Fall 2017

Cooperative Capital Contractarianism: A Response to David Gauthier's Morals by Agreement

Nathan Davis
University of Iowa

COOPERATIVE CAPITAL CONTRACTARIANISM: A RESPONSE TO DAVID GAUTHIER'S MORALS BY AGREEMENT

by

Nathan Davis

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in the Philosophy

_____

Diane Jeske
Thesis Mentor

Fall 2017

All requirements for graduation with Honors in the
Philosophy have been completed.

_____

Carrie Figdor
Philosophy Honors Advisor

Cooperative Capital Contractarianism:
A Response to David Gauthier's *Morals by Agreement*

by

Nathan Davis

A thesis submitted in fulfillment of the requirements
for graduation with Honors in Philosophy

Diane Jeske
Thesis Mentor

The goal of this paper is to evaluate a few alternative responses in the defense of rational egoism. I begin by exploring Galucon's challenge to Socrates in Plato's Republic as a goal to measure various theories against. Next, I discuss Thomas Hobbes' theory of contracts and the necessity of a commonwealth and his "fools" objection, as presented in *Leviathan*. Then, I turn to David Gauthier and examine in depth his argument for constrained maximization in *Morals by Agreement.* I contend that in an attempt to provide a satisfactory response to both Glaucon and the Fool, Gauthier opens himself up to several major problems. I conclude by providing my own view, Cooperative Capital Contractarianism, and expanding on how it can effectively respond to the problems facing Gauthier's view.

Glaucon's Challenge/Ring of Gyges

In Plato's *Republic,* Socrates discusses with several interlocutors what justice is. For this paper, I am concerned with the argument proposed by Glaucon in Book II. Previously in Book I Thrasymachus argued that justice was "the advantage of the stronger." Justice, Thrasymachus claims, are the rules that rulers establish for their own gain, it is more beneficial for the actor to be unjust than just.[1] Ultimately for Thrasymachus, being just is something we do simply because we cannot get away with being unjust. Socrates quickly gets Thrasymachus to abandon his claim but Glaucon is not ready to accept Socrates' view that "it is better in every way to be just than unjust."[2]

Glaucon begins his argument that being just is not always better, for the agent that is acting justly, by defining three different types of goods.[3] These types of goods are as follows:

1. Goods we desire for, and only for, their own sake.
    - Joy and harmless pleasures.

[1] Plato, *Republic,* Translated by G.M.A. Grube, Indianapolis/Cambridge: Hackett Publishing Co., 1992, 15.
[2] Ibid., 33.
[3] Ibid.

2. Goods we desire for their own sake and their consequences.
    - Knowledge, sight, and health.
3. Good we desire for, and only for, their consequences.
    - Physical training and medicine.

After establishing these three types of goods, he asks Socrates to which category justice would belong. Socrates claims that justice is in the greatest category of goods, those goods that are desired for both their own sake and for their consequences or type 2. At this point, Glaucon sets out to establish three different points: what most people consider justice and where it came from, people that do justice do it out of necessity[4] not because justice is valuable for its own sake, and the life of a successfully unjust person is better than that of a just person. After he has done this, he wants Socrates to praise acting justly *by itself* or, in other words, after it has been removed of all its consequences. This is so that people cannot be said to pursue justice just for the good consequences it brings about.

According to Glaucon, people believe that to do injustice is good and to suffer it is bad, but it is far worse to suffer it than to do it.[5] This has forced people that lack both the power to do injustice and the power to avoid suffering it to make agreements with others to refrain from doing injustice. This is justice[6], a middle ground between the best option, doing injustice without suffering it, and the worst, suffering injustice without being able to do any injustice yourself. These agreements are regularly made into laws defining what is just. This leads Glaucon to the claim that those who do justice only do it because they lack the ability to do injustice. People "value [justice] not as a good but because they are too weak to do injustice with impunity."[7] It

[4] Necessity here is being defined as a necessity of rationality. To be unjust, generally, carries all sorts of consequences and as such is worse for the actor then being just. Insofar as the actor is rationally pursuing his or her own benefit, it is irrational in most cases to be unjust.
[5] Plato, 34.
[6] Justice, as referenced here by Glaucon is not his own definition of justice but what he takes to be the common understanding of justice. This idea can be understood as following laws of your land and contracts you have made.
[7] Plato, 35.

would be "madness" for a man that has the power to act unjustly to make an agreement to refrain from it. He goes on to claim that if two men, one just and the other unjust, are given freedom to do whatever they wished, both would inevitably act unjustly.[8] For the purposes of this paper, we take human nature to be in line with rational egoism:

> Rational egoism – I have reason to do something if and only if doing it supports my own interests.[9]

Glaucon's point is that if people were free to commit injustice without consequence, they would invariably do so because being unjust would further their self-interests and thus is a rational action.

This freedom to do injustice without consequence is illustrated by Glaucon with a story about the ancestor of Gyges of Lydia.[10] During a storm an earthquake split the ground open. In the hole a shepherd ends up finding a golden ring that when turned on his finger made him invisible. Using this ring, the shepherd seduces the king's wife, kills the king, and takes over the kingdom. According to Glaucon, this ring would give its wielder the power of a god, and no one would remain uncorrupted by it. As Glaucon mentioned before, were two men to be given these rings, one just and the other unjust, they would end up acting the same. "No one believes justice to be good when kept private, since, wherever either person thinks he can do injustice with impunity, he does it."[11] Without such a ring, men must agree to not do each other injustice, but with a ring no such compromise needs to be made.

---

[8] Ibid.
[9] It should be mentioned that we are concerned with rational egoism and the justification for our actions. This is in contrast with psychological egoism and the motivation of one's actions. Even if people do not always act on their justifications they would simply be acting irrationally. It is not why they act that we are concerned with, in this paper, but what justification they have for doing so.
[10] Plato, 35.
[11] Ibid., 36.

For Glaucon's final point he wants to show that the life of the unjust man is better than the life of the just man. To do this he begins with the unjust man. This man must be perfectly unjust. In doing so, he not only gets the benefits from being unjust, but his excellence at being unjust allows him to develop the reputation of being just regardless. Now for the just man to be completely just, we must remove the benefits associated with being just so that we are certain he is not being just for its consequences but for the sake of being just alone. To further test him, not only should we remove the reputation and honors of being just but give him the reputation of an unjust man.[12] What would these two men's lives look like? The unjust man with the reputation of being just would have all of the benefits of taking advantage of others, winning any contest, and becoming wealthy. He also has all the benefits of a just reputation: he can marry into any family, makes contracts with anyone he wishes, and many others. The just man on the other hand will be "whipped, stretched on a rack, chained, blinded with fire, and, at the end, when he has suffered every kind of evil, he'll be impaled…"[13] At the end of all of this the just man will realize it is better to seem to be just than actually to be just. Glaucon's challenge to Socrates is to provide a definition of justice that, were one to have the ring of Gyges and thus the ability to perform injustice with no bad consequences, it would still be rational to act justly.

There appear to be a couple general responses to Glaucon's challenge. First, it could be argued that one's interests might include being just. The issue here is that we are trying to find a rational reason to be just for self-interested people. By including being just as part of one's interest, we are essentially saying "the reason I should do what is in my interest, is because it is in my interest." Another response to Glaucon would be to simply reject rational egoism and argue that there are things other than self-interest that provide justification for one's actions. I

[12] Ibid., 37.
[13] Ibid.

personally find rational egoism compelling, so for the purposes of this paper, I will set aside this concern in favor of more fully exploring theories of rational egoism.

In what follows, I will explore two theories of contractarianism, that of Thomas Hobbes and of David Gauthier. Both theories provide justification for justice or in their cases honoring ones contracts on the basis of rational egoism. For simplicity, in the rest of the paper self-interest will be defined as what promotes one's own pleasure.[14] From this framework, I will describe various strengths and weaknesses of both theories. Ultimately, I aim to answer what, if any, response that rational egoism has to Glaucon's challenge and Hobbes' fool.

Hobbes and the Fool

The first of two systems of contractarianism that I will lay out is that of Thomas Hobbes. In Hobbes' book, Leviathan, he explains the state of nature of humans and how out of nature there arise contracts which should be kept by those that make them. He also addresses what is famously known as the "fool's objection," which will be an important objection for each of our contractarianism systems.

