

Impartial Contractarianism: The Unwritten Contract and the Impartiality Problem

Wijsgerig Leeronderzoek
Naam: Frank van Hal
Student Nummer: 3674363
Begeleider: dr. Katrien Schaubroeck
Faculteit: Geesteswetenschappen
Departement: Wijsbegeerte
Datum van Inleveren: 26-6-2013

Impartial Contractarianism: The Unwritten Contract and the Impartiality Problem

Within philosophical ethics contractarianism, the idea that morality is based on a social contract, is an old tradition, dating back to antiquity. In Plato's *Republic*, Glaucon suggests that morality is based on an agreement between people not to inflict injustice on each other.¹ However the most famous account of is probably Thomas Hobbes' *Leviathan*, where Hobbes proposes that in order to escape a state of nature, where each person is constantly under threat of assault, deception or death by others, people form a social contract with each other. As part of this contract they instate an absolute monarch, or sovereign as Hobbes calls this figure, to enforce the contract through the use of fear and force against those who would undermine it. Despite its age contractarianism is still adhered to today. In contemporary philosophy influential contractarian theories have been provided by philosophers like John Rawls, David Scanlon, David Gauthier, and Nicolas Southwood.² Modern day contractarianism has a variety of branches inside it, the two most important streams of thought being Hobbesian contractarianism, which tends to argue that the social contract serves to implement morality as a way of constraining people from causing each other harm, and Kantian contractarianism in which the social contract is a hypothetical one based on a Kantian conception of people as rational agent and members of a kingdom of ends.

Despite its venerable history and contemporary adaptations, contractarianism is not free of criticism both from without and within. In this paper I will address one criticism leveled from within contractarian thought against Hobbesian contractarianism: the Impartiality Problem.³ This problem states that by trying to base itself on an instrumental conception of rationality, with a focus on individual gain, Hobbesian contractarian theories are incapable of providing an adequate account of morality's demand for the moral agent to be impartial as to who deserves to be treated moral. My ultimate goal in this paper is to propose a Hobbesian or at the very least Hobbesianesque contractarian theory that I believe is capable of dealing with the impartiality problem in an adequate way. I will do this by showing how individuals that can only be included into morality through impartiality can be included, not as fellow individuals but as fellow members of a moral society.

In order to reach my argument I will explain the issues of impartiality by first describing one version of Hobbesian contractarianism, as provided by Gauthier, which is probably the most influential contemporary account of Hobbesian contractarianism. Following this I will describe the problem of impartiality, including what it entails and why it is important, through a description of Southwood's contractarian thesis, which levels the impartiality objection against Gauthier's theory. After describing Southwood's objections to Hobbesian contractarianism, I will provide some arguments against both Southwood and Gauthier, as well as some remarks on their respective theories. Finally I will propose my alternative interpretation of a Hobbesian social contract theory and explain why taking persons as a fellow member of society rather than mere individuals can justify moral impartiality in Hobbesian contractarianism. My in this venture will be to at least remove one obstacle for the plausibility of Hobbesian social contract theories in general, and my own in particular.

1 Plato, *Republic*, 358b-359b. I owe this point to David Gauthier, see David Gauthier, *Morals By Agreement*, (Oxford: Clarendon Press, 1986), p10.

2 Although Rawls' contractarian theory is politically philosophical rather than ethical in nature, the influential status it has as a contractarian theory makes it worth mentioning.

3 The impartiality problem is also known as the impartiality objection, I will use the term "impartiality problem" in this essay however, treating the terms as interchangeable.

1. Gauthier and Modern Hobbesian Contractarianism.

Early on in *Morals by Agreement* Gauthier makes the claim that morality can only be based on self-interest and can be valid only insofar as it serves self-interest.⁴ This claim coincides with this aim to make a moral theory based on his conception of rationality, even if this moral theory deviates from every day morality.⁵ Gauthier's conception of rationality is one based on economic models of minimax: It is rational for an individual to try and get a maximal gratification of preferences for a minimal cost.⁶ A result of this is that impartiality, the notion that morality is impartial to non-moral aspects such as differences in wealth or power,⁷ is not particularly necessary according to Gauthier, at least not at a fundamental level. Any obligations are derived from the benefits of mutual agreement to cooperation after all, which is only possible if the persons I make the agreement with are capable of being part of a cooperation that benefits me.⁸ Gauthier's rejection of impartiality as important for a moral description seems to go against the intuition that, when properly understood, morality is impartial to such non-moral factors. However Gauthier seeks to provide an account of which moral constraints we could rationally accept, rather than trying to make an account of morality that is compatible with every day moral intuition.⁹

1.1: Gauthier and Rationality

To leave the description of rationality at such a preliminary stage would be unfair towards Gauthier, who spends the majority of the book refining what it means to be rational through maximization and investigating the consequences this has for morality. If we are to have an accurate view as to what extent Gauthier is capable of accurately describing the impartial aspects of morality, we should have a better look at how exactly Gauthier envisages rationality to work.

Gauthier bases his theory on actual beliefs rather than hypothetical notions of informed preference. He gives the example of Gertrude from Shakespeare's *Hamlet*, who drank a cup of poison, expecting it to be wine. This, Gauthier states, was not an irrational thing of her to do, since she did not know that she was drinking a cup of poison, instead believing she could sate her desire for wine by drinking the contents of the cup.¹⁰ Indeed, Gauthier states that to suppose that people might have preferences beyond their beliefs is implausible. This is because our preferences are based on our beliefs.¹¹ The Doctor's choice to try apples after regenerating for the 11th time was based on the belief that apples tasted good.¹² This was true for his 10th incarnation, but the 11th Doctor found that this belief did not hold for his new shape. He had no way of knowing that he would consider apples to be rubbish, until he tasted one. It is however necessary for preferences to hold up against both reflection and experience.¹³ If the Doctor had been able to reflect that he would no longer like apples, it would have been irrational for him to have tried an apple. It would have similarly been irrational for the Doctor to have a preference for eating an apple if he already knew that he, at least in this regeneration, hated them, because he had tasted them before.

4 Gauthier, *Morals by Agreement*, pp5-6.

5 Gauthier, *Morals by Agreement*, p6.

6 Gauthier, *Morals by Agreement*, p7.

7 I will discuss impartiality a bit further when discussing Southwood's Impartiality Objection later on in this essay.

8 Gauthier, *Morals by Agreement*, p17.

9 Gauthier, *Morals by Agreement*, p6.

10 Gauthier, *Morals by Agreement*, p29.

11 Gauthier, *Morals by Agreement*, pp29-30.

12 "The Eleventh Hour", *Doctor Who*, Youtube, BBC, 2010, <URL=<http://www.youtube.com/watch?v=IQuCP0GJHf4>> (viewed April 23, 2013)

13 Gauthier, *Morals by Agreement*, p33.

A reason given by Gauthier for choosing preferences over interests is because preferences and interests do not always coincide. To express an interest, according to Gauthier, is nothing more than expressing a preference for a certain state of affairs. If this is not admitted, Gauthier states, there would be a “divergence between expressed and revealed preferences.”¹⁴ A further argument Gauthier provides in support of his claim that we should judge whether or not something is rational is based on preferences, rather than interests is the idea that if interests are indeed based on subjective preferences, they themselves cannot be subject independent and objective.¹⁵

Similarly Gauthier rejects happiness or satisfaction as a criterion for rationality. His first issue with a concept like happiness as the base for rationality is that it seems next to impossible to define happiness in any measurable way.¹⁶ But even if this were possible, Gauthier foresees two problems that follow from this: Either satisfaction of preferences and enjoyment do not always coincide, in which case there might be situations where enjoyment does not hold up as the preferred thing after being reflected upon. Or preferences are solely concerned with enjoyment so the amount of enjoyment I expect to receive from something correlates with how much I would prefer it, regardless of other factors.¹⁷

Another, final form of rationality Gauthier rejects is prudential rationality. He does this on the grounds that it, even after reflecting on a preference not to do the prudential thing right now, it does not mean we will not do it later. Based on his concept of rationality as being preference based he thus rejects prudence as being a relevant criterion for judging someone irrational.¹⁸ This however does not mean that a prudent person is necessarily irrational, rather Gauthier insists that a preference that holds up against reflection while simultaneously not being prudential is still rational.¹⁹ In addition Gauthier attacks the notion of the seeing the future as rationally neutral, which is important for both interest-based, and prudence-based notions of rationality. Gauthier's reasoning behind rejecting future neutrality is based on idea that a person only identifies with her future self insofar her current preferences allow it, regardless of the contents of these preferences.²⁰

