

Punishing and Atoning: A New Critique of Penal Substitution

Brent G. Kyle

I. Introduction

It is central to the Christian faith that humans are separated from God by sin but reconciliation is made possible through the life, death, and resurrection of Christ. But how exactly do Christ's life, death, and resurrection reconcile us with God? Answering this question is the job of an atonement theory. A variety of such theories have been proposed, but there is much debate among theologians and philosophers as to which, if any, is correct.

Nevertheless, one particular atonement theory—the doctrine of penal substitution—has gained many adherents. This doctrine claims that it was good, or perhaps required, for God to punish in response to human sin, and that Christ received this punishment in our stead. By virtue of Christ receiving our punishment, we are reconciled with God.

The doctrine of penal substitution is accepted by a number of church traditions, theologians, and philosophers. Take evangelicalism. One of the most influential evangelicals in North America, J.I. Packer, claims that the belief in penal substitution “is a distinguishing mark of the world-wide evangelical fraternity” (1974, 3). Consider also the Reformed tradition. A version of penal substitution is stated within *The Belgic Confession*, a doctrinal standard in many Reformed churches:

We believe that God, who is perfectly merciful and just, hath sent his Son for to take upon him that nature, in which the disobedience was committed: that so in it he might give satisfaction for, and suffer the punishment of sin by his severe sufferings and Death (De Brès 1689, 6).

The doctrine of penal substitution is also stated within the influential *Heidelberg Catechism*, which is honored by many Reformed churches as well as the Presbyterian Church USA:

It is his righteousness which requires that sin committed against the supreme majesty of God be punished with extreme, that is, with eternal punishment of body and soul.

The catechism then goes on to stress that Christ fulfills this requirement (Péry 1963, 35-6). In addition to winning support among many churches, the penal substitution view has also won over great theologians, like Martin Luther¹ and John Calvin (1977, 504, 509-510), as well as contemporary philosophers, like Oliver Crisp (2008; 2009) and Steven Porter (2002; 2004). The gist of their views will be explained in sections II and III.

Despite its prominence, I believe there is a fundamental problem with the doctrine of penal substitution. Whereas most criticisms focus on whether it would be *moral* for God to punish the innocent Christ (e.g. Lewis 1997), my focus is different. I argue that God did not punish Christ *at all*. The doctrine's central factual claim—that Christ was punished by God—is mistaken. In the final section, I show that my criticism of penal substitution does not apply to a closely-related atonement theory, the Anselmian satisfaction view.

II. Preliminaries.

What is the doctrine of penal substitution? In general, atonement theories begin by claiming that human beings are encumbered by the problem of sin. According to many views, we have wronged or unjustly taken something away from God by sinning, and this has resulted in a distorted relationship between us and God. But this distorted relationship must get corrected *somehow*, and there are many different atonement theories that purport to explain how this occurs. But the penal substitution view claims that the distorted relationship between God and humanity was corrected (at least partly) in virtue of Christ voluntarily meeting a specific penal requirement that God has imposed in response to sin. This penal requirement is the requirement that there must be punishment for sin (e.g. physical and spiritual death or even eternal damnation). Indeed, this requirement has been fulfilled, according to the doctrine of penal substitution. But it is not *us* who have fulfilled it. Christ met the requirement in our stead. By enduring the cross, Christ was punished on our behalf.

¹ For this characterization of Luther's view, see Grensted (1920, 198-209).

The doctrine of penal substitution involves two central claims that need to be explained. I will call them the ‘normative’ and ‘factual’ claims. The normative claim states a *reason* for God having imposed a penal requirement in response to sin. This normative claim comes in at least two different versions—a strong and weak version. Calvin’s strong version of the normative claim states that God was necessitated to punish in response to sin. This necessity to punish may be said to follow from the fact that God is a “righteous Judge” (Calvin 1977, 504). However, the weak version of the normative claim, proposed by Stephen Porter, holds that God has in fact decided to punish in response to sin although he was not necessitated to do so. This leads one to ask: Why has God decided to inflict punishment for sin? And Porter argues not only that it’s permissible for God to punish in response to sin, but also that “it is good that God punish sinners...” (2002, 602). The supposed goodness of punishment provides a reason for God to punish, and also explains why God has decided to punish even though he was not necessitated to do so.

How do we pick between these two versions of the normative claim? My overall argument will apply equally to either version. So, for present purposes, I shall bracket Calvin’s stronger version of the normative claim and focus instead on Porter’s weaker version:

Normative Claim: It is good for God to exact punishment in response to human sin.

Since God is perfectly good, the normative claim would initially lead us to expect that God will punish *us* for our sin. But this is where the factual claim becomes important. The factual claim states the substitutionary element of this doctrine of atonement. Christ has taken our place in receiving the punishment for sin. And he has done this voluntarily. The factual claim states,

Factual Claim: God exacted punishment on the innocent Christ, who voluntarily received it in the place of sinners.

According to the factual claim, Christ has satisfied our penal requirement by receiving punishment instead of us. Together, the factual and normative claims are central to the doctrine

of penal substitution. They provide an account of the means through which we are reconciled with God.

My concern in this paper is solely with whether the factual claim is true. I argue (in sections IV and V) that the factual claim, under its proper interpretation, is mistaken. The basic idea behind my argument is this: In order to punish someone, one must believe the recipient is responsible for an offense. But God surely did not believe the innocent Christ was responsible for the offense of human sin. So, the factual claim is mistaken.

Before we get to that argument, we must first concern ourselves with how to interpret the factual claim. As I will show, the factual and normative claims are ambiguous due to fact that the term ‘punishment’ has two meanings. But the doctrine of penal substitution requires that these claims be interpreted in accord with only one of the two meanings.