What is the state of nature as described by Hobbes? He begins by explaining that humans are all, for the most part, equal in strength and ability. Some of us may be stronger or smarter but not by enough that the weaker could not, through either some weapon or teamwork, kill the stronger.[15] Naturally there will be some resources that many different people desire but only some can have. Because of this limited resource, men will be at war with each other constantly striving to kill or subdue the others for personal gain, safety, or reputation. In this state of war,

[14] I focus on hedonistic pleasure as what is valuable because I find it plausible. Though this should be defended, I leave it now out of a lack of time for exploring different theories of what we value.
[15] Thomas Hobbes, *Leviathan*, A.P. Martinich ed., Oxford: Clarendon Press, 2012, 93.

where every man is against every other man, there is no justice or injustice.[16] Each man is thus limited to what he can produce on his own and protect from the others. Not only is he prevented from obtaining any benefit from cooperation, but large amounts of resources must be devoted to protect his life and what little he does have.[17]

Now, Hobbes claims that by nature we have rights[18] to anything and everything that aids in the preservation of our life. In the state of war there is nothing that is not of aid, so we have a right to everything including other people's bodies. Because of this, as long as this right of every person to everything remains, there is no security to life, let alone the fruits of your labors. As such, Hobbes states that it is a general rule of reason or law of nature, "that every man ought to endeavor peace, as far as he has hope of obtaining it; and when he cannot obtain it, that he may seek and use all helps and advantages of war."[19] This rule of reason contains two components: to seek peace and to defend ourselves by any means if necessary. Now even though men seek peace, it is the second rule of reason that brings about the possibility of peace. The second rule is "that a man be willing, when others are so too, as far forth as for peace and defense of himself he shall think it necessary, to lay down this right to all things; and be contented with so much liberty against other men as he would allow other men against himself."[20] In other words, in order to gain the benefits of peace, men agree to give up their right to certain things such as each other's bodies and property.[21] There are a couple specifications that Hobbes makes here that are

[16] Justice here is the making and honoring of contracts. This will all be defined in detail shortly.
[17] Hobbes, *Leviathan*, 95-6.
[18] Hobbes use of rights is essentially that if we have a right to something it means that it is not wrong to do anything we wish with it. If we have rights to another's body, we can kill them and there is nothing wrong in doing so. We do not however have any correlative duties. This means, even though someone may have a right to another's body, it does not mean that other person has a duty to allow them to do what they will with their body.
[19] Hobbes, *Leviathan*, 99.
[20] Ibid.
[21] It should be specified that ultimately this is settling on less than they want. Were someone to be able to completely dominate another, as in the case of the ring of Gyges, they would not have to settle by agreeing to give up any rights.

important to the overall argument. First, there are some rights that cannot be given up. Whenever someone gives up a right or transfers it to another, they do so for their own benefit. As such, it cannot be the case that people give up their right to their own lives because this could not be in their benefit.[22] Second, if someone transfers or gives up a right, they *ought* not to go back on what had been previously agreed upon, and it would be unjust to do so.[23] The agreement to give up rights was originally based on reason and on the benefit that would come from that agreement. To reverse that decision would be contrary to reason and thus absurd.

From here, Hobbes goes on to discuss different types of agreements. For our purposes in this paper, I will focus on the differences between contracts and covenants as Hobbes defines them and discuss when these different types of agreements are valid. A contract is simply the mutual transferring of rights.[24] If in a system that already has respect for property I agree to trade you food for money, this would be a contract. A covenant is a special contract in which one party delivers goods or services at a time prior to the other.[25] For example, George and Selma are neighboring farmers. They have crops that are harvested at different times of the year and recognize that by working together to harvest each set of crops they will each gain more in their yield of crops than they would doing it alone without requiring additional effort. Because the crops are harvested at different times, George fulfills his contract to help Selma and then Selma has a covenant to help George at a later date when his crops are ready. Because covenants are merely contracts that are fulfilled at a later date, from this point forward I will use contract in both cases for simplification.

---

[22] Hobbes, *Leviathan,* 106.
[23] Ibid., 100.
[24] Ibid., 101.
[25] Ibid.

Knowing the definition of a contract is only half the battle because a contract can easily be invalid. What makes a contract invalid? Why are we concerned with the validity of contracts? In the light of rational egoism, invalid contracts provide us no benefit and thus only valid contracts rationally ought to be adhered to. Hobbes' third law of nature is "that men perform their [valid] covenants made,"[26] because without honoring our contracts we would revert back to the state of nature that is all men at war with each other, thereby losing any potential benefit from cooperation. Therefore, Hobbes' definition of justice is following through with valid contracts and covenants you have made. That is all well and good, but what invalidates a contract? At first, one might think that a contract made out of fear should be invalid, but according to Hobbes this is not the case. When there is no law to prevent such a contract, it is a valid contract. Take the above example with the farmers George and Selma. George happens to have a stockpile of weapons and Selma does not. George could refuse Selma's arrangement of helping each other and demand that Selma help him harvest his crops regardless. If she does, he will allow her to keep her own crops and her life. What a deal! Even though this situation is obviously skewed due to an imbalance of power, this is still a valid contract where George has received help with harvesting, and Selma keeps her life when it could have been taken. Hobbes does make two stipulations, however, where a contract would not be valid. The first is the case of making a contract to give up your life. As mentioned previously, someone cannot give up their own right to life because this could not be in their benefit. As such, a contract where someone agrees to be killed is ultimately not to their benefit, which gives them no reason to follow through. If they have no reason to follow through, there is reason to believe they will not and thus the contract is invalid.

[26] Ibid., 108.

The other way a contract is made invalid is crucial to Hobbes' overall argument. A contract is made invalid if there is reason to suspect that the other party will not follow through with their part of the contract.[27] The problem with this is that in the state of nature there appears to always be reasonable suspicion that the other will not follow through making all contracts invalid. To get around this problem, Hobbes suggests that what is needed is a commonwealth or some type of coercive force to enforce contracts that are made. This coercive force eliminates the fear of non-compliance by imposing "terror of some punishment greater than the benefit they expect to gain by the breach of their covenant…."[28] Essentially the coercive force enacted by the commonwealth makes it irrational to break your contracts. This allows for contracts to be made and justice, in the form of honoring contracts, to emerge. In order to better explain how coercive force alters what is rational, I will briefly discuss some fundamentals of game theory and then return to Hobbes and his commonwealth.

Prisoner's Dilemma

Within game theory[29] there is a famous scenario called the Prisoner's Dilemma, henceforth PD. PD illustrates a major problem with systems of cooperation and social interaction in general. Before getting into the specifics of this, it is important to understand the basics of game theory. Within game theory there are actors, actions, strategies, and outcomes.

---

[27] Ibid., 105.
[28] Ibid., 108.
[29] Most of the concepts on game theory in this section are being pulled from what I learned through lectures given by Douglas Dion in Intro to Political Analysis, Fall 2016. No notes were used and any additionally referenced material will be cited as such.

Actor: (Also referred to as a player) is someone who has an action to make that will impact the current scenario. There can be a single player or more than one, and players can act either at the same time or individually.

Actions: Each player has a set of actions to choose from, and each action has an outcome or set of outcomes. If there are two players who each have two possible actions, then a single player's action has two possible outcomes, one for each of the other player's possible actions.

As mentioned in the section on Glaucon, I am presenting these scenarios through the lens of rational egoism, and as such each player is choosing actions based on what will give the most pleasure/happiness. Now when evaluating which action the player has good reason to choose, the rational decision is more or less straightforward when there is only a single player. It can be illustrated as follows:

Player A	Action 1	Outcome 1
	Action 2	Outcome 2

Again, the best action is the one with the outcome that provides more happiness/pleasure to the player. I say *more or less straightforward* because oftentimes deciding which outcome is greater can be difficult. To give an idea of why this is let's look at the following scenario:

Ryan	Give Information	Outcome 1
	Stay Silent	Outcome 2

In this scenario, Ryan, a soldier that has been captured by his enemies, is being tortured for information. He can choose to either give information or stay silent. Based on rational egoism, Ryan is going to make decisions based on his own self-interests, but what brings us pleasure can be complicated. If he chooses to give information, the torture stops, but if he ever

gets home, he will be labeled as a traitor and he could lose his family, spend time in jail, or worse. If he stays silent, eventually either he will die, or they will give up. If he is rescued, he will be hailed as a national hero, and all sorts of rewards will follow. Which is the right decision? I will spend more time discussing this concept later, but for the rest of this section I will focus on simple outcomes, ones where only immediate results from an action are considered and not the future impacts of that decision.

In comparison with single player scenarios, adding even one additional player significantly complicates what action is rational. When you add a second, player the scenario can be illustrated as follows:

	Player B		
Player A		Action 1	Action 2
	Action 1	(A1,B1)	(A1,B2)
	Action 2	(A2,B1)	(A2,B2)

Here we can see sets of outcomes, one for Player A and one for Player B, each associated with a set of actions. Player A's outcome is always listed first for ease of reading the chart. Outcomes can be measured in many different ways such as money earned, years in prison, and simple values. The reason this becomes so complicated is that each player no longer decides on one action with one outcome but must also consider what the other player might choose.

Now, there are an infinite number of different sets of actions and outcomes that we could evaluate but for now I want to specify four: selling pies, going to a movie, one-way attraction, and the prisoner's dilemma. There are two specific types of outcomes that should be discussed before evaluating the following scenarios. The first is the Nash Equilibrium, henceforth NE.

Nash Equilibrium: An outcome where neither player would individually benefit from changing his/her action, provided all other players keep their actions the same.[30]

There is not always a Nash equilibrium, but there can also be more than one. The next type of outcome that is important is an optimal outcome.

