

Hobbes' foundation for peace and property

by

Michael Shaun Christopher Cust

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Political Science

Waterloo, Ontario, Canada, 2008

©Michael Shaun Christopher Cust 2008

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

I defend Hobbes' foundation for peace and property. His foundation for peace and property is his major argument for why society's moral order (i.e. collection of rules of interaction) should be based on the principles of non-interference and exclusive use of material objects. His foundation is that in the absence of both a recognised moral order and government, it would be rational, or felicity maximising, for individuals to agree to a moral order constituted by peace and property. The cogency of his foundation depends on the accuracy of the second of the two steps of his state of nature thought experiment. In the first step, he formulates the state of nature by defining it as a social state of affairs with no government, by arguing that, as a consequence of there being no government, there would be no recognised moral order, and by assuming there would be relievable scarcity. In the second step, he theorises that interactions this anarchic state of affairs would be periodically violent. Also, the second step is informed by his theory of human nature, that is, his theory of the major characteristics common to all humans. Given that his foundation's cogency is subject to the accuracy of the second step of his state of nature thought experiment and that the second step of his thought experiment is informed by his theory of human nature, my defence of his foundation involves arguments in favour of his theory of human nature, his state of nature, and his foundation. I first contend that the six characteristics that compose his theory of human nature are true. I next argue that the second step of his state of nature thought experiment – his theory that state of nature interactions would be periodically violent – is accurate. Lastly, I argue that his foundation is true, that it would be felicity maximising for individuals to agree to a moral order based on peace and property in the absence of government and a recognised moral order. To make my argument, I construct a hypothetical bargain between individuals in the state of nature where they choose between Hobbes moral order based on peace and property and the sort of moral order most contemporary political philosophers would propose as alternative (e.g. one based on general non-interference and a redistribution requirement) as their improvement over the state of nature. I argue that individuals would choose the former over the latter as their improvement because the former is purely mutually beneficial while the latter is only partly mutually beneficial.

Acknowledgements

I would like to thank my partner Jill Isenberg for her love and support during the composition of this thesis and throughout my master's program more generally. She, more than anyone, is the reason I was able to overcome several illnesses and times of much stress to complete it.

I would also like to thank my supervisor, Dr. Jan Narveson, for his intellectual tutelage and his patience. He carefully guided my reading of Hobbes and answered countless questions by email.

I would like to thank my second reader, Dr. Richard Nutbrown, for helping narrow my topic, for some insightful secondary sources, for instilling in me a serious interest in the history of political philosophy, and for standing in my corner when I needed him most.

I would like to thank my grandmother for providing eleven months of lodging so that I could focus on writing this thesis.

I would like to thank my parents for their love and support.

I would like to thank Lorne for his support and for his advice on how to approaching writing and deal with stress.

I would like to acknowledge the gracious support I received from the Ludwig von Mises Institute. They twice hosted me as a Rowley Visiting Fellow, once in the fall 2005 term and once in the summer 2006 term. Their gifts allowed me to complete the residual course work that remained from when I was sick as well as conduct research towards this thesis.

Lastly, I would like to thank Liberty Fund, the Institute for Humane Studies, and The Fraser Institute for the conferences they invited me to. These gatherings provided much fruit for thought.

Table of Contents

List of Figures	vii
List of Tables.....	viii
Chapter 1 Introduction.....	1
1.1 Hobbes' foundation for peace and property and his political theory	1
1.2 Overview of my analysis and defence of Hobbes' foundation.....	3
Chapter 2 Theory of human nature	5
2.1 Humans are agents of desire	5
2.2 Humans are rational.....	6
2.2.1 Prudence.....	7
2.2.2 Reason	7
2.2.3 Judgement	8
2.2.4 Deliberation	9
2.2.5 Rational choice theory and Hobbesian rationality	10
2.3 Humans act to maximise felicity	14
2.4 Humans are fundamentally equal physically	15
2.5 Humans are naturally amoral	15
2.6 Humans possess limited altruism	15
Chapter 3 State of nature	17
3.1 Formulating the state of nature	17
3.2 Hobbes' theory of anarchic interactions	21
Chapter 4 Hobbes' foundation for peace and property: mutual benefit.....	33
4.1 Definition of law of nature.....	34
4.2 First law of nature.....	36
4.2.1 Peace includes property	36
4.3 Second law of nature	37
4.3.1 Forgoing the absolute liberty to use all material objects	38
4.3.2 The limit on forgoing the absolute liberty to use all material objects.....	40
4.3.3 By social contract	40
4.3.4 To establish a moral order based peace and property	41
4.4 Arguing in favour of Hobbes' foundation.....	42
Chapter 5 Conclusion	49

References	50
Works Consulted.....	51

List of Figures

Figure 1: First argument of Hobbes' theory of anarchic interactions

Figure 2: Second argument of Hobbes' theory of anarchic interactions

Figure 3: Third argument of Hobbes' theory of anarchic interactions

List of Tables

Table 1: Prisoner's Dilemma

Chapter 1

Introduction

1.1 Hobbes' foundation for peace and property and his political theory

In this thesis, I defend Hobbes' foundation for peace and property. His foundation for peace and property is the primary argument he puts forward for why society's moral order should be based on the principles of peace and private property. That is, it is his fundamental reasons for why society's system of rules of interaction should be constituted by the rule of non-interference and the rule of exclusive use of material objects. His foundation is that in the absence of both a recognised moral order and government, it would be rational, or felicity maximising, for individuals to agree to a moral order constituted by the principles of peace and property. So, the reason why society's moral order should be composed of the principles of peace and property is that these principles are mutually beneficial; that is, they best serve our long-term well-being given that we need the agreement of others who are similarly concerned with their long-term well-being. In this introduction, I will provide an overview of how I will defend his foundation. In order for the reader to adequately appreciate my overview, I must first briefly analyse the five parts of his political theory (of which his foundation is one part). My brief analysis will consist of identifying the five parts of his political theory, explaining what is meant by each part, and describing the function each part serves in his political theory.

I will begin my brief analysis by identifying the five parts of his political theory. Once I have identified the five parts, I will then explain what is meant by each part and outline the function each part performs in his political theory. The five parts of his political theory are: 1) his normative moral question and his normative political question, 2) his theory of human nature, 3) his state of nature, 4) his foundation for moral order based on peace and property, and 5) his foundation for an absolute sovereign. I will now explain what is meant by each part and outline the function each part serves in his political theory. Let me begin with his normative moral question and his normative political question. I will consider each question separately beginning with the former. To appreciate what is meant by his normative moral question, we must first understand what normative moral questions are. Normative moral questions are ought questions concerning rules of interaction, such as should individuals be made to aid others in overcoming their addictions? Understanding what normative moral questions are, we can now appreciate what is meant by his normative moral question. His normative moral question is 'what moral order ought govern society?' That is, what system of rules of interaction should regulate our interfaces? The function of his normative moral question in his political theory is to elicit a response. To appreciate what is meant by his normative political question, we must first understand what normative political questions are. Normative political questions are ought questions relating to government, such as what surveillance powers should the state be granted? Understanding what normative political questions are, we can now appreciate what is meant by his normative political question. His normative political question is 'is government justified and if so to what extent?' That is, should there be an organisation with a monopoly on force in this territory and if so what powers should it possess? The function his normative political question serves in his political