III. Punishing vs. Exacting Forms of Punishment.

It is possible to distinguish between two different senses of the term ‘punishment’—a primary sense and a secondary sense. In the primary sense, ‘punishment’ refers to a particular act whereby one intends to harm someone in response to an offense. In a moment, I will provide a necessary condition for punishment in this primary sense of the word. But I first want to provide a rough explication of a secondary sense of the word ‘punishment’, and to say why this secondary sense cannot be featured within the correct interpretation of the normative and factual claims.

In its secondary sense, the term ‘punishment’ is synonymous with the phrase ‘form of punishment’. To say that something is a punishment, in this secondary sense, is simply a looser way of saying that it is *a form of* punishment. Consider some types of treatments that we feel perfectly comfortable calling ‘punishments’—being fined, imprisoned, sent to the electric chair, placed on probation, and forced to do community service. Even obsolete treatments can rightly be call ‘punishments’, such as being hanged, flogged, beheaded, or crucified. We also mustn’t

forget the punishments within non-legal contexts, such as when parents ground or spank their children. The list goes on. These treatments are naturally referred to as punishments, though, if we were precise, we would call them ‘forms of punishment’. In this secondary sense of ‘punishment’, the word refers not to any particular action but to a general *type of treatment*. Let us define the secondary sense roughly as follows:

A type of treatment T is a punishment, in the secondary sense, if and only if T is widely recognized as a way of punishing.

In the definiens, ‘punishing’ is used in the primary sense mentioned earlier. So, naturally enough, the secondary sense is defined in terms of the primary sense. No doubt, this definition is rough, but it will suffice for present purposes.² As stated, the account seems able to explain why the above mentioned treatments are correctly called ‘punishments’ while other treatments (e.g. tickling, hugging, etc.) are not correctly so called.

It is important not to confuse the two senses of the term ‘punishment’. So, I will henceforth use the more precise phrase ‘form of punishment’ to express this secondary sense, and I will reserve the terms ‘punish’ and ‘punishment’ for the primary sense.

The two senses of ‘punishment’ can come apart in a very intuitive way—that is, there are cases where one sense of the word will apply while the other does not. For example, a sadist can inflict a *form of punishment* (e.g. flogging) onto a masochist, even though the sadist doesn’t really *punish* the masochist. Here we have the secondary sense without the primary sense. But the opposite is also possible—there are cases in which the primary sense applies but the secondary sense does not. In other words, one can be *punished* without undergoing something that could be

² Since T needs to be *recognized* as a way of punishing, there also needs to be a population that recognizes it as such. Which population should that be? The natural way of dealing with this question is to say that punishment, in the secondary sense, is relativized to a certain population, and that this population can change based on which population speakers are dealing with. For our purposes, we can relativize the account to our general population. But it is worth noting that the account would yield different results if we shifted to other possible populations—perhaps tickling is a punishment relative to Martian population. I believe this relativization is a desirable feature of the account. If we found out that Martians widely recognized tickling as a way of punishing, we would not plainly admit that tickling is a punishment after all; nor would we say that the Martians are wrong to recognize tickling as a punishment. We would instead say that tickling is a punishment *for Martians*, but not *for us*.

rightly called a *form of punishment*. For example, when nineteen-year-old Ryan Wood was late for his work at a sporting goods store, his employers punished him as follows:

Ryan was obliged to spend the morning, from 9:45am to 12 noon, dressed as a mannequin and taking his place in the window of the shop (*Emigrant Online: News for the Global Irish Community*).

Being required to pose as a mannequin is not a form of punishment, though Ryan Wood was punished nonetheless.³ So, although the concept of *punishment* and the concept of a *form of punishment* are naturally associated, they are quite distinct.

Assuming there are two meanings of the term ‘punishment’, let me now make two interpretive points regarding the doctrine of penal substitution.

The first is that the normative claim would be highly implausible if its occurrence of ‘punishment’ expressed the secondary sense. To see why, let us first reflect on what reasons are typically provided in favor of the normative claim.

The normative claim takes root in the common belief that there is a fundamental link between the goodness of an authority, on the one hand, and whether or not she punishes the guilty, on the other.⁴ More specifically, there is some temptation to think, as Steven Porter does, that it is generally good for authorities to punish the guilty. However, even if this is true, there appears to be no similar reason for thinking that it is generally good for authorities to exact forms of punishment onto the guilty (as opposed to equally bad treatments that would not count as forms of punishment). Consider what Porter says in support of the “moral worth or fittingness” of punishing wrongdoers. In support of this, he cites the familiar utilitarian ends of punishment, such as deterrence, rehabilitation, and prevention. And he also cites a few “intrinsic ends” that seem to justify punishment. Porter explains these intrinsic ends as follows:

For to demand that a wrongdoer suffer the loss that he deserves takes the harm done with due moral seriousness; it treats the wrongdoer as a responsible moral agent; and

³ Certainly it is possible for this sort of treatment to become a form of punishment, if it came to be recognized as a way of punishing. But I am assuming (plausibly) that it is not in fact recognized as such.

⁴ Of course, this is only a putative link. Some people argue that legal punishment is always impermissible. For example, see Boonin (2008) and Barnett (1977, 279-301).

it expresses the value of the victim as well as the value of the personal relationship involved (2004, 234).