Optimal Outcome: One in which no alternative outcomes would make some player(s) better off and without hurting anyone else.[31]

Essentially, an outcome is optimal if there is no other outcome that everyone would rationally agree to switch to. There can be multiple optimal outcomes for each scenario. Let's take a look at the following examples.

S0: Selling Pies	P2: Jake		
P1: Lindsay		A1: Agreed Price	A2: Lower Price
	A1: Agreed Price	($500,$500)	($200,$800)
	A2: Lower Price	($800,$200)	($250,$250)

In the first example we have Lindsay and Jake who are two pie bakers at a state fair. If they can work together by setting an agreed upon price they can avoid any price undercutting from the other player and both bakers will make more money, $500 for the one day fair. If, however, one of the two breaks their agreement and lowers his/her pie prices, that baker will have significantly more sales because his/her pies are cheaper. As a result, they make more profit, $800 and the other makes less from reduced sales, only $200. Now, if both players decide to undercut the other they both leave with only $250. This example has both a NE and an optimal outcome; the issue is that they are different outcomes. The optimal outcome is both players

[30] David Gauthier, *Morals by Agreement*, Oxford: Clarendon Press, 2006, 65-6.
[31] Ibid., 76.

honoring the agreed upon price. Though a player could gain more by breaking the agreement, the other gets less, and thus it is a non-optimal outcome. The NE, on the other hand is when both players break the agreement. We can see this because if both players choose to lower their prices, neither would benefit from changing their choice. To do this would simply benefit the other player and hurt the player making the change. Though the NE is a worse outcome at $250 each as compared with the optimal outcome of $500 each, it is the natural resting point of this scenario.

I should explain what I mean by natural resting point. Take a look at the scenario we have above. When the fair starts, both players have agreed to a specific price. Based on rational egoism, both players want the most amount of money for themselves; knowing that the other has agreed to keep their prices stable, if one breaks their agreement that baker makes $300 more than otherwise. But both players know this, and they also know if the other breaks the agreement and they honor it, they only get $200. If they are convinced the other is likely to break the agreement, they would be better off also breaking the agreement and getting $250 rather than $200. This is why the NE is the resting point. Without trust or some way of altering the outcomes, both players have reason to break their agreement and thus reason to think the other will. If one breaks their agreement, the second player is better off breaking theirs as well. When the NE and optimal outcomes do not line up, agreement is the hardest, but there are some types of scenarios where agreement is far easier. The types of scenarios are as follow:

Type 1: NE that is the same as an optimal outcome.

Type 2: Optimal outcome(s) with no NE.

Type 3: Optimal outcome that is different than the NE.

I will now give an example of each of these types. The first two will not use money or any specific outcome but simple values on a scale of 1-4. Think of it as units of pleasure, 4 units is better than 1 unit. Thus the higher the value equals the better outcome.

S1: Going to a Movie	P2: Fred		
P1: Todd		A1: Horror	A2: Action
	A1: Horror	(3,4)	(1,1)
	A2: Action	(2,2)	(4,3)

In this Type 1 scenario we have Todd and Fred who are friends. They both prefer to go to a movie together but each have their own preference on the type of movie they see. Todd prefers action movies while Fred likes horror. Both would rather see any movie with their friend than see one alone but would prefer they agree to go to their favorite type of movie. In this scenario we have two NE outcomes that also happen to be optimal: (A1,A1) and (A2,A2). The optimal outcomes are where they go and see a movie together and because seeing a movie together is more important than seeing the type they prefer. Once they agree on a type of movie, neither would want to switch. For example, if they both choose to go to a horror but then if Fred were to switch, he would get 1 unit of happiness instead of 4. Were Todd to change from going together, he would get 2 instead of 3. The same would go for if they had chosen an action film but the values would be swapped, Todd would trade 4 for 1 and Fred 3 for 2. This is an ideal cooperation scenario because neither person has a reason to switch their own action, and as a group they have nothing to gain from a different strategy.

S2: One-Way Attraction	P2: Sally		
P1: Todd		A1: Go to Party	A2: Stay Home
	A1: Go to Party	(4,1)	(2,3)
	A2: Stay Home	(1,4)	(3,2)

The Type 2 scenario has Todd once again who has fallen in love with Sally. The problem is Sally does not share Todd's feelings. Todd would rather be where Sally is but is not much of a party goer. Sally, on the other hand, loves parties but is unable to enjoy herself when Todd is there due to his repeated unwanted advances. In this scenario all of the outcomes are optimal but there are no NE outcomes. If we start with Sally going to the party and Todd staying home, he would decide to go to the party. If Todd comes to the party, Sally would decide to stay home. If Sally stays home, then Todd would want to stay home; and if Todd stays home, then Sally would want to go to the party. Around the circle we go. There is no NE because someone always would benefit from changing their action provided the other stayed the same. Every outcome, however, is optimal because there is no set of actions that is better for one player without be worse for the other.

This case demonstrates some difficulties with cooperation, but there are strategies that can help here. Let's look at a couple different options. If, because the two cannot decide whether to go or not, they just decide to flip a coin each night then each would have a 50% chance of going or staying. This means that each outcome has a 25% chance of occurring:

25% Todd goes, Sally stays
25% Todd goes, Sally goes
25% Todd stays, Sally stays
25% Todd stays, Sally goes

First we take the outcome values and multiply them by their probability. From this we get the following (probabilities of doing a given action are given out of 1):

S2: One-Way Attraction	P2: Sally		
P1: Todd		A1: Go to Party (0.5)	A2: Stay Home (0.5)
	A1: Go to Party (0.5)	(1,0.25)	(0.5,0.75)
	A2: Stay Home (0.5)	(0.25,1)	(0.75,0.5)

Now if we add the values together, it gives us the average value of randomly choosing to go to the party or stay home. Both players get 1+0.5+0.75+0.25 which equals an average value of 2.5. What if instead of making random decisions, Todd decides to just go to every party hoping to see Sally? It would look like this:

S2: One-Way Attraction	P2: Sally		
P1: Todd		A1: Go to Party (0.5)	A2: Stay Home (0.5)
	A1: Go to Party (1)	(2,0.5)	(1,1.5)
	A2: Stay Home (0)	(0,0)	(0,0)

Under this arrangement Todd would have an expected value of 2+1+0+0 which equals 3. However, Sally only has an expected value of 0.5+1.5+0+0 which equals 2. This outcome is better for Todd, but surely Sally would just stay home as soon as she knew Todd was always going to the party. Because every outcome is optimal and there is no NE, Sally and Todd will eventually return to randomly choosing their actions, and ultimately anything but randomly choosing their actions will be better for one or the other. Thus, someone will then want to modify their strategy.

S3: Prisoner's Dilemma	P2: Betty		
P1: Todd		A1: Keep Quiet	A2: Confess
	A1: Keep Quiet	(-4,-4)	(-20,0)
	A2: Confess	(0,-20)	(-16,-16)

This final scenario is the famous Prisoner's Dilemma and a type 3 scenario like the pie bakers scenario. Todd and Betty are partners in crime. They have just been captured for

committing a small robbery but the police suspect them for a murder. They are isolated and interrogated. The police make the following claim. “We know you committed the robbery and for it you’re going to jail for 4 years. If you give us information to put your partner away for the murder, we will let you off the robbery charge (what a deal!). If however your partner confesses, you will get 16 additional years for murder.” The outcomes are as follows. If both Todd and Betty keep quiet, they will both get only 4 years. If they both confess, they will both get 16 years. If, however, one confesses and one keeps quiet, then the one that confesses will walk free and the other will get 20 years. This scenario is interesting because it has an optimal outcome and a NE like with the movie goers, but these outcomes are different. In a world where you cannot trust your fellow criminal, the only rational outcome is the NE or both people confessing. Both Todd and Betty get 16 years. The optimal outcome, however, would be neither confessing and each receiving only $1/4^{th}$ the amount of time at 4 years. The problem is that when the optimal outcome and NE do not line up, the optimal outcome is hard to obtain. If they both go in with the agreement to keep quiet, both Todd and Betty have good reason to confess. They know that if they confess and the other doesn’t, they get 0 rather than 4 year. If they keep quiet and the other confesses, the quiet one goes down for 20 years. It seems they have little reason not to confess. Because of this attitude, both are likely to confess, and thus both get 16 rather than 4 years. This optimal outcome seems out of reach, but is it? That is something I will discuss in detail later in this paper.

