Religious Studies (2016) **52**, 223–245
doi:10.1017/S0034412515000098

Anselm of Canterbury on the fall of the devil: the hard problem, the harder problem, and a new formal model of the first sin

WILLIAM WOOD

Oriel College, Oxford, OX1 4EW, UK
e-mail: william.wood@oriel.ox.ac.uk

Abstract: **The fall of the devil poses two distinct philosophical problems. Only one of those problems has received sufficient scholarly attention. The hard problem asks how the devil's choice to disobey God can be both suitably free and morally significant. The harder problem asks how it can be subjectively rational. Explaining the former does not suffice for explaining the latter. Drawing on the thought of Anselm of Canterbury, I develop a model of the first sin that uses the framework of consumer preference theory to show how Satan's act of disobedience can be free, morally significant, and subjectively rational.**

The fall of the devil poses the problem of how to explain the very first sin, the initial act by which evil comes to mar God's wholly good creation. On the traditional Christian account, Satan and the other angels were created completely good, in an environment that was completely good, and with intellects and wills that functioned exactly as designed by God. Nevertheless, Satan rebelled against God, an act so grave that he was justly condemned to Hell, where he suffers the torment of eternal separation from his creator. What can account for such an inexplicable choice? Given Satan's pre-fallen cognitive and volitional strengths, his sinful choice seems utterly perverse: an act of existential self-harm heightened to an almost infinite degree.

Though less frequently addressed, the problem of the first sin poses a challenge to traditional Christian theism that is no less grave than more widely discussed problems like the general problem of evil, the problem of divine hiddenness, the problem of other religions, and so forth. If we are forced to admit that God is responsible for the first sin, then God cannot be perfectly good. Alternatively, if we can only explain the first sin by positing some unintended defect in the

created intellect or will, or some unintended source of evil that already infects creation, then God no longer seems like a sovereign, all-powerful, creator.

It is easy to misunderstand the precise nature of the problem, and so it is easy to embrace various proposals that initially seem like solutions but really are not. One might think that the problem of the first sin is merely another instance of the problem of evil. That is, one might suppose that the question at hand is simply 'Why would God allow the devil to sin?' But the problem of the first sin requires us to justify *Satan's* choices, not God's. The relevant question is why Satan would choose to sin, not why God would allow him to sin. Christians traditionally explain the devil's sin by appealing to his pride, and to the ancient commonplace that Satan – and later Adam – fell because he loved himself more than God. Yet this suggestion, however venerable, merely redescribes the problem. Granting that Satan disobeyed God because he loved himself too much, the question remains: unless Satan's cognitive and volitional faculties are defective, how could he ever come to believe that he, a finite creature, is more worthy of love than his perfectly good creator? Satan's pride pushes the problem of the first sin back a step, but does not solve it (Brown (1978), 322; Willows (2014), 259).

In fact, the fall of the devil poses two distinct philosophical problems, and only one of those problems has received sufficient scholarly attention. The first problem is about the nature of free will and how Satan's choice to disobey God can be free, motivated, and morally significant. This is a hard problem to solve, since Satan's fall presents the riddle of free will in especially sharp relief (as I discuss below). Nevertheless, the riddle itself remains an instance of a standard philosophical puzzle that admits of standard solutions. The fall of the devil also poses an even harder problem, however. The harder problem is the problem of showing how the first sin can be subjectively rational, rational from Satan's own point of view. Once we have shown that it is possible for Satan to reject God freely, we still must explain why Satan – a completely good, properly functioning rational agent – *would* make such a choice, even if he could. Existing philosophical accounts of Satan's fall rarely address this problem at all. When they do, they tend to obscure its real force because they assume an account of Satan's preferences and desires that is imprecise and unhelpful. In order to make progress on the harder problem of Satan's fall, we need sharper conceptual tools.

Though they may seem out of place in the philosophy of religion, the tools of consumer-preference theory are well suited to the task of modelling rational behaviour. I aim to use these tools to model the fall of the devil and show how his choice to disobey God can be subjectively rational. First, however, I motivate my approach by disambiguating the hard problem of Satan's fall from the harder problem. Next, I present Anselm of Canterbury's account of the first sin. I turn to Anselm because he comes closest to solving both the hard problem and the harder problem. Although he gives us the building-blocks of a viable solution, he does not set them in a sufficiently precise conceptual framework, one that allows a fine-grained analysis of both the initial conditions of creation and

the relative strength of Satan's desires and preferences. I then argue that consumer preference theory offers just such a framework. Finally, I use that framework to develop my own model of Satan's fall, one that captures Anselm's key insights and shows how Satan's sinful choice to disobey God can be rational from his own point of view.

The hard problem: how can the devil's choice be free and morally significant?

Most commentators, ancient and modern, have treated the problem of the fall of the devil as a problem about the nature of free choice. They have accordingly sought to solve the problem by appealing to Satan's creaturely freedom even to reject his creator. On this telling, Satan's sin is simply a free choice against God, and once we have explained how that choice counts as both free and morally significant, there is nothing left to explain.

It turns out to be quite difficult to explain how Satan's choice against God could be both free and morally significant, however. The hard problem derives from the following dilemma: do Satan's desires, dispositions, and motivations causally determine his will, and therefore necessitate his sinful choice? Suppose the answer is yes. Then we face a further problem: on the assumption that a sinful choice must be caused by sinful desires, how did Satan acquire any sinful desires in the first place? Since he is a creature, Satan's initial desires, dispositions, and motivations must have been given to him directly by God at the time of his creation. Since God is the author of Satan's desires, this horn of the dilemma threatens to make God responsible for his fall. Alternatively, suppose that Satan's desires, dispositions, and motivations do not causally determine his will. Then we can safely absolve God of determining Satan's sinful choice, but that choice then seems utterly inexplicable. By hypothesis, Satan would have to make an evil choice even though all his desires, dispositions, and motivations were good. Such an event seems more like an accident that happened to Satan, rather than a deliberate choice for which he is responsible: a case of very bad moral luck. Thus the dilemma: either God is responsible for the devil's fall, or else the devil's freely chosen act of disobedience is utterly inexplicable. If the devil's choice is uncaused, it is unintelligible; but if it is caused, it can only be caused by God.

The harder problem: why would the devil choose to disobey god?

Most treatments of the fall of the devil – ancient, medieval, and modern – have focused on the previous dilemma. In my view, however, the devil's fall poses an even harder problem: the problem of how Satan's choice could be subjectively rational, rational from his own point of view. This harder problem also concerns Satan's moral motivation. Suppose that we give an account of free will on which

Satan's sinful choice had no prior determining cause. Suppose too that we successfully explain how Satan's choice nevertheless reflects his motivations and desires, and therefore counts as a genuine expression of his moral agency, rather than as a case of bad moral luck. Even so, we still must explain why a completely good, properly functioning rational agent *would* choose to disobey God, even if he can. A successful appeal to Satan's free will explains how his sinful choice is metaphysically possible, but it does not explain why it is subjectively rational. This is the harder problem of the fall of the devil.