1.2 Gauthier's Morality from Market Failure

Gauthier sees our ability to maximize our preferences as a market situation, which in an idealized and undistorted situation is a perfectly competitive market. Gauthier sees such a perfectly competitive market as creating a natural harmony between the interests of individuals, thus avoiding constraints as is the case in a prisoner's dilemma. Although Gauthier considers the perfectly competitive market to be preferable, this does not mean he claims that it is always possible.²¹ Indeed morality is a necessity as a result of the market failure that would result from a situation where the actual market of preferences displays imperfections.²² The perfectly competitive market would, according to Gauthier, be free of force and fraud and the actions of others would, for the purpose of calculating our own choices be fixed.²³ The ultimate goal of the perfectly competitive market is, according to Gauthier, to reach an optimal equilibrium in which no one can become better off

14 Ibid.

15 Gauthier, *Morals by Agreement*, p34.

16 Gauthier, *Morals by Agreement*, p35.

17 Gauthier, *Morals by Agreement*, pp35-36.

18 Gauthier, *Morals by Agreement*, p37.

19 Gauthier, *Morals by Agreement*, p37.

20 Gauthier, *Morals by Agreement*, p38.

21 Gauthier, *Morals by Agreement*, p84.

22 Ibid.

23 Gauthier, *Morals by Agreement*, p85.

without making somebody else worse off.²⁴ Gauthier rejects a utilitarian solution here however, deeming it irrational for people to give up their individual goods in favor of the maximization of goods overall. Merely the fact that nobody in an optimal equilibrium can be made worse off without making somebody else better off, once the optimal equilibrium has been reached, does not mean, nor should it in Gauthier's eyes be made to mean, that the maximum amount of possible goods has been achieved.²⁵ This is mostly because utilitarianism is concerned with maximizing the utility for a group as a whole, regardless of utility of individuals, whereas Gauthier's market focuses on maximizing utility for individuals regardless of the utility for the whole.

Despite Gauthier's insistence that to be rational means that we try to maximize the satisfaction of our preferences, and the way he incorporates economic theory as a means of describing the rational cause of action, this does not mean that we automatically strive for what seems the best possible result in a straight forward way, without regard for the actions of others. Gauthier spends an entire chapter of his book demonstrating that sometimes compromises have to be made, to facilitate an equilibrium that leads to better odds of getting a preferred result in the form of an equilibrium.²⁶ He is also keenly aware that the ideal of a perfectly competitive market is, at least in many cases, unachievable in reality, which results in the need for morality.²⁷ In order to maximize our preferences it can therefore be necessary to constrain ourselves in order to guarantee better returns. Considering this statement it would, according to Gauthier, be irrational to agree to a moral code in which we need to hand in part of the goods we bring to the bargaining table.²⁸ Furthermore Gauthier states that those who do not contribute have no say in the bargaining about the course of action to be taken.²⁹ But even when an agreement to be moral is made in order to guarantee that the market is closer to competitive so people do not harm each others interests as readily, there still remains a question on how it is rational to adhere these moral rules if they get in the way of maximization.³⁰ In order to explain this Gauthier distinguishes between constrained maximization and straightforward maximization. Whereas straightforward maximization means acting in a way that would maximize the satisfaction of your preferences as they are right now, constrained maximization means acting in a way so the moral system, which will normally allow you to maximize the satisfaction of your preferences, is maintained.³¹ There are, according to Gauthier, two main reasons for constraining maximization instead of going for straightforward maximization: receiving more benefit from a cooperative strategy than it would be to pursue an individual strategy, or being unable to pursue an individual strategy at the cost of a cooperative strategy because it is impossible to do so undetected.³² To be moral, rather than just acting in accordance with strategy however one must, according to Gauthier, take and have a “genuine interest” in constraining one's maximization.³³

1.3 Gauthier and Impartiality

I have discussed how Gauthier defines rationality which is important for Southwood's critique on the ability of Hobbesian contractarianism to deal with impartiality. Furthermore I have mentioned that Gauthier does not really seem too bothered with impartiality to begin with. However the final chapter of *Morals by Agreement* provides some additional information that is important for

24 Gauthier, *Morals by Agreement*, p89.

25 Gauthier, *Morals by Agreement*, p104.

26 Gauthier, *Morals by Agreement*, pp60-82.

27 Gauthier, *Morals by Agreement*, p84.

28 Gauthier, *Morals by Agreement*, p134.

29 Ibid.

30 Gauthier, *Morals by Agreement*, pp157-158.

31 Gauthier, *Morals by Agreement*, p172.

32 Ibid.

33 Gauthier, *Morals by Agreement*, p189.

understanding Gauthier's approach to impartiality more completely. In this final chapter Gauthier starts by arguing that humans derive purpose from scarcity.³⁴ As a result scarcity is a requirement for human society, which would be “chimerical” without it.³⁵ One way in which cooperation and scarcity coincide is through specialization; humans, being unable to be equally good at everything, achieve greater results by specializing in certain skills and harmonizing these specializations.³⁶ Indeed Gauthier denies that cooperation decreases the value of the fruits of the labor. In fact the entirety of public life, and everything we value in it, requires cooperation according to Gauthier.³⁷ Being moral for the sake of cooperation does not automatically make us care for each other as fellow persons, so Gauthier states, but it does create a ground for cooperating with other persons as other persons, instead of merely cooperating with them because they happen to be part of a cooperation that benefits us.³⁸ However this gives rise to an opportunity for impartiality, since a person who adopts a moral society will adopt the moral rules towards any member of said society, and not merely towards particular individuals in said society.³⁹ Gauthier states that a just moral society should be such that each member can freely act in accordance with their preferences, as long as these do not conflict with its moral laws.⁴⁰ Despite all this, however, Gauthier maintains that the individual takes precedence over society, because rational reasons are primarily based on the preferences of the individual.⁴¹ Gauthier does however state that a completely neutral basis for society is impossible due to the effects of socialization on our abilities, aspects of which can be either hard-wired, i.e. a part of human nature, or soft-wired, i.e. the effect of socialization.⁴² As far as abilities are soft wired Gauthier states that a just society must accommodate people in two ways: it must strive for equality, but not to an impossible extent.⁴³

2. Southwood, The Impartiality Problem and Other Objections to Hobbesian Contractarianism.

It is easy to see why Gauthier has troubles providing impartiality in his theory, but why is that even important? In order to investigate this we will turn to Nicholas Southwood, who raises the impartiality objection against Hobbesian contractarianism in the book *Contractualism and the Foundations of Morality*. Southwood defines Impartiality as all sane humans having moral standing, regardless of wealth, dominance, strength or needs.⁴⁴ This involves universal patient rights, such as duties, obligations and restrictions towards each individual affected morality, ensuring that anyone affected by morality is free from wrongdoing by others. However Southwood describes impartiality as going further, using it as a reason for others to have a right to aid.⁴⁵

Impartiality is important because it is, according to Southwood, important for the extensional character of morality.⁴⁶ A proper explanation of the extensional character of morality, which we will discuss below, is considered by Southwood to be an important aspect of any moral theory. Now

34 Gauthier, *Morals by Agreement*, p333.

35 Ibid.

36 Gauthier, *Morals by Agreement*, p336.

37 Gauthier, *Morals by Agreement*, p337.

38 Gauthier, *Morals by Agreement*, p338.

39 Gauthier, *Morals by Agreement*, p339-340.

40 Gauthier, *Morals by Agreement*, p340-341.

41 Gauthier, *Morals by Agreement*, p349.

42 Gauthier, *Morals by Agreement*, p350.

43 Ibid.

44 Nicholas Southwood, *Contractualism and the Foundations of Morality*, (Oxford: Oxford University Press, 2010), p19.

45 Ibid.

46 Ibid.

that we have seen that impartiality is, according to Southwood, an important part of morality's extensional character, I will discuss why extensionality is important next.

2.1 Southwood: Morality's Intensional and Extensional Character.

Southwood starts by stating that morality has an extensional aspect and an intensional aspect. Morality's extensional character consists of substantive judgements about what is morally permissible, forbidden or allowed.⁴⁷ For example, if two people were to agree that lying is morally forbidden under all circumstances, they agree on morality's extensional character when it comes to the act of lying. Southwood considers Impartiality and Partiality essential aspects of morality's extensional character, and as a result any theory incapable of accounting for morality's impartiality and partiality fails to accurately describe morality.⁴⁸

Morality's intensional character consists of those judgements about moral predicate with which an individual ought to concur if they understand what it means to be moral.⁴⁹ So if you claim, as Kant does, that moral requirements are categorical in nature, you claim an individual can only be said to properly understand moral requirements if they consider non-categorical moral requirements impossible. In other words the extensional character of morality decides what acts are morally permissible, forbidden or obligatory; the intensional character describes what makes an act so.