Back to Hobbes

Now that we have a basic understanding of game theory, we can see that even if both players can see the optimal outcome, they cannot obtain it because of a lack of control and/or trust of the other player. How does Hobbes’ commonwealth use coercive force to change what is rational for

the different players? Let's take a look at the very first scenario discussed in the previous section to see how it is impacted by some coercive force. The original version of the scenario was:

S0: Selling Pies	P2: Jake		
P1: Lindsay		A1: Agreed Price	A2: Lower Price
	A1: Agreed Price	($500,$500)	($200,$800)
	A2: Lower Price	($800,$200)	($250,$250)

Consider that Lindsay and Jake are not alone in their agreement to sell their pies at the same amount but belong to an organization of bakers that has the ability to fine them for lowering their prices. The fine happens to be $600 and so the new scenario would look like:

S0: Selling Pies	P2: Jake		
P1: Lindsay		A1: Agreed Price	A2: Lower Price (-$600)
	A1: Agreed Price	($500,$500)	($200,$200)
	A2: Lower Price (-$600)	($200,$200)	(-$350,-$350)

In this new scenario the choice comes down to either choosing to keep the agreed upon price and getting either $500/$200 or lowering your price to get more sales but after the fine ending with $200/-$350. The best outcome from lowering your prices is worse than the worst outcome from keeping the agreed upon price. Thus, it would be irrational to choose to lower your prices. It is crucial to mention that the coercive force, in this case the fine, only works to aid in both players following through on their contract if both conditions of what I will term *dual awareness* are met. Dual awareness' conditions are:

1: both players know of the fine and believe it will be enforced

2: both players know the other is aware of this as well.

If either of these two conditions are not met, then there is reasonable suspicion that the other will not follow through, and thus the contract would be invalid. If both conditions are met, however, then in this case Lindsay and Jake can trust each other, given the other is rational, and their contract with each other is valid.

Next let's look at an even more extreme example. Take the case of Todd and Betty in the prisoner's dilemma. Again for reference, below is the original scenario with no commonwealth or coercive force.

S3: Prisoner's Dilemma	P2: Betty		
P1: Todd		A1: Keep Quiet	A2: Confess
	A1: Keep Quiet	(-4,-4)	(-20,0)
	A2: Confess	(0,-20)	(-16,-16)

For the second version of this scenario, instead of being totally isolated, the two belong to a group of criminals with their own set of "laws" and leaders that enforce those laws. It might help to think of the group as similar to the mafia. They have all made an agreement to never confess to the police and recognize that if they do confess they will be killed by the other members of their group. The new set of choices and outcomes appears as follows:

S4: Prisoner's Dilemma W/ Commonwealth	P2: Betty		
P1: Todd		A1: Keep Quiet	A2: Confess
	A1: Keep Quiet	(-4,-4)	(-20,dead)
	A2: Confess	(dead,-20)	(dead,dead)

If Todd and/or Betty keep quiet in this version, they could risk getting 20 years in prison but they can trust that the other will also keep quiet because 20 years is significantly better than being killed. Because their group has applied this coercive force, which greatly changes the

outcomes related to breaking their contract, it is only rational to be just and keep the agreement (provided both conditions of dual awareness are met). This is the point of the commonwealth; it allows for contracts to be made where otherwise they would be impossible.

At this point we should address Hobbes famous fool's objection:

> "The fool hath said in his heart, there is no such thing as justice; and sometimes also with his tongue, seriously alleging that every man's conservation and contentment being committed to his own care, there could be no reason why every man might not do what he thought conduced thereunto; and therefore also to make or not make, keep or not keep covenants was not against reason when it conduced to one's benefit."[32]

Essentially the fool is suggesting that even though it might be in one's interest to follow through on your contracts most of the time, there could be situations where it is ultimately in your interest to break your contracts. What might a scenario such as this look like? Let's modify the baker's example from before:

S0: Selling Pies	P2: Jake		
P1: Lindsay		A1: Agreed Price	A2: Lower Price (25% Jail)
	A1: Agreed Price	($500,$500)	($200,$10,000,000+1% chance of 10 years in jail)
	A2: Lower Price (25% Jail)	($10,000,000+25% chance of 10 years in jail, $200)	($250+25% chance of 10 years in jail, $250+25% chance of 10 years in jail)

Now, if either defect they have a 25% chance of being charged and sentenced to jail for 10 years. Also, if they are the only one to defect, they catapult their business into a multi-million dollar bakery business. Because the coercive force is being applied by humans, there is no guarantee that it will actually happen. As such, with a low risk of 10 years in jail, the reward of

[32] Hobbes, *Leviathan*, 109.

over 10 million dollars might be worth that risk. If we decided to lower the risk or the negative consequence, the player would need even less to make it rationally worth breaking their contract.

Hobbes has a few potential responses to this objection. The first would be to claim that in the situation described just now, the contract between Lindsay and Jake would be invalid. This is because if the risk is great enough to outweigh the reward, then both players have a reason to not follow through with their agreement even with the coercive force in play. This would invalidate the contract. But the failure of the coercive force is due to humans not being perfect. If Hobbes intends to require that there never even be a hint of potential benefit from breaking a contract, then this eliminates most if not all opportunities for agreement. This would prevent any commonwealth from having the force to enable the creation of valid contracts thus leaving us stuck in the state of nature.

A more reasonable solution is to argue that situations like this are extremely rare, and we often go wrong in our calculations of risk. It might seem that there is only a 25% chance of going to jail, but what if the odds are 75%? This may sway the rationality of breaking the agreement. Also, the outcomes are often far more complicated than simply a chance of jail time, and as the outcome gets more complicated, our ability to calculate risk gets worse. Another complication in the calculations is the social consequences of breaking contracts. Hobbes states:

> "He, therefore, that breaketh his covenant and consequently declareth that he thinks he may with reason do so, cannot be received into any society that unite themselves for peace and defense but by the error of them that receive him; nor when he is received be retained in it without seeing the danger of their error…."[33]

To allow someone who makes these kinds of risk calculations into society is an error on the rest of society, and once someone is discovered as someone who breaks their contracts it is only

[33] Ibid., 110.

rational for those in his society to remove him from it. So by breaking contracts, you risk any potential benefits and protections from cooperation and society as a whole.

The other consideration is that the situations where there would be enough benefit to justify the risk of being exiled from any future contracts would be extremely rare if existent at all. How many times in your life have you been presented with an opportunity to make millions of dollars with only a low risk of a really negative reward? Personally, I have yet to be presented such an occasion, and that is likely the case for most others.

There is another concern for Hobbes' position. How does a commonwealth emerge out of nature? It appears that there would need to be some initial agreement to get any sort of society going but that agreement would be invalid and thus irrational to follow through on. It would be irrational because there is no coercive force in place yet. This brings up another problem. Though Hobbes can offer the previous responses, they feel a bit lacking. There are still many different scenarios where there is no coercive force present and Hobbes view provides no method for gaining the benefit of contracts in these cases. It is unrealistic to think all of our interpersonal interactions would or even could be governed by a sovereign. If there is no sovereign control, there can be no valid contracts and therefore, under Hobbes view we would miss out on the benefits of cooperation in at least some, if not most, of our day to day lives. There may be ways Hobbes could respond to these objections, but instead I will present another contractarian view, presented by David Gauthier in his *Morals by Agreement*, which I think better addresses these concerns.

Gauthier and Morals by Agreement

Gauthier's goal is to show that, "in certain situations involving interaction with others, an individual chooses rationally only in so far as he contains his pursuit of his own interest or advantage to conform to principles expressing the impartiality characteristic of morality."[34] As I mentioned in the previous sections, in cases similar to the prisoner's dilemma there are differences between the resting point of the player's actions, the NE, and the optimal outcome. Gauthier claims that we can gain the benefits of the optimal outcomes through rational constraints on self-interest. Morality, as defined by Gauthier, are these impartial rational constraints and he claims, "individual[s], reasoning from non-moral premisses, would accept the constraints of morality on his choices."[35] In other words, rational and self-interested people would be willing to adopt impartial constraints on the pursuit of their own self-interest.

<u>The Market</u>

The first important concept to Gauthier's overall argument is that of a perfectly free market. Though he spends a fair amount of time discussing the intricacies of a market system, it is not as important to the overall point of this paper. For this reason I will only touch on it briefly to show his definition of a market, and what part the market plays in his overall argument.

According to Gauthier, a free market has two main components, private ownership and private consumption. These together give the free market an absence of externalities and ultimately the removal of "circumstantial uncertainty and strategic calculation."[36] When circumstantial uncertainty and strategic calculation are removed, there is no longer a need for morality, and hence a free market is also a morally free zone. What does all of this mean?

---

[34] Gauthier, *Morals by Agreement*, 4.
[35] Ibid., 5.
[36] Ibid., 85.

Private ownership, as a feature of the economic system Gauthier is describing, is comprised of two parts: individual factor endowments and free individual market activity.[37] Individual factor endowments are the means of production that each individual of the market has. For example, this could be ownership of a field to produce crops or access to a kitchen for baking pies. Individual market activity provides that each individual can use their production factors and products as they please either for production, consumption, or exchange. Private consumption on the other hand is composed of private goods and mutual unconcern.[38] To say a good is private is to indicate that its consumption can or must be exclusive to a single individual or group. Food is a private good, once it is eaten, it is gone. A house is also a private good; it may benefit a group, but the amount of people it can benefit is limited. Clean air, on the other hand, is a public good because it provides benefit to everyone; its goods cannot be exclusive. Finally, mutual unconcern requires that individuals not take a concern in the interest of those with which they make exchanges.