A crude analogy: suppose a wealthy benefactor offers you the following 'dilemma'. You must either accept a huge sum of money with no strings attached, or else throw yourself out of a twenty-story window. Even if you are free to jump out of the window, why would you? More to the point, if we heard a story about someone who chose the window, we would surely not be satisfied with an explanation that appealed only to the brute fact that his choice was free. Nor would we accept the vacuous explanation that 'he just wanted the window more'. We would instead demand an explanation that presents this bizarre choice as intelligible and rational to the agent in question. Even a less loaded analogy makes the necessary point. Suppose the choice is not between an obvious good and an obvious evil, but between two genuine goods. Yet on the assumption that one of the goods in question is the good of obeying God, the sovereign creator and the source and summit of all value, then whatever the other good is, we would expect it to seem minor in comparison. The analogous choice might not be like the choice between the fortune versus the window, but it would be just as weighted. It would be like the choice between, say, the fortune versus a really nice piece of cake.

Accounts of Satan's fall that successfully address the hard problem can still founder on the harder problem. Consider, for example, Scott MacDonald's well-known account of 'primal sin' (MacDonald (1999)). On MacDonald's account, the first sin results from a failure of practical reasoning. The fallen angels choose some genuine, though lesser, good that they actually desire, and, as a result, they turn away from God and fall. They fail to attend to their belief that God is the highest good, and so that belief does not factor into their practical reasoning. We do not need to explain the fact that the angels fail to attend to their belief that God is the highest good, according to MacDonald, because 'failing to attend' is not a positive action, and therefore not the sort of thing for which a positive explanation is required. The fallen angels remain responsible for their sin, because they could have based their practical reasoning on their belief that God is the highest good, even though they did not. Their sin is uncaused, but it is not utterly random or unmotivated, because it does reflect their genuine desires. The fallen angels do desire the lesser good – indeed, the desirability of the lesser good is more salient to them at the time of their choice.

MacDonald's account is intriguing, but it leaves crucial questions unanswered. We would expect such failures of practical reasoning in fairly low-stakes situations.

For example, MacDonald offers the case of a man who goes for coffee with a colleague because he fails to attend to the fact that he had previously agreed to take his children to the park (MacDonald (1999), 128–129). Absent a much fuller explanation of Satan's motivations, which MacDonald does not offer, it is exceedingly difficult to see how Satan could similarly fail to attend to the goodness of God. Surely an unfallen angel, with a properly functioning intellect and will, would not so thoroughly fail to attend to God's goodness that he is able to commit a sin that warrants eternal punishment. Such an explanation of Satan's choice is like explaining the behaviour of a routine museum-goer who tries to steal the *Mona Lisa* from the Louvre by saying that he sees it, wants it, and then just happens not to attend to the fact that it is heavily guarded and that stealing is wrong. Absent further details, although this story is logically possible, it is so vanishingly improbable that we would not posit it as an explanation of the behaviour of an agent who is even remotely rational. We need to do better.

Anselm solves the hard problem

Anselm of Canterbury addresses the fall of the devil in a short dialogue called *De casu diaboli* (On the Fall of the Devil; henceforth: *De casu*).[1] Anselm inherits from Augustine a specific – and problematic – account of Satan's fall. Augustine's explanation for the origin of evil, in summary, is that even creatures created good have a tendency to fall back into nothingness unless they are continually given extra grace by God. Augustine's initially puzzling claim, in Book 12 of *The City of God*, that the first evil choice has no efficient cause but only a 'deficient cause' amounts to the claim that Satan's fall is 'caused', by the fact that God did not give Satan the gift of perseverance-in-the-good.[2] On Augustine's account, the fall of the devil is therefore caused not by anything present or really existing, but by a lack, an absence: the absence of divine grace. God gave the gift of perseverance to some angels, but not to Satan and his cohort, and as a result, some angels persevered in goodness, but Satan and his cohort did not. Augustine insists that God does not thereby became responsible for the devil's fall, since Satan's choice still satisfies all of Augustine's own (broadly compatibilist) conditions for freedom and moral responsibility: Satan's choice is not compelled by anything external to him and is caused by his own desires. Although Augustine has his defenders, I side with those who argue that his account renders the devil's sin intelligible only at the cost of making God responsible for his fall.[3]

Anselm improves on Augustine's account of the first sin with respect to both horns of the hard problem's central dilemma. First, he explicitly argues that God offered the gift of perseverance to all the angels, not just to the ones who remained good. Because God does not withhold grace from any of the angels, God does not even indirectly cause their fall. Second, Anselm gives a better account of the angels' initial inclinations, desires, or dispositions-to-will, which allows him to present a more persuasive account of the devil's sinful choice.

Anselm insists that God offered the gift of perseverance to all the angels, not just to the ones that ultimately chose to persevere in the good. He further insists that all the angels, including Satan, wanted to persevere. At first blush, this is a difficult line to defend, given that the devil did not actually persevere. If God wanted the devil to persevere, and offered him the gift of perseverance, and if Satan also wanted to persevere, how could it be the case that he did not persevere? In response, Anselm draws a series of clever distinctions. Strictly speaking, he argues, one can only be said to 'give' a gift when that gift is also accepted. Suppose that John tries to give Tom a horse. If Tom refuses to accept it, then even though we can say that John *offered* Tom the horse, we cannot say that John *gave* Tom the horse. So in the strict sense, it is correct to say (as Augustine and the wider tradition insist) that Satan fell because he lacked the gift of perseverance. It is even correct to say that he lacked it because God did not give it. Yet God did not give it only in the sense that they did not choose to accept it (*De casu* 3). Still, if Satan and the fallen angels wanted to persevere, why did they reject the gift of perseverance?

According to Anselm, the devil rejected the gift of perseverance because he preferred some other, incompatible good, and so he gave up the gift of perseverance in order to acquire that good instead. Anselm does not say what the incompatible good is, but it is clear that it is a highly desirable good that confers a considerable advantage on those who possess it. Indeed, according to Anselm, God later rewards the good angels, who persevered, with that same good. On Anselm's account, then, the fallen angels did not nihilistically reject goodness as such, nor seek some apparent good that was essentially undesirable or unfitting. Rather, they desired a genuine good that really would benefit them considerably, and moreover a good that God really did want to give them – eventually, but not yet (*De casu* 4–6).

Anselm still must explain the fact that Satan preferred the forbidden good. After all, his initial desires, dispositions, and motivations were given to him directly by God at the time of his creation. It seems that Satan can prefer the forbidden good only if God created him to prefer it – which, once again, threatens to make God responsible for the fall. To address this problem, Anselm argues that God gives every rational agent two fundamental inclinations or dispositions-to-will: an inclination for justice (*affectio justitiae*) and an inclination for benefit (*affectio commodi*). On Anselm's (broadly libertarian) account of free choice, an action must be self-initiated in order to count as free, which means that it cannot be caused by any external agent, including God (*De libertate arbitrii* 2).[4] Because God gives created agents two fundamental inclinations, however, they are able to initiate self-caused actions: they can decide for themselves how to weigh the relative value of different goods, and thereby decide which goals to pursue. These two inclinations can conflict, and so it is possible for an agent to desire, and pursue, personal advantage beyond the restraints of justice (or the reverse).