theory is to elicit a response. I will now explain what is meant by his theory of human nature and outline the function it serves in his political theory. His theory of human nature is his theory of the major characteristics common to all humans. In particular, he assumes six such characteristics: i) humans are agents of desire, ii) humans are rational, iii) humans act to maximise felicity, iv) humans are fundamentally equal physically, v) humans are naturally amoral, and vi) humans possess limited altruism. To allow the reader to comprehend the function that his theory of human nature fulfils in his political theory, I must first undertake my explanation of what is meant by his state of nature and my outline of the function that it serves in his political theory. His state of nature is a thought experiment that he performs in two steps. In the first step, he formulates the state of nature by defining it as a social state of affairs with no government, by arguing that, as a consequence of there being no government, there would no *recognised* moral order, and by assuming that there would be relievable scarcity. In the second step, he theorises how interactions in this anarchic state of affairs would play out. His theory involves three arguments. The first argument is that anarchic interactions would be characterised by war, that is, periodic violence. The second argument is that this periodic violence would result in short lives and no productive activity, and, further, this result would mean that each individual would face the prospect of a low long-term felicity. The third argument is that each dweller's prospect of a low long-term felicity would move them to establish a moral order based on peace and property. The function of his state of nature in his political theory is providing a method for developing answers to both his normative moral question and his normative political question. I will explain how his state of nature enables him to develop an answer to his normative moral question after I put forward the function of his foundation for a moral order based peace and property below. The reason that I present this explanation after I put forward the function of his foundation is because the reader must understand the function of his foundation to appreciate it. Further, I will explain how his state of nature allows him to develop answer his normative political question after I consider the function of his foundation for an absolute sovereign. The reason that I put forward this explanation after I present the function of his foundation is because the reader must comprehend the function of his foundation to appreciate it. Now that I have undertaken my explanation of what is meant by his state of nature and my outline of the function it serves in his political theory, the reader can now comprehend the function that his theory of human nature serves in his political theory. The function that his theory of human nature serves is informing the second of step of his state of nature, or his theory of anarchic interactions. That is, how he thinks anarchic interactions would unfold is based on the five major characteristics common to all humans that he identifies. I will now outline the function that his foundation for a moral order based on peace and property serves in his political theory. I will not explain what is meant by it because I established its meaning in the opening paragraph. Its function in his political theory is to provide a reply to his normative moral question. Now that the reader appreciates the function of his foundation, I can explain how he uses his state of nature to develop an answer to his normative moral question. He uses his state of nature to develop an answer to it in that his reply's (i.e. his foundation's) cogency is dependent on the accuracy of the first argument of the second step of his state of nature. That is, that it would be rational for dwellers to agree to a moral order based on peace and property is dependent on state of nature interactions being periodically violent. I should note that it would still be rational for them to agree to such a moral order if interactions developed peacefully. It is just *a fortiori* rational for them to agree given that interactions are periodically violent. This is because the moral order based on peace and property

directly addresses the cause of their low long-term felicity (i.e. a short life and no productive activity). I will now explain what is meant by his foundation for an absolute sovereign and outline the function that it plays in his political theory. His foundation is his major argument for why members of society should grant government complete power to control them. His foundation is that having agreed to a moral order based on peace and property, individuals in the state of nature would agree to establish a government with absolute power to enforce it. They would do so because moral order can be successfully enforced over time only by a government with absolute power. Were they to choose enforcement by private means or by a government of only limited powers, they would risk returning to the periodic violence of the state of nature. So, members of society should grant government absolute power to control them so that they do not risk periodic violence resembling the state of nature. The function of his foundation is to act as a reply to his normative political question. Now that the reader appreciates the function of his foundation, I can explain how he uses his state of nature to develop a reply to his normative political question. He uses his state of nature to develop a reply (i.e. his foundation) in that the extent to which his reply is convincing is dependent on the accuracy of the first argument of the second step of his state of nature. That is, that it would be rational for individuals in the state of nature to agree to establish an absolute state to enforce moral order is contingent on state of nature interactions being periodically violent. If state of nature interactions were peaceful, there would be little reason for dwellers to agree to set up a kind of state that could potentially do great harm to them.

1.2 Overview of my analysis and defence of Hobbes' foundation

Having so far briefly analysed the five parts of his political theory, I can now present my overview of how I will defend his foundation. I will defend his foundation by analysing, and making arguments in support of, his theory of human nature, his state of nature, and his foundation. The reason that I provide an analysis of these three parts of his political theory in addition to my arguments in favour of them is that a proper understanding of them is required to both formulate and appreciate compelling arguments in favour of them. The reason that my defence of his foundation includes arguments in favour of his theory of human nature and his state of nature and not just his foundation is that his foundation rests on his state of nature which in turn rests on his theory of human nature. Recall that in my brief analysis of the five parts of his political theory above, I showed that the cogency of his foundation is dependent on the accuracy of his theory of anarchic interactions (i.e. the second part of his state of nature) and that his theory of human nature informs his theory of anarchic interactions. Let me now explain my defence of his foundation in greater detail. My defence has three parts. In the first part, which I undertake in chapter 2, I analyse, and make arguments in favour of, his theory of human nature. In particular, I outline the six assumptions that compose his theory and I argue that each of them is true. In the second part, which I will carry out in chapter 3, I analyse the first step of his state of nature thought experiment (i.e. his formulation of the state of nature) and I analyse, and make arguments in favour of, the second step of his state of nature thought experiment (i.e. his theory of anarchic interactions). My analysis of the first step of his state of nature experiment will involve arguing that he formulates his state of nature by defining it as a social state of affairs with no government, by arguing that, as a consequence of there being no government, there would no *recognised* moral order, and by assuming that there would be relievable scarcity. For my analysis, and

arguments in favour, of the second step of his state of nature, I will explicate the three arguments that compose his theory of anarchic interactions and argue that each is strong so, on the whole, his theory is accurate. I should note that after I analyse the first argument of his theory – that anarchic interactions would be characterised by periodic violence – and contend that it is strong, I consider two prominent sets of game theoretic objections to his argument. The proponents of both sets of objections view anarchic interactions as an iterated series of prisoner’s dilemma games. I show how he could successfully reply to both sets of objections. In the third part, which I perform in chapter 4, I analyse and argue in favour of his foundation. My analysis of his foundation will consist of arguing for a specific reading of his foundation, viz. that it is felicity maximising for dwellers to socially contract to establish a moral order based on the principles of peace and property. My defence of his foundation will involve a hypothetical negotiation between individuals in the state of nature where they choose between a moral order based on the principles of peace and property and the sort of moral order most contemporary political philosophers would propose as alternative (e.g. one based on general non-interference and a redistribution requirement) as their improvement over the state of nature. I will demonstrate that individuals would choose a moral order based on peace and property instead of the sort most contemporary political philosophers would propose as alternative because the former is purely mutually beneficial and the latter is not.

Chapter 2

Theory of human nature

Let me begin with the first part of my defence of Hobbes' foundation for peace and property, namely, analysing, and making arguments in favour of, his theory of human nature. His theory of human nature, as was mentioned in the introduction, is his theory of the characteristics common to all humans. As I pointed out, he assumes six such characteristics: 1) humans are agents of desire, 2) humans are rational, 3) humans act to maximise felicity, 4) humans are fundamentally equal physically, 5) humans are naturally amoral, and 6) humans possess limited altruism. In this chapter, I briefly outline each characteristic and argue that each is true. Also, I perform one other task. After I outline the second characteristic he assumes (*viz.* humans are rational) and argue that it is true, I compare it to rational choice theory. The reason that I undertake this comparison relates to the two sets of game theoretic objections to the first argument of his theory of anarchic interactions that I consider in §3.2 below. For these two sets of objections to pose a meaningful challenge to the first argument of his theory, it must be the case that their assumptions are compatible with those of his first argument of his theory. The reason for undertaking the comparison is that it shows that the assumptions of the two sets of objections and those of the first argument of his theory are compatible. The comparison shows that the assumptions of both are compatible because the assumptions of the two sets of objections are captured by rational choice theory and the assumptions of the first argument of his theory are the six characteristics that compose his theory of human nature. The reason that the comparison is between only the second assumed characteristic of his theory of human nature and rational choice theory instead of all six characteristics of his theory of human nature and rational choice theory is because rational choice theory makes claims relevant only to the second assumed characteristic. I should note that we know that the assumptions of his first argument are the six assumed characteristics that compose his theory of human nature because the assumptions of his first argument are those of his theory of anarchic interactions and the assumptions of his theory of anarchic interactions are, as was alluded to in the introduction, the six assumed characteristics that compose his theory of human nature.