Now when all four of these aspects hold, there can be no externalities. Externalities are positive or negative impacts on individuals not part of an exchange. Gauthier uses the example of lighthouses to illustrate the importance of externalities. If a handful of ship-owners agree to set up a lighthouse to aid in navigation near a major port, all ship-owners that use the port benefit from the construction of the light house. There is no ability of those that construct it to limit the use of the lighthouse. In this way, when either supply or demand changes, the other is not impacted which is important to the free market. For example, as more ship-owners start to use the port, though demand for a lighthouse may increase, a single light house is sufficent no matter how many ship-owners are using the port. The importance of externalities is that they prevent the

[37] Ibid., 86.
[38] Ibid., 87.

crucial direct interaction between supply and demand. Since a free market has no public goods, it also has no excess goods. As demand increase or decreases, so does supply in order to meet that demand. This results in one being able to be made better off without making someone else worse off, and thus all outcomes are optimal.[39] The other result of no externalities is that, in conjunction with both private ownership and private consumption, we can view individuals actions as free from circumstantial uncertainty and thus from a need for strategic calculation.[40] This is due to three reason: individuals will always act in their own interest, individuals cannot be impacted by exchanges that they were not a part of, and all outcomes are optimal. When all three of these conditions hold, one must simply choose the outcome that is in their best interest, and others will do the same.

How is this important to Gauthier's overall argument? Gauthier is developing a system where there is no need for impartial constraint, so he can then differentiate between that system and why we need those constraints in the world we actually live in. He thinks the above restrictions to a market system can ultimately remove this need because "since the market outcome is both in equilibrium and optimal, its operation is shown to be rational, and since it proceeds through the free activity of individuals, we claim that its rationality leaves no place for moral assessment."[41] In other words, in this free market everyone is acting independently for their own interest. Because every outcome is optimal, no one would agree to alter their behavior. If this all holds, then there would be no cases of the prisoner's dilemma where parties would gain by cooperation. As such, honoring ones contracts, or morality as we understand it for this paper, has no place in a free market system.

---

[39] This is the definition we are using to understand optimal outcomes as presented in the section on the prisoner's dilemma.
[40] Gauthier, *Morals by Agreement*, 85.
[41] Ibid., 94.

In moving forward we will examine various situations that are everyday occurrences and yet are not possible within a market system and show how rationally one should compromise in order to gain benefits unavailable to non-cooperators.

Justice and Bargaining

At this point, Gauthier moves on to discuss his conception of justice and various aspects of bargaining. Gauthier defines justice as, "the disposition not to take advantage of one's fellows, not to seek free goods or to impose uncompensated costs, provided that one supposes others similarly disposed."[42] Ultimately, Gauthier aims to argue that it is rational to be just and to develop such a disposition, but more on that in a bit. We must first understand how Gauthier conceives of bargaining.

The first way in which normal societies differ from a market system is the variable nature of the supply of goods.[43] Because there is an ebb and flow in the supply and we have an awareness of this fact, we begin to view others in different ways. First, we start to view others as competition for those goods. Second, we see others as a means to increase production of the supply through cooperation. There are some goods, in fact, that are only possible through cooperation such as harvesting several large fields of crops throughout the year. Normal, everyday maintenance may be accomplished by a single person, but the necessity to harvest the entire field in a short window requires cooperation. This understanding of the variable nature of supply in conjunction with an awareness of externalities and the self-interested nature of others results in a rational need for cooperation. In what follows, I will explore what Gauthier takes to be "the conditions for rational co-operative choice, or rational agreement on an outcome."[44]

---

[42] Ibid., 113.
[43] Ibid., 114.
[44] Ibid., 117.

The first condition of rational agreement on an outcome is that the outcome is optimal.[45] This is fairly straightforward based on our understanding of optimal outcomes. Let's take a quick look again at the scenario with Todd and Fred to see more directly how Gauthier's rational agreement relies on optimal outcomes (numbers have been slightly modified from before to more clearly illustrate the point).

S1: Going to a Movie	P2: Fred		
P1: Todd		A1: Horror	A2: Action
	A1: Horror	(3,6)	(1,1)
	A2: Action	(2,2)	(6,3)

As previously discussed, the two optimal outcomes are when both Todd and Fred can agree to go to a specific movie. But because there are two optimal outcomes, how do the two friends decide which is rational to agree upon? This is where I need to bring up Gauthier's ideas of the initial bargaining position, cooperative surplus, and the principle of minimax relative concession.

> Initial bargaining position: The value each player would get without any cooperation.[46]
> Cooperative surplus: The greatest additional value each player can get through cooperation as opposed to non-cooperation.[47]
> Principle of Minimax Relative Concession (PMRC): "In any cooperative interaction, the rational joint strategy is determined by a bargain among the cooperators in which each advances his maximal claim and then offers a concession no greater in relative magnitude than the minimax concession."[48] In other words, each asks for the most they could get and then each settle on the largest claim, or least concession that is required for an agreement to be reached.

How do these ideas play a role in determining what is rational for Todd and Fred to agree to? First, what is the initial bargaining position for both Todd and Fred? The best outcome each

[45] Ibid., 117.
[46] Ibid., 130.
[47] Ibid.
[48] Ibid., 145.

could hope to gain without cooperation is randomly choosing which to go to and hoping the other is there. If each assign a probability of 0.5 to each choice, the outcome chart would look as follows:

S1: Going to a Movie	P2: Fred		
P1: Todd		A1: Horror (0.5)	A2: Action (0.5)
	A1: Horror (0.5)	(0.75,1.5)	(0.25,0.25)
	A2: Action (0.5)	(0.5,0.5)	(1.5,0.75)

When we add these outcomes together, we get an average expected value of 3 for both Todd and Fred. This is the initial bargaining position and what each could expect without being able to agree to a movie. What are each of their Cooperative Surpluses? This is essentially the most each could hope for beyond their initial bargaining position. For Todd this would mean going to exclusively action films with Fred and for Fred it would be both going exclusively to horror films. These two sets of actions would result in one of the two getting a value of 6 or a cooperative surplus of 3. Now the PMRC says it would be rational for each to advance their maximal claim for 6, going only to their preferred type of movie. They would then offer to concede a relatively equal amount or 1.5 and both would get an expected value of 4.5. This is ultimately accomplished by them agreeing to alternate which movie they go to. This is only an equal concession, next I will show an example with a relative concession.

For this example imagine there are two business women, Caitlyn and Ashley. They have the opportunity to enter into a joint business venture but on unequal grounds. If Caitlyn were to be given all the surplus from their cooperation, she would earn herself $1000. Ashley, on the other hand, would only get $100 from getting all of the surplus. Now in this case they do not need to give up half of their surplus but only 30% to reach a relative concession. This can be illustrated as follows:

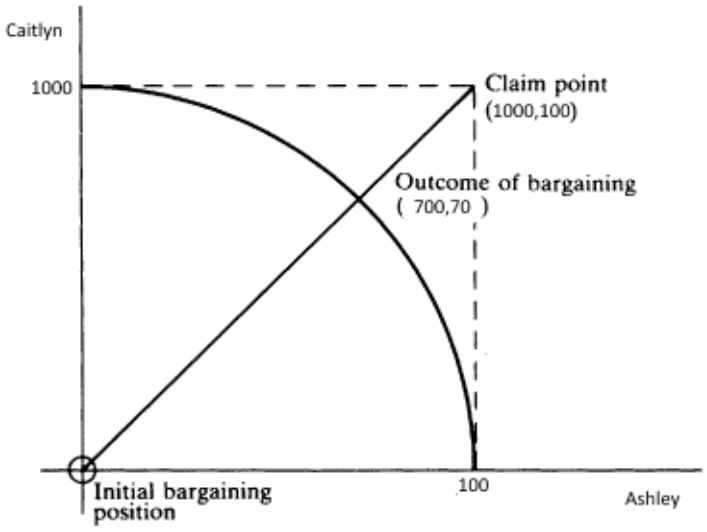

Figure 1[49]

Some might be concerned that Ashley would be upset by Caitlyn getting $700 while she only gets $70. Ashley might argue that Caitlyn should give up a little more so that each end up with closer to the same amount. The problem with this line of thought is that Caitlyn can argue the opposite. She might argue that Ashley is only giving up $30 while she is forced to give up $300. In her mind, Ashley should give up just a slight bit more, so that their concessions are closer to equal. This is why Gauthier argues it is rational to accept a relative concession. Though you are not gaining the entire surplus, that is not a realistic scenario. If the other person stands to gain nothing, they will not participate, nor would you if the situation were reversed. By using relative concessions, rational self-interested individuals can reach an agreement that is deemed fair by both parties and by reaching an agreement reap the partial surplus of cooperation.

There are two major concerns here. The first is that some might not be okay with a relative concession. Perhaps Caitlyn, being self-interested, refuses to take an equal relative concession and is only willing to take $900 which leaves Ashley with approximately $24.

