these valued ends requires suffering as a constituent. One cannot purposely enter into a process of overcoming adversity without actually encountering adversity as a part of that process. On this view, the suffering does not *cause* a valuable outcome but instead *constitutes* (in part) some valued end.

There are several justifications for punishment that follow this general strategy. Following Kant, many philosophers take it as a given that happiness should be proportionate to virtue. This is one way of supporting the claim that deserved suffering has value, since the suffering of the transgressor is somehow appropriate to her vice. Likewise, some would say that deserved suffering has intrinsic value or that deserved suffering is just and thus constitutes a moral good. Others point to the communicative function of punishment whereby punishment communicates censure for transgression. On this latter view, the value of punishment is constituted by the relationship between the punishment and the crime that came before it (cf. fn. 6). The point I want to make is simply that all such justifications of punishment presuppose the following principle: R—The value of punishment is not (entirely) derived from its consequences.

As I argue above, eternal punishment in hell is not necessary to secure any good consequences. Moreover, if good consequences can be brought about some other way, then hell is not justified as a means of bringing about those good consequences. Thus, if hell is justified, then punishment must have some unique value aside from its good consequences and a value that cannot be achieved by other means. That is, punishment (or perhaps suffering due to punishment) needs to be intrinsically good, or perhaps good in relation to what came before the

act of punishment (e.g. the transgression). I believe that this kind of *non-derivative* value for punishment is exactly what people are trying to capture when they say that punishment is deserved. What they mean is that *in relation to the transgression*, punishment is somehow fitting or good or that punishment as a response to transgression is intrinsically good. If we define retributivism as the claim that punishment can be deserved (in a way that is not reducible to the consequences of punishment), then R is one way of capturing the essence of retributivism.

Explaining retributive inclinations

What I now want to show is that the primary evidence for R is undercut when we consider the evolution of punishment. To see this, we must first appreciate that our own inclinations to act and judge in accordance with R is our primary evidence for believing it.

Consider Michael Moore's strategy for justifying retributivism:

I take seriously the sorts of particular moral judgments that...thought experiments call forth in me and in most people I know... for example, Dostoevsky's Russian nobleman in *The Brothers Karamazov*, who turns loose his dogs to tear apart a young boy before the mother's eyes; imagine further that circumstances are such...that no [good consequence would be achieved] by punishing this offender...Question: should...the offender be punished, even though no other social good will thereby be achieved? The retributivist's 'yes' runs deep for most people.⁸

Moore concludes that this is the best way to justify a principle like R:

As even the gentle Alyosha murmurs in Dostoevsky's novel, in answer to the question of what you do with the nobleman: you shoot him...The only general principle that makes sense of the mass of particular judgments like that of Alyosha is the retributive principle that culpable wrongdoers must be punished. This, by my lights is enough to justify retributivism.⁹

⁸ M. S. Moore, *Placing Blame: A Theory of the Criminal Law* (Oxford: Oxford University Press, 2010), 163.

⁹ *Ibid.*, 188.

I suspect that Moore is right and that the majority of evidence in support of R will be our inclinations to judge and act in accordance with it, as manifested in “the mass of particular judgments” that we are inclined to make.

So where do these inclinations come from? They are present not only in moral punishment, but are also observed in what I call “personal punishment,” whereby a person retaliates in order to repay a personal offense (as opposed to a moral offense).¹⁰ For instance, those who seek revenge often believe and act as if revenge has value, even if payback does not actually pay. This has been demonstrated in a variety of economic games in which *irruptive motivational states* like anger cause people to perform in less than optimal ways. For instance, in anonymous one-shot games, people forgo real monetary gains in order to repay perceived offenses (e.g. by diminishing the gains of a competitor who acted unfairly).¹¹ By all appearances, irruptive motivational states function to outweigh immediate gains or override practical reasoning in favor of a costly, punitive response. Either way, they appear to interrupt the ordinary functioning of self-interest and deliberate choice to produce vengeful but counterproductive behaviors. There is some evidence that these irruptive motivational states lead to the development of more cool-headed retributive inclinations that also seem to support principles like R.¹²

¹⁰ See also L Zaibert, “Punishment and Revenge,” *Law and Philosophy* 25, no. 1 (2006): 81–118. There, Zaibert argues that there is no in principle distinction between revenge and retribution.

