Social Evil*

Ted Poston
University of South Alabama

Abstract: Social evil is any pain or suffering brought about by game-theoretic interactions of many individuals. This paper introduces and discusses the problem of social evil. I begin by focusing on social evil brought about by game-theoretic interactions of rational moral individuals. The problem social evil poses for theism is distinct from problems posed by natural and moral evils. Social evil is not a natural evil because it is brought about by the choices of individuals.   But social evil is not a form of moral evil because each individual actor does not misuse his free will. Traditional defenses for natural and moral evil fall short in addressing the problem of social evil. The final section of this paper discusses social evil and virtue. I argue that social evil can arise even where virtue is lacking. Further, I explore the possibility of an Edwardsian defense of social evil that stresses the high demands of true virtue. The conclusion of this paper is that social evil is problematic and provides new ground for exploring the conceptual resources of theism.

Discussion on the problem of evil assumes that there are two classes of evils: natural evil and moral evil.[1] Richard Swinburne divides "the world's evils in the traditional way into moral evils and natural evils."[2] Swinburne characterizes moral evils as "those brought about by human intentional choice, or knowingly allowed to occur by humans, together with the evils of their intentional bad actions or negligence."[3] Natural evils, according to Swinburne, are "all other evils, such as bad desires that we cannot help, disease, and accidents."[4] Alvin Plantinga provides a similar division of evils. He writes, "In addition to 'natural' evils such as earthquakes, tidal

---

[1] The distinction between two classes of evil goes back, at least, to Augustine. On natural evil see Augustine's *De Ordine* and on moral evil see *Confessions* and *On Free Will*.
[2] Swinburne (2004), 236; cf. van Inwagen (2006), 9.
[3] Ibid., 236.
[4] Ibid.

waves, and virulent diseases there are evils that result from human stupidity, arrogance, and cruelty."[5] William Rowe's famous Bambi and Sue cases represent, respectively, natural and moral evil. The Bambi case involves the prolonged suffering and death of a fawn; the Sue case relates the horrible beating, rape, and murder of a five-year old girl.[6] Both these forms of evil exercise the conceptual resources of theism to explain how an omnipotent, omniscient, and morally perfect being might allow these evils to occur. But there is another form of evil that has not received the attention of philosophers working on the problem of evil. This form of evil I call *social evil*.

At the outset I want to be clear about the nature of evil and its various types. For the purposes of this paper, I assume that evil is any instance of pain or suffering and that types of evils are individuated by the processes that bring them about. Moral evil is an instance of pain or suffering brought about by the direct agency of a person, and natural evil is an instance of pain and suffering brought via the operation of laws of nature on matter. The nature of direct agency is difficult to analyze. I think of direct agency via moral responsibility. A person exercises direct agency in performing an action when they have full responsibility for that action. Even though action may require background conditions such as the operation of laws of nature, a person can still be held fully accountable for her act. Thinking about natural evil in terms of responsibility, the "blame" for natural evil lies with impersonal forces. If John is killed by an avalanche, the reason for his death lies in inanimate forces that produced the avalanche. Social evil is an instance of pain or suffering that results from the game-theoretic interactions of many individuals. When a social evil occurs, responsibility for the outcome lies with no particular

---

[5] Plantinga (1974), 8.
[6] The original "Bambi" case comes from Rowe (1979) and the "Sue" case is in Rowe (1988). William Alston (1991) dubbed these cases "Bambi" and "Sue" and, following Alston, the terminology has stuck.

person and no impersonal force of nature; rather it lies with a group of people, each of whom may be morally in the clear. Until the last section of this paper, I focus on the most problematic form of social evil, social evil that occurs because of the game-theoretic interactions of rational moral individuals. It is widely acknowledged in the literature on game theory that rational well-intentioned agents can bring about horrible social outcomes. Russell Hardin (1995) provides an arresting example of this by offering a game-theoretic analysis of violent group conflict. Once we see the game-theoretic machinery in play, it is hard to resist the thought that much evil in our world is the unintended result of collective agency among individually rational participants. My goal in this paper is to introduce and explain social evil and the problems it poses for traditional defenses. Furthermore, I discuss the prospects for a theistic diagnosis of social evil. I intend my discussion to be a starting point for a deeper reflection on the nature of social evil and the theoretical lessons it has for theism. The concept of social evil has wider applications to ethical theory. Reflection on social evil shows that, in at least some cases, an individual's actions are not the result of a bad will but rather the result of an individual being caught in a tragic game-theoretic scenario. An understanding of social evil suggests that there are far fewer moral monsters that we might otherwise think.

I. What is social evil?

I begin with a clear case of pain and suffering that results from the game-theoretic interactions of rational, well-intentioned individuals. Suppose you are a resident of Los Angeles and the greater Los Angeles area is facing a serious water shortage. The reservoirs in northern California are running dry; the Los Angeles and San Gabriel rivers are bone dry; even Oregon's plentiful lakes and rivers are ominously diminished. Without a significant decrease in overall

water consumption, the Los Angeles area will run out of an adequate water supply. City planners foresee the possibility of severely restricting residential water use. However, if most residents significantly decrease their water consumption—by not watering lawns, washing cars, or letting the tap run unnecessarily—the Los Angeles area will manage until the winter rains come. Obviously, it is in the best interest of all that *most* everyone follows this advice. But this represents a considerable cost to *each* person. If, for example, you decrease your water usage, your carefully cultivated garden and fruit trees will wilt and die. This is a hefty burden to pay. However, if no one decreases his water usage, each will pay an even greater cost. Yet you realize that if most everyone decreases his water consumption, then you may continue your normal usage without any ill consequence. Moreover, because the benefit of decreased water usage requires a very large number of participants—well over a million homeowners—your own individual contribution does not affect whether or not the benefit is realized.