In order to give an accurate account of morality, an ethical theory must be capable of accounting for morality's intensional character.⁵⁰ Southwood considers two platitudes essential for a proper understanding of the intensional character of morality: The first platitude concerns morality's normativity; morality must be minimally normative, providing any moral agent that has a moral duty to do something with a reason to follow this moral reason.⁵¹ However this minimal normativity is not enough according to Southwood, who adds the assumption that morality must be categorical and binding in its normativity.⁵² This means that it is to be independent of the agent's preferences or interests and inescapable by departing from society.

The second platitude Southwood assumes to be necessary for morality to be accurately described by a moral theory concerns Morality's objectivity. In order to be sufficiently objective Southwood believes morality needs to be non-subjective and universal.⁵³ The former means that a moral claim concerns something that is actually the the case, regardless of the agent's personal tastes. The latter means that moral facts are the case regardless of the moral system they occur in.⁵⁴

2.2 Southwood's criticism of Hobbesian and Kantian contractarianism.

Southwood turns to the Hobbesian contractarianism and the impartiality problem by posing that Hobbesian contractarianism, is incapable of sufficiently describing morality's extensional and intensional character. If Hobbesian contractarianism is to provide an adequate description of morality's extensional, Southwood argues, it would have to be capable of explaining why every

47Southwood, *Contractualism and the Foundations of Morality*, pp8-9.

48Southwood, *Contractualism and the Foundations of Morality*, pp17-19.

49Southwood, *Contractualism and the Foundations of Morality*, p8

50Southwood, *Contractualism and the Foundations of Morality*, p12.

51Southwood, *Contractualism and the Foundations of Morality*, p14.

52Southwood, *Contractualism and the Foundations of Morality*, p15.

53Southwood, *Contractualism and the Foundations of Morality*, p16.

54Southwood, *Contractualism and the Foundations of Morality*, p16-17.

person has the same rights, regardless of differences in for example wealth or power.⁵⁵ Southwood continues by stating Hobbesian contractarianism is incapable of this on the grounds of its instrumental origins. What principles are accepted in the moral code are, according to Southwood, dependent on the interests of the individuals that has to accept these principles. If they lack benefiting from a principle due to large differences with the people that benefit from it, a person would reject it as a ground to act morally, according to Southwood.⁵⁶

Even if a Lockean proviso⁵⁷ is assumed for creating the contract, Southwood is skeptical about the Hobbesian contract's ability to be impartial. Southwood's reason for doubting the proviso's effectiveness stems from two issues he sees with the sort of maximizing rationality he believes Gauthier to be using. For the proviso to work two conditions must be met. The first condition states that if it is rational for an individual to cooperate with another individual in absence of the provision, it would also be rational to do so with the provision in place. The second reason, which Southwood believes Gauthier's account of rationality to be incapable of, states that if it is rational for two individuals to cooperate under the proviso, and the proviso is in effect, they ought to cooperate.⁵⁸ This is not the only issue Southwood has with the Lockean proviso as a solution for the impartiality problem however. If the Lockean proviso as stated before is too strong for Hobbesian contractarianism, it can also be too weak for it, so Southwood states. The proviso is, according to Southwood incapable of compensating for radically unequal endowments. Nor does it prevent radically unequal endowments from playing a role if they are the result of one individual harming the other without the aim to benefit from it. The Lockean proviso is, according to Southwood, of such a nature that the strong do not have to take the weak into account.⁵⁹

A final reason Southwood provides for Hobbesian contractarianism being incapable of providing for impartiality is that impartiality cannot be based on altruism. This is because using altruism as a basis for impartiality does not give reasons to people who do not have altruistic preferences, unless they require cooperation from sufficiently powerful altruistic people.⁶⁰ Gauthier, Southwood observes, explicitly denies the possibility of this option, but even a Hobbesian that does allow for altruism faces a problem. This problem is, according to Southwood, that it would be too optimistic to assume altruism can provide general impartiality; it is often limited to a person's direct surroundings, and those exceptions, the truly altruistic, are too few to persuade the truly powerful to accept principles of impartiality.⁶¹ But, so Southwood argues, even if the global altruists were sufficiently powerful to allow impartiality be incorporated, it would only be incorporated as long as the global altruists are both powerful enough and altruistic enough, if either were to disappear, so would the rights of the powerless. The rights of the powerless would then be derivative and contingent, which would make contractarianism lose its appeal, which is, in Southwood's view, situated in the fact that those affected by morality get a say in its contents.⁶²

The impartiality problem, however, is not Southwood's only objection to Hobbesian contractarianism. I will briefly touch upon another objection, he raises, as part of the further

55Southwood, *Contractualism and the Foundations of Morality*, pp42-43.

56Southwood, *Contractualism and the Foundations of Morality*, p44.

57 The Lockean Proviso, as Southwood uses it here, bans any power or goods that have been derived through actions that were harmful to others from being considered part of a person's initial position upon entering into a social contract. For example if you and I were to enter into a contract, and we both had 2 apples, it would (provided we adhered a Lockean proviso) not be irrational for me to part with my apples if I had obtained them by stealing them from your stockpile.

58Southwood, *Contractualism and the Foundations of Morality*, pp46-47.

59Southwood, *Contractualism and the Foundations of Morality*, p47.

60Southwood, *Contractualism and the Foundations of Morality*, pp47-48.

61Southwood, *Contractualism and the Foundations of Morality*, p48.

62Southwood, *Contractualism and the Foundations of Morality*, p49.

discussion of Southwood's deliberative contractualism. This aforementioned second objection to Hobbesian contractarianism consists of the claim that it lacks the kind of normative force that Southwood assumes to be necessary to get the intensional character of morality right.⁶³ In particular Hobbesian contractarianism lacks both the authority and other-regardingness that Southwood considers essential for a moral theory. Since Hobbesian reasons are agent based, rather than agent directed, other agents are incidental to the actions of any individual. The rights of others are dependent on the necessity of the interest that an individual has in adhering to those rights, making these rights too weak to compel the agent into taking them into consideration if the agent has no interest in doing so.⁶⁴ Altruism is, just like before, incapable of serving as a reason to add agent-directedness, since this possibility is, according to Southwood, denied by Gauthier.⁶⁵ Southwood is skeptical of the power of constrained maximization, claiming that it would be irrational to adhere to when it is harmful.⁶⁶ But even if one were to adhere to constrained maximization for the sake of maintaining it, Southwood still foresees a problem for Hobbesian contractarianism. The problem is that a person who merely adheres to constrained maximization for the sake of maintaining the system, and prevent being ostracized from it, would fail to satisfy an important moral intuition. "do not mention me saving you, I just do not want to get ostracized." lacks the other-regardingness that people expect of morality.⁶⁷

Similarly Hobbesian contractarianism is, according to Southwood, incapable of explaining concepts like guilt or blame as justified when it is directed to third parties; i.e. towards those who harm others without it affecting us negatively. Guilt is essentially other-regarding as it stems from harming another person, without it harming ourselves.⁶⁸ Southwood believes, blame similarly stems from other-regardingness, because, according to him, the instrumental rationality used by Hobbesian contractarians is incapable of explaining why persons who are not directly affected by an action are justified or rational in blaming others for wrongful behavior.⁶⁹ There are two reasons why Southwood believes Hobbesian contractarianism to be incapable of accounting for blame. The first is that Hobbesian contractarianism would be incapable of explaining why wrongdoers deserve whatever punishment they deserve. The second is that instrumental reasons seem incapable of providing reasons for blame.⁷⁰

Having discussed Southwood's criticism of Hobbesianism, I would like to briefly touch upon his criticism on Kantian contractarianism, and his own theory of deliberative contractualism. Kantian contractarianist theories are, according to Southwood guilty of two fundamental flaws in the arguments that support them. The first flaw is that they are subject to an explanatory circularity, relying on some presuppositions derived from morality in order to explain morality.⁷¹ The second flaw consists of the claim that even if Kantian contractarianism were capable of avoiding circularity it would still lack "fundamentality", that is to say, it still presupposes certain normative claims without further supporting these on an explanatory foundation.⁷²

63Southwood, *Contractualism and the Foundations of Morality*, p34.

64Southwood, *Contractualism and the Foundations of Morality*, p36.

65Southwood, *Contractualism and the Foundations of Morality*, pp36-37.

66Southwood, *Contractualism and the Foundations of Morality*, p38.

67 Southwood, *Contractualism and the Foundations of Morality*, p38.

68 Southwood, *Contractualism and the Foundations of Morality*, pp39-40.

69Southwood, *Contractualism and the Foundations of Morality*, p40.