¹¹ E.g. Francine Espinoza, Alexander Fedorikhin, and Joydeep Srivastava, “Anger in Ultimatum Bargaining : Emotional Outcomes Lead to Irrational Decisions,” 2000, 1995–97; Joydeep Srivastava and Francine Espinoza, “Coupling and Decoupling of Unfairness and Anger in Ultimatum Bargaining” 489, no. December 2008 (2009): 475–89, doi:10.1002/bdm.

¹² Michael A. Milburn, Miho Niwa, and Marcus D. Patterson, “Authoritarianism, Anger, and Hostile Attribution Bias: A Test of Affect Displacement,” *Political Psychology* 35, no. 2 (August 19, 2014): 225–43, doi:10.1111/pops.12061.

The nature of these phenomena makes them difficult to explain. Given that people forgo monetary gains (among other things) merely to avenge or “repay” offenses, it is difficult to explain why people view these as a worthwhile aim. That is, it is difficult to explain the desirability of revenge in terms of other benefits that people reasonably aim to achieve.¹³ Similarly, it is difficult to see how revenge could be a product of learning. Children exhibit so-called reactive aggression at a very young age, and this kind of retaliatory behavior can persist even when it is actively discouraged and is socially detrimental. Finally, cultural explanations of personal and moral punishment are not likely to work out. Norms of revenge exist in a vast majority of the cultures that anthropologists have studied, making culture an unlikely source of these norms.¹⁴ Moreover, as Robert Frank notes,

Most cultures not only do not encourage the pursuit of vengeance, they take positive steps to curtail it. Contrary to impressions, the biblical reference, “an eye for an eye, a tooth for a tooth,” is not an exhortation to seek revenge, but a plea to restrain it to the scale of the original provocation. We may safely presume that, where a cultural norm attempts to restrain a given behavior, people left to their own devices would tend to do more of it. Thus, it hardly makes sense to offer cultural conditioning as the explanation for why we see such behavior in the first place.¹⁵

In sum, the most obvious psychological explanations fail. Moreover, they fail in ways that suggest an evolutionary explanation. For instance, the universality of revenge makes it likely that it will be explained by common biological inheritance (from ancestral populations) rather than by common incentive structures (that might guide learning) or cultural inheritance.

Evolutionary models suggest that revenge and retribution are evolutionary adaptations.¹⁶ Adaptations are traits that survive a given selection regime because of their favorable

¹³ Thanks to Curtis Holtzen for pressing me to clarify this point.

¹⁴ Martin Daly and Margo Wilson, *Homicide* (Transaction Publishers, 1988).

¹⁵ R. H. Frank, *Passions within Reason: The Strategic Role of the Emotions*. (New York: Norton, 1988), 39.

¹⁶ See e.g. Frank, *Passions within Reason*; Herbert Gintis et al., “Strong Reciprocity and the Roots of Human Morality,” *Social Justice Research* 21, no. 2 (March 4, 2008): 241–53, doi:10.1007/s11211-008-0067-y; Nicolas

consequences. For instance, traits that cause an organism to leave a greater number of descendants (or rather, copies of its genes) in subsequent generations are more likely to persist in a population. If so, the trait exists because of its consequences (more descendants or gene copies) for the organisms that possess that trait. Retribution and revenge may have been selected in just this way, specifically because they deter certain forms of behavior in the future. For example, if someone has a reputation for punishing offenses even in the face of immediate costs, this can have certain long term advantages. People are less likely to cross a person who is irascible in this way, thus irascibility can deter bad treatment and may enable those who possess it to leave more offspring in future generations.¹⁷

If as I have suggested, retribution and revenge are inclinations to judge and act in accordance with R and if they evolved because of their favorable consequences, it follows that they are not good evidence for R. Given that they evolved because of their consequences, they are not a good indicator that punishment has value aside from its consequences. Let me flesh out this inference more clearly. Consider Moore's argument above. His idea is that only a principle like R would "make sense" of the particular judgments that we make or are inclined to make. In other words, if someone is inclined to judge that a particular set of actions are morally right or wrong, then she should take the inclination as an indication of the truth (or accuracy) of some principle that is necessary to make sense of the particular judgments.