The Los Angeles water shortage case is a simultaneous move game. A simultaneous move game is one in which you (and everyone else) act in ignorance of what the other people do. In this case let us assume that you are rational and blameless. You do not suffer from a failure to realize that you are in this kind of circumstance. Moreover, you do not suffer from a moral fault; you do not want to harm anyone and you do not want your action to bring about a worse circumstance. Furthermore, while you could reason in a self-interested manner, you do not. You reason from states of value in the world. Your garden is an item of beauty and it is worth preserving. In this situation you realize that your decision whether to continue to water affects only whether an item of value is preserved, i.e., your beautiful garden. Taking into consideration all the relevant factors, the best option for you is not to conserve water. But many other Los Angeles residents face similar situations in which the best option for each is not to conserve

water. If most everyone is rational and blameless, each individual will play his best strategy and the collective result will be unintended disaster. The disaster that results is a social evil.

This is a standard form of a multiplayer prisoner's dilemma, also known as "the tragedy of the commons."[7] Cases of this sort are widespread. Achieving adequate healthcare, decent education, effective inoculations, safe freeways, fertile fishing waters, and pristine national parks all require the cooperation of a sufficiently large group of individuals. In these cases the goods achieved and the evils avoided require solving a multiplayer prisoner's dilemma. Because each individual's dominant strategy is to depart from the cooperative act (i.e., to defect), governments aim to avoid the destructive logic of these games by, among other things, imposing significant penalties for defecting. While this is a pressing practical problem, there is a theoretical problem for theism that has not been addressed in the literature to date.

To guide our discussion I present a standard two-person prisoner's dilemma.[8] We can consider a two-person dilemma as the smallest case of a social evil, a case in which an unintended and worse outcome results for each player enacting his best strategy. Consider the following game.

Two-Person Prisoner's Dilemma

		Player II	
		Defect	Cooperate
Player I	Defect	2,2	4,1
	Cooperate	1,4	3,3

---

[7] See G. Hardin (1968).

[8] I give a multiplayer prisoner's dilemma below in section IV. At this point in our discussion, we will gain no appreciable advantage by using the more complex multiplayer PD.

In this game Player I has two options: either defect (maintain his water usage) or cooperate (decrease his water usage). Player I achieves his *best* result by defecting given Player II's cooperation, and he achieves his *worst* result by cooperating given Player II's defection. The numbers 4-1 represent an *ordinal* ranking of the players' preferences. An outcome 4 is preferred to 3 which is preferred to 2 which is preferred to 1. Ordinal rankings reflect a series of ordered preferences, and they do not reflect the *strength* of preference. With an ordinal ranking you cannot infer that Player I's highest ranked option, 4, is twice as desirable as his second ranked option, 2. Player II's rankings are symmetrical with Player I's. Thus, for Player II his highest ranked option is where he maintains his current water usage and Player I decreases water usage.

What should each Player do? Given Player I's situation, he can always improve his outcome if he defects. To see this suppose Player I cooperates. Then if Player II defects, Player I would do better by *defecting*. He would move from an outcome of 1 to an outcome of 2. If Player II cooperates, then, just as in the other situation, Player I would do better by *defecting*. He would move from an outcome of 3 to an outcome of 4. In this situation Player I's option of defecting *strictly dominates* cooperating because regardless of what Player II does, Player I can do better by defecting. Player II, of course, should engage in the same reasoning and so Player II should defect. Thus, these two rational well-intentioned players achieve an outcome neither wants, viz., mutual defection. Both players would prefer mutual cooperation to mutual defection but the logic of the game lands rational, well-intentioned players in mutual defection.

The water shortage case illustrates the painful logic of group action. Each participant realizes the structure of the game they are in and each participant is blameless. But *given* the structure of the game, it is rational for each player to defect. Note that because of the nature of group action, each individual player does not bring about a bad state of affairs. In large multi-

6

player prisoner dilemmas the benefit (or detriment) is achieved regardless of what any one

individual does. In more realistic games—like the water shortage case—the benefit (or

detriment) is insensitive to at least one tenth of a percent of the total number of players. For

example, if a multiplayer prisoner's dilemma is realized with 10^6 players then the social outcome

will obtain regardless of what 10^3 players do. If 1,000 players don't make a difference to this

game your decision surely does not make a difference. Consequently, each individual player's

choice does not affect the outcome of the game. Moreover, in large multiplayer games each

player may know that defection will bring about a Pareto improvement. A Pareto improvement

to an allocation of resources is a redistribution of those resources that makes someone better off

without making anyone worse off. In a large multiplayer prisoner's dilemma, any change in any

one individual's strategy doesn't affect anyone else, so a player can know that defection will be a

Pareto improvement. We might say that the problem of social evil is that the road to hell is

paved with Pareto improvements.

I assume that this discussion is sufficient to motivate the thought that social evils are

distinctive because they occur as the result of game-theoretic interactions among many

individuals and *not* on account of some individual's choice (as in the case of moral evil) or some

natural process (as in the case of natural evil). Let us now examine challenges to the idea that

social evil is a distinctive kind of evil.

II. Three objections to the distinctness of social evil

Is social evil really distinct from natural and moral evil? In one sense I take it as obvious

that social evil is distinct from natural and moral evils. Natural evils are instances of pain and

suffering brought about by natural processes like earthquakes and tidal waves. Moral evils are

instances of pain and suffering brought about individual human choice. If Sam loathes Bill and breaks his nose, then this is moral evil. Yet in another sense, one may very well wonder whether social evil is distinct from moral and natural evil. Might not social evil really be the problem of limited resources? Might not social evil be a subtle form of natural evil, viz., human stupidity? And might not social evil be a subtle form of moral evil? In this section, I respond to these questions.