2.1 Humans are agents of desire

Let me begin with the first characteristic of Hobbes' theory human nature, humans are agents of desire. By it, he means that desires arise in us and we act to fulfil them. In explaining this characteristic, he answers three related questions: i) what are desires and aversions, ii) how do they arise in us, and iii) how do we act to fulfil them?

In answer to the first question, he defines desires as tiny motions of some of the matter that composes our bodies. The combined movement of these motions is in the direction of whatever causes them. Aversions, by contrast, are tiny motions of some of the matter in our bodies that together move away from whatever causes them (2002: 38). Desires make us want to act on whatever is causing them, for example by consuming the thing that causes them. By contrast, aversions make us want to leave the presence of whatever is causing them.

In answer to the second question – how do desires arise? – he theorises that desires and aversions have two sources: either they are a) inborn (e.g. the desire for food, excretion, exoneration, etc.) or they proceed from b) things external to us with which we have experience (2002: 39). He develops a theory of external-source desires, but not one of internal-source desires. In the following paragraph, I will briefly consider his theory because it provides some clarification of his definition of desire.

Hobbes contends that external-source desires arise when external objects act on our sense organs. External objects “presseth the organ proper to each sense, either immediately, as in the Taste and Touch; or mediately, as in Seeing, Hearing, and Smelling” (2002: 13). Objects press our organs when “so many several motions of the matter” that composes them produces “divers motion” of the matter that composes the organs for “motion, produces nothing but motion” (2002: 14). This motion is “continued from the Eyes, Eares, and other organs to the Heart” (2002: 40). Once there, if the motion is “a corroboration of vital motion” – that is, a support to our “reflexive or physically compulsory motive force” (Frost 2001: 34) – a desire is produced (Hobbes 2002: 14). By contrast, if the motion is a hindrance to the vital motion, an aversion to the object is formed.

I will leave a substantial analysis of Hobbes’ definition of desire and his account of desire formation to more able philosophers of biology and mind. I will do so because, from our vantage point, it does not matter what desires actually are, it just matters that people have them and that they are moved by them. It is enough for our purposes that we are each aware that we have wants, or will have wants, and that we act, or will act, to fulfil them. We do not need a detailed theory of internal springs and whistles.

2.2 Humans are rational

In terms of the first characteristic his theory of human nature (*viz.* that we are agents of desire) we have yet to consider his answer to the third question: how do we act to fulfil desires? In answer to this question, he puts forward the second characteristic of his theory of human nature: humans are rational. His theory of rationality, which I will now consider, outlines both the component parts of our rational faculty and explains how we use those parts to act to fulfil our desires. He identifies four components of our rational faculty: a) prudence, b) reason, c) judgement, and d) deliberation. He argues that we use them to determine which actions to undertake to best satisfy our desires via a three-step process. In the first step, we use prudence, reason or judgement to determine both possible actions we can undertake and the consequences, or outcomes, of those possible actions. In the second step, we again use prudence, reason or judgement. We use them to establish the level of either desire satisfaction or aversion that each of the consequences of each possible action will cause. In the third step, we choose to do that action whose consequences collectively represent the greatest increase in desire satisfaction. In what follows, I provide an overview of his explanation of each component of rationality. When I consider his explanations, I also outline how he thinks each type of rationality is used in its respective step(s) of the three-step desire satisfaction process. That is, when I consider his explanations of prudence, reason and judgement, I also provide an account of how he thinks each of them is used in the first and second steps of the desire satisfaction process. And, when I outline his explanation of deliberation, I show how he thinks it is used in the third step of the desire satisfaction process.

2.2.1 Prudence

Hobbes defines being prudent as follows: “[if] a man desires to know the event of an action... he thinketh of some like action past, and the events thereof one after another; supposing like events will follow from like actions” (2002: 22). Prudence, then, is determining the outcomes over time of a possible action by reflection on similar past actions. As mentioned, he contends that it is valuable in the first two steps of desire satisfaction. In terms of the first step – determining which actions are possible and what their consequences will be – his definition of prudence informs us that it is useful for the second action, determining the effects of possible actions. We can also infer from his definition that prudence is likely useful for the first action as well. We can figure out which actions are possible by recalling similar past situations and the actions we undertook in them. In terms of the second step – figuring out the level of desire satisfaction or aversion each of the consequences of a possible action will cause – prudence enables us to determine the amount of desire satisfaction or aversion that will be caused by each consequence by thinking of similar past consequence and the level of mental well-being or harm that they caused. There are also two observations that Hobbes makes regarding prudence that are worth noting because each is important to one of his later arguments. The first observation is that prudence is universal in distribution and quality – all humans possess it and all are equally proficient in its use (2002: 87). This observation is important to his later argument regarding how interactions in anarchy would play out. In fact, his argument depends on the truth of this claim. The second observation is that the predictive accuracy of one’s prudential faculty is dependent on how much experience one possesses (2002: 52). Those with more experience are more prudent. This observation is important to the reply I construct on his behalf to the game theoretic claim that anarchic would turn out differently than he contends. His reply turns on this observation being true. I will contend that both observations are true when I argue that the second characteristic of his theory of human nature, that humans are rational, is true. I will make that argument presently.

2.2.2 Reason

Reason, for Hobbes, is what we now consider basic, or introductory, logic. He defines it as “Adding and Subtracting... of the Consequences of general names agreed upon for the *marking* and *signifying* of our thoughts.” (2002: 32). By this definition, he has in mind a four-step reasoning process. In the first step, we perspicuously define the names that pertain to the subject we wish to analyse. We do so to avoid absurdity, for “the errors of Definitions multiply themselves, according as the [reasoning process] proceeds” until we end up with “false and senseless tenets” (2002: 28). The names we define for our analysis are to be general. General, or universal, names are names that denote “many things, for their similitude in some quality, or other accident” (2002: 26). He contrasts general names with particular, or proper names – names that are “singular to one only thing” (ibid). Particular names are not our concern, because we seek general rules of cause and effect (2002: 34). In the second step of the four-step reasoning process, we express true propositions using the general names we have defined. Propositions are true when the general name of ‘greater extent’ in the proposition – i.e. the name that signifies more things – signifies all that is signified by the general name, or particular name, of ‘lesse extent’ – i.e. the name that signifies fewer things (2002: 27). For example, ‘all humans are animals’ is true because animals – the name of greater extent – signifies all that is

signified by humans – the name of the lesser extent. In the third step of the four-step reasoning process, we construct arguments. This step involves taking the true propositions expressed in step two and putting them into arguments as premises. The aim is to determine their consequences.

Consequences are propositions that must be true if the premises of the argument are true (Blackburn 1996: 77). That is, they are the *conclusions* of *valid* arguments. In the fourth step of the four-step reasoning process, we construct additional arguments. We do so by taking the consequences found in step three and use them as premises in a new set of arguments. These new arguments produce a new set of consequences. The new set of consequences is then used as the premises in yet another set of arguments which generate yet another set of consequences. This process of argument construction is continued, “till we come to a knowledge of all the Consequences of names appertaining to the subject in hand.” (Hobbes 2002: 35). That is, we continue constructing arguments until we find a set of consequences that can no longer be used as premises in a new set of argument. This could occur either because we reach one final consequence, or because the new consequences do not imply consequences when put together as premises. The consequences found are then reduced to general rules of cause and effect, or conditional predictive rules, that tell us which effects will follow from a certain set of causes (2002: 34). These rules have the following form: if cause A, then effects x, y, z.