Baumard, Jean-Baptiste André, and Dan Sperber, "A Mutualistic Approach to Morality: The Evolution of Fairness by Partner Choice," *Behavioral and Brain Sciences* 36, no. 01 (February 1, 2013): 59–78, doi:10.1017/S0140525X11002202.

¹⁷ Indeed, vigilante revenge is most common in conditions in which deterrence has a high value, such as conditions in which there is no centralized law enforcement, in which wealth is portable, and in which there are few effective ways to monitor or prevent transgressions like robbery and adultery. E.g. Richard E. Nisbett and Dov Cohen, *Culture of Honor: The Psychology of Violence in the South* (Boulder, CO: Westview Press, 1996).

Nevertheless, the inclinations to judge in accordance with R can only be a good indication of the truth (or accuracy) of R if the principle is true and if there is a non-accidental relationship between the inclinations and the principle. For instance, suppose that the principle is true, but then imagine that retributive inclinations were the outcome of a demiurge flipping a coin to decide whether humans would have retributive inclinations or not. In that case, it would only be by chance that retributive inclinations co-occurred with the truth of the principle.¹⁸

So what kind of non-accidental relationship must exist to make the inclinations a good indicator of R? There are two possibilities. On the one hand, there could be some causal relationship that produces a correlation between the inclinations and the principle; or, on the other hand there could be a constitutive relationship that produces the correlation. Both of these relationships are asymmetric dependencies that can produce non-accidental co-occurrences/correlations. For example, there is a correlation between smoking and lung cancer precisely because smoking *causes* lung cancer. On the other hand, there is a correlation between the redness of the chimney and the redness of its bricks precisely because the bricks *constitute* the chimney.

There are three possible dependencies that could explain the correlation between retributive inclinations and the retributive principle R.¹⁹ First, the truth of R could depend on the inclinations. Second, the inclinations could depend on the truth of R. Third, both could depend on some further state of affairs. Though there is not space to make the argument in detail here, I

¹⁸ Imagine: had the demiurge flipped heads rather than tails, humans would have instead had an inclination to “let bygones be bygones”. Thus, in the imagined example, it would only be by chance that R is true AND humans have retributive inclinations, and we would not trust our retributive inclinations (or the resulting intuitions) as indicators of R if we knew about their determination via the coin toss. Thanks to Matthew Hill for pressing me to clarify this point.

¹⁹ Cf. Fred I. Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1999).

have argued elsewhere that the evolutionary etiology of the inclinations makes each of these possibilities highly implausible.²⁰

If this argument is sound, then retributive inclinations are not good evidence for R. Given that the justification for punishment in hell requires the truth of R, it becomes difficult to see how punishment in hell (as traditionally conceived) could be justified. That is, unless there is some independent evidence for R.

Adding to the evidence base

Thus far, for simplicity I have been neglecting a significant part of the Christian evidence base. I can see at least two ways in which a Christian could draw on this evidence base to evaluate R. First, one could assess R *indirectly* by looking for additional evidence whether retributive inclinations are a good indicator of the truth of R. For instance, human psychology is not merely a random product of evolution. Rather, human evolution has been guided in some way that is consistent with the biblical claim that humans were created. Perhaps God created us with retributive inclinations precisely because they are a reliable route to the formation of true beliefs about punishment, such as belief in R.

I do not believe the notion of creation by *itself* plays this role. This is because we are imperfect creations. Given our imperfection, there is no guarantee that all our inclinations to