One objection to the distinctness of social evil is that social evil is really a form of natural evil and so an adequate defense for natural evil will carry over to social evil. The water shortage example involves a severe drought; droughts, like earthquakes, are natural evils. Even so, the extent of the pain and suffering of the Los Angeles drought depends upon the collective response of a great number of people. If most everyone cooperates, the drought will not drastically affect the Los Angeles area. The reason a drought can be so devastating lies not merely in the severity of the natural conditions but also in the collective response of many individuals. A group is accountable for the severity of the drought even though no particular person is accountable. So even though there is a natural evil in the background, the problem in the foreground is social evil.

The present objection is that any form of social evil is a form of natural evil on account of a scarce resource. In the standard cases of multiplayer prisoner's dilemma there is a limited resource that needs to be distributed over a large population. But the present objection infers an evil from a scarcity of resources. This move is dubious. What reason do we have for thinking that some pain or suffering will result from the mere fact that a resource is limited some pain or suffering will result? Even in normal circumstances, water is a limited (i.e., a scarce) resource. Everyone can't run the tap all day long. The fact that a resource is limited doesn't imply that any

pain or suffering results. It may only imply a limited resource that has no Pareto-efficient allocation. A Pareto-efficient allocation is one that admits of no Pareto improvement. As we saw in the water shortage case, there is always a Pareto improvement to the allocation of water resources; there is always a reallocation of water related benefits and burdens in which some individual is better off and no one else is made worse off. Any individual can continue to use water thereby increasing overall value in the world in some respects without making the world any less valuable in other respects. In general, the structure of the tragedy of the common cases always permits a Pareto improvement. If the commons is a grassy green then one could always permit one more sheep to graze without affecting the overall outcome. To summarize: a situation without any Pareto-efficient allocation of resources need not be one in which there is suffering and so such a situation need not be an instance of natural evil.

Another response to the objection that social evil is a natural evil on account of limited resources is that social evils can arise from resource *abundance*. Many harms to our environment are caused by society overusing a certain resource. I am morally in the clear driving my car to work, but the result of millions of individuals driving is an overabundance of carbon which results in global warming, thereby causing some pain or suffering. Similarly, John is morally in the clear to fertilize his yard, yet the result of many individuals doing the same thing is that too much fertilizer gets into our lakes and rivers, ruining our recreational fishing grounds.

A second objection to the distinctness of social evil is that it is a subtle form of natural evil, specifically human stupidity. If an individual defects while fully realizing that everyone else faces the exact same reasoning, then the person is stupid. But why should we think that the individual defector is stupid? One reason is that the individual does not realize that his defection

brings about a worse state of affairs, viz., one confederate fewer. But this reason rests on a

serious misunderstanding of the logic of multiplayer prisoner's dilemmas. The individual

defector does not *bring about* a worse state of affairs by bringing it about that there is one fewer

confederate. This is because the social benefit will be achieved (or not) regardless of what an

individual person does. That is, for any n, n ± 1 confederates do not change the outcome of the

game. If one million confederates realize the outcome then one additional or one fewer

confederate will not change the outcome. To suppose that an individual's choice affects the

outcome of a multiple player game is a gross misunderstanding of the logic of such games.

The principle that *for any n, n ± 1 confederates do not change the outcome of the game* is

true of cases in which the threshold of confederates required to achieve the good and avoid the

evil is vague. In cases of vagueness, one cannot apply the principle iteratively without at some

point losing knowledge about whether the threshold is met. In the Los Angeles water shortage

case, let us suppose that three million confederates will achieve the social good but one hundred

thousand will not. One can gradually diminish the numbers from three million participants to

one hundred thousand, but at some point in this series it becomes vague whether that number of

participants will achieve the good. Still, it remains true that one knows that small changes, i.e.,

plus or minus 1, will not affect the outcome to be achieved.

Readers familiar with the vagueness literature will recognize that the epistemicist claims

that this margin of error principle is false.[9] According to epistemicism, vagueness arises because

of ignorance. We do not know the position of the cutoff, the position at which *n*+1 number of

grains makes a heap while *n* does not make a heap. Consequently, the epistemicist claims, we

mistake our ignorance about cutoffs for there being no cutoff (i.e., for the margin of error

---

[9] See Williamson (1994).

principle being *true*). There is much to be said in favor of epistemicism: it is elegant, upholds classical logic, and handles higher-order vagueness. But it is false. Epistemicism is primarily developed as a semantic theory designed to get sharp truth-conditions for vague terms. But in our case we are interested in causal influences. And given the nature of a multiplayer prisoner's dilemma any single player does not affect the overall outcome. Relatedly, if epistemicism is true, then game theory must be revised since the claim that there is no Pareto-efficient allocation of a resource rests on a margin of error principle.

A third objection to the special nature of social evil is that it is a form of moral evil. Specifically, an individual defector is morally to blame for defecting. While there *might* be something to this claim (though see sections III & IV below), it will require serious argumentation that goes beyond standard characterizations of moral evil. On the traditional view, moral evil is a form of pain and suffering that results from the agency of another person. Typically, this involves the *misuse* of free will.[10] Clearly, in a multiplayer prisoner's dilemma an individual defector does not bring about any (relevant) pain or suffering. The individual defector is not an outcome-cause of the effect. The effect will occur regardless of what the individual does. Just as one raindrop does not make France fertile, so one person's decision in a large multiplayer prisoner's dilemma does not affect the overall outcome.