And this is exactly what Satan did. Satan wanted and chose some forbidden good that he believed would benefit him more than the gift of perseverance (*De casu* 4, 12–14). He desired personal advantage more than justice, and when he freely acted on that desire, he disobeyed God, and fell.

According to Anselm, the forbidden good is a genuine good that really does confer advantage, and so it is entirely proper that Satan's *affectio commodi* would incline him towards it. Similarly, Satan's *affectio justitiae* properly inclines him toward obeying God and accepting the gift of perseverance instead. The nature of created freedom is such that neither inclination can determine an agent's will, and so Satan is free to choose which one to follow. His choice is an uncaused act of brute libertarian freedom, to be sure, but the choice is genuinely *his*, and not just an event that befalls him, or a case of bad moral luck, because that choice really does express Satan's own desires and motivations.

Having established that Satan's choice is free and morally significant, Anselm ends his dialogue on an apophatic note. Why did Satan choose the forbidden good instead of the gift of perseverance? 'Simply because he willed it. For there was no other cause by which his will was in any way incited or attracted. Instead, his will was its own efficient cause . . . and its own effect' (*De casu* 27). Contemporary interpreters are inclined to agree with Anselm, and to treat Satan's sinful choice as no more or less mysterious than any other act of libertarian free will (see, for example: Rogers (2008), 97–107; Timpe (2012); Willows (2014)). On this line of argument, once we have shown that Satan's choice is free, and expresses his genuine preferences and desires, then no further explanations are available, even in principle.

For example, Katherin Rogers writes that:

> The devil brought about his choice by his own conscious efforts. But the same would have been true had he chosen otherwise. The difficult spot is the moment of preference. We can point to the reasons for choosing each option. But there is no antecedent cause or explanation for the preference of one over the other ... [Anselm's account of free choice] entails the mysterious position that the created free will is the originator of its morally significant choices. But mystery is not logical impossibility. (Rogers (2008), 104, 106)

In short, according to Rogers, Satan's choice of the forbidden good is explained by the fact that he preferred it. Since he also desired the gift of perseverance, he could have preferred, and chosen, that good instead. It is a mystery that he preferred the former and not the latter, but once we have shown that Satan's choice is logically possible, our work is done (see also Visser and Williams (2009), 189–191; Timpe (2012), 203; Willows (2014), 267).

Anselm on the harder problem

But our work is not done. After all, the unfallen angels are paradigmatically rational agents, with intellects and wills that function exactly as designed by God.

Even if we can show that Satan's sinful choice is logically possible, we still must show how it can be subjectively rational – a choice that a completely good, properly functioning, rational agent would ever make, even if he is free to make it. In other words, we still must address the harder problem of the devil's fall.

It is clear that Anselm recognizes the harder problem. In *De casu* 21–25, he patiently establishes that the angels could not have known that God would actually punish them for sinning, although they did know that they would deserve punishment. If Satan had been certain that God would punish his sin, then 'he, as one who both willed and had happiness, could not have spontaneously willed that which would make him wretched' (*De casu* 23). In other words, if Satan had known that God would punish him, then his choice to sin would be subjectively irrational, since he would be knowingly exchanging happiness for misery. Anselm presents Satan as a rational agent who weighs up the costs and benefits of sinning, and decides that the benefits outweigh the costs. In short, Anselm assumes that Satan's choice is not just logically possible, but subjectively rational.

Anselm seems to think that his solution to the hard problem also suffices for the harder problem. After all, he has established that Satan chooses the forbidden good because he genuinely wants it, and so we might think that, by that very fact, he has also shown that Satan's choice is subjectively rational. Yet recall the example above, of the choice between the fortune and the piece of cake. If a man chooses the cake, we would not be satisfied with the explanation that because he wants both the fortune and the cake, either choice is equally rational for him. Even if he would like to have both, he surely wants the fortune much more than the cake – and if he does not, then that fact must be explained as well. 'He chose the cake because he wanted the cake' is true but uninformative. To be sure, there are cases when we would accept it as an explanation – if the choice were between roughly equivalent goods, like cake and ice cream, perhaps – but there are also cases when we would not. The fall of the devil is a case of this sort.

It is not enough to say, as Anselm does, that Satan desires both the gift of perseverance and the forbidden good, and then chooses the latter over the former. Desire is not binary, such that one either desires a good, or one does not, *tout court*. In order to rationalize Satan's choice of the forbidden good, we must account for the relative strength of his desires, and how they contribute to his overall preferences. When we try to account for the relative strength of Satan's desires, however, we find ourselves once again faced with the threat that God is responsible for his fall. After all, Satan's initial dispositions and desires – and therefore also his preferences – were given to him directly by God. Did God create the angels to prefer the forbidden good, only to punish those who subsequently act on their God-given preferences? To avoid this uncomfortable question, suppose instead that God created the angels to prefer the gift of perseverance. In that case, why would Satan choose the forbidden good? His choice would be massively irrational from his own point of view, since it would be a choice against his own preferences. Finally, consider the compromise suggestion, that God created the

angels with a roughly equal preference for both goods. This option seems to reinstate the problem of bad moral luck. God created the angels so that they teetered on a moral precipice and then ruthlessly punished those who happened to incline one way instead of the other. Faced with three unacceptable options, we once again find ourselves in a quandary.

A formal framework for modelling the devil's fall

Anselm recognizes the force of the harder problem, but he does not quite solve it. Nor have his contemporary defenders so far been able to come to his aid.[5] On my reading, however, Anselm does present us with all the insights we need to solve the harder problem, and when we translate those insights into a better formal framework, we can see the way forward. We need a framework that allows us to see clearly the real logical structure of the devil's fall. The ideal framework would capture the initial conditions at the time of creation, as well as the competing desires and incentives that led Satan to make his sinful choice. It would also show us how Satan's choice could be subjectively rational even though it is contrary to God's will, and therefore objectively irrational.

Fortunately, such a framework is available. Economists use a variety of conceptual tools to model the decisions and preferences of rational consumers. We can borrow some of these tools to model the devil's fall. In the rest of this section, I first introduce the relevant tools and concepts: indifference curves, budget or consumption constraints, and graphs that show what happens when consumers seek to maximize their subjective satisfaction (utility). I then use these tools to model a broadly Anselmian account of the fall of the devil.[6]

Consider two people, Tom and John, who have a fixed amount of money to spend on burgers and hot dogs. Tom and John enjoy both burgers and hot dogs, but their tastes diverge: Tom prefers hot dogs to burgers, whereas John prefers burgers to hot dogs. Note that these preferences are relative, however: even though Tom prefers hot dogs, he also likes burgers, and vice versa for John. Although Tom and John like both hot dogs and burgers, because they only have a fixed amount of money, they cannot consume as much as they would like of both, and must consume less of one in order to consume more of the other. Moreover, as they each decide how many burgers and hot dogs to buy, there will naturally be some possible combinations to which they are indifferent. Perhaps Tom has no preference between 12 hot dogs and 7 burgers, on the one hand, and 13 hot dogs and 5 burgers, on the other. Similarly, John may have no preference between a package of 10 burgers and 2 hot dogs, versus 7 burgers and 10 hot dogs. (Even though John prefers burgers to hot dogs, he is still willing to consume three fewer burgers in order to get eight additional hot dogs.) Let us further suppose that Tom and John each have $100 to spend, and that burgers cost $10 each, but hot dogs cost only $5. In addition to their preferences, their choice of which specific combination of burgers and hot dogs to