²⁰ See Isaac Wiegman, “Anger and Punishment: Natural History and Normative Significance” (Washington University in St. Louis, 2014); Isaac Wiegman, “The Evolution of Retribution : Intuitions Undermined,” *Pacific Philosophical Quarterly*, forthcoming. Here is the argument in brief. Suppose that the inclinations (I) evolved because of their deterrent effects (D). Once we know this, we know that the truth of principle R cannot be the cause of D, which is the direct cause of I (as in this graph, $R \rightarrow D \rightarrow I$). This is because Inclinations to punish could have deterrent effects regardless of whether punishment has non-derivative value. Likewise, it would appear that punishment could have non-derivative value regardless of whether retributive inclinations have deterrent effects. Thus, it is unclear how the deterrent effects of the inclinations could possibly be a common cause of (or constitutive base for) the truth of R and the inclinations. R says that punishment has non-derivative value, but it seems impossible that this truth could be an effect (causal or constitutive) of the deterrent value of the inclinations. Finally, it is unclear how retributive inclinations themselves could cause (or constitute) the truth of R (as in this graph, $D \rightarrow I \rightarrow R$), since according to R, punishment has value even when it does not have deterrent effects.

believe and act are epistemically reliable or virtuous (or practically or morally virtuous, for that matter). For instance, we have an ingrained tendency to infer hidden causes from observable patterns in the world.²¹ This tendency is likely to be distinctively human,²² and without it, a vast range of human knowledge, scientific or otherwise would surely not be possible. However, this tendency also leads primitive people to believe in ghosts and fairies and all sorts of hidden supernatural causes that do not actually explain observable patterns in the world. Likewise, retributive inclinations might be a beneficial product of design (e.g. as a defense against exploitation) while also leading to false beliefs about the value of punishment. Good design does not necessarily lead to reliable processes of belief formation across every domain.

The other way of vindicating R is more direct, by finding additional reasons in support of R itself.²³ One way to support R is just to point out (as I did above) that the traditional view of hell requires its truth and that one has independent reason to accept the traditional view of hell. If so, one need only provide scriptural support for the traditional view of hell. Since I have independent doubts about the traditional view of hell, and since doubts like these have been sufficiently discussed elsewhere,²⁴ I will leave this approach aside for the remainder of this essay.

²¹ E.g. Alison Gopnik and Henry M Wellman, "Reconstructing Constructivism : Causal Models , Bayesian Learning Mechanisms , and the Theory Theory" 138, no. 6 (2012): 1085–1108, doi:10.1037/a0028044.

²² D. J. Povinelli and N. G. Ballew, *World without Weight: Perspectives on an Alien Mind* (Oxford University Press, 2012).

²³ One might suspect that I am giving short shrift to the first possibility, since we are not only created by God but also in God's image. However, given the criticism above (that design does not imply epistemically virtuous processes of belief formation), creation in God's image can only vindicate R if God's image includes retributive inclinations that are themselves an indicator of the truth of R. Any demonstration of this would provide direct support for R anyway. Thus, one cannot vindicate R indirectly by appealing to creation in God's image.

²⁴ See e. g. W Crockett, *Four Views on Hell*, 1996; EW Fudge and RA Peterson, "Two Views of Hell: A Biblical and Theological Dialogue," 2000; JL Kvanvig, *The Problem of Hell*, 1993.

Here is another direct approach to vindicating R. Christian scripture suggests that God is morally perfect and that God is also wrathful toward sin,²⁵ which if true would directly support R (independently of retributive inclinations). Consider a particularly compelling example:

For if we willfully persist in sin... [there only remains] a fearful prospect of judgment, and a *fury of fire* that will consume the adversaries. Anyone who has violated the law of Moses dies without mercy... How much worse *punishment* do you think will be *deserved* by those who have spurned the Son of God... and *outraged* the Spirit of grace? For we know the one who said, "*Vengeance is mine, I will repay.*" ... (Hebrews 10:26-31, NRSV, emphasis mine)

Hebrews 10 portrays God as a God of vengeance, who punishes "those who go on sinning deliberately" merely because their sinning deserves a reaction of repayment or vengeance. In other words, God is presented as pursuing punishment as a reaction to sin and as an end in itself, suggesting that God would react in this way even if there were no other valuable outcomes for which punishment were necessary. Importantly, the punishment suggested here seems to refer to hell.

Moreover, the pursuit of punishment is portrayed elsewhere in connection with God's wrath (cf. the "fury of fire" in Hebrews 10).