One response to this line of argument is that it mistakes the logic of rational choice with the logic of moral responsibility. A prisoner's dilemma represents the interaction of preferences between various agents. If one agent has a dominant strategy, one cannot infer from dominance alone that an agent is morally in the clear to enact that strategy. In particular, to be morally just an agent's action must be universalizable; it must be possible for everyone to act on the agent's

---

[10] Though see Robert Adams (1985) "Involuntary Sins."

maxim.   But clearly, so the objection goes, an individual's strategy to defect is not universalizable in a multiplayer prisoner's dilemma. If a defector acts on the maxim *I will defect to bring about a better world*, then this is not universalizable because if everyone acted on that maxim a much worse world would result. Thus, the objection is that defection is immoral because it doesn't pass the universalization requirement.

This response may appeal to those with Kantian moral sensibilities. The problem with this reply, though, is that universalizability is not a necessary condition for moral permissibility. Suppose you are considering what kind of life you should live. Should you be a surgeon by which you may save many lives or should you develop your talent as an artist through which you may enrich the lives of many? Each decision is made against the backdrop of a diverse population with different aims and goals. A life as a surgeon assumes that many people are not doctors, and a life as an artist assumes that not everyone is an artist. One cannot simply universalize the maxim *I will become a surgeon to save the lives of many* because if everyone became a doctor that specialization would collapse.

This objection to universalization shows that if the requirement is to be plausible at all, it must allow that diversity is represented within the universalization test. To test the moral appropriateness of one's choice to become a doctor one must universalize the maxim assuming that other people choose different lives to lead. But once we allow for diversity in the universalization test, the original objection that defecting is blameworthy falls apart. For one can represent in the universalization test that one's action will not influence what the group will do. Even more strongly, one can represent that one knows this to be the case. Once this is represented then the defector's maxim passes the universalization test. There is nothing morally

inappropriate about defection when one knows that one's action will bring about a better world without making anyone worse off.

Another possible attempt to explain the moral wrongness of defection is rule utilitarianism. According to this view, an act is morally good if and only if it is in accord with the optimific rule, the rule that has the best consequences. Clearly, if everyone cooperated then we would achieve the full benefits of cooperation. So, the objection goes, complete and full cooperation is the optimific rule, and in this connection defection is unjust.

The problem with this reply is that there is no unique optimific rule in a multiplayer prisoner's dilemma. For any number of confederates, one less confederate will achieve the same social utility. Moreover, an individual's defection will bring about more value in the world. An individual can know all this and act to bring about a better world without adversely affecting anyone. Rule utilitarianism does not imply that an individual defector has done anything amiss.

Perhaps, though, if we examine Kant's second formulation of the categorical imperative—to treat people as an ends, never as a means—we can discover why it is morally wrong to defect in a multiplayer prisoner's dilemma. Perhaps, in defecting one treats people as a means for one's action rather than treating people as ends. The suggestion is that the individual in the Los Angeles water case reasons to the conclusion to continue to water by not properly taking into account the autonomy of other individuals. But this suggestion fails. Granted, the language of 'defection' suggests that one is defecting from a group consensus and thus uses the group to further one's own aims, but the reasoning of the individual in the water case doesn't proceed like this. The individual in the water case recognizes that regardless of what people do, he can bring about an improvement without harming anyone else. It's entirely consistent with this reasoning that the individual treats each other person as an end. After all, he may very well

be concerned to uphold each other individual's dignity by not interfering with their ends. And, of course, there's nothing he can do in the situation to further their ends.

Some people persist in thinking that there is an attenuated sense of 'bring about' in which the individual brings about a worse state of affairs. But, to repeat a point made above, this is a failure to realize the logic of a multiplayer prisoner's dilemma. An individual doesn't affect the social outcome. Further, an individual's change of strategy doesn't affect the probability that the social outcome is achieved. Removing one grain of sand from the beach doesn't affect the probability that the beach has enough sand to make a sand castle. A related but confused objection is that an individual's act of defecting represents a cost itself, so the individual is not faultless for defecting. This objection is confused because it changes the nature of the game. A prisoner's dilemma is a mathematical object that represents moves and preferences. If one thinks that defection itself is a cost, then that should be reflected in the system of preferences. Either the new game will be a prisoner's dilemma or not. If it is, then the problem of social evil will arise; if it is not, then the problem may not arise.[11]

III. Standard Defenses

In this section, I examine standard defenses to determine how they might handle social evil. I argue that social evil is difficult to incorporate within standard theistic responses to evil. I consider three defenses: the value of natural laws, a soul-making defense, and the free will defense. At the outset I should make it clear that I am not merely interested in the logical compatibility of social evil and theism. Alvin Plantinga has conclusively shown that there is no logical problem of evil, and, as I explain below, Plantinga's story can be extended to show that

---

[11] I leave open the question of whether prisoner dilemmas are a necessary condition for social evil.

there is no logical incompatibility between theism and the existence of social evil. Even so, granting the compossibility of social evil and theism, the distinctive problem social evil poses for theism remains. What the theist needs to offer is a reason for thinking that God would permit the destructive logic of multiplayer prisoner's dilemmas when we already have enough natural and moral evils to secure the kinds of goods discussed in the standard defenses. The theist need not offer anything amounting to a theodicy, but she should be able to sketch a not completely implausible explanation for why God would allow multiplayer prisoner's dilemmas. In this connection, we can set aside a skeptical theist position.[12] The skeptical theist eschews the task of offering explanations-why for evil and instead focuses on our epistemic limitations to discern a God-justifying good, if one exists. If skeptical theism is plausible, then it has the resources to handle most any evil we are acquainted with. Skeptical theism has been developed with an eye to moral and natural evils, but it can easily be extended to social evil. Even so, the concept of social evil forms a fruitful area of investigation since the extant discussion on the problem of evil focuses on values that are applicable only to natural and moral evils. The problem of social evil should lead the theist to search out new kinds of values that might justify a perfect being in permitting this destructive logic.