[49] Ibid., 139. This is a modified chart to suit my example.

Ashley's choices are to either not get any benefit from the cooperation or settle for approximately $24, far less than with a relative concession. This is a concern, but Gauthier claims that Ashley would find another to make a more equitable arrangement ultimately leaving Caitlyn with nothing. In an ideal system with many people to cooperate with, this may be possible but in reality it is far less likely. What if Ashley is stuck with no other partners for cooperation? Would it be rational to cooperate? It would seem that with Gauthier's system the stronger willed individual will benefit as the other will be forced to give up more than a relative concession. If this is true then the PMRC does not appear to be rational for a self-interested utility maximizer to follow. I will temporarily leave this concern and return to it a little later on.

The other major concern is that once a bargain has been rationally reached by two individuals, is it actually rational for them to follow through on that agreement? In the next section I will explore Gauthier's response that it is, in fact, rational to become a constrained maximizer who honors their agreements.

Constrained Maximization

When evaluating the rationality of following through on agreements Gauthier discusses two different approaches people can take. They are:

> Straightforward Maximizer: This person considers each agreement individually, making his choice based on his expectations for the actions of the other player with his aim being maximizing personal value.[50]
> Constrained Maximizer: This person "is ready to co-operate in ways that, if followed by all, would yield outcomes that she would find beneficial and not unfair, and she does co-operate should she expect an actual practice or activity to be beneficial."[51]

The straightforward maximizer has, as the name suggests, a much more straightforward decision to make. Based on how they view potential outcomes, they choose which action

[50] Ibid., 167.
[51] Ibid.

provides them the most value based on what they expect others to do. Now, the SM is not stupid and takes into consideration not only immediate consequences but also potential future consequences of his actions. Regardless of this, he may find himself faced with situations where, as suggested by Hobbes' fool, it is in his benefit to break his contracts and does so.

One such example is a special, no outside consequence version of the prisoner's dilemma. Generally when laying out the prisoner's dilemma, all that is accounted for is the years in jail that result from the decision. There are, however, many other potential consequences such as reputation, ability to make deals in the future, etc. that are also at risk with any decision made. This is because often we want to evaluate a strategy or plan of action over several instances of the scenario in order to determine its long term effectiveness. If someone is found to always honor their agreement no matter what, the other player would be foolish not to take advantage of that irrationality and confess to the police every time. On the other side, if it is shown that someone will never cooperate, then cooperation is foolish. If a straightforward maximizer were to know they were going to repeat the prisoner's dilemma over and over with the same person, they could try a couple different things. First, they know in any single scenario it is best to confess to the police when the other keeps silent. Here is the scenario again for reference:

S3: Prisoner's Dilemma	P2: Betty		
P1: Todd		A1: Keep Quiet	A2: Confess
	A1: Keep Quiet	(-4,-4)	(-20,0)
	A2: Confess	(0,-20)	(-16,-16)

For repeated scenarios this takes on a different aspect. First, by not keeping their agreement, SMs risks losing the ability to make another contract with the other. Second, they risks losing the trust of the other person thus making it far more likely they will break the

agreement in future. Over a long set of scenarios, this makes breaking their contract quite costly and most often not worth it. What is interesting is how a SM would respond in the special version of the prisoner's dilemma mentioned previously. In this situation there are no external impacts of any decision. All that is at stake is the number of years in prison. No one ever finds out what decision you made so it does not impact your reputation, your ability to make future agreements, or anything else. In this case the SM has no reason to cooperate because nothing outside of this scenario is at risk. Confessing is always better for them, regardless of what the other does. If the other confesses, the SM would have been worse off having kept quiet, and if the other keeps quiet, the SM again is better off having confessed. The problem with this strategy is that in a world of SMs the other is also going to confess and both end up with 16 years in prison rather than the 4 years from dual cooperation.

This is one of the major differences that Gauthier points out between SMs and CMs. He claims that constrained maximizers are able to gain the benefit of dual cooperation even in prisoner's dilemma scenarios with no external consequences. What exactly are constrained maximizers, and how are they able to get that benefit? The decision making process of a CM is very different from that of the SM. Rather than basing the decision off of what is individually best based on other decisions, the CM bases her decision on the shared benefit/fairness of a cooperative action and her expectation of the other's willingness to cooperate.[52] If there is an outcome that is beneficial and close to fair for all involved and the CM expects the other to cooperate, she will do so as well. Gauthier makes an even stronger claim by stating that the CM does not simply make the decision at the time but must have a disposition to cooperate, provided that she estimates the other will cooperate. This disposition is what makes someone an

[52] Ibid., 169.

acceptable partner for a contract.[53] How does this play out in the prisoner's dilemma with no external consequences? In a world filled with CMs both cooperate and only get 4 years. This is far better than in a world filled with SMs where they both get 16 years, but what about a world with both CMs and SMs?

In a world that is a mix of both CMs and SMs the ability of either to do well depends on the ability to detect the difference between CMs and SMs and the ability of SMs to hide their disposition. It would be easy to detect the difference if everyone were transparent in what type of person they were, but this is unrealistic. Instead if people are only translucent, meaning others have the ability to detect correctly more often than incorrectly, Gauthier claims the CMs still are better off.[54] If people are only translucent, then sometimes CMs will miss out on cooperation with other CMs from miscalculating and sometimes they will cooperate with SMs. SMs, on the other hand, will sometimes try cooperating with other SMs which will hurt them. Gauthier discusses this at length but concludes that even with the form of translucency we are talking about, it is of benefit to be a CM. He concludes this because as the percent of CMs increases it becomes more and more profitable to be a CM. This is due to the reduced risk associated with sometimes mistaking another for being a CM when they are a SM. Even when there is an even number of CMs to SMs there is incentive to improving detection, which helps CMs much more significantly than SMs.

Even if we grant to Gauthier that we can detect what someone's disposition is, there still seem to be a couple major problems with saying that the rational choice is to be a constrained maximizer. First, where exactly is the constraint? It appears that CMs are making rational choices when considering all of the impacts of being a CM. Also, there appears to be an

[53] Ibid., 173.
[54] Ibid., 176-7.

additional type of person who could get the benefits of both the CMs and the SMs. In response to the first concern, Gauthier states CMs are rationally choosing to develop the disposition for constrained maximization. Though developing this disposition is rational, once the disposition is developed it can lead to some irrational actions. He equates this to the idea of following through on threats. When the threat is originally issued its purpose is to modify behavior of another. If the threat fails, it may actually be worse to follow through on the threat for the one that issued it. Gauthier's point is that being known as someone that is disposed to irrationally follow through on failed threats puts you in a better position when issuing those threats.[55] For the case of the CMs, it is thus rational to develop a tendency to do even sometimes irrational things because others will see that and know they can trust you, opening up the opportunities only available to CMs.

Now, for the next concern of an additional type of person in addition to CMs and SMs. As we have seen previously there are very few instances where a SM and CM would differ in their actions. This is largely because SMs ability to recognize the future risks associated with any action leads them to cooperate in most instances. The main example given by Gauthier of where CMs and SMs differ is in the special prisoner's dilemma case with no external consequences. It is there that they act differently, where their actions do not impact their future in either reputation and/or ability to make future agreements. But what if there was another type of person, one who in all the cases except these special ones behaved like a CM. How would anyone know otherwise? Their whole life they behave like a CM and for all intents and purposes are one. Then they notice their chance where there will be no future consequences of their actions. Though not as clear, this could also be a situation where, as Hobbes' fool suggests, there is so much on the

[55] Ibid., 185.

line, say 10 billion dollars, it is worth the risk or possible consequences. Let's go back to the situation where the person knows there will be no future consequences. If this is the one place they act like a SM, I suggest that this would provide more value than being a CM. Let's call this person a momentary straightforward maximizer (MSM). How would this play out in that single no future scenario? We can even grant complete transparency over the type of person they are which should benefit CMs, but remember the MSM has lived their whole life as a CM so they would appear to be a CM in this scenario.

	CM	SM	MSM
CM	KQ(-4),KQ(-4)	CON(-16),CON(-16)	KQ(-20),CON(0)
SM	CON(-16),CON(-16)	CON(-16),CON(-16)	CON(-16),CON(-16)
MSM	CON(0),KQ(-20)	CON(-16),CON(-16)	CON(-16),CON(-16)

Now take an average of the totals for each type of person and we are left with:

CM: (-4-16-20)/3 = -13.3
SM: (-16-16-16)/3 = -16
MSM: (0-16-16)/3 = -10.66

These numbers are assuming an even distribution between CMs, SMs, and MSMs. As the number of CMs increases, MSMs only do increasingly better. They are able to portray themselves as a CM thereby gaining all of the normal benefits of CMs but they are able to take advantage of actual CMs. Unlike CMs they are also not taken advantage of by SMs and other MSMs in this rare situation with no future consequences. Some might say, "But what about their reputation and ability to make future interactions?" We already took that off the table with this kind of special scenario, there are no additional consequences either good or bad, outside of years in prison, for anyone involved. So in this case, sure, it is better to be a CM over a SM, but it seems foolish to not be a MSM over the other two.