Those who worship the beast and its image, and receive a mark on their foreheads or on their hands, 10 they will also drink the wine *of God's wrath*, poured unmixed into the *cup of his anger*, and they will be *tormented with fire and sulfur*... (Revelation 14:9-10, NRSV)

²⁵ See e.g. Exodus 20:17, Deuteronomy 9:7, Ezekiel 25:17, Psalm 7:11, Psalm 75:8, Nahum 1:2-6, Mark 3:5, Luke 12:5, John 3:36 Romans 1:18, Romans 2:5, 1 Thessalonians 1:10, Revelation 19:11-21

A naïve interpretation of these passages suggests the following divine psychology: in reaction to sin, God experiences an irruptive motivational state that motivates God to pursue punishment independently of its consequences.

If we take these passages at face value, R is doubly reinforced. First, if God is morally perfect and is also motivated to pursue punishment independently of its consequences, then it is conceptually necessary that punishment really does have some moral value that is not derived from its consequences. Otherwise, God simply would not pursue it in these ways. Importantly, if one accepts that God has irruptive motivations to punish, then one is actually forced to accept R (and probably the traditional view of hell also) by conceptual necessity (when conjoined with God's moral perfection). Second, if God is morally perfect and possesses retributive inclinations like ours, then it would appear that these inclinations (together with their epistemic role of supporting principles like R) are part of God's image and are not an unintended byproduct of God's design.

However, the attribution of irruptive motivational states to God is theologically problematic. This is because an important part of the Christian tradition presents God as *unchanging* (rather than being influenced by momentary passions), *active* (not the passive recipient of emotional disturbances), *invulnerable* to the influence of "external" emotional disturbances (like those humans experience when in the grip of anger or wrath), and *simple*, meaning that there is no division between divine will and divine passions (as suggested by the very nature of irruptive motivational states).²⁶ In any case, these are some of the reasons why many early and medieval theologians did not take these passages at face value. Moreover, I

²⁶ For a detailed discussion of the history of these claims in relation to God's emotional states (or lack thereof), see A. P. Scrutton, *Thinking Through Feeling: God, Emotion and Passibility*, 2011, chap. 1.

suspect that most contemporary theologians would accept one or more of these claims about the nature of God.

At the very least, we are not *obligated* to take these passages at face value, and vindicating R in this way carries with it significant costs. A more plausible way to vindicate R is to suppose that God persistently values punishment as an end in itself, or equivalently, that God's wrath is "bloodless."²⁷ Moreover, this motive for punishment is understood as 1) a persistent aspect of God's character (preserving God's immutability and simplicity); and 2) unmediated by irruptive motivational states (preserving God's activeness and invulnerability). According to this interpretation, descriptions of God's fury and wrath are like exclamation marks on statements about God's persisting desire to punish sinfulness. If this interpretation is correct, then R is reinstated. Like the previous interpretation, God's moral perfection and his desire to punish (as an end in itself) suggest that God would not treat punishment as an end in itself were it not actually valuable as such.

The difficulty with this line of thought comes from a contestable supposition (shared with the naïve view): that the purpose of passages like these is to tell us something about God's nature and what God values. If we accept this *informative* view of passages like these together with the bloodless wrath interpretation (as opposed to the naïve reading), the language about God's wrath seems superfluous. On the naïve reading, the function of wrath is to explain or make intelligible (to us) God's pursuit of punishment as an end in itself. Why does God punish as an end in itself? Because God is angry. Why is God angry? Because of our sin. Anger, fury, and wrath are the causal intermediates between our sin and God's punishment. But if we then assume that God is

²⁷ Joel Potter raised this possibility at the annual meeting of the Wesleyan Philosophical Society and also suggested this label for the motive to punish.

not subject to irruptive motivational states (as the bloodless wrath interpretation suggests), then divine wrath and fury are stripped of their ostensible role in these texts. We are left with a God who values punishment as an end in itself, but inexplicably so (or at least, not for the reasons presented in these passages). The attribution of wrath seems entirely unnecessary to serve the communicative purpose of informing us of God's nature and values. We would be just as well off simply being told that sinners deserve punishment and that God will punish them in accordance with their desert (and perhaps, whether or not punishment is necessary to achieve some good consequence). Why the misleading attributions of wrath and fury?