A. The value of natural laws

Richard Swinburne argues that the problem of natural evil is lessened by the value of natural laws.[13] A universe with recognizable regularities allows persons to successfully predict the consequences of their actions. If I want to help you by offering you nourishment, it is valuable to know that if I offer you bread it won't kill you. Similarly, if I want to study the

---

[12] For an explanation and defense of skeptical theism see Bergmann (2001).
[13] See Swinburne (2004), 245ff.

behavior of atoms in very cold environments, it will help to know that there are reliable ways to bring about freezing temperatures. Because a world with laws is beneficial in this respect, it may be a consequence that some natural evils occur. For instance, a world in which substances have stable properties may imply that in some cases the stable properties of substances harm individuals (e.g., a tree falls and breaks Joe's leg). Or, a world in which environments have stable properties may imply that at boundary zones violent storms occur (e.g., a cold front meets a warm front).

Regardless of what one makes of Swinburne's claim, it is clear that this will not help with social evils. Social evils arise because of the collective agency of rational well-intentioned individuals. This requires that individuals can effectively reason about their options, which in turn requires observable regularities in the world, but that's true of moral evil as well. The crucial difference between social evil and natural evil is that social evil occurs because of human agency. Thus, the appeal to natural laws it does not handle social evil.

B. Soul-Making

A different defense appeals to the value of certain kinds of character traits—patience, fortitude, courage, and compassion—and then argues that these valuable traits require evils. Compassion requires suffering; courage requires injustice. The soul-making defense argues that a world that contains the great goods of character must also contain the great evils of character as well.   John Hick provides a nice summary of this defense:

> The value-judgement that is implicitly being invoked here is that one who has attained to goodness by meeting and eventually mastering temptations, and thus by rightly making responsible choices in concrete situations, is good in a richer and more valuable sense than would be one created *ab initio* in a state either of innocence or of virtue. In the former case, which is that of the actual moral achievements of mankind, the individual's goodness has within it the strength of

temptations overcome, a stability based upon an accumulation of right choices, and a positive and responsible character that comes from the investment of costly personal effort.[14]

Does the soul-making defense offer promise for handling social evil? An initial hurdle for the soul-making defense is that social evils are not cases in which an individual *brings about* a worse outcome. There is no causal connection between an individual's choice and the outcome that is realized by the group. This verdict is especially clear if we apply the distinction between outcome causation and aspect causation. An outcome cause is a difference maker to whether or not the effect occurs. An aspect cause is a difference maker to the effect *occurring as it does*. In a large multiplayer prisoner's dilemma each individual's action is not a difference-maker to the outcome being achieved or not. It is only in the most attenuated sense that an individual's action is an aspect cause of the effect, i.e., by contributing to the group, the individual causes the effect to occur as it does with $n+1$ confederates rather than n. So, there is very little room for an individual in such a case to achieve valuable character traits. Perhaps, though, a defender of this move will stress the value of solidarity or the value of self-inflicted loss even when those choices do not have any larger social consequences. But it's doubtful whether this move could be sustained because there is no causal connection between what an agent does and what social outcome is achieved. If an individual knows that her choices will have no negative effects while also increasing value in the world, it's hard to see how the individual could be blamed for doing what she does.

There is another problem with the soul-making defense. The soul-making defense is deeply individualistic. It focuses on the value of an individual's own character traits. Thus the soul-making defense doesn't have the conceptual resources to explore the value of traits of

---

[14] Hick (1977), 255-6.

societies, traits that do not reduce to traits of individuals. Exploring the value of traits of societies seems to be a productive area for reflection on social evil. Perhaps further investigation on the value of certain types of society will uncover reasons God may have to permit social evils. But unfortunately our current discussion of evil is largely informed by individualistic assumptions to the extent that we lack the language and perspective to discuss societal goods and evils.

C. The Free Will Defense

The free will defense focuses on the immense value of free will and its irresponsible use in generating horrendous evils. A free will defense is a story about why God may permit moral evil. A crucial part of the free will defense is that persons are responsible for actions that they freely bring about. But as we've seen above, social evil lacks the feature that an individual brings about the evil. Social evils result from the collective agency of individuals, not from any particular individual's choice. Another way to see this is that social evils do not result from any specific individual *misusing* her free will. If an individual in a multiplayer prisoner's dilemma enacts her best option, she is not thereby stupid or wicked. As explained above, individuals in prisoner's dilemmas need not be either irrational or mean. In the water shortage case, I stipulated that each individual was rational and blameless. They each know the situation, and they don't want to harm anyone by their actions. But given the situation each person's best play is defection. In this case, defection is not morally unjust. Thus, the value of an individual's free will doesn't account for social evils because there's no sense to be made of a culpable misuse of free will in this case.

Alvin Plantinga, pursuing a line by Augustine, considers the possibility that natural evil is the result of the free action of non-human spirits.[15] It's possible that, as Plantinga says, "Satan rebelled against God and has since been wreaking wherever havoc he can. The result is natural evil."[16] Given this possibility, the free will defense shows that there is no logical inconsistency between theism and natural evil. In the past, some free agent performed a culpable act and one of the enduring results of that act is natural evil. A similar move may be considered in connection with social evil. It is possible that the occurrence of the destructive logic of a multiplayer prisoner's dilemma is due to the past misuse of free will. Following Plantinga's Augustinian lead, it's possible that "Satan rebelled against God and has since been wreaking wherever havoc he can. The result is social evil."