As I pointed out, Hobbes thinks that reason is useful in the first and second steps of the desire satisfaction process. In the first step – determining which actions are possible and what their outcomes will be – it is valuable because it allows us to formulate general conditional predictive rules that tell us which effects will follow from which types of actions. In the second step – determining the amount of desire satisfaction or aversion caused by each consequence of a possible action – it is useful because it provides us with the means to build general predictive rules that communicate what level of mental well-being or harm follows from which outcomes. We should note that he makes a contention regarding reason that is important for both his later argument regarding how interactions in anarchy would unfold and his reply to the game theoretic argument that anarchic interactions would unfold differently than he maintains. His contention is that all people can “reason alike, and well, when they have good principles” (2002: 35). However, he does not think that everyone can formulate such rules. As for the above-mentioned four-step reasoning process by which general rules are created, only “very few have [it], and but in a few things” (2002: 87). I will argue that his contention that all can reason with general principles, but that not all can construct those general rules, is true. I will do this when I argue that the second characteristic of his theory of human nature, that humans are rational, is true. I will do so presently.

2.2.3 Judgement

Hobbes defines judgement as “the last Opinion in search of the truth of Past, and Future” (2002: 47). He means that when we are thinking about whether something did or did not happen, or whether something will or will not happen, the thought we settle on is our judgement. As I mentioned, he thinks that judgement is valuable in the first and second steps of the desire satisfaction process. In terms of the first step – determining which actions are possible and what their outcomes will be – it enables us to decide on which actions we can do and what consequences will follow from them. In terms of the second step – figuring out the amount of desire satisfaction or aversion each consequence

of a possible action will cause – it allows us to conclude just what the amount of desire satisfaction or aversion for each consequence will be.

2.2.4 Deliberation

Hobbes defines deliberation as follows,

When in the mind of man, Appetites [i.e. desires], and Aversions... concerning one and the same thing, arise alternately; and divers good and evil consequences of the doing, or omitting the thing propounded, come successively into our thoughts; so that sometimes we have an Appetite to it; sometimes an Aversion from it... the whole summe of Desires, Aversions... continued till the thing be either done, or thought impossible, is that we call DELIBERATION. (2002: 44)

Later in the same chapter (i.e. VI), while setting down his definition of apparent good, he elaborates on his definition of deliberation,

...because in Deliberation, the Appetites, and Aversions are raised by the foresight of the good and evill consequences, and sequels of the action whereof we Deliberate; the good or evill effect thereof dependeth on the foresight of a long chain of consequences, of which very seldome any man is able to see to the end. But for so farre as a man seeth, if the Good in those consequences, be greater than the Evill, the whole chaine is that which Writers call *Apparent*, or *Seeming Good*. And contrarily, when the Evill exceedeth the Good, the whole is *Apparent*, or *Seeming Evill*: so that he who by Experience [i.e. prudence], or Reason, the greatest and surest prospect of Consequences, Deliberate best himselfe... (2002: 46)

To understand deliberation, we must first grasp what Hobbes means by good and evil consequences. To grasp what he means by good and evil consequences, we must appreciate his definitions of good and evil. That is, we must comprehend his theory of value. His theory is this: “whatsoever is the object of any mans Appetite or Desire; that is it, which he for his part calleth *Good*: And the object of his Hate, and Aversion, *Evill*” (2004: 39). Value for Hobbes, then, is subjective. As such, good consequences are the outcomes of a possible action that we think will result in a certain amount of desire satisfaction. Accordingly, evil consequences are the effects of a possible action that we think will cause a certain amount of aversion. Knowing what good and evil consequences are, the definition of deliberation becomes clear. The definition presents deliberation as involving three stages. The first stage follows on the heels of the second step of the desire satisfaction process. In the second step of the desire satisfaction process, we use prudence, reason, or judgement to determine the amount of desire satisfaction or aversion each consequence of a possible action will result in. In the first stage of deliberation, for each consequence of a possible action we think will result in a certain amount of desire satisfaction, a desire to undertake the possible action of a strength that corresponds to the anticipated amount of desire satisfaction from the consequence arises in us.

Similarly, for every consequence of a possible action that we think will cause in us a certain amount of aversion, an aversion towards undertaking the possible action of a potency commensurate with the predicted amount of aversion of the consequence arises in us. In the second stage of deliberation, of the possible actions we can do in a given situation, we will develop the strongest desire to do that action whose consequences, collectively, represent the greatest amount of desire satisfaction (or the least amount of aversion). In the third stage, we choose to do that action that we have the strongest desire to do. As noted above, Hobbes thinks that deliberation is valuable in the third step of the desire satisfaction process. In that step, we choose to do that action whose consequences collectively represent the greatest quantity of desire satisfaction. As I have just described it, deliberation is clearly the method by which we undertake this step.

Having considered the four component parts of the second characteristic of Hobbes' theory of human nature – prudence, reason, judgement, and deliberation – I will now argue that the second characteristic of his theory is for the most part true. That is, I will contend that it is mostly true that humans are rational. Consider: almost all humans use language. Almost all humans can do simple math. Almost all humans use these skills in the service of their desires. On the face of it, then, the second characteristic of his theory is mostly true. My defence of this second characteristic must also include a defence of one specific claim that he makes regarding human rationality. That claim is that prudence is a universal, or at least near universal, trait. I take it as obvious that most people can reflect on past situations to figure out which actions are possible in their present situation, remember the consequences of similar past actions in order to determine the consequences of present possible actions, and think of the desire satisfaction or aversion caused by similar past consequences to predict whether the consequences of present possible actions will result in desire satisfaction or aversion. So, this claim is true. Worth noting, I have two reasons for defending this claim. The first reason is that Hobbes uses this aspect of rationality as a premise in the first argument of his theory of anarchic interactions. This is significant because the cogency of his foundation and my ability to defend it rests on the accuracy of his theory anarchic interactions which in turn rests on the truth of the six characteristics of his theory of human nature. Therefore, if in his theory of anarchic interactions, he focuses on one specific characteristic of his theory of human nature, I should explicitly defend that characteristic as true. The second reason I defend his assertion regarding prudence is that it is important to the reply I formulate on his behalf to the second set of game theoretic objections to the first argument of his theory of anarchic interactions.

2.2.5 Rational choice theory and Hobbesian rationality

Before we proceed with my outline of the remaining three characteristics of Hobbes' theory of human nature, I would like to examine the relationship between his theory of rationality and rational choice theory. My examination involves two undertakings. First, I will present a short overview of rational choice theory, including its two main variations, utility theory and expected utility theory. Second, I will argue that Hobbes' theory of rationality is fundamentally similar to rational choice theory by making four comparisons between his theory and utility theory. My reasons for comparing his theory to utility theory instead of expected utility theory, and instead of both utility theory and expected utility theory, will be put forward when I make my argument below. The reason that I carry out my examination is because in my later analysis of his theory of anarchic interactions in §3.2, I put

forward two sets of game theoretic objections to the first argument of his theory that suggest different outcomes to anarchic interactions than does his first argument. For the arguments to pose a meaningful challenge to his first argument, it must be the case that the assumptions upon which they are based are compatible with the assumptions on which his first argument is based. If the assumptions are not compatible, the demonstration will be ineffective. So what if a theory based on different assumptions suggests a different outcome than does Hobbes' first argument? (Of course, it is possible for a criticism based on different assumptions to be effective if we argue that its assumptions are right and that his assumptions problematic, yet we are, in chapter 2, as will be shown immediately, arguing that the assumptions on which his first argument is based are true.) The assumptions on which the two sets of game theoretic assumptions are based are represented by rational choice theory. The assumptions on which his first argument are based are the same as those of his theory of anarchic interactions. The assumptions on which his theory of anarchic interactions is based are the six characteristics of his theory of human nature. The reason that I demonstrate the compatibility of rational choice theory with only his theory of rationality, and not the other characteristics of his theory of human nature, is because rational choice theory makes claims relevant only to his theory of rationality. It has no implications for the other characteristics of his theory of human nature (or, at least, it does not contest them).