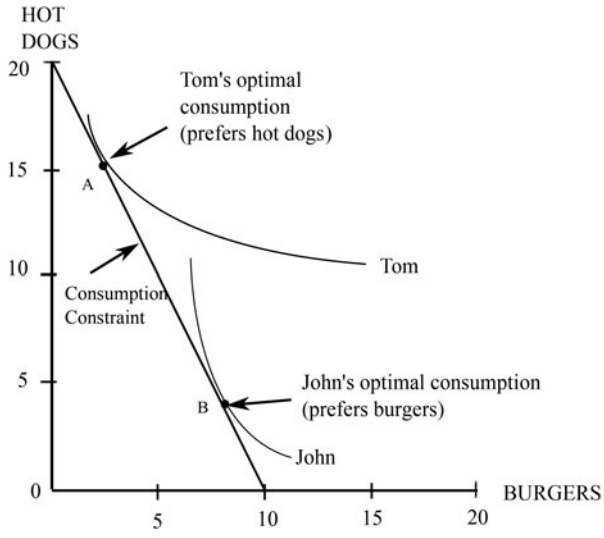

Figure 1. Indifference curves and budget constraints.

consume will be constrained by the fact that they each only have a fixed amount of money to spend, and by the fact that burgers cost twice as much as hot dogs.

Figure 1 represents a plausible model of this scenario. The diagonal line connecting the X and Y axes represents the *consumption constraint* that Tom and John face. They both have only $100, and so given the cost of burgers and hot dogs, if they each buy 10 burgers, they cannot buy any hot dogs at all, and if they each buy 20 hot dogs, they cannot buy any burgers at all. The points on the consumption constraint reflect combinations of burgers and hot dogs they can afford. The line that captures the consumption constraint is fairly steep, which reflects the fact that hot dogs (on the Y axis) are cheaper than burgers, and so they can buy more of them with the same amount of money. If burgers were cheaper than hot dogs, the consumption constraint line would be flatter.

The downward sloping curves labelled 'Tom' and 'John' are called *indifference curves*. These curves reflect combinations of burgers and hot dogs for which Tom and John have no preference. The two curves have different shapes because they reflect each person's relative desire for hot dogs and hamburgers. Tom's indifference curve is flatter, because he prefers hot dogs: he is willing to give up a large number of burgers (moving leftward on the X axis) in order to gain just a few more hot dogs (moving upward on the Y axis). John's preferences are the reverse, and so his indifference curve is very steep: he is happy to give up many hot dogs for a few more burgers.

The points labelled 'A' and 'B' in Figure 1 capture the optimal point of consumption for Tom and John. Formally, A and B are the points at which their respective indifference curves are tangent to the consumption constraint. Less formally, because their indifference curves capture what they want to

consume, given their preferences, and the consumption constraint captures what they are able to consume, given their resources, it follows that their optimal choice will be to consume at the point at which the two are aligned. They each consume at the precise point A or B because of the shape of their respective indifference curves, which in turn present a nuanced picture of the relative strength of their desires for hot dogs and burgers. Recalling the discussion of Anselm, the key point for my own constructive purposes is that Figure 1 conveys a much richer set of information than just the binary insight that Tom prefers hot dogs whereas John prefers burgers, let alone the bare claim that Tom and John each desire both goods.

Indifference curves reflect several general assumptions about consumer preferences that are worth spelling out. First, and most crucially, is the assumption that *more is better*. Given two baskets of goods, consumers will prefer the basket that offers them more goods. Every consumer has an infinite array of indifference curves, and will always prefer a higher indifference curve to a lower. Second, indifference curves are normally convex with respect to the origin, which means that as consumers purchase ever more of a given good, they derive ever less pleasure from each additional unit of that good. (As I discuss below, this assumption does not always hold.) In Figure 1, for example, Tom and John's indifference curves flatten as they extend further from the Y axis, which reflects the fact that they derive diminishing returns from each additional hot dog or hamburger. Third, and summarily, consumer preferences are reflexive, transitive, complete, and continuous.[7] It follows that unless there has been some change in a consumer's preferences, his own indifference curves cannot intersect.

Now consider Figure 2. Figure 2 presents three indifference curves for both Tom and John, and also captures the effect of a drop in the price of burgers. When burgers become cheaper, Tom and John can use the same amount of money to purchase even more burgers, and so their consumption constraint rotates outward along the X axis. Even though the price of burgers has changed, Tom and John still have the same underlying preferences as before. As a result of the drop in price, John, who prefers burgers, chooses to consume at B*, where he buys many more burgers, and a few more hot dogs. By contrast, Tom, who prefers hot dogs, consumes at A*, where he purchases a few more hot dogs and a few more burgers. Note that Tom and John now consume on new, higher, indifference curves, reflecting the fact that they are better off as a result of the drop in price, since they can consume more of both goods. Still, their new indifference curves have the same shape, because their preferences have not changed. A consumer's preferences do not change when goods become cheaper or more expensive, although his or her optimal point of consumption does change. Although his actual purchase is determined by the consumption constraint, John will always prefer baskets of goods on curve J3 to J2, and J2 to J1. The same goes for Tom, who will always fare better on indifference curves to the north-east of T1. In general, baskets of goods on indifference curves to the

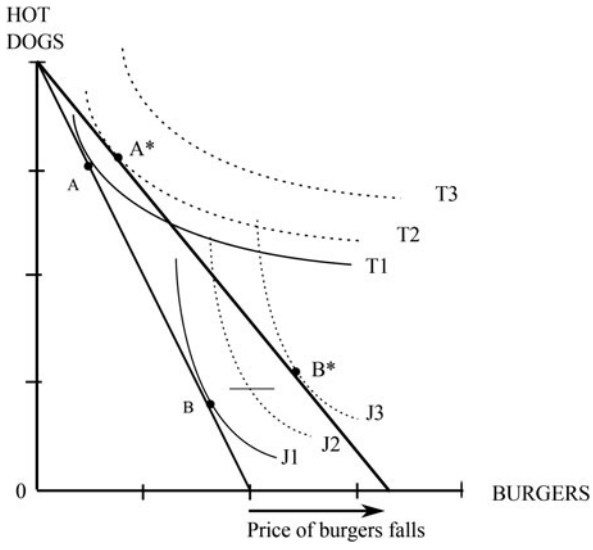

Figure 2. Higher indifference curves are preferred.

north-east of a given curve offer more goods, and therefore, by assumption, more satisfaction.

This framework allows us to capture several important insights: that resources are finite, that desires are not mutually exclusive and have different strengths, and that some goods are more costly than others. An ideally rational agent will take all of these factors into account when he decides what to consume. Furthermore, we can use the framework to test how an agent will react to changes in the underlying conditions. The shape of an agent's indifference curves tells us how he will react to changes in the consumption constraint, as we saw with Tom and John. Alternatively, if an agent's preferences themselves change, we can model that change by constructing new, differently shaped, indifference curves.