Supposing Porter is right in citing these various ends as a justification for *punishing*, we can see that such features do not apply similarly as justifications for exacting a *form of punishment*. That is, the utilitarian and intrinsic ends do not seem to support the claim that it is in general good or morally fitting for authorities to exact forms of punishment onto the guilty (as opposed to equivalent, though unorthodox, treatments). Regarding the utilitarian ends, there seems no reason to deny that forcing an employee to pose as a mannequin, for example, would deter and prevent employee lateness just as well as a regular form of punishment, such as deducting pay.⁵ And regarding the intrinsic ends, it is clear that a punishing authority does not need to use a form of punishment in order to treat the wrongdoer as a responsible moral agent or to express the values mentioned by Porter. It even seems possible that an unorthodox treatment will accomplish these ends more effectively. For example, instead of grounding a misbehaving teen, a mother may require her daughter to spend Friday evenings visiting her grandmother. Thus, even if we grant that the utilitarian and intrinsic ends can establish that it is good for authorities to *punish* the guilty, those very ends do not also establish that it is good for authorities to exact *forms of punishment* onto the guilty (as opposed to equivalent, unorthodox treatments).⁶

Moreover, there is reason to think it is sometimes *bad* or *morally unfitting* for an authority to exact a form of punishment (as opposed to an equivalent, unorthodox treatment). There is a deep contingency in what counts as a form of punishment. It depends (at least partly) on what kinds of treatments we recognize as ways of punishing. But it is highly plausible that some conventional forms of punishment are immoral, such as those that involve torture or degradation. So, we cannot accept the claim that it is in general morally fitting or good for an authority to exact forms of punishment.

⁵ Would requiring a person to pose as a mannequin rehabilitate that person? It's not clear, but it certainly could be *as effective as* a regular form of punishment, such as pay deduction.

⁶ Certainly authorities are morally required, when punishing, to exact a degree or kind of harm that "fits" the offense, but that does not mean they should choose a form of punishment for exacting the harm as opposed to some equally bad, though unorthodox, treatment.

To sum up this first interpretive point, I've noted that there is a possible interpretation of the normative claim where 'punishment' takes its secondary sense:

(NC-Secondary) It is good for God to exact a *form of punishment* in response to human sin.

But the normative claim is rather implausible when interpreted in this way. It appears that, if God were to punish, there would be a wide range of possible treatments that would be equally good for God to exact, some of which are not forms of punishment. Moreover, it might even be morally bad for God to exact some forms of punishment (e.g. degrading forms of punishment). So, by contrast, the interpretation where 'punishment' has its primary sense is the more plausible way of understanding the normative claim:

(NC-Primary) It is good for God to *punish* in response to human sin.

Most likely, this is the interpretation intended by the proponents of penal substitution.

My second interpretive point can now be made. If the normative claim is to be interpreted as (NC-Primary), then the factual claim needs to "match" that interpretation. In other words, the factual claim also needs to be interpreted according to the primary sense:

(FC-Primary) God has *punished* the innocent Christ, who voluntarily endured this in the place of sinners.

Here, the word 'punished' is used in the primary sense so as to denote a particular act of God intending to harm Christ in response to human sin. The alternative interpretation of the factual claim, where we incorporate the secondary sense, would not be sufficient to account for the way in which Christ reconciles us with God. Such an interpretation would be,

(FC-Secondary) God has exacted a *form of punishment* onto the innocent Christ, who voluntarily received it in the place of sinners.

I am not concerned to assess whether this claim is true. I only wish to point out that the combination of (NC-Primary) with (FC-Secondary) is not sufficient to account for Christ's atoning work. After all, (NC-Primary) leads us to expect that we will be punished for our sin, and (FC-Secondary) provides no reason whatsoever to give up that expectation. Simply put,

(FC-Secondary) does not reveal any way in which Christ has fulfilled the penal requirement that (NC-Primary) leads us to expect from God. Thus, (NC-Primary) and (FC-Secondary) are a very poor combination, since they do not account for the way in which Christ reconciles us with God. The penal substitution theory needs (FC-Primary)—the claim that God has punished Christ in response to sin—in order to account for our reconciliation with God. I shall therefore assume that (FC-Primary) is the correct interpretation of the factual claim.

There is one more ambiguity in the factual claim that has yet to be clarified. According to the customary understanding of the trinity, God and Christ are *distinct persons* who are of *one being*. So, ‘God’ and ‘Christ’ can either refer to two distinct persons or they can refer to the one and the same being. How then should we understand these terms within the factual claim? On the traditional way of understanding the factual claim ‘God’ and ‘Christ’ refer to distinct persons, the former to God the Father and the latter to Christ the Son. Here, the factual claim refers to the act of one person, God the Father, who punishes another person, Christ the Son.⁷ For the most part, I assume this traditional interpretation of the factual claim. However, I will later address the possibility of an unorthodox interpretation, where ‘God’ and ‘Christ’ refer to one and the same being.

I will proceed by regarding (FC-Primary) as the correct interpretation of the factual claim, and will largely assume that ‘God’ and ‘Christ’ refer to distinct persons. And I shall now argue that the factual claim is false under this construal. Although Christ underwent a horrific *form of punishment*, he was not *punished* by God in the primary sense of the term.

IV. A Necessary Condition for Punishment.

There is a necessary condition for the primary concept of punishment that falsifies the factual claim of the penal substitution theory. To start off, let’s assume that punishment involves the imposing of harm onto someone. This assumption is helpful in that it gives us a springboard for

⁷ See Swinburne (1989, 151-52) for confirmation of this as the traditional interpretation.

discussing various examples.⁸ It is certainly possible that both Ryan Wood and the masochist were harmed (e.g. humiliated, pained, etc).⁹ What then accounts for the fact that Ryan Wood was punished whereas the masochist was not? What makes up the difference between these two cases?

An initial suggestion is that, unlike the case of Ryan Wood, there was presumably no offense or crime in the masochist case. There are many instances of harm which would not count as punishment because there is no offense. For example, take a father who comes home from a bad day at work and spanks his child just to “let off steam,” knowing full-well that his child did *nothing* wrong. Although this may be abusive, it certainly does not count as punishment. Even though the father exacts a *form of punishment* that harms his child, he is certainly not *punishing*. Punishing always involves an offense in some way or other.

But that by itself is not enough. It also seems clear that the punishing authority must at least *believe* that there was an offense. For instance, we could suppose that the child in the previous example had skipped school unbeknownst to his father. And, just as before, the father had a bad day and spanked his child purely to “let off steam,” knowing nothing about the child’s misbehavior. This too does not seem to be punishment. It seems that the authority must (at least) believe there was an offense.