Unresolved Concerns

There are still a few concerns that I have left unresolved at this point for Gauthier's argument. I will now bring up each briefly and provide what I believe is Gauthier's best response to each concern. The first concern I wish to return to is that of the principle of minimax relative concession. I had mentioned that there was a concern with Caitlyn being more demanding of the cooperative surplus than Ashley and that leaving Ashley with a tough decision. This problem can be made even worse if we take the approach of Gregory Kavka and his "Inequality Glutton."[56] This is a person who unlike Caitlyn does not demand more than an equal share but actually gets more utility due to his psychology simply by getting more of a share than the other. Because of this, Kavka claims, "his maximum claim would correspond to a much higher share of the material benefits of cooperation than would the maximum claim of those who have little or no preference for having much more than others."[57] The end result of this would be that equal relative concession ultimately leaves this glutton with more than most. It rewards him for his psychological drive towards inequality and punishes those that are happy with equality. Surely this is not something Gauthier would want to accept for his view, but how could he respond? He has already stated that each person must "not [take] an interest in the interest of those with whom they exchange."[58] To avoid the problem of the glutton, Gauthier could simply add the requirement that people not care about their standing in regards to others.[59] Though this solves the problem of the glutton, it certainly does not appear to reflect the reality we all find ourselves

[56] Gregory S. Kavka, "Reviewed Work(s): Morals by Agreement by David Gauthier," *Mind*, New Series, 96, no. 381 (1987): 119.
[57] Ibid.
[58] Gauthier, *Morals by Agreement*, 87.
[59] Kavka, "Reviewed Work(s): Morals by Agreement by David Gauthier," 119.

in. Even if you don't desire more than other for the sake of simply having more, I'm confidant that most readers will at least know someone that is like this.

The solution to the inequality glutton is not the only thing that doesn't appear to reflect the world we live in. Gauthier also mentions that being a constrained maximizer relies quite heavily on being able to detect the difference between SMs and CMs at least more often than not. Is this the reality we find ourselves in? Are people more likely to be able to tell if someone is trying to deceive them then not? I don't think so. Again Gauthier has a way out. Perhaps it is only rational to be a CM if one has developed this ability to detect people's dispositions. In all other cases, it would be best to be a SM as to not be taken advantage of. The problem is that the benefit of being a CM in a world that is translucent rather than transparent is that as the number of CMs increases, the greater risks one can take with their cooperation and still be better off being a CM. However, we have just gutted the number of people that can rationally be CMs. It appears that Gauthier must either contend that a good amount of people are capable of detecting other's dispositions, which would need support, or provide an explanation of why being a CM is beneficial even if it is only a small percentage of the population.

The final concern I will bring up in this section is that of MSMs and problems they cause for Gauthier's view. As previously mentioned, MSMs perform as well as CMs all of the time except in the special cases with no future consequences or with extraordinary benefits for breaking a contract. It is in these cases that MSMs outperform CMs, calling into question the rationality of developing the disposition for being a CM. There are two main routes out of this problem for Gauthier. The first is to accept that it might be rational to be a MSM, but he then must either deny that people are capable of living their whole life as a CM while essentially being a SM in disguise or claim that others would be able to notice if someone were to try. Both

of these are fairly weak arguments as the world we live in contains many shady people who often get away with their deception providing very real examples of the contrary. In addition, being a MSM does not require regular deception but merely a recognition of the benefit that would come from defection in very rare situations. They are different from SMs in that they are not regularly looking to break contracts in everyday life. They recognize how rare the situation is where defection is beneficial and only break their contract when and if that moment ever arrives.

The other way Gauthier could respond is again to accept that it is rational to be a MSM over a CM. He would then need to adjust his view to focus on MSMs rather than CMs. The major issue with embracing MSMs as rational is that it prevents Gauthier from providing a response to Glaucon's challenge that Glaucon would find acceptable. If we recall from the start of this paper, Glaucon wants a definition of justice that will apply in all scenarios, one where justice is always better then injustice. Gauthier has attempted to provide this with the PMRC and Constrained Maximizers. If he grants that for the self-interested, it is rational to be a MSM over a CM; he appears to have given up his response to Glaucon and the Fool. His view no longer is applicable in all cases, especially with the ring of Gyges. In fact, even in far less severe of situations, it would be rational to follow self-interest and break your contract as suggested by the fool.

Though Gauthier can respond to the previously stated concerns, his responses seem to require a moderate alteration of his overall argument or to give up important components such as a counter to the fool. His view also seems to lack support from the experiences that we all have every day. This in itself does not discredit his view as we might simply be irrational in our actions. I do, however, have a view that I think can address some of the major concerns with

Gauthier's view while also lining up better with the experiences we all have on a day to day basis. I will now explore that view.

Cooperative Capital Contractarianism

In order to clearly explain my theory I should briefly set forward a few scenarios.

1. George and Selma are farmers that have agreed to help each other harvest their crops. George helps Selma first, and even though she has already been helped by George, Selma also helps George later.
2. A new coworker asks to borrow $10 for lunch. You have only known this person (let's call him Dom) for two days. You loan him the money with the agreement he will repay you tomorrow.
3. Matt, was recently asked to loan his sister $4000. In their numerous smaller agreements his sister has always followed through on her end.
    a. At the recommendation of Matt's other sister he declines the loan request.
    b. At the recommendation of Matt's other sister he grants the loan request.
4. Two versions of the prisoner's dilemma.
    a. A traditional prisoner's dilemma with a husband and wife criminal duo. In this version, neither partner confesses to the police.
    b. The special prisoner's dilemma where there are no future/external consequences. This time both spouses confess.

The question at hand is: are any or all of these choices rational and why or why not? I contend and aim to prove that all of these decisions are rational. So far in this paper we have seen two different views that aim to provide a way for rational egoists to obtain the very real benefits of cooperation. Hobbes required an external force, and Gauthier required a development of a disposition for cooperation and the ability to recognize a similar disposition in others. What I propose instead is the idea of cooperative capital. Cooperative capital is essentially the trust that is developed through repeated cooperative interactions with others. It is best explained through an example. Let's look at example 1 from above. In our original farmer example we have both George and Selma. These two have a long standing relationship as neighbors and fellow

farmers.[60] They started building their cooperative capital years ago by loaning small tools and pet sitting while the other was on vacation. The current situation they are, as we previously understood it, looks like:

S5: Crop Harvest	P2: Selma		
P1: George		A1: Help Harvest	A2: Don't Help
	A1: Help Harvest	(3,3)	(1,4)
	A2: Don't Help	(4,1)	(2,2)

George has already helped Selma, so it would seem that it is in Selma's interest not to help George. The reality is that the scenario actually looks like this:

S5: Crop Harvest	P2: Selma		
P1: George		A1: Help Harvest	A2: Don't Help
	A1: Help Harvest	(3+gain CC, 3+gain CC)	(1+gain CC, 4+lose CC)
	A2: Don't Help	(4+lose CC, 1+gain CC)	(2+lose CC, 2+lose CC)

Though Selma can choose to not help George and benefit this once, she loses the ability to make future deals with George. If, however, she spends the effort to help him thereby reducing some of her potential utility, she gains an increase in cooperative capital from George as his trust in her grows. This is far more valuable in the long run. Let me show how that is the case. Say this scenario repeats itself six years in a row. The first two years Selma plays along, but the third year she is tired and decides to break her promise:

[60] There are many parts of a relationship that could bring value, even to a rational egoist. For this theory we are only concerned with value associated with ability to make contracts and trust that the other will honor their end of the contract.

George's Action	Selma's Action	George's Payout	Selma's Payout
Help Harvest	Help Harvest	3	3
Help Harvest	Help Harvest	3	3
Help Harvest	Don't Help	1	4
Don't Help	Don't Help	2	2
Don't Help	Don't Help	2	2
Don't Help	Don't Help	2	2
Totals		13	16

What if Selma had just kept helping all of those years? The result would be:

George's Action	Selma's Action	George's Payout	Selma's Payout
Help Harvest	Help Harvest	3	3
Help Harvest	Help Harvest	3	3
Help Harvest	Help Harvest	3	3
Help Harvest	Help Harvest	3	3
Help Harvest	Help Harvest	3	3
Help Harvest	Help Harvest	3	3
Totals		18	18

It appears that it is in Selma's interest to maintain her ability to make contracts with George. There are two potential concerns here. First, couldn't Selma just refuse to help on the sixth and final year? I simply choose six years at random but in reality they would not know when the last year would be. Second, even if once they have made numerous agreements, they can trust the other based on the other's desire not to lose their cooperative capital, how do they ever make their first contract to start building CC? In order to address this second concern, let's take a look at example 2. A new coworker, Dom, asks to borrow $10 for lunch. You have only known him for a couple days, but he promises to pay you back $15 tomorrow. Is it rational to loan him the money? Obviously it does depend on your current financial situation, but let's say that $10, while not nothing, is not of major consequence. In this case I argue that it is rational to loan him the money simply for the potential of developing cooperative capital with him. If he

breaks his promise, it's not a big deal. If he follows through, you have learned you should be able to trust him, at least for minor agreements. One time is not enough to completely trust someone but that is not what we are aiming for. Let's say next time you are asked to borrow $20. Again you agree because of the trust or CC you developed last time. From $20, it goes to $25, then to $40, and then to $60. Each time you are repaid on time and with interest. Now, he is asking for $100. Back when you first knew him, you never would have loaned him $100, but you do now and loan him the money rationally. This is because you have something to gain and now Dom has something to lose. You want to get the interest from the arrangement and know that Dom benefits from his ability to get these loans from you. That ability to be trusted is granted by the cooperative capital you and Dom share. We ended up rationally loaning $100 by starting from an initial desire to simply develop CC with Dom in order to potentially gain the benefits from cooperation later on. You increased your CC with Dom little by little, each time trusting that he valued his ability to make those agreements with you, based on his CC with you, more than the desire to walk away with your $100.