A 'market' for perseverance and the forbidden good

We can use this framework to model the fall of the devil. In this section, I discuss the interpretative choices we must make in order to construct the model. I then derive some general conclusions about Satan's sinful choice. In the next three sections, I model that choice. Throughout, I try to remain exegetically faithful to Anselm's account in *De casu*.

Recall that, on Anselm's account, the devil falls when he chooses some forbidden good instead of the gift of perseverance (*De casu* 3–4).[8] Anselm is careful to present the forbidden good as a genuine good. Moreover, even though Satan abandons the gift of perseverance in favour of the forbidden good, at the time of his choice, he also desires perseverance. His desire for the forbidden good is stronger,

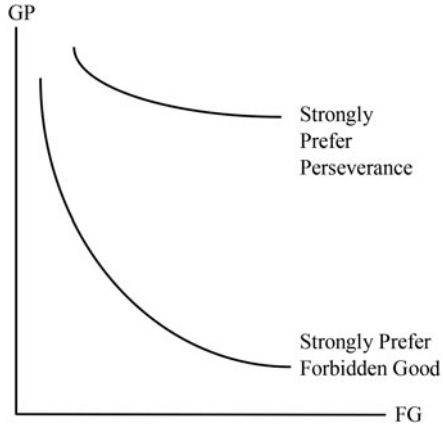

Figure 3. God determines the initial shape of the angelic indifference curves.

relative to his desire for perseverance, but in an absolute sense, he does desire both goods.[9] Similarly, although the good angels desire the gift of perseverance more, they also desire the forbidden good. In order to explain the devil's fall, then, we need to do more than simply appeal to the fact that he 'desires' the forbidden good and then freely chooses in accordance with that desire. All the angels desire both the gift of perseverance and the forbidden good. In order to understand how Satan's sinful choice could be subjectively rational, we need a more finely grained analysis.

Anselm tells us that 'it is quite obvious that every rational thing exists in order that it might love something more or less, or reject it altogether, according as its rational discernment judges that the thing is more or less good, or not good at all' (*Monologion* 68). With the framework of preference theory, we can formalize this insight and apply it to Satan's fall. The angels must choose between two competing goods: the forbidden good (FG) and the gift of perseverance (GP). Let these two goods determine the X and Y axes, respectively.[10] The shape of the angels' initial indifference curves reflects how strongly the angels prefer either good. Given that the indifference curves represent the angels' initial motivational set, God determines their shape at the point of creation. On the assumption that God creates the angels with a strong preference for perseverance, their indifference curves will be relatively flat. On the assumption that God creates the angels with a strong preference for the forbidden good, their curves will be steep (Figure 3).

Now consider Figure 4. Indifference curves alone can show how strongly the angels prefer one good over another, but in order to model their choices, we need to include a consumption constraint. God determines the initial slope of the consumption constraint, just as he determines the initial shape of the angels' indifference curves. The model therefore allows us – indeed, forces us – to lay out our assumptions about the initial conditions of creation with great precision. It also allows us to explore the consequences of modifying those assumptions. It

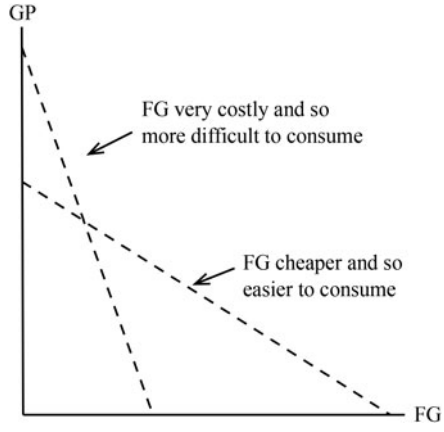

Figure 4. God also determines the initial 'cost' of consuming both goods.

is easy to construct a graph that presents the fall as maximally unlikely – which would also maximally exonerate God. Alternatively, it is also easy to construct a graph on which the angels are created just on the precipice of falling, which makes the fall easier to understand, but also threatens to make God seem more responsible for it.

Formally, the consumption constraint connects the X and Y axes and reflects the 'cost' of either good. Yet what does 'cost' mean in this context? Or, to ask the same question in another way, what do the angels 'pay' when they choose either the forbidden good or the gift of perseverance? Whatever the answer is, for the model to work, it must be a finite resource that the angels possess at the time of their creation. Moreover, they must be able to allocate that resource to either good without being able to allocate all of it to both goods at the same time. Finally, for the model to be useful, the relevant resource must be something that is causally or logically connected to the angels' choice of each good. (In other words, we need to be able to infer *Satan chooses the forbidden good* from *Satan allocates more of his resources to the forbidden good.*) As long as we have reason to believe that something can serve as the relevant resource, we do not actually have to identify it. We could simply speak of the angels' 'resources', letting that word stand in for whatever it is that plays the needed role. We could even embrace the artificiality of the model and treat it like a thought experiment. We can imagine that God creates a 'market' for perseverance and the forbidden good, and gives the angels 100 'credits' each to 'purchase' either one. What matters is that the model helps us understand the devil's choice, even if the model itself is an artificial construct.

It is possible to be more specific, however. Resources like concentration, effort, or willpower seem like the sorts of things that creatures have in finite supply, and that they spend when they try to acquire various goods. It also makes sense to say that some goods are relatively 'expensive' or relatively 'cheap' in that they require greater or lesser effort or willpower to acquire. Another possibility, and one that is

well-grounded in the Christian tradition, would be to identify attention or attentive awareness as the relevant resource. Creaturely attention is finite and restrictive in the relevant sense: the more attention I pay to the book I am reading, the less attention I pay to the music playing in the background (see also MacDonald (1999), 121). Attention seems like a form of consumption in its own right, and it also seems appropriately connected to love, desire, and choice. It seems natural to suggest that the more I focus my attention on a particular good, the more I love it, or will soon come to love it. It also seems clear that it is easier to pay attention to some goods in comparison with others.

Even after we establish the relative 'cost' of either good, one might still object that the model is inappropriate because Anselm presents the consumption of GP and FG as mutually exclusive. By contrast, a model rooted in consumer preference theory assumes that the angels can consume a mix of both GP and FG. There are at least two ways to answer this objection. First, if we treat effort or attention as the finite resource that the angels expend on either good (as above), then there is no problem in suggesting that they can consume a mix of both goods at once. They would simply pay different degrees of attention to both goods, or expend different amounts of effort on acquiring both goods, or whatnot. Alternatively, we could also solve the problem by stipulatively defining the relationship between angelic consumption in the model and angelic choice outside the model. We could say that the angels make a real-world absolute choice of FG instead of GP (or vice versa) when the model shows that they consume more FG than GP (or vice versa). The model would then capture how close the angels are to the relevant tipping-points, the point at which (outside the model) they would choose absolutely one good over the other.

After we have defined consumption and the consumption constraint, we next need to define what it means in the terms of the model to say that Satan sins or falls. Anselm is very clear that 'the will of every creature ought to be subject to God's will' (*Cur deus homo* [Why God Became Man] 1.11). I take this to mean that creatures sin when they will something contrary to God's will. So the first sin is simply the first act of disobedience – the first time a creature wills something contrary to God's will. The model also allows us to define two other relevant points: the point at which Satan consumes more of the forbidden good than of the gift of perseverance, and the point at which he consumes *only* the forbidden good. We can model Satan's fall even more precisely by including both.