70Southwood, *Contractualism and the Foundations of Morality*, p41.

71Southwood, *Contractualism and the Foundations of Morality*, p61.

72Southwood, *Contractualism and the Foundations of Morality*, p70.

2.3 Deliberative Contractarianism as an alternative.

It is after pointing out these issues with Kantian contractarianism that Southwood begins suggesting his own alternative, which he calls Deliberative Contractualism. Deliberative contractualism states that morality is a set of rules to which we would agree and submit if we were completely reasonable.⁷³ The sort of rationality it requires is a limited form of rationality which Southwood calls deliberative rationality. Deliberative rationality means that an act is rational if it follows from a decision based on deliberation with others and requires that three norms are followed in order to let it exist. These norms require that the agent engages in deliberation, follows certain deliberative norms during the deliberation, and finally that the agent bases his decisions on the results of the deliberation.⁷⁴ The deliberative norms mentioned as the second norm for deliberative rationality, are instrumentally derived norms that are required for proper deliberation to proceed.⁷⁵ The rules that follow from the deliberative contract do not merely require compliance but also acceptance according to Southwood, since acceptance does not merely guarantee moral behavior in the agent but also the agent promoting moral behavior in others.⁷⁶ Southwood does not try to explain the intentional nature of morality though, instead seeking to provide a justification for moral claims in general.⁷⁷ Since morality is universal, Southwood argues it is impossible to escape deliberative citizenship, which is our role in the deliberative contract Southwood proposes. But merely the fact that morality is universal does not entail that it is constant, so Southwood writes, so it might be possible to escape deliberative citizenship, at least in the shape it takes in our present time, by traveling to another time.⁷⁸ Southwood also sees the deliberative contractualism as capable of accounting for blaming and punishment, an ability he does not see Hobbesian contractarianism possessing. Southwood states that by basing the contract on deliberative rationality there is a legitimate expectation among deliberative citizens which be be violated by wrongdoers. Through blaming and punishment, Southwood writes, the severity of a wrongdoing is recognized and addressed.⁷⁹

Although morality is universal according to Southwood, that does not mean that deliberative contractualism makes morality necessarily as it is, rather than contingent. In fact Southwood actually foresees trouble for deliberative contractarianism if there were norms that were necessarily impermissible or obligatory, no matter what.⁸⁰ Southwood denies that an act which is morally contingent cannot simultaneously be morally wrong right now, there is merely the possibility of a possible world in which it is not wrong. This claim, Southwood believes, is perfectly compatible with deliberation, where deliberation decides the moral rightness of a rule or act, but will make it extremely unlikely that people are wronged.⁸¹

Southwood also discusses how his theory handles partiality and impartiality. For partiality Southwood distinguishes two kinds. The first is de facto partiality, which exists because an action that is de facto morally permissibly partial is not forbidden by the contract.⁸² The second kind is de jure partiality, actions that are permitted because of the moral principles behind them. De jure partiality requires a negative right, going beyond merely allowing agents to act partially in

73Southwood, *Contractualism and the Foundations of Morality*, p88.

74 Southwood, *Contractualism and the Foundations of Morality*, p89.

75Southwood, *Contractualism and the Foundations of Morality*, pp90-91.

76Southwood, *Contractualism and the Foundations of Morality*, p104.

77Southwood, *Contractualism and the Foundations of Morality*, p141.

78Southwood, *Contractualism and the Foundations of Morality*, p145.

79Southwood, *Contractualism and the Foundations of Morality*, p148.

80Southwood, *Contractualism and the Foundations of Morality*, pp152-153.

81Southwood, *Contractualism and the Foundations of Morality*, p153.

82Southwood, *Contractualism and the Foundations of Morality*, p156.

circumstances where it is important to be able to do so, but also requiring others to refrain from acting partially in these cases. It is because these factors can be established deliberately that Southwood believes deliberative contractarianism to be capable of explaining partiality.⁸³ For impartiality Southwood claims that it requires patiential that are both universal and non-derivative. He argues that deliberative contractualism is capable of dealing with the issue because a completely amoral situation, comparable to Hobbes' state of nature, would be completely undesirable for anyone with the slightest foresight. Deliberative contractors, who are perfectly deliberately rational, have enough foresight to find the Hobbesian state of nature undesirable.⁸⁴ When it comes to impartiality towards atypical persons, those who have additional needs or are less capable of contributing to society, Southwood argues that since deliberative contractarianism involves all those that are affected by it, it is capable of accounting for rights to atypical persons. They are included as representees to the hypothetical contract, making there rights universal and non-derivative.⁸⁵

A few final points of note are that Southwood's deliberative contractualism is, according to him, not explanatorily epiphenomenal,⁸⁶ and that it avoids circularity, since the norms it requires for its foundation are non-moral, but rather instrumental⁸⁷ That he requires norms at all as in his explanation is not problematic, according to Southwood as he argues that consequentialism equally presupposes certain norms in order to function.⁸⁸

3. Criticism of Southwood and Gauthier

Now that both Gauthier's and Southwood's theories have been explained it is time to have a look at some weaknesses that each displays when it comes to explaining morality, and aspects of said morality such as impartiality. I will first discuss some issues with Gauthier before turning to discussing some issues with Southwood's theory.

3.1 Criticism on Gauthier, on preference based morality, and human interdependence.

Although Gauthier is the most widely known account of Hobbesian contractarian theory, his account of morality appears to have couple of points that seem rather problematic, especially concerning Gauthier's notions of rationality, and his focus on preference based morality. This section serves to show what these issues are, and how these issues affect what we can expect of a moral theory. Both rationality and whether or not morality is subjective will have profound effects on what we can expect of morality; such as whether impartiality exists, and whether it is necessary for a proper description of morality.

As we have seen before Gauthier argues that morality and rationality are based on preferences. During his discussion of the subject he uses the example of an addict to show that it would not be irrational for a person to act in their interest if they have no strong preference to do so.⁸⁹ However this statement implies that addiction is a rational choice rather than a compulsive disorder. Due to the fact that it is compulsive, rationality has nothing to do with what my preferences are while addicted. An addict can be either a willing on an unwilling addict, much as described by Harry Frankfurt, but either way the addict is ultimately subject to her addiction, which is characterized by

83Southwood, *Contractualism and the Foundations of Morality*, p157.

84Southwood, *Contractualism and the Foundations of Morality*, p161.

85Southwood, *Contractualism and the Foundations of Morality*, p169.

86Southwood, *Contractualism and the Foundations of Morality*, p180.

87Southwood, *Contractualism and the Foundations of Morality*, p181.

88Southwood, *Contractualism and the Foundations of Morality*, p183.

89 Gauthier, *Morals by Agreement*, p33.

a strong psychological and biological reaction to certain stimuli, rather than the result of reflection. Although Gauthier might argue that the willing addict would be rational, the simple fact remains that rationality has nothing to do with her course of action. Furthermore this can easily be extended to other forms of behavior based on preference. Even if I have a strong preference to play video games, it does not mean that it would be rational for me to do so if I have an interest in finishing my thesis. Although one could argue, as Gauthier probably would, that my preference to finish my thesis is stronger than playing video games, it does not follow that if I instead choose to play video games, that would be the rational thing to do. People simply are not always rational. Similarly the argument that Gauthier seems to give that interests are subjective because they are based on preferences is simply not true. That I have a preference might be subjective, but any interests that come from my preferences can be determined objectively through instrumental rationality. That a space agency wants to send a rocket into orbit might be contingent and subjective, but once this preference is established it entails several interests, including the development and construction of a rocket capable of reaching orbit, proper calculations to calculate the necessary maneuvers to reach orbit, and trained astronauts, capable of piloting the rocket. Similar to the interests of the space agency, what is right or wrong does not depend on a preference of a moral agent to treat a certain person in a certain way. Instead what is right or wrong is, as I will defend below, dependent on the demands of a moral code, independent of preferences about how we would seek to treat individuals subject to said code. The only point at which preferences come into play is when we decide to be moral.

A further issue with Gauthier's theory is the notion of the market as a morally free zone. The problem with this notion is the same one that plagues other libertarian market theories: in order for the market to function force and fraud must be eliminated from it, but this already requires morality. After all, how can it be wrong to use fraud or force if nothing is wrong? The argument that it restricts productivity and profit derived from people's talents is arbitrary, since how is your talent to harvest and sell resources any more important than my talent to deprive you of those resources through conning or raiding? In order to account for the wrongness of force or fraud in a market situation, a moral code has to be presupposed, for otherwise there would be no reason for the morally free market to be different from the Hobbesian state of nature.