Let me now provide a short overview of the two main variations of rational choice theory, utility theory and expected utility theory. I will summarize utility theory first followed by expected utility theory. Utility theory is characterised by an individual performing four tasks. In the first task, she determines the set of possible actions she can undertake. In second task, she ascertains with certainty the outcome that will follow each possible action. (The assumption behind this second task, that individuals can determine the outcome that follows an action *with certainty*, is what distinguishes utility theory from expected utility theory. As I will show presently, expected utility theory holds that individuals are uncertain about the outcomes that follow each possible action. Therefore, they determine the set of possible outcomes and assign each member of the set a different probability of obtaining (that together sum to 1).) In the third task of utility theory, the individual assigns each outcome a (subjectively determined, non-interpersonal comparable) utility function (e.g. 8, 4, 2, etc.) so that she can establish preferences between pairs of possible actions. Further, her assignment of utility functions to outcomes is rational and therefore meets two conditions. The first condition is completeness. This means that she assigns a utility function to *every* outcome. She does so because otherwise she could not undertake the fourth task, choosing that action that maximises utility. If some outcomes are not assigned a utility function, she could not form preferences between *any* two possible actions. And if she cannot form preferences between any two possible actions, she cannot rank all possible actions. And if she cannot rank all possible actions, it makes it impossible for her to choose the action that will maximise her utility (Hausman and McPherson 2006: 47). The second condition that the individual's assignment of utility functions must meet is that they must represent transitive preferences over the set of possible actions. That is, if she prefers possible action x to possible action y, and possible action y to possible action z, she must also prefer possible action x to possible action z. The reason that her preferences must be transitive is that otherwise she would open herself up to manipulation (2006: 46), which is not in her interest. To understand how intransitive preferences would allow her to be manipulated, imagine that she is a sports fan. She prefers attending baseball games to hockey games and she prefers attending hockey games to football games. However, she also

prefers attending football games to baseball games in spite of her other two preferences. Moreover, assume that she is willing to trade the game tickets she has in her possession and a dollar to obtain her preferred alternative. Also, assumes that she is currently in possession of hockey tickets and that she happens upon a scalper holding a wide variety of sports tickets. She begins by trading her hockey tickets and a dollar to the scalper for baseball tickets. She then trades her baseball tickets and a dollar to the scalper for football tickets. Lastly, she trades her football tickets and a dollar to the scalper for hockey tickets. She is manipulated by the (more astute) scalper because she ends up where she started out minus three dollars, whereas the scalper ends up where he started out plus three dollars. Worse, the scalper could continue these trades with her until he takes all her dollars (assuming that she has more than three). The fourth and final task in utility theory, as we mentioned above, is that an individual maximises her utility by choosing to do her highest ranked possible action, or that action to which she assigns the highest (relative) utility function.

I will now summarise expected utility theory. Expected utility theory is characterised by an individual performing five tasks. The first task is identical to the first task of utility theory. The individual determines the set of possible actions they can undertake. The second task differs from the second task of utility theory in that the individual cannot figure out with certainty which outcome follows from each possible action. Rather, she figures out the set of possible outcomes for each possible action and assigns each possible outcome a (subjectively determined) probability of obtaining (such that the probabilities of the possible outcomes for any possible action sum to 1). The third task is similar to the third task of utility theory. The individual assigns each possible outcome a (subjectively determined, non-interpersonal comparable) utility function. In the fourth task, the individual calculates the expected utility of each possible action. This is done by summing the results obtained by multiplying the assigned probability by the assigned utility function for each possible outcome of a given possible action. Let me clarify this calculation with an example. Imagine that a baseball pitcher is contemplating throwing a screwball (as opposed to some other pitch). His screwball has two possible outcomes, ball and strike. He has assigned a probability of 0.6 and utility function of 5 to strike and a probability of 0.4 and a utility function of 0 to ball. He would then calculate the expected utility of throwing a screwball as follows, $[(5 \cdot 0.6) + (0 \cdot 0.4)] = 3$. In the fifth task of expected utility theory, the individual chooses, or undertakes, the possible action with the highest (relative) expected utility.

I will now argue that Hobbes' theory of rationality is fundamentally similar to rational choice theory. I will do so by making four comparisons between his theory and utility theory. Each of the four comparisons is between one of, or part of one of, the three steps in his theory and one of the four tasks in utility theory. For each of the four comparisons, I argue that the step, or the part of the step, in his theory and the task in utility theory being compared are the same type of action. Also, for each of the four comparisons, I outline the differences between the step, or the part of the step, in his theory and the task in utility theory being compared. I argue that they are still effectively the same action in spite of these differences. Given that I am arguing that Hobbes' theory of rationality is fundamentally similar to rational choice theory, the reader may wonder why I am comparing his theory only to utility theory? Why am I not comparing his theory to utility theory *and* expected utility theory? Further, if I am only going to compare his theory to one variation of utility theory, why do I compare it to utility theory and not to expected utility theory? Let me answer the former two questions first,

followed by the latter question. The reason that I do not compare his theory to both variations of rational choice theory is because if I can show that his theory is similar in all, or nearly all, respects to one variation, I will have shown that it is basically the same as the other variation. This is because, as we saw, both variations are identical save that utility theory assumes that an individual can determine with certainty the outcome that follows each possible action and expected utility theory does not. I will now answer the latter question. The reason that I choose to compare his theory to utility theory and not expected utility theory is because his theory is primarily concerned with the outcomes of actions. It is not, or so it seems, concerned with the further complication that outcomes are only probable. As such, I compare it to the type of rational choice theory that assumes away this further complication. (Recall that utility theory avoids the problem of probability of outcomes by assuming that individuals can determine the outcome of each possible action with a probability of 1.) The result is a cleaner comparison.

The first comparison I wish to make is between the first task in utility theory, an individual determines the set of possible actions they can perform, and the first of the two actions performed in the first step of Hobbes' theory of rationality, an individual determines possible actions. Both actions are of the same type. In both cases, the individual figures out which actions are practicable. In spite of this core similarity, the two actions might be somewhat different. That is because it is not clear whether Hobbes thinks individuals determine all possible actions, like in the first task in utility theory, or simply some. That said, this possible difference does not hinder the effective similarity of the actions. Just because they differ in rigour does not mean they differ in type.