The penal substitution theory has no trouble accounting for what has been said so far, since it’s plausible that God believes (and knows) there is an offense—namely, the offense of sin. However, there appear to be problems for this view if we continue to specify this requirement.

In particular, it’s not enough merely to say that a punishing authority must believe *there was an offense*. It also seems that the punishing authority must believe *that the intended recipient is responsible for the offense* (i.e. to be blamed for the offense). Suppose the angry father was upset

⁸ It is worth noting that the condition I end up proposing in no way relies on this assumption. The assumption is merely useful for thinking about examples.

⁹ Oftentimes the masochist example is used as a problem for the harm requirement, but that would depend on how we envision the example itself. All I am claiming here is that there are ways of envisioning the masochist example such that the masochist is harmed, and this wouldn’t change our opinion about the fact he/she is not punished.

because he had his wallet stolen earlier that day. To his frustration, the thief was never caught. Just as before, however, the father comes home and spans his child purely to “let off steam” knowing full-well that his child was not the person who stole his wallet. This too does not seem to count as punishment even though the father believes there was an offense. This is due to the fact that the father does not believe his son *committed* the offense. So, to be precise, it seems that the authority who imposes the harm must at least believe its intended recipient committed the offense.

However, we must be careful about settling for this version of the condition. A person can be punished for an offense, even though he is not believed to have committed it. For example, a man who pays a hit-man to murder a sworn enemy could be punished for the murder itself. In this case, the man is believed to be responsible for *the fact that the murder was committed*, even though he is not believed to have committed the murder itself. So, a more careful statement of this condition would say that an act of punishment occurs only if the authority believes that the recipient is responsible for the offense having been committed. Furthermore, we can imagine cases in which the recipient of punishment is not *fully* responsible for the offense having been committed. A partner-in-crime may be believed to share only partial responsibility. So, we should only insist that the authority believes the recipient is *at least partly* responsible for the offense having been committed.

This seems necessary, but there may be a temptation to strengthen this condition even further. Certainly not all *believed-to-be* offenders are *actual* offenders. Instead of merely requiring that the authority *believe* the recipient is responsible we might insist on the stronger claim that the authority *know* the recipient is responsible. However, there’s a problem with this strengthened condition. Since knowledge entails truth, this condition entails the logical impossibility of punishing the innocent.¹⁰ But this implication is counterintuitive. There seems to be nothing

¹⁰ A.M. Quinton argues that punishing the innocent is logically impossible since guilt is a necessary condition for punishment (1954, 136-139). Porter (2004, 236) rejects Quinton’s claim that it is logically impossible

logically incoherent about the prisoner’s catchphrase “I’m being punished for a crime I did not commit.”¹¹ This provides *prima facie* reason for doubting the stronger version. Nevertheless, the weaker version will suffice for present purposes.

This completes our search for the desired condition on punishment. We can conclude as follows (let *P* be a person or group of persons, and let *S* be the recipient of punishment):

P punishes *S* for (supposed) offense *O* only if:

- (1) *P* believes that *S* is at least partly responsible for *O* having been committed.^{12, 13}

This condition accounts for the fact that Ryan Wood was punished, while the masochist and the child of the angry father were not. Most importantly, the condition has profound implications for the doctrine of penal substitution. In general, condition (1) could never be fulfilled when *P* is an omniscient being (e.g. God) and *S* is completely innocent (e.g. Christ). Surely God did not believe that Christ was at all responsible for human sin having been committed. That God did not hold such a belief is obviously true. Since he is omniscient, God does not hold any false beliefs. Therefore, he did not falsely believe that Christ was at all responsible for human sin having been committed.¹⁴ So, the factual claim is false on *a priori* grounds. It’s not the case that

to punish the innocent. But, as I mention, there is no need to rely on this strong claim. So, Porter’s rejection does not apply to my argument.

¹¹ I do not take this to be a decisive objection to the proposed condition. It might be that the prisoner’s catchphrase needs to be rephrased as “I’m receiving a *form of* punishment for a crime I did not commit.” Still, since the objection provides some reason for doubt, perhaps a weaker version would be in order.

¹² This condition is similar to a condition proposed in Boonin (2008, 25).

¹³ It’s worth pointing out that condition (1) can be given both *de re* and *de dicto* readings. I lean towards the *de re* reading for the following reason. Suppose an uninformed legal authority punishes a person who is in fact the Prince of Wales, although the authority is completely ignorant of this person’s princely status. We should be able to say truthfully that this authority punished the Prince of Wales. However, in such a case, the authority probably does not believe *de dicto* that the Prince of Wales was at all responsible for the offense. But he would likely hold the *de re* belief—of *that man* who happens to be the Prince of Wales—that he is responsible for the offense in question. So it seems that a *de re* reading of condition (1) might be more suitable: *P* punishes *S* for offense *O* only if *P* believes—of the recipient, who happens to be *S*—that he/she is (at least partly) responsible for *O* having been committed.

The *de re* and *de dicto* readings make a difference for certain substitution cases. In *A Tale of Two Cities* (1992, 443-58), Charles Dickens tells of a man, Darnay, who is accused of treason and sentenced to death, but whose look-alike Sidney Carton faces the guillotine in his stead. What does condition (1) tell us about whether Carton was punished? On the *de dicto* reading, condition (1) tells us that Carton was not punished. The authorities clearly did not believe that *Carton* was responsible for the offense of treason having been committed. They only believed this of Darnay, and had no concern for Carton. On the *de re* reading, however, it is allowed that Carton was punished. Since the Carton/Darnay exchange was covert, it’s plausible that the authorities may have believed—of the person they sent to the guillotine, who happened to be Carton—that he committed treason. So, under the *de re* reading, condition (1) allows us to say that Carton was punished on Darnay’s behalf. And this seems to be an intuitive result of the *de re* reading.