The optimal equilibrium discussed before, in which nobody can be made better off without making anybody else worse off need not be good. If X has nearly all resources and Y and Z are nearly starving, without any hope of getting food from other sources than X, then it does not follow that it would be wrong for X to be made slightly worse off if it prevents Y and Z from being made worse off, i.e. dying from starvation. An optimal equilibrium as supposed by Gauthier does not provide an optimal or even a good distribution, or use, of goods.

Gauthier's claim that it is impossible to rationally part with goods that have been acquired in a pre-moral context when establishing the contract is also incorrect, provided that the goods are handed in for the sake of a cooperation that ultimately benefits the contributor. For example it would be rational for a person to hand in his weapons for the sake of peaceful cooperation, if this peaceful cooperation would be more beneficial for him, for example because it lets him work on his farm more effectively because he now has others protecting him. Nor would it be irrational for people to give up goods when a social contract is renegotiated, especially not if it leads to a better overall effect in the cooperation: the factory owner, having earned his riches as a result of the factory, might find it rational to give up some of his profit or even property in order to prevent his workers from creating unrest, or merely to improve the amount of goods sold by increasing the input in the economy, thus creating more profit. Nor would it be irrational to give up goods acquired through a process that created inequality, unforeseen in the original contract. Although it is not problematic

for Gauthier it is obvious that his claim that only those who contribute to the contract are allowed to participate in it is incomplete. After all, some, or, depending on the interpretation, all people, can only participate if certain conditions are met.

Indeed Gauthier's expectations of the importance of cooperation seem to underestimate how much we rely on it. Humans, being social animals, as even a cursory observation will tell us, seem to rely heavily on cooperation to get anything but the most basic survival done. Corporations tend to be more profitable than one man stores, a large workforce can build more impressive monuments than a lone builder, and a group of programmers can more easily create an impressive piece of software than a lone developer. Indeed it is safe to assume that in human endeavor as a whole, there are very few, if any, situations where a lone individual can get more done than a cooperating team. The only obvious exception might be some oppressive forms of government, but it can be wondered to what extent these are cooperative. It can easily be claimed that such a government is parasitic, rather than cooperative. Indeed wealth and power are more dependent on public life than Gauthier implies. In order to be rich, one requires a market that attributes value to something. Robinson Crusoe, who Gauthier uses as an example of a pre-moral individual, can consider himself rich when he has sufficient goods to easily survive, and perhaps find himself comfort, whereas in our own society a person who can have a pool inlaid with diamonds would be considered richer than a person who can only afford a pool inlaid with sapphire, let alone "the regular people" who often cannot even afford a pool. Similarly in the computer game Minecraft Iron would normally be considered more valuable than gold, simply for the fact that iron is more useful, as it is more plentiful and can be used for more things. Gold might however become more valuable on multiplayer server with an economy where it is used as currency.

3.2 Criticism of Southwood's Assumptions on Impartiality and Universality.

Southwood's reasoning is similarly not devoid of reasons to argue against it. The first issue being impartiality itself. The way that Southwood phrases impartiality and the universal potential rights it involves sounds rather Kantian in nature, meaning that for it to be accepted Kantian presuppositions need to be made first. Southwood explicitly admits that he relies on rather Kantian assumptions regarding the intensional and extensional character of morality.⁹⁰ This could be crippling for the notion of impartiality as being important for the extensional aspects of morality a whole, after all why would a moral theory have to accept a notion it has no way of proving? The primary redeeming factor would in such a case be that it seems important in moral intuitions. Such a response would merely shift the question around however, as many philosophers find intuition to be lacking when it comes to the explanatory and justificatory arguments for morality. Though intuition might not be suitable as basis for arguments, it is not without value. Indeed, moral intuitions can serve as a means of measuring the plausibility of a moral theory; a moral theory capable of doing justice to moral intuitions is, after all, more plausible than one that goes against them. Using moral intuition in such a way means that some moral intuitions might turn out to be irrational and plain wrong, but as a whole they ought not to be completely excluded as irrelevant factors for determining the plausibility of a moral theory. It seems then that Southwood's demand for universal impartiality might be too strong for what moral intuition can provide, but at the same time impartiality likely cannot be excluded as being fully irrelevant, at least not without good arguments for such a claim.

Another issue with Southwood's theory is the demand that morality is both objectively knowable and universal. Although I do not deny that morality is objective, the claim that morality is objective

⁹⁰ Examples include morality's universality, objectivity and other-regardingness. See Southwood, *Contractualism and the Foundations of Morality*, p15 & p17.

is not uncontroversial in ethical and meta-ethical debates. In order to demonstrate that it is indeed objective Southwood will have to provide strong arguments in favor of such a claim, without relying on presuppositions or intuitions.

Universality provides a bigger challenge for moral theories however, especially if they propose a universal content, as Southwood seeks to do. Different societies throughout time and space hold on to different moral codes. It would be rather arrogant to presume certain moral claims to be better than others without an objective basis for such claims, while any claim based on objective facts would only apply to those societies to which these facts apply. Southwood argues, as we have seen, that being affected grants people deliberative citizenship, allowing them to partake in the deliberation used to establish the rules of the deliberative contract. However this does not lead to rules that are necessarily universal, after all there are people that live far away, that are unaffected by moral rules that are in effect right here and now. Yet these people can affect each other, thus giving rise to their own morality right there that might not account for moral situations specific for myself, and with no need to do so because the moral rule in question is incapable of affecting me. This would severely blunt the ability of deliberative contractualism to create a universal morality, since I can be perfectly deliberatively rational without requiring a need to consult people that I cannot affect or be affected by. In order to provide a sound conception of Morality's intensional character Southwood would have to provide a strong arguments for his assumptions.

Southwood's claims against Hobbesian contractarianism, although they have some cogency, seem to be based on a needlessly implausible interpretation of instrumental rationality. As we've seen with the discussion of Gauthier given above, the interpretation of instrumental rationality as subjective is mistaking. It also seems that Southwood sees Hobbesian contractees as being needlessly selfish, as if the contract serves merely as a guideline while still maintaining a maximizing rationality for individual acts. As a result, the Impartiality problem as posed by Southwood relies on the question why individuals would accept certain principles towards other individuals, while this is already enclosed in accepting to partake into the social contract; each individual has the same duties towards each other contractee because of a contract that benefits them, not because following the duties benefits him all the time. The impartiality problem is therefore not a matter of why an individual would assume certain moral rules that are part of a contract, but rather why we ought to include such rules in the contract to begin with. From a practical point of view this would allow for an altruistic interpretation of human nature to satisfy as a justification for impartial contracts, but this answer would be philosophically unsatisfactory.

A further issue with Southwood is that he presupposes some rights as being necessary results from a moral theory. As mentioned above, whether or not I have a certain right under a Hobbesian contract is decided by the fact of other individuals having an interest in accepting the set of rules that includes those rights.⁹¹ This means that I cannot claim any rights of impartiality in a contract that simply does not account for such rights. The moral intuition, regarding innate rights, that Southwood relies on is just too weak for the argument he is trying to make. It also means that unless a society consists almost solely of the two extremes of the truly powerful and the truly powerless, the former will generally have reasons to submit to rules that require them to be impartial, since such rules would safeguard them against the somewhat powerless and somewhat powerful, who can affect the truly powerful. The result here is that the truly powerful would have to decide whether it is beneficial for them to be in society at all, rather than having to decide whether it is beneficial to cooperate with some individuals in it. If stated like this derivative rights need not be as flimsy as

⁹¹ This does not contradict the above paragraph in the slightest, as it still holds that once they do accept the set of rules that grants these rights they need to follow said set of rules and not just those rules that benefit them.

Southwood implies, as those that are subject to them do not merely derive their rights from the individuals that care about them, but also from the rights of every other individual in similar circumstances within the same society. The only issue that could remain in such a case would be a scenario in which the truly powerless are even incapable of affecting the somewhat powerless, once again making it an issue of why the society as a whole would make rules that were impartial towards the truly powerless.

Southwood's argument that Gauthier cannot account for the proviso relies on a notion that maximizing rationality and constrained maximization are incompatible. But this is a straw man argument that does not sufficiently take into account the nuances that Gauthier applies to the maximizing rationality. As I have shown in my discussion of Gauthier's arguments, Gauthier goes to great lengths to demonstrate that constrained maximization is in many cases more rational than straightforward maximization, countermanding Southwood's primary argument against the Lockean proviso as providing at least some impartiality to Gauthier's theory, even if this impartiality is not as strong as Southwood demands.