The second comparison I wish to draw the reader's attention to is between the second task of utility theory, an individual determines with certainty the outcome of each possible action, and the second of the two actions undertaken in the first step of Hobbes' theory of rationality, an individual figures out the consequences, or outcomes, of each possible action. Both actions are of the same type. Both involve the individual figuring out what will follow a possible action. They are, however, different in two important respects. First, they differ in the number of outcomes that the individual establishes follow each possible action. In the second task of utility theory, an individual determines with certainty that one outcome will follow after each possible action is performed. In second of the two actions undertaken in the first step of Hobbesian rationality theory, by contrast, an individual establishes that multiple outcomes will occur over time after each possible action is undertaken. Recall that Hobbes speaks of 'long chains of consequences.' Second, the actions differ in terms of the extent to which the individual can determine the outcome(s) that follows each possible action. This is because the respective theories they are parts of differ in the amount of rationality attributed to the individual. In utility theory, the individual is assumed perfectly rational and in possession of perfect information. As such, she can determine the outcome of each possible action with certainty and act in accordance with this information. In Hobbesian rationality theory, by contrast, individuals are assumed to differ in their reasoning and prudential faculties. They can figure out all the outcomes over time that will follow a possible action if their reason or prudence is great enough. However, the weaker their reason or prudence, the fewer the number of outcomes they can determine into the future. The two differences between the two actions, number of outcomes and the extent to which outcomes can be determined, does not detract from the fact that they are both actions of the same

type. This is because the differences do not change the common characteristic that unites them – determining what follows a possible action.

The third comparison I wish to put forward is between the third task in utility theory, assigning the outcome of each possible action a utility function, and the second step in Hobbes' theory of rationality, determining the desire satisfaction of each consequence of a possible action. Both actions are of the same kind. Both actions are an evaluation of the level of subjective well-being of the outcome(s) that follows each possible action. That said, there is an important difference between the two actions that we should make note of. The difference is that the second step of Hobbes' theory is quite likely not nearly as sophisticated as the third task of utility theory. There is no evidence to suggest that Hobbes thinks that the desire satisfaction levels the individual assigns to consequences of actions should be complete and transitive. However, this difference does not change the fact that the individual is performing effectively the same action in both the second step of his theory and the third task of utility theory.

The fourth comparison I wish to present is between the fourth task of utility theory, choosing the action with the highest utility function, and the third step in Hobbes' theory of rationality, choosing that action whose collective consequences have the greatest amount of desire satisfaction. Both actions are of the same type. Both involve the individual selecting the possible action that it is thought will provide the greatest increase in subjective well-being. There appears, however, to be an important difference between the two actions. The fourth step in utility theory appears to assume a libertarian account of choice. That is, it seems to assume that the individual chooses the possible action they do using their will, free of any other causes. Hobbes' third step, by contrast, appears to assume a compatibilist account of choice. As we saw in §2.2.4 above, he thinks that the individual chooses the action she desires most. In contending as much, he appears to be suggesting an overlap between the will and a cause external to the will, desire. This difference does not detract from the fundamental similarity of actions in that it the different accounts of choice do not take away from the fact that the purpose of each act is essentially the same.

2.3 Humans act to maximise felicity

I will now outline the remaining three characteristics of Hobbes' theory of human nature, beginning with the third characteristic of his theory. The third characteristic of his theory is that the primary end for which humans act is maximising felicity, or desire satisfaction. We can infer this from his definition of felicity. He defines felicity as “*Continuall successe* in obtaining those things which a man from time to time desireth” (2002: 46). That is, we act now to experience the greatest possible amount of desire satisfaction over time by obtaining material objects that will satisfy our desires as they arise. Admittedly, this third characteristic of his theory can be considered a part of the second characteristic. This is because the end result of the three-step desire satisfaction process is choosing that possible action that offers the greatest amount of desire satisfaction over time. The reason that I treat the third characteristic as distinct from the second is that my overview of the second characteristic focuses more on the three-step process itself, yet it is important to separately identify and emphasise the reason for undertaking that process.

I will now argue that the third characteristic is true. Observation tells us that most people act in the present to obtain material objects that will increase their future desire satisfaction. The most conspicuous example of this phenomenon is that most people spend a significant amount of their time working. That is, they exert effort now so that they can enjoy greater future desire satisfaction than they otherwise would. Those who fail to exert effort in the present for future benefits, like the homeless, generally report being less happy than those who do.

2.4 Humans are fundamentally equal physically

I will now outline the fourth characteristic of Hobbes' theory of human nature, namely, humans are fundamentally equal physically. He argues, "...as to the strength of body, the weakest has strength enough to kill the strongest, either by secret machination, or by confederacy with others, that are in the same danger with himselfe" (2002: 87). Certainly this fourth characteristic is true. Most every human possesses the real ability to kill every other. For example, a tiny grandmother can kill a football player in the prime of his career by slitting his throat while he sleeps or by poisoning his food.

2.5 Humans are naturally amoral

I will now outline the fifth characteristic of Hobbes' theory of human nature. The fifth characteristic of his theory is that humans are amoral by nature. He argues that "Justice and Injustice are none of the Faculties neither of the Body nor Mind. If they were, they might be in a man that were alone in the world, as well as his Senses, and Passions" (2002: 90). That is, humans are naturally amoral because we lack an innate mental or physical capacity that tells us when an action is right or wrong. If we actually had such an internal capacity, it would be with us when we are isolated from others as are our other mental and physical capacities.

This characteristic appears to be true if we consider young children. They seem to have no inborn moral sense. What moral sense they do have appears to develop as a product of their parents' coaching and learning by watching others.

2.6 Humans possess limited altruism

I will now outline the sixth and final characteristic of Hobbes' theory of human nature. The sixth characteristic of his theory is that humans possess limited altruism. By limited altruism, I mean that humans are primarily motivated by, and act to satisfy, desires that concern their own person and their immediate family and intimate friends. For example, humans often desire satiation for themselves and their children and they will act to obtain foodstuffs for themselves and their young to satisfy this desire. The main implication of limited altruism is that humans are rarely, if ever, motivated by, and act to satisfy, desires that concern perfect strangers or society as a whole. That Hobbes assumes that humans possess limited altruism can be inferred from at least three passages in chapter XIII of *Leviathan*. The first two passages offer evidence that he thinks humans are motivated by, and act to satisfy, desires that primarily concern themselves. The third passage offers evidence that he thinks that humans are motivated by, and act to satisfy, desires that concern their immediate family and

close friends. Let me first demonstrate that the first two passages offer evidence that he thinks that humans are motivated by, and act to satisfy, desires that primarily concern themselves. Each of the first two passages represent one of the reasons that constitute the first argument of his theory of anarchic interactions, viz. that interactions would be periodically violent. (It should be noted that I will analyse the first argument of his theory of anarchic interactions in greater detail in §3.2 below.) The first passage is the reason competition. It reads, "...if any two men desire the same thing, which nevertheless they cannot both enjoy, they become enemies; and in the way to their End... endeavour to destroy, or subdue one another" (2002: 87). He is contending, in effect, that scarcity will lead to violent interactions. For this to be true, it must also be true that individuals are mostly motivated by, and act on, desires that concern their person. For, were they not – that is, were they fundamentally motivated by, and acted on, desires that concern strangers or society as a whole – they would likely let the person they are in conflict with take the object of contention. The second passage is the reason of the desire for glory. He writes, "every man looketh that his companion should value him, at the same rate he sets upon himself: And upon all signes of contempt, or undervaluing, naturally endeavours... to extort a greater value from his contemners, by dommage; and from others, by the example." (2002: 88). For our purpose, what matters is that he attributes the desire for glory to each individual. That is, he assumes that each individual wishes his peers to perceive him as better than both themselves and others. This shows that he thinks that humans are mostly motivated by, and act on, desires that concern their own person because it suggests that he thinks that individuals are motivated by, and act on, a desire that places themselves above others. I will now demonstrate that the third passage from chapter XIII offers evidence that he thinks that humans are motivated by, and act to satisfy, desires that concern their immediate family and close friends. The third passage is his reason for why the small family governments of the North American natives of his day were stable. It reads, "...the savage people in many places of *America* [enjoy] the government of small Families, the concord whereof dependeth on naturall lust..." (2002: 89). So, the small family governments of North American natives of his day survived because humans are naturally motivated by, and act on, desires that concern their immediate family and close friends.