¹⁴ Although the *de re/de dicto* distinction makes a difference for some substitution cases (see footnote 13), it doesn’t make a difference for the one where Christ substitutes for humanity. It’s clearly not possible that an

God punished the innocent Christ in the place of sinners. It is not the case that God punished Christ *at all*.¹⁵ Certainly Christ underwent the most horrific *form of punishment* in his crucifixion. But, as I've suggested, that is an importantly different matter.

I have so far assumed the traditional interpretation of the factual claim, where 'God' and 'Christ' refer to distinct persons. But, according to the doctrine of the trinity, there is also a sense in which God and Christ are of the same being. So, it is possible to interpret the factual claim as referring to an instance of self-punishment—God punishes God's self.¹⁶ This interpretation is unorthodox. But does it avoid the objection? No. The factual claim is falsified by condition (1), even under the present interpretation. When *P* and *S* are identical, condition (1) says that the punisher must believe that *he himself* is responsible for the offense having been committed. But, for the same reasons as before, God surely does not believe that God is responsible (i.e. to be blamed) for human sin.¹⁷ So, God does not punish God's self. Thus, even if we interpret the factual claim in this unorthodox way, it is still falsified by my view that condition (1) is necessary for punishment. For the remainder of this essay, I will assume the traditional interpretation of the factual claim, where 'God' and 'Christ' refer to distinct persons.

V. Potential Rejoinders.

How might the proponent of penal substitution respond? Initially, one might contest my claim that condition (1) is not fulfilled by God with regard to Christ. This response might take root in Calvin's claim that "the guilt that held us liable for punishment has been transferred to the head

omniscient God could have held the *de re* belief—of the innocent Christ—that he was (even partly) responsible for human sin having been committed. Nor is it possible for an omniscient God to hold the corresponding *de dicto* belief.

¹⁵ Incidentally, we can see that even the relevant *human authority*, Pontius Pilate, did not punish Christ in the primary sense of the term. The textual evidence clearly states that neither Pilate nor Herod believed Christ was guilty. See Luke 23:4, 23:14-15, and 23:22 in *The Holy Bible: New Revised Standard Version* (1998).

¹⁶ The possibility of self-punishment is sometimes used to present a challenge to certain definitions of 'punishment' (for discussion, see Boonin 2008, 23-4). However, this possibility only presents a problem for the common assumption that the punisher must be an authority over the recipient, which is impossible if the punisher is identical to the recipient. Nonetheless, I do not rely on this assumption, and do not regard the possibility of self-punishment as a problem for condition (1).

¹⁷ Some of God's actions, as creator and sustainer, are causally necessary for human sin to occur. But this is not the relevant sense of 'responsibility', which, as noted parenthetically, is tied up with blameworthiness.

of the Son of God” (1977, 509-10). If this transference is regarded quite literally, then a proponent of such a view may want to insist that Christ really was guilty of sin and that God believed this to be true.

But this is a very perplexing rejoinder. I’m not sure we can make any sense of a literal transmission of guilt. And some proponents of penal substitution, such as Oliver Crisp, agree with me on this. Crisp writes,

... I cannot make any sense of this view as it stands. It seems to me that the person who has committed a particular crime is guilty of having committed it, and that liability to guilt cannot simply be transferred from one individual to another (2008, 218).

This quote echoes my own confusion about Calvin’s view. But, let’s suppose for the sake of argument that we can make sense of this literal transmission of guilt. Even so, there is an important tension between this transmission of guilt and a crucial part of the atonement theory. The penal substitution view takes Christ’s innocence as one necessary feature that makes him fit to atone for human sin. If Christ was not innocent, then he would also need to be atoned for and we could not be reconciled with God through him. But, insofar as we understand this transference of guilt as literal (i.e. in such a way as to make condition (1) come out true), then we are being inconsistent with this crucial claim about Christ’s innocence. On the other hand, if we wish to avoid inconsistency by understanding Calvin’s claim non-literally, then the original problem arises in that condition (1) goes unsatisfied.

A more promising strategy for defending the penal substitution view is to argue that we are not required to accept condition (1) as a necessary condition for punishment. There are (at least) three different ways of doing this.

The first is that one may state putative counterexamples to condition (1). Let me just mention a common way of proposing counterexamples and then indicate a general strategy that is often available in defense of condition (1). Many proposed counterexamples will be ones that do not meet condition (1), but indeed involve a person who clearly receives a form of

punishment. The fact that the person receives a form of punishment is then taken to suggest that the person is indeed punished in the primary sense of the term. For example, suppose a man is sent to the electric chair under the command of a corrupt dictator, even though this man was not believed to be responsible for any crime. We might imagine that the man was executed purely because he knows certain information which, if publicized, would cause an uprising. This does not fulfill condition (1), but, since he received a form of punishment, this may suggest that he was thereby punished. However, in response we must recall what the masochist case reveals: *One can receive a form of punishment without being punished.* This is a general reply available in response to many putative counterexamples. I can freely admit that this man received a form of punishment at the hands of the corrupt dictator, but this concession is consistent with saying that he was *not* punished. As long as this type of response is consistent, it may plausibly apply to very many putative counterexamples.

There's also a second strategy for trying to maintain the penal substitution view in face of my critique. Instead of outright rejecting condition (1), the proponent of penal substitution might merely claim that my argument in favor of condition (1) is inconclusive. Recall the initial case that supported condition (1)—the angry father case. This father had his wallet stolen and, to his frustration, the offender was never caught. He then went home and spanked his child, purely to “let off steam,” knowing full-well that his child had done nothing wrong. Earlier I claimed that the father clearly did not punish his son. But perhaps this intuition is not shared.