3.3 Criticism of Southwood's Approach to Who affects Whom.

Southwood's claim that the appeal of contractarianism comes from the idea that those affected by morality contribute to it appears to be inaccurate as well. After all animals and children have moral rights, but it can hardly be argued that they contribute to the moral code. If Southwood's claim that a contractarian theory relies on its ability for those affected by it to define it were true, animals and children would need to have an equal say in it as human adults. Impartiality seems to derive its importance exactly because not everybody affected by the contract forms it. In addition to this, most people are already subject to a moral code the moment they enter a society, regardless of their ability to influence it before entering the society; i.e. If I were to emigrate to another society, the moral code they have there has come to pass without my inference, even though it affects my rights while living there. Instead it seems that the main strength of contractarian theories comes from its ability to justify morality, and our reasons for abiding to it through the way it affects us. The claim that agent based rights remove justification once it is not in my interest to adhere these rights is a point that Southwood makes too quickly. The contract is what gives me a reason to accept a set of rules, and due to me accepting a system, that when adhered works in my favor by increasing cooperation, means that any duties and rights that flow from the system are valid as long as I accept the system as a whole. The mere fact that paying for an insurance is not in my favor is not a reason not to pay, as paying for my insurance means I am insured should something happen. As we have seen, Gauthier provides a similar reasoning for constrained maximization as being rational, further adding that people usually cannot just get away with acting in a way that is not cooperative. Indeed Southwood seems to suggest that instrumental rationality makes the contract an epiphenomenal ruse to hide a selfish act-utilitarian "moral" system.

As discussed above, Southwood sees an inability to explain concepts like guilt and blaming appropriately as a weakness of Hobbesian contractarianism. However this might just be a case of Southwood asking too much, or rather the wrong thing, of an account of morality. To demand that a moral theory has to explain moral feelings, is asking ethics to provide an account of moral psychology, instead of it providing an account of morality and its justification. Describing actual moral psychology is a task for psychology as a science, with the philosophical apparatus being ill equipped to discover the exact functioning of the human psyche. This does not mean that moral theories are incapable of making certain assumptions about the roles of certain psychological states on morality. It merely means that moral theories ought not be incompatible with the existence of moral feelings, even if such feelings prove irrational or epiphenomenal to morality. Being incapable

of explaining blaming and guilt as rational is not a fault for a moral theory like Hobbesian contractarianism, as it can easily see such emotions as being morally useful while simultaneously considering them explanatorily irrelevant. This means that a moral system might be able to justify such feelings, for example on the grounds that they help enforce the social contract, but that does not mean that this has to be a causal explanation of why we have such feelings to begin with, beyond assumptions about the moral psychology that a human being has. The issue of guilt is furthermore not problematic on the ground that it is other-regarding, as Southwood claims it to be. Gauthier does not exclude the option of moral internalization, and as a result would be quite capable of dealing with the issue. However as mentioned above guilt as moral emotions can be explained equally well through psychological means, with the only demand for a moral theory being that it can at least leave the option of moral feelings open, regardless of how much they contribute to the justification of the moral theory. When regarded as this, a lack of other-regardingness as justification in a moral theory needs not be crippling for a Hobbesian contractarian theory.

4. The Unwritten Contract: Hobbesian Contractarianism as Morally Accurate.

There are more objections to both Gauthier and Southwood, but these objections are best expressed in the context of an alternative Hobbesian contractarian theory. As a result I will now propose an alternative Hobbesian contractarian theory that I believe to be capable of dealing with the challenge of Impartiality problem, as well as some of the issues posed by both Gauthier and Southwood, by demonstrating that it can accurately, or at least plausibly, describe both the intensional and extensional characteristics we expect from a moral code. In order to achieve my goal of answering the impartiality objection and showing through a plausible Hobbesian social contract theory, I will use Southwood's method of discussing it, paying attention to morality's intensional and extensional aspects. Unlike Southwood and Gauthier however I will try to refrain from making substantive claims about morality, especially where it does not concern intuitions that are not broadly, if not universally, shared between different approaches to morality. Instead I will focus on trying to define a foundation of morality in general.

4.1 What it means to be moral.

In order to propose a plausible moral theory it is a good idea to look at what morality is and why it is important. In fact the theory I will propose seeks to do exactly that; to describe morality and its essential normative aspects, rather than trying to phrase a substantive moral framework. In other words I will try to extensively describe morality's intensional character and foundations in such a way that it allows for the extensional aspects, such as partiality and impartiality, to be filled in in a plausible way, while avoiding making too many claims about the substantive elements of said extensional character.

An important aspect of morality, regardless of its contents, is that a moral code entails certain rights, duties, and similar constraining factors in order to decide whether or not an act is permissible, in other words that it has an extensional character. But even if we determine that morality has an extensional character, we are left with the question what constraints are rational to assume. My claim is that a moral system, no matter what sacrifices it requires from those subject to it, is only rational if these sacrifices or constraints yield something else. As I have argued before, humans are social animals that require cooperation to do more than basic attempts at survival. But in order for cooperation to succeed certain rules are required, preventing harm to the cooperative venture, allowing each to achieve their goals within the cooperation. Therefore I claim that morality is the set of rules required to let a group function.

Now the word 'group' has a broad meaning which can describe anything from a pair of people to pretty much the entirety of individuals in existence. Although the claim that each group⁹² has its own moral code might seem odd at first, it is not contradictory to how morality works, and it can even explain moral partiality. Groups are not mutually exclusive, a person can, at the same time, have a group of friends, a family, a group of colleagues, they can be a member of a religious community, an inhabitant of a city, a citizen of a country, etc. These groups can demand different things of the individual and provide specific benefits in return. An agent might have obligations to her friends and family to help out in times of need, to her church not to blaspheme and to her nation to pay taxes and take up arms in its defense. A person's allegiance to different groups can even cause conflicts in such situations, requiring them to decide what group they value more in specific situations. An example of such a moral conflict might be the choice between being a member of an orthodox parish which considers homosexuality to be a sin, and a friend who happens to be a homosexual in which case the agent would have to decide between whether they value their friendship more than they value adhering that particular parish' beliefs. It is only when a specific group demands of the agent that they owe allegiance only to the group and no other groups that things can become problematic, and in such a case we can speak of a cult. The stereotypical religious cult is an example of this, forcing those who join it to forsake their allegiance to any government, friends and even family that do not share the cult's beliefs.

For a group to function two things are necessary: A) The group must be maintained in such a way that it can continue to function. And B) The members that make up the group must not be harmed in their functioning in the group. This leads, in combination with the above paragraph, to the consequentialist conclusion that what is morally permissible for a group is that which is not harmful to the functioning of the group or its members.⁹³ However, although this consequentialist theory gives us an objective ground from which to judge what is morally right or wrong for a specific group, such judgments require an exact knowledge of what actions will have what effects. Here we run into an issue for human beings, since, although they are often capable of making some a priori estimates on how things will turn out, they cannot be certain until it has happened. Furthermore, history is rife with examples of people misjudging what is best for their society and its members. In other words consequentialism alone can only provide an explanation of how acts can be judged, it does not provide us with a moral code to adhere, or a reason to adhere it; i.e. consequentialism alone fails to provide a moral theory that is normatively adequate, which, as we have seen above, Southwood is willing to grant. It is here that Hobbesian contractarianism comes into play.

4.2 The Social Contract

In order to better explain what the role of contractarianism is in this theory, let us turn to a moral skeptic. The moral skeptic might ask a question what reason he has to be moral. Surely he sees that people are social beings and that they get a lot done in a cooperative setting, but that does not answer the question why he should be moral. To answer this question we must ask the moral skeptic several questions: the first is what he gains from being part of society, or any other group. Second he must ask himself whether he would prefer a situation where this group does not exist. This latter question might appear a bit severe, but it is very much an implication of moral behavior. Moral behavior as I have noted before exists in order to let a group function, and a group must as a result

92 At least those groups that are cooperative in nature, rather than just a collection of individuals that happen to share similar characteristics.

93 This way of phrasing it also prevents issues like torturing people for the amusement of others becoming or appearing to become morally permissible, since this does not prevent harm, it merely creates benefit. In other words it allows for the establishment of a Lockean proviso. However the fact that people may not be harmed in their functioning also means that it is permissible to ask sacrifices for the greater good, while simultaneously limiting what sacrifices can be demanded.

be maintained. Maintaining moral behavior requires both that the vast majority of agents act minimally moral and that those who do not act morally are subject to sanctions in order to minimize damage to the group. How large the amount of people that acts in accordance with morality has to be, and what sort of rules they must adhere to, depends on the nature of the group. Where a murder probably will not cause a society to collapse, a friendship is far less likely to survive it. Since each agent, as an agent, is first and foremost responsible for her own actions, and she has to make sure these actions are not harmful to the groups she a part of, no matter her motivations for these actions. In other words, it is up to the preference of the individual to decide whether she wants to be a part of a group, but once this preference is decided it becomes in her interest to act in accordance with the group in a way that facilitates the functioning of the group. This means that once an agent decides to act morally, her actions must become group-directed, rather than being merely agent-directed. This normativity is not merely reason giving to the moral skeptic, but reason giving in such a way that it constrains her ability to maximize her own interests all the time, in favor of maintaining the group, which provides a greater net benefit.