The initial conditions of creation and the trajectory of the fall

According to Anselm, God created the angels as self-determining agents who can freely imitate God by choosing justice for its own sake. In order to be able to choose justice for its own sake, they must also be able to abandon justice (*De casu* 18).[11] Anselm's account suggests a model of the initial conditions of creation like Figure 5 (below).

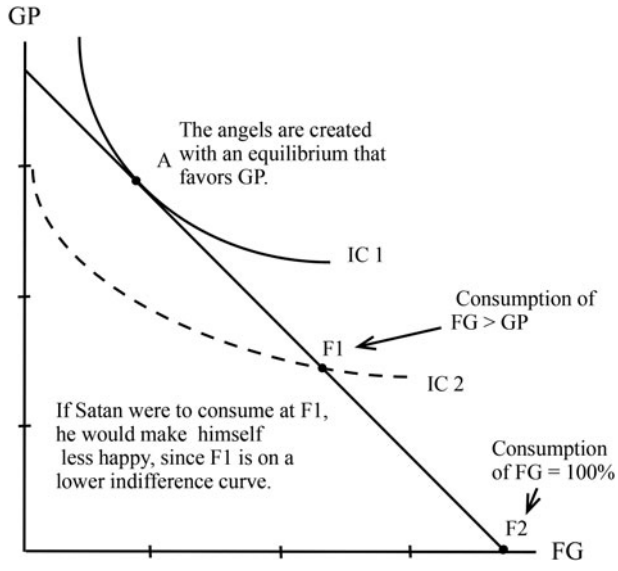

Figure 5. The initial conditions of creation and the trajectory of the fall.

The angels are created good, with a balanced desire for both GP and FG (IC 1). GP and FG, in turn, are roughly equivalent in cost. Nevertheless, because God is good, he creates the angels at an initial point of equilibrium that favours GP (point A). Consider now points F1 and F2 on Figure 5. At F1 the angels would consume more FG than GP, and so F1 is the first point at which Satan could be said to be making an overall choice for FG. At point F2, however, Satan would consume only FG. In terms of the model, at F2 he completely rejects the gift of perseverance because he does not value it at all.

Straightaway, we can draw several conclusions from Figure 5. First, note that it is *possible* – in the sense of logically or metaphysically possible – for the angels to choose to consume anywhere along the consumption constraint. (By definition, the consumption constraint is the line that captures all the combinations of the two goods that are available to a consumer given his resources.) Thus, even at the initial point of creation, it is possible for Satan to choose to consume at F1 or even F2. It is possible, but why would he? Given Satan's overall motivations and preferences at the time of creation (which, along with those of the other angels, is expressed by IC 1), Satan has no reason to consume at F1, even though he is free to do so. Point F1 lies on a lower indifference curve, and so were Satan to consume at F1, he would be freely choosing to be less happy than he could be. Although it is also available to him, F2 lies on an even lower curve and is accordingly an even more subjectively irrational choice.

Since most attempts to explain the fall of the devil ignore the harder problem altogether, they do not go beyond the scenario captured by Figure 5. Figure 5 thus presents the harder problem in sharp relief. In order to explain the devil's

fall, we must do more than merely establish that Satan freely chooses the forbidden good. Absent new information or some shift in their underlying environment that disturbs the initial equilibrium, we would expect that the preferences of the angels would remain stable, with the result that they consume at point A.

Satan falls after God restricts consumption of the forbidden good

Figure 5 presents the angels' initial desire for the forbidden good. Still, as constructed, Figure 5 does not yet reflect the fact that God has explicitly forbidden the good in question. Point A on Figure 5 is the point at which the angels' intrinsic (created) desire for FG is tangent to the consumption constraint. The consumption constraint, in turn, establishes the maximum amount of both goods that the angels can consume given their resources. But according to Anselm, God withholds the forbidden good, and he orders the angels not to consume FG at the level of their intrinsic desires and resources.

God's command might seem like an arbitrary test that puts the angels in a position of greater moral hazard. On Anselm's account, however, God is not arbitrarily testing whether the angels will obey his command. Rather, God is offering them a wonderful gift: the opportunity to become genuine, self-determining, moral agents, instead of beasts who unreflectively satisfy their desires as they arise. This gift puts the angels in a position of greater moral hazard only in the sense that, by definition, becoming a moral agent entails moral risk. Anselm writes that the forbidden good is 'an extra something' (*illud plus*) that the angels 'were able to attain, which they did not receive when they were created, in order that they might advance to it by their own merit' (*De casu* 6; see also Rogers (2008), 58–59). They 'advance to it by their own merit' by obeying God's command to refrain from consuming it. When the good angels choose GP over FG, they obey God, and show that they prefer justice to private benefit. They show that they love justice for its own sake and in so doing, they imitate God, who is Justice (Anselm, *Monologion* 68). Conversely, when the bad angels choose FG over GP, they show that they prefer private advantage to justice.

Figure 6 presents this scenario. Figure 6 features two different constraints: the initial consumption constraint at the time of creation (the solid line), which reflects how much of each good the angels are able to consume, given the absolute level of their resources, and a new 'obedience constraint', which reflects the fact that God wants the angels to consume less FG than they can. In effect, when God explicitly forbids the good in question, he asks the angels to act as if it were more expensive to consume (see Rabin (1995); Caplan (2000), 191–211). The good angels obediently respond by consuming less FG and more GP, even though their underlying preferences do not change: they consume at A*, along the new obedience constraint, even though it lies on a lower indifference curve. They do not desire FG any less when they restrict their consumption of it. They simply obey God's command to eschew it. In so doing, they forgo benefit and

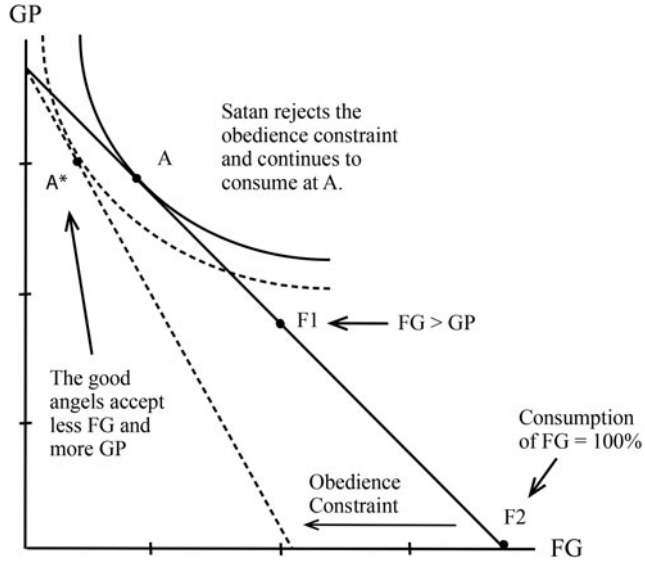

Figure 6. God restricts the forbidden good; Satan disobeys God's command.

will justice for its own sake, exactly as God wants them to do. (Recall that for Anselm, when we freely obey God, we will justice for its own sake.)