I will now argue that it is true that humans possess limited altruism. The most conspicuous example that they do is how most people spend their after tax income. They usually spend the vast majority on themselves and their immediate family. Often only a tiny amount, if any, is ever spent on strangers or society as a whole.

Chapter 3

State of nature

I will now undertake the second part of my defence of Hobbes' foundation, namely, analysing the first step of his state of nature thought experiment (i.e. his formulation of the state of nature) and analysing, and making arguments in favour of, the second step of his state of nature thought experiment (i.e. his theory of anarchic interactions). I will begin by analysing the first step. Once I have done so, I will analyse, and make arguments in favour of, the second step.

3.1 Formulating the state of nature

In this section, I will analyse the first step of Hobbes' state of nature thought experiment, his formulation of the state of nature. I will argue that he formulates the state of nature by defining it as a social state of affairs that is characterised by no government, by arguing that, as a consequence of there being no government, there would be no *recognised* (i.e. accepted) moral order, and by assuming that there would be relievable scarcity. Specifically, I will provide evidence that confirms two claims that together constitute my argument. The claims are: a) he defines the state of nature as a social state of affairs with no government and he argues that, as a consequence of there being no government, there would be no recognised moral order and b) he assumes that the state of nature would be a social state of affairs with relievable scarcity.

Let me begin by advancing the evidence for claim (a). The evidence is three items: i) his attribution of the right of nature to each individual in the state of nature in his chapter on the first two laws of nature, ii) six phrases from his chapter on the state of nature, and iii) his argument that moral order would not be recognised in the state of nature. Item (i) completely confirms claim (a). It demonstrates that he defines the state of nature a state of affairs that possesses no government and that he argues that, as a consequence of there being no government, there would be no recognised moral order. Item (ii) partly corroborates claim (a). It establishes that he defines the state of nature as a social state of affairs without government. Item (iii) partly validates claim (a). It shows that argues that, as a consequence of there being no government in the state of nature, there would be no recognised moral order. I will present and analyse the three items in the order they are numbered.

Let me start, then, with item (i) – Hobbes' attribution of the right of nature to each state of nature dweller. As I mentioned, it fully confirms claim (a) in that it is proof that he defines the state of nature as a social state of affairs without government and that he argues that, as a consequence of there being no government in the state of nature, there would be no recognised moral order. Item (i):

THE RIGHT OF NATURE... is the Liberty each man hath, to use his own power, as he will himselfe, for the preservation of his own Nature; that is to say, of his own Life; and consequently, of doing any thing, which in his own Judgement, and Reason, hee shall conceive to be the aptest means thereunto. (2002: 91)

The first thing to note in the above passage is that he is not using the term 'right' in the sense in which most contemporary political theorists do. Most contemporary political theorists use 'right' to mean a claim by one person which has correlative duties for all other persons. The claim can be to anything, an object, an action, a service, etc. And the correlative duties are either positive or negative. A negative duty means that a non-right holder has an obligation not to act (in some way) with reference to that which the right holder lays claim. A positive duty means that a non-right holder has an obligation to act (in some way) with reference to that which the right holder lays claim. To make better sense of the contemporary notion of right, consider two examples. First, Smith has a negative right to a dock in Vancouver. This means that she has a claim to the dock such that others have a duty not to moor their boats at it (or otherwise use it) without her permission. Second, Smith has a positive right to a dock in Vancouver. This means that she has a claim to a dock in the area known as 'Vancouver' such that others must build, buy, or take one for her.

Hobbes uses 'right' in a much different sense than the contemporary sense. By it, he means the "liberty to do, or to forbear" (*ibid.*). That is, he is referring to a sphere of action wherein the right holder enjoys the unrestricted choice to do or not do those actions included in the sphere. What is interesting about his notion of right is that, unlike the contemporary notion, there are no correlative duties -- negative or positive -- on non-right holders. So, if someone enjoys a Hobbesian right, there is no negative duty on others not to interfere with the choices in their sphere of action, or, if for some reason they cannot exercise some of the choices within their sphere, there is no positive duty on others to help them do so.

Knowing what Hobbes means by 'right,' we can now determine the meaning of his 'right of nature.' He is arguing that people in the state of nature would have the liberty to do or not do whatever actions they think would best achieve self-preservation. That is, the sphere of action wherein they enjoy the unrestricted choice to do or not do those actions included in the sphere would include any action they think would best satisfy their present or future desires -- the assumption being that we preserve ourselves by maximizing our desire satisfaction. Since, it is possible that any combination of actions could, by each state of nature dweller's own estimation, satisfy a maximum amount of present or future desire, the sphere of action would include all (possible) actions. As such, we can say that the right of nature means that each dweller would enjoy absolute liberty of action to best satisfy their desires, present and future.

Understanding what Hobbes means by right of nature and knowing that he attributes it to each state of nature dweller, we have our evidence that defines the state of nature as a social state of affairs wanting a government and that he argues that, as a consequence of there being no government, there would be no recognised moral order. For, an individual can enjoy absolute freedom of action only where there is no monopolistic coercive power and where there are no accepted rules governing interaction.

I will now present and analyse item (ii), the six phrases from his chapter on the state of nature. I will advance the phrases in the order they appear in his chapter. As I previously noted, they partly confirm claim (a) – the assertion that he defines the state of nature as a social state of affairs with no government and that he argues that, as a consequence of there being no government, there would be no recognised moral order – in that they are evidence that he defines the state of

nature as a social state of affairs with no government. The first two phrases are antecedents of conditionals. The consequents of the conditionals are inconsequential, as it were, so I do not consider them. The first phrase reads "where an Invader hath no more to feare, than an other mans single power" (2002: 87). He means a condition where there is no government to protect potential victims of aggression from potential aggressors. He thinks that this antecedent is true of the state of nature. That is, he stipulates of the state of nature that there would be no state to protect potential victims of aggression from potential aggressors. The second phrase is "where there is no power able to over-awe them all" (2002: 88). He means a state of affairs where there is no person or group with a monopoly on force, i.e. no established government. He thinks that this antecedent is true of the state of nature. That is, he stipulates of the state of nature that there is no government. The third, fourth, and fifth phrases are descriptive propositions. He thinks that each proposition is true. The third phrase describes state of nature dwellers. It reads, "them that have no common power to keep them in quiet" (*ibid.*). That is, state of nature dwellers are individuals who live without a government that maintains peace among them. As such, the third phrase implies that Hobbes stipulates of the state of nature that it wants government. The fourth and fifth phrases describe the state of nature. The fourth phrase reads, "the time when men live without a common Power to keep them all in awe" (*ibid.*). This phrase has effectively the same meaning as the second phrase, viz. a state of affairs where there is no person or group with a monopoly on force. Because this phrase is a description of the state of nature, we can conclude that he is defining the state of nature as a social situation without government. The fifth phrase reads, "the time, wherein men live without other security, than what their own strength, and their own invention shall furnish them withall" (2002: 89). He means that the state of nature is a condition where humans would be without any other protection than their own physical strength and ingenuity. The 'other protection' Hobbes has in mind is government. Therefore, he is defining the state of nature as a social state of affairs that lacks government. The sixth phrase, like the first and second phrases, is an antecedent of a conditional. It reads, "where there were no common power to feare" (2002: 90). That is, a state of affairs where there is no government. He thinks the antecedent is true of the state of nature. Wherefore, he stipulates of the state of nature that it is a state of affairs without government.