The above explanation, although providing a form of normativity that I expect to be intensionally sufficient,⁹⁴ does not yet provide us with a moral contract. The contract mostly comes into its own when looking at the substantive aspects of a moral code. Just as each group requires that its members adhere to the rules required to make it function, it relies on its members to make these rules. Each member of the group that is capable of contributing to the cooperation can express their interpretations on what is harmful or beneficial for the group, and once a certain critical mass is reached on what is permissible it can be incorporated as a moral rule in the cooperation. This is how the social contract is formed, as a result of cooperation between members who contribute to the norms. Once the normative system is established each member faces the choice of confirming to the rules, leaving the cooperation or facing the consequences.

It might be objected that it is not always easy to leave a certain group. This, however, does not have to be problematic, since I have never claimed that it needs to be easy. A person might isolate themselves in an uninhabited area, move to another, more suitable group, or even commit suicide, but it is up to the person whether the costs of leaving the group weigh up to the costs of remaining in a group and minimally conforming to it. Another objection is that this implies that moral codes are established solely by a majority. This objection can be answered in two ways however: The first response is that moral codes are not necessarily established through a majority but rather by a critical mass, the exact nature of which requires a sociological study that goes well beyond the confines of this paper. This critical mass might even be subject to more partiality, allowing some individuals to exert more influence in it than others, it might also require unanimity or near unanimity. This however leads to the second point that states that the contract merely establishes what sort of substantive assumptions are made about what is harmful or beneficial to the group, as well as a as which rules flow from this, what is ultimately moral however is what actually turns out to be moral for the group, which, as I have claimed before, is only knowable a posteriori. Since people can change their minds, the contract is mutable and norms and moral rules will shift as new norms reach critical mass, while old norms lose it.

As noted above the exact moral code is defined by the actual members of the group, and it follows that it affects all members of that group in the same way. After all the basis of morality demands that harm to the functioning of members of the group, or the group as a whole, is avoided. In a case where discrimination or other situations where morality is highly partial between several groups, is morally accepted this implies that rather than having a single group the situation consists of two different groups that are in contact with each other. This might sound a lot like moral relativism,

⁹⁴ In other words, it provides the right kind and the right amount of normativity for morality as I describe it in this paper.

but I seek to give a description of morality in general, without making substantive claims about moral contents. The group does not consider itself to be immoral, even if an outsider considers the division they create between themselves and other groups as based on morally irrelevant assumptions. Since the contract exists between actual members that can affect each other, there are constraints on how universal morality can and has to be, not unlike what I have noted about Southwood above. It would not be wrong to torture a unicorn or to decapitate the king of France, simply because the simple facts that unicorns do not exist and that France is a republic, so it has no king. Similarly, since the Spanish and the Arawak did not have any contact or even ability to contact each other prior to Columbus' first voyage in 1492, they did not have any ability to be moral towards each other. It was only after the Spanish came into contact with the Arawak that cooperative moral norms could be, and indeed were, established by forming a group of people that shared the same lands who kept the peace by adhering certain norms. In other words they established an Unwritten Contract, at least until the destruction of Fort Navidad at the hands of the Taino Arawak. This also demonstrates that the social contract does not merely occur between individuals, but can also occur between groups, demanding that members of either group treat members of the other group in accordance with the moral duties established between the groups.

4.3 Impartiality: dealing with extensional aspects of morality in a non-substantive theory.

The issue of impartiality, although partially covered by the fact that, once included in a group, each member of the group has to prevent harm to the functioning of the group and its members, still lingers though. We have yet to see reasons why we would include atypical persons such as the severely handicapped into a group to begin with. There are also issues concerning the inclusion of those who cannot partake in the contract, but to who or which rights might still be attributed, such as animals. And finally we face an issue of contracts between groups, after all what reason does a stronger group have to respect a weaker one? These are of course not the only issues the Unwritten Contract faces so far. I will briefly discuss some examples, together with some other issues faced or brought up by Southwood and Gauthier below. By doing so I will attempt to show that the Unwritten Contract can accurately describe the demand that morality is not to be selfish, but has to give those subject to it legitimate expectations towards moral agents; in other words that the Unwritten contract can place plausible constraints on both partiality and impartiality.

As discussed above one of the reasons Gauthier chooses preferences over interests is because the two do not always coincide. I fully admit that there is no interest if there is not at least some preference for said interest. However the mere fact that interests are derived from preferences does not make them explanatorily epiphenomenal or obsolete. As argued by Peter Railton in his essay on "Moral Realism", interests provide an objective standpoint from which ethics can be derived.⁹⁵ Railton argues, much in the vein of what I have stated above, that interests can be derived through an instrumental approach, such as with the example of the rocket scientist discussed above. In this respect I differ from Gauthier as described above, making my theory capable of accounting for an objective morality which independent of individual preferences, at least insofar as preferences lead to instrumental interests. The only way in which morality is dependent on preferences is that an agent needs to want to conform to society, with morality itself being derived from the instrumental interests that flow from wanting to belong to said society. That these interests and other preferences can diverge is not as problematic as Gauthier makes it to be, as shown by the irrationality of addiction as displayed above.

The Unwritten Contract and its consequentialist elements provide a way in which interests can be

⁹⁵ Peter Railton, "Moral Realism", *Foundations of Ethics, An Anthology* (ed. Russ Schafer-Landau and Terence Cuneo), (Malden: Blackwell Publishing, 2007), pp186-204.

judged; interests can correspond with situations in which individuals can either function at least as well as before or be hampered in their functioning. This means that the maximization of preferences as displayed by Gauthier, at least the straightforward maximization of preferences that Southwood ascribes to Hobbesian contractarianism does not have to be morally right. By basing morality on interests we avoid the risk of being accused of going for straightforward maximization, as the only maximization of preferences that is allowed by an interest-based morality is constrained maximization. Ultimately morality is uncaring about whether or not we achieve what we prefer, as long as we do not damage the ability of others to function in the group.⁹⁶

Several issues remain when it comes to impartiality though. Despite being objective the rights are still morally contingent. As discussed above this is not necessarily problematic, as even Southwood, who seeks a universal morality, rejects the notion of morality needing to avoid contingencies at all costs. But unlike Southwood's deliberative contractualism the consequentialist elements of an Unwritten Contract are capable of dealing with taboos that are not, or at least less contingent; under the condition that these taboos concern acts that are detrimental for any form of cooperation.

As mentioned above Southwood mentions two types of partiality with which his theory of deliberative contractualism is capable of dealing, *de jure* and *de facto* partiality. Southwood's approach seems to have one issue however, he does not seem to be able to account of partial duties, special obligations. The Unwritten Contract on the other hand is quite capable of both accounting and limiting partial duties. As I have argued above, the Unwritten Contract accounts for partiality through the existence of overlapping groups, each of which attribute its members rights and duties that accumulate together. I have also argued that a group can have a subdivision of groups in it. Combining both of these factors we are capable of accounting both for obligations of partiality and impartiality. For example though all inhabitants of a country can be expected to have obligations towards family and friends they associate⁹⁷ with, politicians form a special group within society with special obligations to not benefit their friends and family over other citizens, insofar as their duties as politicians are concerned.

The partiality within groups can even provide a solution against being overly impartial, a situation discussed by Southwood.⁹⁸ Southwood discusses a situation not unlike the Singer Solution, which demands not only that we are impartial, but spend as much of our spare time and resources as possible on improving the life of the impoverished.⁹⁹ Since we are part of several groups which each have their own moral demands in order to let them function, we cannot merely spend all of our time and resources improving the lives of strangers completely unknown to us. If we were to focus all our resources on aiding the largest possible amount of humans, other groups we are part of such as friendships, families, religious congregations, etc. would suffer as a result. A member of such groups could no longer function as a member if they were to follow the Singer Solution, essentially making those who do follow the Singer Solution part of a cult of humanity. Furthermore such a

96 Professor Katrien Schaubroeck pointed out to me that this interpretation of interests over preferences might result in paternalism becoming morally acceptable, a situation that is objectionable to both Kantian thinkers and many defenders of negative freedom. For paternalism to be morally objectionable however, several substantive claims about morality have to be assumed, such as, for example, the substantive conception of agency in Kantian thought. As stated above however, I do not seek to make substantive claims about morality in this paper. It is however safe to assume that if a substantive claim about morality were to defend paternalism, it would find paternalism to only be permissible insofar it would not inhibit the functioning of the society or its members about which this claim was made.