How plausible is this response? It bears acknowledging that if it is possible that natural evil occurs because of Satan's activity then it is possible that the occurrence of the destructive logic of multiplayer prisoner's dilemmas is also the result of Lucifer's actions. After all, one might contend, wouldn't the prince of darkness delight in bringing about games in which rational moral individuals collectively produce disaster? Yet I confess that I find this response completely implausible. Given what we know about natural processes—ocean temperatures, weather patterns, plate-tectonics—we know how many natural evils occur. Similarly, given what we know about how society organizes itself—the division of labor, the need for protection and access to resources—we understand how multiplayer prisoner's dilemmas occur. There is no need to introduce a non-human agent to explain the occurrence of natural and social evils. Consequently, the Augustinian / Plantingian line does not provide a plausible solution to this problem.

---

[15] Plantinga (1974), 192.
[16] Ibid.

A different possibility for the free will defense is that prior to human sin God prevented the occurrence of multiplayer prisoner's dilemmas.[17] Perhaps prior to the misuse of free will there were no scarce resources, or individual's incentives and options never realized the destructive logic of certain games. But upon the misuse of free will, God removed these protections. So even though social evil presently arises from no wrongdoing, the occurrence of multiplayer prisoner's dilemma is the tragic result of *previous* sin. This strikes me as the best option for a free-will defense. But it stands in need of further elaboration and defense. One issue is that even if it's true that God prevented the occurrence of multiplayer prisoner's dilemmas prior to human sin, we need some additional account of what might justify God in permitting this destructive logic to occur now. As I've argued, horrible social outcomes can result from the collective choices of rational well-intentioned individuals, individuals who use their free will responsibly. It seems a harsh consequence for *previous* human sin that the proper use of free will would now result in evil. So the possibility that multiplayer prisoner's dilemmas arise because of past human sin does not exonerate the free will defense.

IV. Social Evil and Virtue

To this point I've argued for the claims that social evil is distinct from natural evil and moral evil and also that standard theistic defenses do not handle social evil. In this section, I have two goals. First, I weaken the assumptions I made regarding social evil. I have argued above that social evil arises by the collective action of rational, well-intentioned individuals in multiplayer prisoner's dilemmas. This is but one form of social evil. Below, I substantiate the claim that social evil can arise from the collective action of rational *blameworthy* agents in

---

[17] Peter van Inwagen (1988) suggests this possibility for natural evils. Thanks to … for extending this suggestion to social evils.

multiplayer prisoner's dilemmas. This can occur when game-theoretic scenarios make it very

difficult to avoid doing the wrong thing. In these kinds of cases an individual can be strongly

tempted to perform an act that, apart from the wider game-theoretic scenario, would be a

peccadillo. But because of the destructive game these individuals find themselves in, small sins

add up to horrendous evils.

The second goal of this section is to explore the possibility of an Edwardsian response to

social evil. Jonathan Edwards claims that true virtue consists in love for being in general.[18] A

truly virtuous person does not love merely a limited system—including himself—but loves every

being and seeks the good of all. A truly virtuous person will perform the act that is best for all

even if that act requires shunning his own private good. Since God has a reason to bring about

truly virtuous people, one might reasonably think that the value of true virtue can provide God

with a reason for permitting the game-theoretic machinery that produces social evil. The ensuing

discussion will show that the Edwardsian defense solves a two-person prisoner's dilemma but it

does not solve a multiplayer prisoner's dilemma.

A. Immoral man and very immoral society[19]

Our initial examination of social evil investigated instances of pain and suffering brought

about by the collective action of rational well-intentioned agents. The key feature that

distinguishes social evil from natural and moral evil is that the pain and suffering that occurs

does not arise from the direct choice of any individual nor from the result of natural processes. It

is the cumulative effect of very many choices within the multiplayer prisoner's dilemma that

---

[18] Edwards (1960).

[19] I'm told by a reliable source that Reinhold Niebuhr later thought that this would be a more apt
title for his famous book *Moral Man and Immoral Society*.

produces pain and suffering. Where each individual enacts their best strategy, the structure of the game the actors are in produces horrendous outcomes.

This description of social evil does not assume that the individuals are blameless. Social evil requires only that game-theoretic machinery produce an amount of pain and suffering that is disproportionate to the individual choices in the game. If a game is realized in which each individual is more self-interested than not and each chooses his best strategy, a horrendous outcome can occur that is not mirrored in the actions of each individual. In these cases the engine of society produces a magnitude of pain and suffering disproportionate to the fuel of individual animosity.

Examples of this kind of social evil are all too common. Russell Hardin in his book *One for All* provides a game-theoretic account of several well-known conflicts: Yugoslavia, Northern Ireland, Somalia, Rwanda, and the nationalist movement in Quebec.[20] Each conflict shares a basic structure. The roots of the conflict lay in norms of group identifications, norms that govern group inclusion and exclusion. These norms are important because there are significant benefits realized by being a member of the group. In 17th century England, for example, identification with the aristocracy carried the promise of better jobs, education, and social mobility. But identification requires norms of exclusion, norms that distinguish one group from the other. In 17th century England, one powerful norm of exclusion was the duel; an aristocrat was required on pain of loss of honor to risk injury and death at the smallest of offenses (e.g., disagreement over the merits of John Donne). Once the norms of identification are in place, each individual's incentive structure favors identifying with a group. This leads to competition between groups for access to benefits. Often this competition leads to conflict and even violence. But once violence

---

[20] See Hardin (1995), 155-177.

is realized, this provides the tipping point of group conflict in which escalating reprisals are rewarded. In many cases, this leads to open war. There is, thus, this basic structure: norms of group identification and consequent benefits from identification; competition; conflict; tipping point; system in which reprisals are rewarded; and escalating violence.