I think we can better answer this question by changing the assumptions with which we approach this scripture. Instead of assuming that the role of wrath and pursuit of punishment in these scriptures is to *communicate* something about God's nature and what God's values, perhaps it is to *invoke certain responses in us*, responses like awe, respect, submission, and perhaps even worship. By comparison with the informative views above (the naïve and bloodless wrath readings), this *evocative* view does a better job of explaining why it is that these passages mention both God's wrath toward and God's punishment of sin. We are confronted both by the nature of our actions and character (the serious nature of which is conveyed by their "deservingness of punishment") and brought to submission by images of God's disapproval and wrath. The intended effect of this confrontation is that we are compelled to take on certain attitudes toward our sins *and* toward God.

The idea is that God accommodates human understand by evoking responses in ways that we can understand. Anger and punishment are pan-cultural phenomena that are ingrained in the

human psyche by evolutionary forces as a response to weighty transgressions.²⁸ Thus, it makes sense to call on these powerful and primitive motives to instill respect for, and awe toward God on the one hand and repentance for and aversion to sin on the other. We do not have to literally believe that our sins deserve punishment or that God is angry at our sins to be exhorted in this way (though these passages probably serve their function best if questions of literalness are not broached). We need only *imagine* God's attitude toward our sins and the *apparent (to us)* fittingness of punishment in response to them.

Of course, authenticity matters for evoking these responses. If a parent were to feign anger toward a misbehaving child, one suspects that this would not have desirable effects. Nevertheless, parental authenticity matters for reasons that do not apply to scripture or to divine wrath. First, scripture is presented through human intermediaries. That is, the language of wrath accommodates our concepts and gains authenticity through human conveyance. The reader of the passage can invoke their own sense of anger to convey (to herself or to others) the necessary emotional state, which is authentic insofar as the reader can truly deploy her own capacities to view sin as offensive and thus to "simulate" an angry reaction to it.²⁹ Due to human conveyance, the practical risks of God "faking" anger do not arise in this case.

Second, it is not possible for God to feign emotion in the same way that humans sometimes do. Feigning an emotion is usually set in opposition to an authentic manifestation or experience of emotion. Thus, if an organism is not capable of manifesting or experiencing a certain emotion, then the conditions for authenticity and disingenuousness shift slightly. To see

²⁸ See e.g. Daly and Wilson, *Homicide*; Aaron Sell, "Applying Adaptationism to Human Anger: The Recalibrational Theory," 2005.

²⁹ For a review of neurological evidence that perception of emotional expressions is mediated by simulation, see Alvin I Goldman and Chandra Sekhar, "Simulationist Models of Face-Based Emotion Recognition" 94 (2005): 193–213, doi:10.1016/j.cognition.2004.01.005.

this, suppose that there were a race of aliens without faces and without an analogs of human anger (perhaps their ancestors were subjected to different evolutionary pressures). Now, if these aliens ever made contact with humanity, they would eventually be confronted by the pervasiveness of human nonverbal communication via facial expressions. In their superior wisdom, they might decide to don dynamic masks that conveyed human facial expressions of anger under contextually appropriate circumstances (e.g. when making a threat or when responding to something to which they take offense). Moreover, we can imagine this aiding their attempts to communicate with us (e.g. by helping us to see more directly what an alien finds offensive or when an alien is making a threat). If this occurred, no one would say that the aliens were always feigning anger whenever angry expressions flickered across their masks. Even though they are incapable of manifesting the emotion with the full authenticity of a human, it would be incorrect to say that they were always faking it.³⁰

I think the case may be similar for God. If, for instance, there is no division between God's will and passion, then God cannot truly be in the grip of an irruptive motivational state like anger. Thus, like the aliens, God's use of the language of anger (e.g. to exhort or to convey offense) cannot be disingenuous in the same way that human beings can fake anger. Thus, there is going to be a clear sense in which expressions of divine anger in scripture do not constitute faking.