97 I say associate with mostly because it would be odd to presuppose duties towards family members that have broken off contact.

98 Southwood, *Contractualism and the Foundations of Morality*, p161.

99 For the Singer solution, see Peter Singer, "The Singer Solution to World Poverty", *The Utilitarian*, <URL=<http://www.utilitarian.net/singer/by/19990905.htm>> (viewed 17-5-2013)

Singer Solution goes beyond our moral duties not to harm others in their functioning, presupposing a substantive moral duty towards all people, not merely those we affect.

This leads us to a point where the impartiality problem becomes glaring. How can an Unwritten Contract account for impartiality at all if all moral duties are partial in nature? In the next few paragraphs I will argue that impartiality is not as stringent a criterion for moral accuracy as Southwood assumes. I have already discussed that moral rules within groups are essentially impartial when it comes to factors irrelevant to the functioning in a group. I have also argued that for morality to work it needs not be universally impartial, but it must, at the very least, not discount impartiality as irrational. This leads us to the issue of impartiality boiling down to something I have claimed before: the issue is not why individuals would accept a moral code that demands impartiality, but why the moral code should accept a substance of impartiality at all. For those who, through a small amount of aid, can create a net benefit for society are not problematic. If a person with a light handicap can function normally in a society when they have access to certain facilities, such as laws against discrimination, tools that aid them, and affordable healthcare, not helping them will cause more harm to a society, whereas helping them will not. After all, if they do not have this aid they cannot function in the society, being parasitic to the situation, whereas with aid they can contribute as much as other members.

The greater issue comes for accepting principles of impartiality towards atypical persons that can never, and have never been able to, function in society. There are several options available, each relying on a substantive choice by a society to sacrifice some means to grant atypical persons rights. These options are not mutually exclusive, and a society can rely on multiple of these options together to rationalize a choice for impartiality. The first solution is prudential in nature: since not everybody is capable of functioning to the same amount, you need contingency plans for those who cannot function as well as others. In some cases atypical persons will only be able to function insofar as they live without suffering, which would grant them rights against suffering and neglect. If stated like this, the moral imperative behind the Unwritten Contract would be changed to "The group and its members are not to be harmed in their functioning so far as they can function." Which would allow for affective rights not only towards atypical persons but also towards animals. This sort of prudential reasoning would also account for people who become incapable of fully functioning, or who come from outside of the group, since such people could be a danger to the society or its members if they were able to affect it without moral constraints in place.

A second option is affective in nature: Atypical persons incapable of functioning in a society are not only incapable of functioning themselves, but also inhibit the functioning of others. They require care, nourishment, attention and sometimes even restraining, which takes away from the ability of their caretakers to function in society. Unless these atypical persons opt-out of society, it is in the interest of the caretakers to be supported in their role through rights to care and support for their atypical charges, a minor sacrifice spread across a group does not inhibit the functioning of the members of the group, but prevents the caretakers from suffering in their functionality in the group. The mere fact that these people are dependent of the group in order to function does not deny them their rights; any role that is derived from cooperation is dependent on that cooperation. A rich CEO depends on his company, the economy and that which allows both these factors to exist in order to function as a CEO. If he were stranded outside of society, say somewhere on the African savannah, he would quickly find that nature and the predators that reside there are particularly uncaring about wealth, social status and the support of shareholders.

A third option, which is closely related to the second one is that all rights are derivative; each person's rights in a group are dependent on them being a part of said group. If a total stranger were to call me with the question why I have not visited them in hospital yet, I could quite rightfully reply that since they are neither a friend, a family member or even an acquaintance I am under no

obligation to visit them; though I could equally well visit them out of some altruistic preference to alleviate the loneliness of this person that is in a hospital. A result of this third solution would be that as long as a critical mass of people wanted to grant a right on altruistic grounds, or even because they required something from someone that would benefit from such rights, there would be no reason for such a right to be worth any less than any other rights. In such a case Southwood's objection against derivative rights would be void, since it would imply all rights are void. It can easily be argued that any contractarian theory, be it Hobbesian, Kantian or deliberative would be hard pressed to avoid this third option, since, provided they are not circular, they all derive the rights people have from a social contract, be it between actual or hypothetical people. All of the three options discussed above however lead to contractarianism being capable of providing and limiting partiality in such a way that the extensional character of morality can be sufficiently explained, even if full impartiality, as supposed by Southwood, remains out of its reach.

One issue that plays when discussing impartiality in a contractarian is that of harm to society, and it is easy to see monetary harm as an important example of this. To show that monetary harm is not the most important form of harm a society can suffer, let us turn to a common argument used against egalitarian economic theories like communism. The argument in question states that if each person gets exactly the same reward for their work, regardless of the work being done people will lose motivation to do certain work. This harm done to the group would in such a case not be monetary in nature, since the amount of money made is assumed to be roughly the same for the sake of this argument. The issue is that a lack of motivation is harmful for the functioning of the group. As a result it might be possible that certain harms will cause less damage to functioning than others. A more exact analysis of what harms cause which damage to the functioning of the group and its members is beyond the scope of this paper however, and might well be too dependent on the situation to have a definitive answer available.

Conclusion

Although the Unwritten Contract is not capable of making impartiality a requirement for all moral codes, it does show that impartiality does not have to be rejected either. If my theory is correct however, impartiality is not necessarily required for a proper explanation of morality's extensional character, rather it depends on the substantive nature of the moral code. This, as has been discussed above, also has implications for the universality of morality. Southwood's claim that morality is universal would hold if what is right for a group is indeed dependent on contingent facts about what rules the group requires for itself and its members to function. The argument that there are multiple overlapping groups that each have their own moral codes further enhances the need to reject complete impartiality, since it makes partial duties much more prominent. On the other hand the need for groups to function, also means that partial duties, as far as they concern others in the same group, still require some degree of impartiality.

We have seen that Southwood and Gauthier each have their own reasons to respectively accept and reject the impartiality problem. Gauthier, basing his moral theory on the maximization of preferences with minimal constraints where possible, rejects impartiality as unnecessary for morality. As discussed above, this does not mean that he argues for complete pursuit of self-interest at the cost of morality however, as he argues that constrained maximization allows for cooperative ventures that allow a greater maximization of preferences for all those involved. Southwood on the other hand argues for a universal moral code based on deliberation with all those involved, which requires complete impartiality for those involved. To this end Southwood criticizes Hobbesian contractarianism which is incapable of providing such a universal moral code.

I have then argued that Gauthier's reliance on the rationality of preferences, as well as a, from a non-moral standpoint, arbitrary reliance on non-forceful talents as opposed to talents involving force or fraud, provide a problem for his theory's ability to provide an accurate account of morality. I have also argued that Southwood's arguments against Gauthier and Hobbesian contractarianism in general rely on a wrongful interpretation of instrumental rationality and the presupposition of morally substantive claims. Finally I have argued that by focusing on a moral theory that uses the functioning of the group and its members as the basis for morality we can provide an account of morality that can both describe different moral codes, and deal with impartiality by showing both its uses and its limits.

It seems that the impartiality problem is not as crippling for Hobbesian contractarianism as initially assumed. Impartiality, though useful is not unlimited in its force and its constraints will have to be taken into account in order to provide a plausible account of morality, both when discussing it in a descriptive non-substantive way, and when discussing particular moral codes.

Bibliography

- Í *Doctor Who*. 2010. "The Eleventh Hour". BBC. <URL=<http://www.youtube.com/watch?v=TQuCP0GJHf4>> (viewed April 23, 2013)
- Í Gauthier, David. 1986. *Morals By Agreement*. Oxford: Clarendon Press.
- Í Plato. 2004. *Republic*. Translator: Reeves. Indianapolis: Hackett Publishing Company, inc.
- Í Railton, Peter . 2007. "Moral Realism", *Foundations of Ethics, An Anthology*. Editors: Russ Schafer-Landau and Terence Cuneo. Malden: Blackwell Publishing. pp186-204.
- Í Singer, Peter. "The Singer Solution to World Poverty". *The Utilitarian*. <URL=<http://www.utilitarian.net/singer/by/19990905.htm>> (viewed 17-5-2013)
- Í Southwood, Nicholas. 2010. *Contractualism and the Foundations of Morality*. Oxford: Oxford University Press.