In response to this resistance, we can propose a simple revision to the angry father case so as to “tease out” the right intuition. We can replace the form of punishment with an unorthodox treatment that involves an equivalent level of harm. This alternative treatment would be one that is *not* generally recognized as a way of punishing. That is, it would not be a form of punishment. For example, we can simply retell the angry father case so that the child receives *ridicule* from his father rather than a spanking. Intuitively, the lack of a form of punishment makes this revised case seem very much *unlike* punishment. This sort of revision, in

which we replace the form of punishment with an alternative treatment, is perfectly acceptable given what we've learned from the Ryan Wood case: *One can be punished without receiving a form of punishment*. So, if one has the intuition that the original angry father case is an instance of punishment, then, unless one is confusing punishment with the exacting of forms of punishment, the original intuition should survive the above revision. The revised case, however, does not at all seem to be a case of punishment.

A third general strategy, which I will deal with at length, is to argue for a rival account to supplant condition (1), one that is friendlier to the penal substitution view. A proponent of penal substitution may argue that there are other conditions on punishment which, instead of (1), are able to provide the right results with respect to all the cases. So far, we have four cases on the table. The proponent of penal substitution needs to argue for a condition that predicts that the child of the angry father and the masochist are not punished whereas Christ and Ryan Wood are.

In order to pursue this strategy, we should first agree to a weaker version of (1), one that will at least rule out the masochist case (without excluding Christ's treatment or Ryan Wood's). Something like the following will do the trick:

P punishes person *S* for (supposed) offense *O* only if:
(1) *P* believes that there was such an offense as *O*.

Once this condition is accepted, we can *add on* some further necessary condition in order to rule out the angry father case. What should this further condition be?

One feasible candidate might be a reprobative requirement.¹⁸ This requires that the authority somehow expresses disapproval when punishing. To see the motivation for this requirement consider an example, due to David Boonin (2008, 23), of a gang who requires new members to break the law—say, by stealing a car—in order to be eligible for initiation into the gang. We can imagine that an authority within the gang inflicts painful treatment on the new

¹⁸ The phrase 'reprobative requirement' comes from Boonin (2008, 23).

member—say, by branding or tattooing him—precisely because the new member committed a crime. Clearly this infliction of harm does not count as punishment, even though the gang authority believes there was an offense. The salient feature in the gang case is that the authority is expressing *approval*, whereas punishment seems to involve an expression of *disapproval*.

But how should we state this reprobative requirement? Take the following as two preliminary attempts:

- (2a) *P*'s performing the act expresses disapproval of *S*.
- (2b) *P*'s performing the act expresses disapproval of *S*'s supposed offense *O*.

These certainly will not do. Whereas these versions of the reprobative requirement would correctly exclude the gang case, they do not help the penal substitution theorist at all, since Christ's treatment would not satisfy these conditions. It does not seem true that, in allowing Christ's death and suffering, God was expressing disapproval of Christ or of Christ's supposed offense.¹⁹ So these versions do not help, but there may be another way of stating the reprobative requirement without presenting further problems for the penal substitution view.

As far as I can tell, there is only one version of the reprobative requirement which, together with (1'), would account for all the cases we've discussed so far:

- (2c) *P*'s performing the act expresses disapproval of the type of offense under which *O* is classified.

This condition would correctly rule out Boonin's gang example, in which the gang authority expresses *approval*. It also appears to account for the lack of punishment in the angry father case. The father had his wallet stolen by a thief, and spanking his child clearly does not express disapproval of thievery. Moreover, the doctrine of penal substitution would not be troubled by this condition, because (2c) does not involve an authority disapproving of the *recipient* of punishment (who, in this case, would be Christ). The authority only expresses disapproval of the type of offense (e.g. sin), which is perfectly consistent with the penal substitution view.

¹⁹ Some might disagree with this claim. That is, some might say that God allowing Christ's suffering and death was an expression of disapproval of Christ. However, such people are now faced with the challenge of showing that the angry father's action toward his child does not also satisfy this condition. That is, in whatever way (2a) or (2b) does not exclude Christ's treatment, I suspect it also won't exclude the child's treatment.

There are problems, however. Other possible cases reveal that (2c) is not necessary for punishment. Consider an individual who would be expected to have great respect for the law, such as the mayor of a city. Imagine, strangely, that the mayor is convicted of marijuana possession. Furthermore, suppose that the local judge who tries the case is a former beatnik who believes that marijuana should be legalized. Consequently, this judge *normally* lets marijuana offenders “off the hook” with a less-than-stern warning and a “wink-wink.” However, since the mayor is expected to have higher regard for the law, the judge instead decides to deliver a moderate jail sentence for the mayor’s offense. Obviously, the mayor is punished in this example. Yet, the judge’s act does not express disapproval of the type of offense under which the mayor’s offense is classified—it does not express disapproval of marijuana possession. If the offender were anyone else, besides the mayor, the judge would have given him only a warning.²⁰ So, (2c) does not seem necessary. Thus, the reprobative requirement seems unhelpful upon further inspection.

Can the proponent of penal substitution look elsewhere? Perhaps (1') should be combined with a different condition that can account for the angry father case, without ruling out Christ’s treatment or Ryan Wood’s. In the pivotal case of the angry father, I construed the father as being frustrated and angry. Maybe it is necessary that the punishing authority is dispassionate with respect to the offense and the recipient of punishment.²¹

We need not go into detail about how exactly this condition is to be stated, because we can already see serious problems. First of all, it is unclear how this condition would help the traditional penal substitution theory of atonement, since this theory claims that the punishment for sin involves the bearing of God’s wrath. This punishment, at least on some understandings, is construed as the manifestation of anger and retribution for sin.²² But second, there does not

²⁰ And we can imagine that the judge makes it explicit, when handing down the mayor’s sentence, that a harsh sentence is being delivered, not because of the offense itself, but because it was the mayor who committed it.

²¹ For an argument for this sort of condition, see Gendin (1967, 236).