Now with both the cases of Selma and Dom, we saw the impact that keeping or breaking a contract can have on your CC with the person you made the contract with. This in turn alters your ability to make additional or increasingly beneficial contracts with them in the future. There is, however, another way that your choice of whether to honor a contract or not can impact you greatly. Let's take a look at example 3. Matt is fairly well off and has been asked for a loan of $4000 by his struggling sister, Sara. Matt has made her a handful of smaller loans in the past and has always received his money back. As such, Matt and his sister have developed quite a bit of CC together, and he is strongly considering loaning her the money. Now there are two ways this could go. First, Matt receives a call from his other sister, Tanya. She is concerned after hearing

that Matt has been asked for a loan. She informs him that their sister owes her $6000 and hasn't been making any payments like agreed. Over the years, Matt has had many cooperative interactions with Tanya and has a large amount of CC with her as well. As such, her concern and recommendation against loaning their sister the money in effect reduces the CC that Matt had with his other sister. Sara's reputation was damage with Matt because of her broken agreements with Tanya. This has multiple interesting implications. First, maintaining your CC with another is crucial because a loss of CC with one person could mean a loss of CC with anyone that person knows. On the other side of things, maintaining a good level of CC with lots of people protects you from potential risks like the $4000 loan. The other way this scenario could go is the exact opposite. Instead of Tanya suggesting against loaning to Sara, she could speak highly of her and convince Matt to trust Sara. Sara's good CC with Tanya is helping her get a loan from Matt when she might not have otherwise.

The final example, number 4, is the classic game theory scenario of the prisoner's dilemma. This time I will use a husband and wife crime duo. The two have years of cooperation and built up CC with each other. They have be caught in the past for lesser crimes and been brought in for questioning but never at the same time before. They have always stayed silent, but is it rational to do so now? It would look like this:

S3: Prisoner's Dilemma	P2: Betty		
P1: Todd		A1: Keep Quiet	A2: Confess
	A1: Keep Quiet	(-4,-4)	(-20,0+Loss of 20 Years of CC)
	A2: Confess	(0+Loss of 20 Years of CC,-20)	(-16+Loss of 20 Years of CC,-16+Loss of 20 Years of CC)

Because these two have been cooperating for so many years, they have a lot to lose. They can enter into almost any agreement and not have to worry about the other defecting because the large amount of CC at risk and the potential for extremely beneficial future contracts far outweighs a couple years in jail. In this way, the cooperative capital operates very similarly to Hobbes' commonwealth, except it is each person's own recognition of the value of their shared CC that is making it rationally to keep their agreement.

But what about the special prisoner's dilemma where there are no future or external consequences. The reality is that in this scenario, loss of CC cannot be a potential risk as a part of the outcomes because that would be an external consequence. It might seem confusing with the given example, but Todd and Betty's reputations cannot be hurt by any action in the special prisoner's dilemma, and as such the scenario defaults to its original form with only years in jail at stake:

S3: Prisoner's Dilemma	P2: Betty		
P1: Todd		A1: Keep Quiet	A2: Confess
	A1: Keep Quiet	(-4,-4)	(-20,0)
	A2: Confess	(0,-20)	(-16,-16)

This makes it only rational both Todd and Betty to confess. It does not matter how many years they have had together, how many past cooperative ventures they have had, when there are no future consequences, no amount of CC can provide aid to cooperation.

How does Cooperative Capital Contractarianism (CCC) deal with the major concerns plaguing Gauthier's view? The first concern I wish to address is that of the inequality glutton. In the current system, people choose who they wish to cooperate with. An inequality glutton is

going to want more than an equal share of the cooperative surplus, and as such will be limited on who he is able to make contracts with. Not only will any interaction with others not end well, but that information will spread, and the glutton will struggle to find anyone to cooperate with.

The second concern for Gauthier was people's ability to detect what other's dispositions are. This is not a concern for those that use cooperative capital. There is no need to determine what kind of person someone is because we don't start off offering to loan someone $4000 without knowing them. Sure the Nigerian prince that emailed me might be really nice and actually want to help me out, but if it seems too good to be true, it probably is. With relational contractarianism we begin small, risking as little as possible while still enough to start to build CC. It takes a lot of past cooperation building CC before people are able to rationally agree to risky cooperative actions with greater payoffs. These last two concerns also show that relational contractarianism better reflects the reality of the world we live in. Unlike Gauthier, who might have to defend rationally loaning $4000 if you think they are a constrained maximizer, my view starts small and builds to larger cooperative actions over time.

The final concern is with providing a response to the Fool and Glaucon. Here it seems I am stopped in my tracks. Glaucon would not be satisfied with CCC. This is because it denies that it is always best, to honor your contracts but I contend Gauthier's view is similarly committed. There are cases, like the special version of the prisoner's dilemma, where one would be rational driven to break their contract. For Gauthier, these cases make it rational to be a MSM over a CM, thereby forcing him to give up the ability to justify cooperation in the special prisoner's dilemma.

I accept the inability to give Glaucon a satisfactory response as a consequence of embracing CCC. What I will say, however, is that these scenarios are few and far between. I

would even go so far as to say that most people will never experience a scenario where there are no future/external consequences in their entire lives. What is more likely, but still fairly rare is a situation where enough is on the line to be worth risking all of the CC you have with everyone in your life. Perhaps you are offered a private island with billions of dollars to spend on whatever you want. It might then be worth the risk of losing the ability to have cooperative interactions with everyone you know. Hobbes calls the man a fool who says it is in one's interest to sometimes break their contracts. I contend that it is only foolish to not act rationally when doing so demands that, in cases with no external consequences, we break our contracts.

Bibliography

Dion, Douglas. Intro to Political Analysis, The University of Iowa, Iowa City, Iowa, Fall 2016.

Fumerton, Richard A. *Reason and Morality: A Defense of the Egocentric Perspective*. Ithaca: Cornell Univ. Press, 1990. 18-60.

Gauthier, David. *Morals by Agreement*. Oxford: Clarendon Press, 2006.

Gauthier, David. "Why Contractarianism." *Contractarianism and Rational Choice.* Peter Vallentyne, ed. Cambridge: Cambridge University Press, 1991. 15-30.

Harman, Gilbert. "Convention." *The Nature of Morality.* New York: Oxford University Press, 1977. 103-14.

Harman, Gilbert. "Moral Relativism Defended." *The Philosophical Review* 84, no. 1 (1975): 3-22.

Hobbes, Thomas. *Leviathan*. A.P. Martinich, ed. Oxford: Clarendon Press, 2012. 93-119.

Kavka, Gregory S. "Reviewed Work(s): Morals by Agreement by David Gauthier." *Mind*, New Series, 96, no. 381 (1987): 117-21.

Kavka, Gregory S. "Right Reason and Natural Law in Hobbes's Ethics." *The Monist* 66, no. 1 (1983): 120-33.

Kraus, Jody S., and Jules L. Coleman. "Morality and the Theory of Rational Choice." *Ethics* 97, no. 4 (1987): 715-49.

Moore, Margaret. "On Reasonableness." *Journal of Applied Philosophy* 13, no. 2 (1996): 167-78.

Plato. *Republic*. Translated by G.M.A. Grube. Indianapolis/Cambridge: Hackett Publishing Co., 1992.

Pollock, John L. "How Do You Maximize Expectation Value?" *Noûs* 17, no. 3 (1983): 409-21.

Rawls, John. "Two Concepts of Rules." *The Philosophical Review* 64, no. 1 (1955): 3-32.

Scanlon, T.M. “Contractualism and Utilitarianism.” *Utilitarianism and Beyond.* Amartya Sen and Bernard Williams, eds. Cambridge: Cambridge University Press, 1982. 103-28.

Timmons, Mark. “The Limits of Moral Constructivism.” *Ratio* 16, no. 4 (2003): 391–423.