The model shows that Satan is free to make this choice: the original consumption constraint is still available to him, and still represents the maximum level of his consumption given his resources. The model also shows that his choice is not unmotivated, inexplicable, or even especially mysterious. In fact, as Figure 6 shows, his choice is subjectively rational: point A lies on a higher indifference curve than A*, and so Satan rightly calculates that he can maximize his happiness by disobeying God. According to Anselm, Satan does not believe that God will punish him for consuming the forbidden good, even though he knows that he would deserve punishment (*De casu* 23). With the fear of punishment removed, Satan correctly judges that he can maximize his happiness – in the short term at least – by continuing to consume at A, contrary to God's command.

By contrast, the good angels decide to forgo short-term satisfaction and to obey God by consuming at A*. Perhaps the good angels respond to the uncertainty of divine punishment differently, and calculate that it is likely that God will punish them if they disobey. Or perhaps they wager that God will reward them for their obedience, and so decide that it is rational to forgo short-term happiness for the promise of greater future reward. We could tell a number of stories to explain why the good angels' choice is also subjectively rational. It is comparatively easy to explain why it is subjectively rational to obey God, the sovereign creator and the source of all goodness.

It is time to take stock. Have I really advanced beyond the bare claim, previously rejected, that Satan chooses FG over GP because he wants FG more? Perhaps I

have only *depicted* the harder problem, but have not made any progress on actually solving it. I believe that I have made progress, on three fronts. First, the model does show a way in which Satan's choice can be both free and subjectively rational – and, after all, that was the challenge at hand. Strange as it may seem, Satan actually does maximize his own happiness by disobeying God, at least until he is punished, and so his choice to disobey God is rational from his own point of view. Second, the model does not rely on any appeal to the brute fact of free will, nor does it treat Satan's choice for FG as mysterious, unmotivated, or inexplicable. Rather, it explains Satan's choice for FG explicitly and precisely: he wants to consume FG at a higher level than the obedience constraint permits; he is able to do so, and by doing so, he maximizes his own happiness. Third, the model explains exactly what distinguishes the good angels from the bad angels, and why the latter sinned while the former did not. The good angels voluntarily chose not to maximize their own short-term happiness, and to obey God.

The fall of the devil in three stages

The precise moment at which Satan falls is the moment of the first sin, the moment at which he first disobeys God. This moment is well-captured in Figure 6 (above). Still, given that the Christian tradition presents Satan's fall as both grave and perverse, and given the dire consequences that follow in its wake, one might worry that if we stop here, the story so far either makes God a cruel tyrant (for punishing a peccadillo with eternal damnation) or else makes the fall a bit too trivial. After all, note that at point A, Satan still consumes more GP than FG, and so, according to the model, in an absolute sense he still prefers justice to benefit. By contrast, according to Anselm, when Satan chooses the forbidden good, he abandons justice as such, and is no longer able to will justly at all (*De casu* 17). We should honour this traditional intuition. We want a model on which Satan's sinful choice seems not just wrong but utterly perverse. It seems that we still have more to say. Consider Figure 7.

Figure 7 presents a fuller picture of the fall of the devil, in three stages.[12] In stage one (as in Figure 6), Satan has rejected the obedience constraint and chosen to consume along the original consumption constraint, at point A. He is able to make this choice because this original level of consumption is still available to him. Just by virtue of disobeying God, Satan already counts as fallen, by definition. Let us suppose, however, that at this point he begins to rejoice in the fact that he appears (to himself) to have thwarted God successfully. He ruminates over just how much he loves FG, which causes his desire for it to grow. As a result, his preferences change, and his indifference curve becomes steeper. His new indifference curve is tangent to the consumption constraint at point F1, where he consumes more FG than GP (IC 2). At this point, for the first time, we can say that in an absolute sense, Satan prefers FG to GP. But let us not stop here. Let us further suppose that once Satan is fully in the grip of his sinful self-regard, his evil desire for the

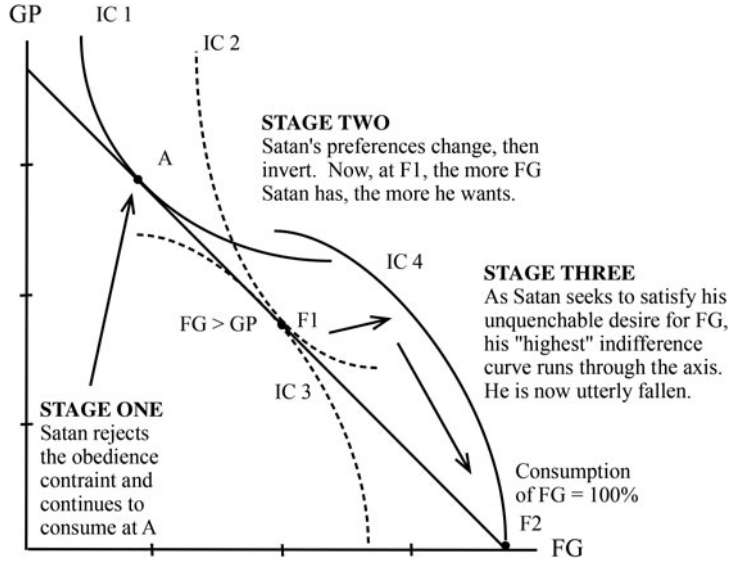

Figure 7. The fall of the devil.

forbidden good becomes self-reinforcing. His preferences shift again, as he comes to love himself, and desire the forbidden good, ever more. This time, however, his indifference curve does not merely become steeper but actually *inverts*, becoming concave with respect to the origin (IC 3).

Economists use concave indifference curves to model addictive goods (see Salvatore (2008), 67, 77–78). With addictive goods, the law of diminishing returns does not hold: we always want more of an addictive good, no matter how much of it we have already consumed. (If I am genuinely addicted to chocolate, then the more chocolate I have, the more I want, and I would not be willing to give up any chocolate to gain any amount of another good.) For obvious reasons, I do not want to say that Satan is 'addicted' to FG. I would say instead that, after the fall, Satan is no longer able to value GP and FG properly at all. After his initial decision to reject God's command, Satan's desires become totally warped and inverted, so that he now regards GP as completely worthless in comparison with FG, and seeks to consume as much FG as possible.

Stage three of Figure 7 presents the inevitable result. With his inverted desires, even Satan's new (evil) equilibrium at F1 is not stable, since 'higher' indifference curves are still available to him. As ever, he chooses to maximize his own happiness, and so consumes at the point at which his highest indifference curve is tangent to the consumption constraint. But since he now wants only to maximize his consumption of FG, and does not want to consume GP at all, his new point of equilibrium is F2, on IC 4, the point at which the consumption constraint intersects the X axis. At this point, Satan consumes only FG and has completely rejected GP. Because his indifference curve is inverted, when Satan seeks to maximize his own happiness,

he now does the very opposite of what God commands. Given his own preferences, he cannot 'rationally' consume GP at all. He is utterly depraved, unable to will justice for its own sake – exactly as Anselm posits and the Christian tradition demands.