²² As Grensted puts it, the “general Penal Theory” states simply that “Justice demands the punishment of sin” and “the attitude of a just God towards the sinner can only be one of wrath” (1920, 204).

even seem to be a necessary condition in this vicinity. A punishing authority could clearly deliver a harsh sentence, perhaps the harshest sentence allowed by law, simply because she felt repulsed and angered by the offender and his heinous crime (e.g. rape, murder, or genocide). But it seems absurd to discount such cases as punishment, simply on account of the fact that the authority is impassioned. To get clearer on this intuition, simply consider a paradigmatic case of punishment, and then imagine further that the punishing authority is deeply angered by the offender or his crime. Surely, this mere addition would not disqualify it from being punishment.

In summary, none of the rejoinders I've discussed seem to rescue the penal substitution view. The original condition (1) seems to be the best account for the cases considered herein. In order for an act to count as punishment, the punisher must believe that the recipient is (at least partly) responsible for the offense having been committed. But it is clear that God, an omniscient being, did not falsely believe that Christ was (even partly) responsible for human sin.

If there are no other plausible options for resisting condition (1), then we must reject the penal substitution view. This theory may seem feasible upon first blush, but I suspect this feasibility trades on a conceptual confusion between *punishing* and *the exacting of forms of punishment*. As I have argued, God did not punish Christ in the primary sense of the term. Certainly, Christ received a horrendous *form of punishment* in his crucifixion, but that is not sufficient to account for the atoning work of Christ. Consequently, the doctrine of penal substitution should be abandoned.

VI. The Anselmian Satisfaction View.

In reaction to the above criticism, one could maintain that Christ's life, death, and resurrection has *somehow* reconciled us with God, but insist that the underlying nature of this reconciliation is an ineffable mystery (e.g. Torrance 1992). The first thing to note is that this reaction does not salvage the doctrine of penal substitution. Indeed, the reaction is fundamentally at odds with this doctrine. The doctrine of penal substitution is an attempt at expressing the underlying

nature of atonement, while the present view claims that this nature cannot be expressed. But second of all, it would be premature to appeal to mystery, merely in light of the failure of the doctrine of penal substitution, because there are other possible theories of atonement that are unaffected by my critique.

I cannot, in this paper, provide a conclusive argument in favor of any one atonement theory. But I will show that a similar theory, the Anselmian satisfaction view, which is also characterized in terms of punishment, can avoid my objection.²³ Thus, as I will argue, the mere fact that Christ was not punished by God does not mean that atonement cannot be characterized (at least partly) in terms of punishment.

The Anselmian satisfaction view is similar to the doctrine of penal substitution in that both views hold that God has imposed a requirement in response to sin, and both views characterize this requirement at least partly in terms of punishment. Here's how Anselm (1962) describes that requirement in *Cur Deus Homo* (henceforth *CDH*). Anselm holds that "it is not right to cancel sin without compensation or punishment". For "if sin is neither paid for nor punished, it is subject to no law". And in that case God would be passing over sin "undischarged" which would not be "fitting" for God to do (*CDH* bk I ch XII). As should be clear, Anselm holds that it is fitting for God to impose the following *disjunctive* requirement:

Anselmian Requirement: Either compensation for sin must be paid *or* there must be punishment.

We may understand 'compensation' as Anselm did, as the making of "restoration in some way satisfactory to the person whom [the sinner] has dishonored" (*CDH* bk I ch XI). Consider a legal analogue to the Anselmian Requirement. Jack is caught vandalizing a person's house with spray paint. As result, the judge requires the following: Either Jack must ensure that the house gets repainted to the owner's satisfaction *or* (if Jack fails to ensure this) he must serve a jail sentence.

²³ Richard Swinburne (1989, ch 10) proposes an atonement theory that closely resembles the Anselmian view.

The Anselmian Requirement is importantly different from the requirement associated with the penal substitution view. According to the doctrine of penal substitution, God imposes a *non-disjunctive* requirement:

Penal Substitution Requirement: There must be punishment for sin.

Under this requirement, there is no option to pay compensation for sin. And so, the only way for Christ to fulfill this requirement is for Christ to be punished.

By contrast, there are two ways in which the Anselmian Requirement can be fulfilled—through compensation and through punishment. According to Anselm, the first disjunct cannot be satisfied by ordinary humans, since any payment they could make is already owed to God:

If in justice I owe God myself and all my powers, even when I do not sin, I have nothing left to render to him for my sin (*CDH* bk I ch XX).

Thus, ordinary humans cannot pay compensation for sin, which makes it seem as if they are doomed to receive punishment. However, according to Anselm, the sinless Christ, who owes nothing, makes payment on behalf of humans. Regarding this “price paid to God,” Anselm writes,

If it be necessary, therefore, as it appears, that the heavenly kingdom be made up of men, and this cannot be effected unless the aforesaid satisfaction be made, which none but God can make and none but man ought to make, it is necessary for the God-man to make it (*CDH* bk II ch VI).

In short, the God-man, Christ, fulfills the first disjunct of the Anselmian Requirement so that humans aren’t doomed to fulfill the second.²⁴ So, the way Christ fulfills the Anselmian Requirement does not involve Christ being punished by God.²⁵ Instead, Christ pays the compensation that we owe to God (and cannot pay) *so that* we are permitted to go unpunished. In this sense, Christ saves us from punishment.

²⁴ Technically speaking, Christ doesn’t *fully* satisfy the first disjunct, since human penance is also required in order to make full satisfaction (*CDH* bk I ch XIX).

²⁵ As noted in footnote 23, Swinburne’s theory is very similar to that of Anselm. One thing to note is that, even though Swinburne speaks as if Christ received our punishment, he says it is “only punishment in the somewhat remote sense that it replaces the punishment which we should otherwise have deserved...” (1989, 162). So, Swinburne does not appear to be committed to the problematic view that Christ was punished by God.