The Croatian War of Independence from 1991-1995 provides a case in point of the dreadful game-theoretic situations that produce horrendous suffering.[21] Croatia was one of six republics in Yugoslavia, and in 1990, faced with the prospects of Serbian dominance in Yugoslav politics, Croatian leaders decided to move for independence. Croatian Serbs, in addition to being a substantial minority in Croatia, constituted a sizable portion of the Croatian military and police force. When the Croatians moved for independence, they faced the prospects of an internal military revolt from the Croatian Serbs. Regrettably, these Croatian Serbs found themselves in a game in which they could do nothing to ensure loyalty to the new Croatian government. Any promise of loyalty to the new Croatian state would be perceived as subterfuge. In this situation, Croatian leaders preempted the possibility of a fifth column by dismissing Croatian Serbs from military and police positions. This move, though favored by the game the Croatians found themselves in, created the tipping point for internal Serbian dissent and led to revolt. The ensuing series of moves and countermoves resulted in the devastating Croatian civil war.[22]

This story highlights the awful logic of some games. Further, this story doesn't assume that each actor is morally in the clear. Arguably, events around the tipping point are replete with bad intentions. But the bad intentions and wrongful actions do not themselves explain the

---

[21] This paragraph is a condensed summary of Hardin's analysis (see pp. 156-163).
[22] Hardin also notes that the Croatian move for independence made group cooperation in Bosnia "virtually impossible" (p. 159).

descent into open war and the carnage that followed. In this case we can clearly see the effects of game-theoretic scenarios in producing horrendous pain and suffering.

In the discussion of the first several sections, I focused on pain and suffering that resulted from the collective action of rational moral agents. Now we have seen that social evil can result when the game-theoretic scenario magnifies the consequences of individual blameworthy action. Sometimes pain and suffering result from collective action even though no one individual has done anything amiss. This is the first kind of social evil. But in other cases, social evil results from collective action when the individuals are to blame for their acts and yet the social situation they are in amplifies the effects of those wrong acts. This is the second form of social evil.

B. An Edwardsian Defense of Social Evil

Jonathan Edwards, in his book *The Nature of True Virtue*, argues that true virtue is love for being in general. The truly virtuous person does not seek to benefit only a limited group of participants but rather seeks the good of every being. Edwards' discussion on true virtue includes a penetrating exposition on the consequences of self-love. He offers what is, in effect, a game-theoretic explanation of apparently moral behavior. He observes that a person's self-interest will motivate them to appear to be altruistic. Consider giving to the needy. Edwards argues that this act may often arise from a limited benevolence. One can be concerned only with the needy in one's own town, or one can give to the needy to identify with the "moral" crowd and then reap the benefits of being a member of that crowd. In this case, adopting Hardin's terminology, sacrificial giving is a norm of exclusion. Edwards underscores that any such act does not arise from true virtue. True virtue consists in love for being in general.

Given Edwards' emphasis on the high demands of true virtue together with the claim that a perfect being desires to bring about truly virtuous individuals, there may be the makings here for an Edwardsian defense of social evil. A case can be made for an Edwardsian defense, but, if my analysis is correct, an Edwardsian defense answers only social evil arising from a two-person prisoner's dilemma; it does not extend to a general solution to the multiplayer prisoner's dilemma. Even though the Edwardsian defense is unsuccessful, it holds important lessons for reflection on social evil.

I begin by examining an Edwardsian solution of a two-person prisoner's dilemma. I will assume that the truly virtuous person will favor the option with highest social utility. In a two-person prisoner's dilemma, I will take social utility to be the sum of the individual preferences. This way of generating a social utility makes the most sense under the assumption that the preferences represent *strength of preference* rather than order of preference. Since nothing hangs on this change, in the following game the reader may assume we are summing across preference strengths. In the matrix below I put the social utility in parentheses. Consider the following game.

Two-person PD with social utility

		Player II	
		Defect	Cooperate
Player I	Defect	2,2 (4)	4,1 (5)
	Cooperate	1,4 (5)	3,3 (6)

By inspection we see that the option with the highest social utility is the one in which everyone cooperates. This option achieves 6 units of social utility. What does this revised game look like for the truly virtuous person? Let us assume that Player I is truly virtuous. Player I faces the following choice: defect or cooperate. If Player I defects, then he realizes either the social utility of 4 (in which Player II defects) or the social utility of 5 (in which Player II cooperates). In either case, Player I would realize more social utility by cooperating. For the truly virtuous person, cooperating is the dominant action. In this way, the truly virtuous person solves the two-person prisoner's dilemma.

Does the Edwardsian position solve the general multiplayer prisoner's dilemma? No. To see this I will first describe the general form of a multiplayer prisoner's dilemma and then describe an instance of this for the truly virtuous. This shows that multiplayer prisoner's dilemmas can arise for the truly virtuous. Consequently, true virtue doesn't lead to the dissolution of multiplayer prisoner's dilemmas.

The multiplayer prisoner's dilemma I describe below assumes that there is a vague threshold for cooperation required to achieve the social good and avoid the social evil. As I argued above in the water shortage cases, it is eminently reasonable that the threshold for providing adequate water supply is vague. Any change of one person's strategy will not make a difference to the outcome achieved. Standard representations of multiplayer prisoner's dilemmas assume that the threshold for cooperation is exact. To provide a suitable matrix for the game I focus on three states: the state that doesn't meet the threshold, the state that is at or exceeds the threshold, and the penumbral state. In the penumbral state, it is unclear whether the social good will be achieved or not. I assume that in the penumbral state there is enough cooperation to

achieve at least some (but not full) benefit. When the threshold of cooperation is met then the full benefit is achieved.