A graver concern is that scriptural expressions of God's anger are making empty threats of punishment. Nevertheless, the evocative interpretation does not carry any such implication. Consistent with this reading, God might still punish sin, but punish in order to secure some good

³⁰ I say they are not "always" faking because one could imagine that the aliens' masks were also under direct voluntary control and that the expressions could sometimes be manifested in order to manipulate and control an audience instead of conveying a response to human actions that are actually perceived as an offense to the aliens.

outcome (e.g. deterrence). In effect, punishment could be justified by the value of enforcing a threat (or by the moral right to do so³¹), where the threat itself is intended to secure the deterrent effect. If so, then punishment in hell is justified, but it is still implausible to suppose that eternal punishment is necessary to achieve this deterrent effect.³²

Importantly, if we reconceive the intended effect of these passages (as evocative rather than informative), then these passages no longer support R. That is, we have no independent reason to suppose that punishment has non-derivative value. There is a clear sense in which God can make use of the language of anger and desert without actually being motivated to pursue punishment as an end in itself. Thus, on this interpretation, R remains without evidential support.

Ultimately, this is the view I favor: scripture does not offer any independent reasons to believe R or to trust our retributive inclinations. Insofar as one accepts this view, it follows that one should not accept the traditional view of hell. If one lacks reason to believe R, then, *ceteris paribus*, one lacks reason to believe that punishment in hell is justified (as traditionally conceived). Thus, one has reason to doubt the Punishment Thesis (that hell is a place of punishment); or to doubt the No Escape Thesis (that people in hell cannot be redeemed); or to doubt the Anti-Universalism Thesis (that in the end, some will be consigned to a place of eternal torment); or the Eternal Existence Thesis. There are voices in the Christian tradition that take each of these doubts as a cause for revision. In denial of the Eternal Existence thesis, many have endorsed annihilationism. In denial of the No Escape Thesis or the Anti-Universalism Thesis, others have adopted universalist positions. Others have begun to envision the rejection of the

³¹ See e.g. W Quinn, "The Right to Threaten and the Right to Punish," *Philosophy & Public Affairs*, 1985.

³² This raises the question of why, on this reading, there are scriptures that seem to claim that hell is eternal. Personally, I have doubts about whether this is the correct reading of any of these passages. For a discussion of these texts and the language therein, see the last section of T Talbot, "Three Pictures of God in Western Theology," *Faith and Philosophy*, 1995.

Punishment Thesis, which is perhaps most central to the traditional conception of hell. Those who reject this thesis usually understand hell as the consequences of an individual's choice to be separated from God.³³ While the landscape is ever changing, these are minority viewpoints, and as I see it, this is the main drawback of this interpretation.

Concluding remarks

My purpose here was to pose a new problem for the traditional view of hell and to lay out a few of the most promising ways I see of evaluating to it. While my preferred approach is among them, my intention is not to give that view a compelling or adequate defense. Instead, I argued that if punishment in hell (as traditionally conceived) is justified, a retributive principle (R) must be true, and once we consider the evolution of retribution, the main reasons to believe R come from scriptures pertaining to divine wrath. This is because there are plausible evolutionary explanations for why human beings would find R compelling. If true, those explanations would undercut the main reasons for accepting it, leaving punishment in hell without a plausible justification (outside of Christian scripture). The retributive principle may receive independent support from Christian scripture depending on how one interprets scriptures having to do with divine wrath and also hell. I raised four different possibilities for evaluating R scripturally and briefly discussed the implications of three of these possibilities. On two these interpretations the language of God's wrath is informative: it communicates information about God's nature. These interpretations either have implausible implications (on systematic grounds) or are unable to fully explain attributions of God's wrath. Finally, I suggested that the language

³³ E.g. CS Lewis, *The Great Divorce*, 1946; E Stump, "Dante's Hell, Aquinas's Moral Theory, and Love of God," *Canadian Journal of Philosophy*, 1986; Kvanvig, *The Problem of Hell*.

of God's wrath may instead be evocative: it evokes certain responses in us. Among these responses are repentance, submission, awe, and worship.

If this is so, then we can begin to envision an alternative to the picture of worship with which we began. On that view, appeasement of God's wrath is a central aspect of worship. But if the purpose of wrathful expressions is to evoke responses in us, then there are many alternatives to appeasement that we might explore. For instance, we can respond with awe and respect for an almighty God who does not tolerate sin and who annihilates injustice and restores God's creation.