Is forced compensation importantly different from punishment? In other words, are the two disjuncts of the Anselmian Requirement importantly different from one another? There are likely to be many differences,²⁶ but the most pertinent to our discussion is that condition (1) is necessary for punishment but is not necessary for forced compensation. More precisely, for one to force *S* to compensate the victim of offense *O*, it is not necessary that one believes that *S* is responsible for *O* having been committed. For example, suppose my grandparents were slave owners before I was born. Because of my grandparents' wrongdoing, I could be required to pay reparations to former slaves and their progeny. This would certainly count as forced compensation, even though no one believes that I am responsible for the offense (after all, everyone knows that I was born after the offense).²⁷ Thus, condition (1) is not a necessary condition for forced compensation. And this is at least one way in which forced compensation differs from punishment.

Most importantly for our purposes, Christ is not disqualified from paying compensation merely by the fact that he was not believed to be responsible for sin. So, my critique of penal substitution, which relies on condition (1) being necessary for punishment, cannot be applied directly, or even indirectly, to the Anselmian satisfaction theory.

In sum, the Anselmian satisfaction view is characterized (at least partly) in terms of punishment, but it does not fall prey to my critique because it does not claim that Christ was punished by God. Thus, my critique of penal substitution does not entail that atonement cannot be characterized in terms of punishment.

²⁶ At least one way in which they differ is that it's not necessary that forced compensation involve an intention to harm the wrongdoer, whereas an intention to harm the wrongdoer is plausibly a necessary condition for punishment (see, e.g., Boonin 2008, 12). Forced compensation necessarily involves an intention to restore the victim to her previous state, but it does not need to involve an intention to harm the wrongdoer (even though harm is foreseen in many cases).

²⁷ Even critics of reparations recognize that this type of situation can arise, and would be considered a form of compensation. For example, see David Horowitz's fourth argumentative point against reparations (2001/2010, 364).

I have not provided a full defense of the Anselmian view—doing so would involve an examination of a wider body of criticisms, which is beyond the scope of this paper.²⁸ I have only shown that Anselm’s view is immune to my critique of penal substitution. This conclusion is important because certain proponents of penal substitution (e.g. Porter 2002) have juxtaposed these two views and have argued that the penal substitution view comes out ahead. I have shown, by contrast, that there are important considerations (overlooked by these proponents) that weigh heavily against the penal substitution view but *not* against the Anselmian view.

The doctrine of penal substitution is unlike any other atonement theory in that it involves the claim that Christ was punished by God. I have argued that this particular claim should be rejected. If I’m right, then the doctrine of penal substitution should also be abandoned.

Brent G. Kyle
United States Air Force Academy

Bibliography

- Anselm (1962) *Cur Deus Homo*, in *St. Anselm: Monologium; An Appendix on Behalf of the Fool by Gaunilo; and Cur Deus Homo*, second edition, trans. Sidney Norton Deane, LaSalle, Ill: Open Court Publishing.
- Barnett, R. (1977) Restitution: A New Paradigm of Criminal Justice. *Ethics* 87, 279-301.
- Boonin, D. (2008) *The Problem of Punishment*, Cambridge, UK: Cambridge University Press.
- Calvin, J. (1977) *Institutes of the Christian Religion* vol. 1, trans. Ford Lewis Battles, Philadelphia: Westminster Press.
- Crisp, O. (2009) Original Sin and Atonement, in T. Flint and M. Rea (Eds.), *The Oxford Handbook of Philosophical Theology* (pp. 430-451), Oxford: Oxford University Press.
- Crisp, O. (2008) The Logic of Penal Substitution Revisited, in D. Tidball, D. Hilborn and J. Thacker (Eds.), *The Atonement Debate: Papers from the London Symposium on the Atonement* (pp. 208-227), Zondervan.

²⁸ See Porter (2002, 598-601) for criticisms of Anselm’s view.

- De Brès, G. (1689) *The confession of faith, of the Reformed Churches in the Netherlands*, Amsterdam: Printed by the widow of Steven Swart.
- Dickens, C. (1992) *A Tale of Two Cities*, New York: Random House.
- Emigrant Online: News for the Global Irish Community (2002) Wexford: Unusual Sanction for Lateness,
http://www.emigrant.ie/index.php?option=com_content&task=view&id=11504&Itemid=25,
 accessed February 13, 2011.
- Gendin, S. (1967) The Meaning of 'Punishment'. *Philosophy and Phenomenological Research* 28, 235-240.
- Grensted, L.W. (1920) *A Short History of the Doctrine of Atonement*, Manchester, U.K.: The University Press.
- Péry, A. (1963) *Heidelberg Catechism with Commentary*, Philadelphia: United Church Press.
- The Holy Bible: New Revised Standard Version* (1998), Oxford: Oxford University Press.
- Horowitz, D. (2001/2010) Ten Reasons Why Reparations for Blacks Is a Bad Idea for Blacks—and Racist, Too! Originally posted on frontpagemag.com, January 3, 2001, reprinted in Boonin and Oddie, *What's Wrong? Applied Ethicists and Their Critics*, Oxford: Oxford University Press, 364-366.
- Lewis, D. (1997) Do We Believe in Penal Substitution? *Philosophical Papers* 36, 203-209.
- Packer, J.I. (1974) What Did the Cross Achieve? The Logic of Penal Substitution. *Tyndale Bulletin* 25, 3-45.
- Porter, S. (2002) Rethinking the Logic of Penal Substitution, in W.L. Craig (Ed.), *Philosophy of Religion: A Reader and Guide* (pp. 596-608), New Brunswick, NJ: Rutgers University Press.
- Porter, S. (2004) Swinburnian Atonement and the Doctrine of Penal Substitution. *Faith and Philosophy* 21, 228-241.
- Quinton, A.M. (1954) On Punishment. *Analysis* 14, 133-142.
- Swinburne, R. (1989) *Responsibility and Atonement*. Oxford: Oxford University Press.
- Torrance, T. (1992) *The Mediation of Christ*, Edinburgh: T&T Clark.