In a multiplayer prisoner's dilemma each player faces the same matrix. The table below represents the matrix for a random player i.

Multiplayer Prisoner's Dilemma with a vague threshold

		Below the threshold of cooperation	*The penumbral state*	*At or above the threshold of cooperation*
Player i	Defect	No cost & no benefit Ordinal rank: 2	No cost & some benefit Ordinal rank: 4	No cost & full benefit: Ordinal rank: 6
	Cooperate	Cost & no benefit Ordinal rank: 1	Cost & some benefit Ordinal rank : 3	Cost & full benefit: Ordinal rank: 5

What should player i do? As before with the two-person prisoner's dilemma we can see that defection is the dominant strategy. If player i cooperates, then, whatever the other players do, i will do better by defecting. Inspect the above table: if player i cooperates then if the threshold isn't met, i does better by defecting; if the penumbral state is realized, i does better by defecting; and if the threshold is exceeded, i does better by defecting. The reason this situation arises is that each individual can bring about a Pareto improvement of resources by defecting. Since an individual doesn't affect the social outcome but does affect some item of value in the world, each individual faces the prospect of a wasted sacrifice and a loss of some item of value by

cooperating. Perhaps, though, if all players were truly virtuous we could avoid this disastrous game. In the following, I argue against this by providing a multiplayer prisoner's dilemma for the saints.

Suppose ten saints each manage their own orphanage. Christmas time is near and it is time to solicit gifts. Each orphanage draws upon a common pool of resources that in normal times is not adequate to meet every need but it is just enough for most. Each orphanage sends out letters to the community asking for donations. Now each saint faces the following decision. Should I *cooperate* by sending out only letters and relying on the donations generated by those letters, or should I *defect* by sending out letters and then making personal phone calls? The strategy to defect should not be thought of as a departure from true virtue. In the case at hand, defection amounts to making that extra call to ensure that Johnny gets the red fire truck he wants. It's implausible to think that the saint that defects from the group is acting out of anything but love. But if the saints each defect what they collectively bring about is a race to request charity. In addition to raising the required effort each year on the part of the saints, it is likely to lead to less overall charity. Besieged by letters, emails, phone calls, and visits, people grow weary of the increasing intrusions. Thus, the collective effect of departure from the standard is to realize an overall worse situation. However, whatever each individual saint does will not affect the overall overcome. If Theresa, at the last minute, makes that extra phone call she will bring about a better situation—Johnny gets the fire truck and all the giving is as it would otherwise be.

The orphanage case illustrates the dreadful logic of a multiplayer prisoner's dilemma. Even if each player is truly virtuous, the painful consequences of enacting each player's best option will not be avoided. Ultimately, the reason the Edwardsian defense fails is that each truly virtuous person can know that her act will bring about a better world but the collective result of

each person making a better world is unintended disaster. Even with the truly virtuous, the road

to hell is paved with Pareto improvements.

Conclusion

Social evil, pain or suffering resulting from the collective agency of rational individuals in

multiplayer prisoner's dilemmas, is a pervasive feature of our world. Yet traditional defenses do

not address it. Traditional defenses have focused exclusively on pain or suffering that results

from either natural processes or from moral agency. Theists should view this problem as an

opportunity to further mine the conceptual resources of theism. Additionally, social evil

provides strong motivation for everyone to be concerned about the structures of society. We

should aim for societal structures that minimize this dreadful logic.

<u>Bibliography</u>

Adams, R. 1985. "Involuntary Sins." *The Philosophical Review* 94:1, 3-31.

Alston, W. 1991. "The Inductive Argument from Evil and the Human Cognitive Condition." *Philosophical Perspectives* 5, 29-67.

Bergmann, M. 2001. "Skeptical Theism and Rowe's New Evidential Argument from Evil." *Nous* 35:2, 278-296.

Edwards, J. 1960. *The Nature of True Virtue*. Ann Arbor, MI: The University of Michigan Press.

Hardin, G. 1968. "The Tragedy of the Commons." *Science* 162, 1243-1248.

Hardin, R. 1995. *One for All: The Logic of Group Conflict*. Princeton: Princeton University Press.

Hick, J. 1977. *Evil and the God of Love*. New York: HarperCollins.

Kuhn, S. 2009. "Prisoner's Dilemma", *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, Edward N. Zalta (ed.), URL <http://plato.stanford.edu/archives/spr2009/entries/prisoner-dilemma/>.

Niebuhr, R. 1932. *Moral Man and Immoral Society*. New York: Charles Scribner's Sons.

Plantinga, A. 1974. *God, Freedom, and Evil*. Grand Rapids, MI: Wm. B. Eerdmans.

Rowe, W. 1979. "The Problem of Evil and Some Varieties of Atheism." *American Philosophical Quarterly* 16, 335-41.

_____. 1988. "Evil and Theodicy." *Philosophical Topics* vol. XVI, no. 2, 119-32.

Swinburne, R. 2004. *The Existence of God*, 2nd edition. New York: Oxford.

van Inwagen, P. 1988. "The Magnitude, Duration, and Distribution of Evil: A Theodicy." *Philosophical Topics* vol. XVI, no. 2, 161-187.

_____. 2006. *The Problem of Evil*. New York: Oxford.

Williamson, T. 1994. *Vagueness*. New York: Routledge.