Conclusion

This model graphically presents a plausible rendering of the fall of the devil, one that also captures Anselm's major insights. God creates all the angels with a specific set of preferences and resources, and he also establishes the relative value of every created good. He creates the angels with a balanced desire for both the gift of perseverance and the forbidden good. Because he does not want them to fall, God initially makes it easier for the angels to choose the gift of perseverance (Figure 5). On the other hand, God does want the angels to become genuine moral agents, who love justice for its own sake, rather than bestial consumers that always pursue their own advantage. And so he orders them to forgo the forbidden good, even though they desire it, and even though, in itself, their desire is good. The good and bad angels respond differently to God's command. The good angels obey, and voluntarily choose to be less happy than they could be, in order to serve God. Satan and the bad angels, however, chafe at God's command (Figure 6). After Satan disobeys God and falls, his desire for the forbidden good quickly becomes perverse and all-consuming, to the point that he is unable to choose any other good at all. He makes himself a voluntary slave, in utter thrall to his own desires, and completely cuts himself off from justice (Figure 7). Satan damns himself.

The formal framework I have offered should prove valuable even to scholars who disagree with the substantive interpretative choices I have made. Indeed, they can use the very same framework to present their own alternative interpretations in a suitably precise and perspicuous way, which can only make fruitful disagreement more likely. Moreover, the framework itself can shift the focus of scholarly attention away from the relatively easy problem of explaining how Satan's sinful choice can be free, and towards the relatively hard problem of explaining how it can be subjectively rational.

On my model, the first sin is neither compelled, nor unmotivated, nor unintelligible. It is not a product of bad moral luck, nor is it even indirectly caused by God. Satan makes an evil choice as he seeks to maximize his own benefit and satisfy preferences that are themselves good. The model preserves Satan's freedom and moral agency, and presents his fall as the result of subjectively rational choices that are objectively sinful and disordered. What the traditional Christian story demands, the model supplies.[13]

References

Anselm of Canterbury (2007) *Anselm: Basic Writings*, Thomas Williams (tr.) (Indianapolis: Hackett).
(2007) *De libertate arbitrii* [On Freedom of Choice].

(2007) *De casu diaboli* [On the Fall of the Devil].
(2007) *Monologion*.
(2007) *Cur deus homo* [Why God Became Man].
Augustine of Hippo (1972) *Concerning the City of God against the Pagans* [*De civitate dei*], Henry Bettenson (tr.) (New York: Penguin Books).
(1993) *On the Free Choice of the Will* [*De libero arbitrio*], Thomas Williams (tr.) (Indianapolis: Hackett).
Babcock, William S. (1988) 'Augustine on sin and moral agency', *Journal of Religious Ethics*, **16**, 28–55.
Barnwell, Michael (2009) '*De Casu diaboli*: an examination of faith and reason via a discussion of the devil's sin', *Saint Anselm Journal*, **6**, 1–8.
Brower, Jeffrey (2004) 'Anselm on ethics', in Brian Davies & Brian Leftow (eds) *The Cambridge Companion to Anselm* (Cambridge: Cambridge University Press), 222–256.
Brown, Robert F. (1978) 'The first evil will must be incomprehensible: a critique of Augustine', *Journal of the American Academy of Religion*, **46**, 315–329.
Caplan, Bryan (2000) 'Rational irrationality: a framework for the neoclassical-behavioral debate', *Eastern Economic Journal*, **26**, 191–211.
Chappell, T. D. J. (1995) *Aristotle and Augustine on Freedom: Two Theories of Freedom, Voluntary Action, and Akrasia* (New York: St Martin's Press).
King, Peter (2012) 'Augustine and Anselm on angelic sin', in Tobias Hoffmann (ed.) *A Companion to Angels in Medieval Philosophy* (Leiden: Brill), 261–281.
Macdonald, Scott (1999) 'Primal sin', in Gareth B. Matthews (ed.) *The Augustinian Tradition* (Berkeley: University of California Press), 110–139.
Rabin, Matthew (1995) 'Moral preferences, moral constraints, and self-serving biases', *Economics Working Papers from University of California at Berkeley*, no. 95-241.
Rogers, Katherin (2008) *Anselm on Freedom* (New York: Oxford University Press).
Salvatore, Dominick (2008) *Principles of Microeconomics*, 5th edn (New York: Oxford University Press).
Timpe, Kevin (2012) 'The arbitrariness of the primal sin', *Oxford Studies in Philosophy of Religion*, **5**, 186–205.
Visser, Sandra & Williams, Thomas (2009) *Anselm*, Great Medieval Thinkers (New York: Oxford University Press).
Willows, Adam M. (2014) 'Augustine, the origin of evil, and the mystery of free will', *Religious Studies*, **50**, 255–269.

Notes

1. References to Anselm are given by chapter and paragraph number as found in Anselm (2007).
2. Augustine discusses the origin of evil in *De libero arbitrio* [On the Free Choice of the Will], written between AD 387 and 395. See especially 2.19–20. Later he revisited the issue in Book 12 of *De civitate dei* [The City of God], written between AD 413 and 427. His views on the nature of grace and freedom evolved between those two treatises, but scholars debate how much they evolved.
3. Prominent critics of Augustine include Brown (1978); Babcock (1988); Chappell (1995). See also Rogers (2008), ch. 2; King (2012).
4. For discussion of Anselm's account of free will, see Brower (2004); Rogers (2008), ch. 3; Visser & Williams (2009), ch. 11.
5. As noted above, Rogers (2008), Visser & Williams (2009), and Timpe (2012) all seem to agree with Anselm that a solution to the hard problem is all that is needed.
6. Readers who are familiar with indifference curves may skip ahead to the next section. For an accessible textbook discussion of consumer preference theory, see Salvatore (2008), ch. 3.
7. The meaning of 'transitive' should be clear. In this context, 'reflexive' means that if goods A and B are identical, then the consumer will be indifferent to a choice between A and B. The assumption that preferences are 'complete' means that the consumer has ranked all possible alternative combinations of available goods. Preferences are 'continuous' in that they are assumed to be infinitely divisible.
8. As Barnwell (2009), 2, n. 3, points out, Anselm 'reifies' the gift of perseverance, treating it like an item that one can accept or reject, rather than a process or a state-of-being.
9. I take it that this is the point of Anselm's example of the miser in *De casu* 3. A miser might prefer some amount of food to some amount of gold, and he might willingly give up some of his gold to obtain food. In

an absolute sense, however, he still *wants* gold – he doesn't cease wanting gold just because he also wants food, and no doubt he would prefer to obtain food and still keep all of his gold.

10. One might worry that Anselm presents the consumption of GP and FG as mutually exclusive. By contrast, a model rooted in consumer preference theory assumes that the angels can consume a mix of both GP and FG at once. I address this later in the article.
11. See also *Cur deus homo* 2.10. Anselm does not make the principle of alternative possibilities a condition on free will per se, but it is an implication of the nature of creaturely free will. For discussion, see Visser & Williams (2009), 178–182.
12. I make no claims about the temporal duration of these stages, and they could even be understood as logical moments of a single instantaneous event.

13. Several friends and colleagues read drafts of this article and made valuable suggestions. I would especially like to thank Brian Leftow, Timothy Pawl, Kevin Timpe, and Gillian Hamnett.