I will now put forward and examine item (iii), his argument that state of nature dwellers would not recognise any moral order, or system of rules of interaction. As I pointed out, his argument partly corroborates claim (a) – the assertion that he defines the state of nature as a social situation with no government and that he argues that, as a consequence of there being no government in the state of nature, there would be no recognised moral order – in that it demonstrates that he argues that, as a consequence of there being no government in the state of nature, there would be no recognised moral order. His argument reads,

The notions of Right and Wrong, Justice and Injustice have there no place. Where there is no common Power, there is no Law: where no Law, no Injustice... Justice, and Injustice are none of the Faculties neither of the Body, nor Mind. If they were, they might be in a man that were alone in the world, as well as his Senses, and Passions. They are Qualities, that relate to men in Society, not in Solitude. (2002: 90)

In his argument he is offering two distinct contentions. First, he is contenting that no moral order of any kind will be recognised in the state of nature because rules of interaction will only be adhered to if they are enforced by a coercive power, or government, and the state of nature lacks a government. Second, he is contending that rules of interaction will only be adhered to if they are enforced by a government because humans do not have an innate tendency to recognise the need for them in that we are without an innate mental or physical capacity that tells us when an action is right or wrong. Had we such a capacity, the value of rules of interactions would be apparent and we might naturally follow them. The first contention of his argument partly corroborates claim (a) in that shows that he argues that, as a consequence of there being no government in the state of nature, there would be no recognised moral order.

I will now put forward the evidence for claim (b), the assertion that he assumes that the state of nature would be a social state of affairs with relievable scarcity. There are two pieces of evidence for this claim. The first is the second of four premises of the second of four short arguments that compose the first argument of his theory of anarchic interactions. (Note that I will analyse each of the four short arguments that compose his first argument in detail in the next section.) The second is one of two inferences of the second of two short arguments that make up the second argument of his theory of anarchic interactions. (Note that I will analyse both short arguments of the second argument of his theory of anarchic interactions in the next section.) The first piece of evidence shows that he assumes that the state of nature would be characterised by scarcity. The second piece of evidence demonstrates that he thinks that the scarcity that the state of nature would be characterised by would be relievable. This is significant because if he assumes that the state of nature would be characterised by scarcity and if he thinks that the scarcity that it would be characterised by would be relievable, then we can conclude that he assumes that the state of nature is a state of affairs with relievable scarcity. The first item reads, "...if any two men desire the same thing, which nevertheless they cannot both enjoy" (2002: 87). He is asserting that the state of nature would involve instances where two individuals would desire the same material object but both could not use it. So, he assumes that the state of nature would be typified by scarcity. The second item is his inference from the periodic violence that occurs in the state of nature that, "...there [would be] no place for Industry; because the fruit thereof [would be] uncertain: and consequently no Culture of the Earth; no Navigation, nor use of the commodities that may be imported by Sea; no commodious Building; no Instruments of moving, and removing such things as require much force..." (2002: 89). This inference implies that he thinks that the scarcity the state of nature would be characterised by would be relievable in that it implies that peace and cooperative activities would create an abundance of those material objects that not all individuals in the state of nature would get to use.

So far, I have analysed the first step of his state of nature thought experiment by providing evidence that confirms that he defines the state of nature as a social state of affairs with no government, that he argues that, as a consequence of there being no government in the state of nature, there would be no accepted moral order, and that he assumes relievable scarcity. I will now analyse and make arguments in favour of the second step of his state of nature thought experiment, his theory of anarchic interactions.

3.2 Hobbes' theory of anarchic interactions

Hobbes' theory of anarchic interactions is composed of three arguments. The first argument is that interactions in the state of nature would be characterised by "a warre, as if of every man, against every man" (2002: 88). By war, he means "not... Battell onely, or the act of fighting; but... a tract of time, wherein the will to Battell is sufficiently known" (2002: 88). His first argument, then, is that interactions in anarchy would be periodically violent and that people would maintain a disposition, or willingness, to be violent. We should also note that his first argument is composed of four short arguments. The first short argument concerns the expectation each person in the state of nature would have of satisfying her desires. It sets the stage for the remaining three short arguments. The remaining three short arguments regard his three reasons for why anarchic interactions would be violent: competition, diffidence, and glory. The reason we should note these four short arguments is because it is necessary to be aware of them when I outline my argument in the proceeding paragraph. The second argument of his theory of anarchic interactions has two parts. First, he contends that the long-term outcome of state of nature interactions would be a short life and the inability to successfully engage in productive activity. Second, he argues that dwellers would, as a result of the outcome, develop two desires, a fear of death and a desire both for those goods that satisfy desires and an ability to procure them by successfully engaging in productive activities. I argue that he means by them a low long-term felicity. The third argument of his theory is that the two desires, or low long-term felicity, would prompt state of nature dwellers to develop a disposition, or willingness, to be peaceful, and then discover the laws of nature. By consult the laws of nature, I argue that he means establish a peace-based moral order.

In this section, I argue that Hobbes' theory of anarchic interactions is accurate because the three arguments that compose it are strong. I make this argument because, as I pointed out in the introduction, the cogency of his foundation, which I later defend, depends on his theory being accurate. In making the argument, I undertake three major tasks. The first major task is analysing and evaluating the first argument of his theory. Within this major task, I perform two minor tasks. The first minor task is analysing the four short arguments of his first argument and contending that each one is strong.¹ Within this minor task, I pick up my contention from the introduction that his theory of anarchic interactions is informed by his theory of human nature. I do this by explaining the close relationship between the premises of his four short arguments and the six characteristics of his theory of human nature.² The second minor task is considering two sets of game theoretic objections to his first argument as well as possible successful replies he could make to each. Both sets of objections turn on the idea that state of nature interactions resemble iterated prisoner's dilemma games. As such, before I consider them, I outline the prisoner's dilemma game and apply it to state of nature interactions. The first set of objections involves Hampton's knowledge problem and de Jasay's reputation argument. Both of these objections concern two major arguments regarding iterated

¹ By the term 'strong,' I mean that there is a greater than 50% chance that the conclusions follow from the premises.

² Incidentally, this explanation helps to show that his four short arguments are strong. Because I earlier argued that the assumptions of his theory of human nature are true, I am able to show that many of the premises of his short arguments are true.

prisoner's dilemma games, viz. the tit-for-tat argument and the backwards induction argument. Therefore, before I consider the first set of objections and Hobbes' possible successful reply to them, I outline the tit-for-tat argument and the backwards induction argument. The second set of objections is my argument that individuals would assume a disposition to peace at the outset of the state of nature and that interactions would play out mostly peacefully. I present my argument and Hobbes' possible successful reply to it after I present the first set of objections. The second major task of my argument for this section is analysing his second argument and arguing that it is strong. The third major task is analysing his third argument and arguing that it is strong.

Let us begin by assessing the four short arguments of his first argument. The first short argument concerns the expectation each state of nature dweller would have of satisfying her desires. It has three premises. The first premise is that humans have ends and they pursue them. By ends, he means satisfaction of their various desires (2002: 87). This premise, then, is essentially the first assumption of his theory of human nature – humans are agents of desire (§2.1). The reader will recall that I defended that assumption, and hence this premise, as true. The second premise is that humans are essentially equal with respect to their rational faculties because they are equally prudent (2002: 87). Please note that this premise relates to his second assumption regarding human nature, that humans are rational (§2.2). Further, recall that in my analysis of that assumption I defended as true the claim that humans are equal with respect to prudence. That is, I defended this premise as true. The third premise is that humans are fundamentally equal in physical strength – i.e. each possesses the ability to kill every other (*ibid.*). This premise is the fourth assumption of his theory of human nature (§2.4). Recall that I defended that assumption as true. The second and third premises together comprise what I will term 'biological equality.' From the facts that humans are agents of desire and biologically equal, Hobbes concludes that humans in the state of nature would develop an equal hope of attaining their respective ends. That is, each would possess an expectation, similar in quality, of satisfying their respective desires. His conclusion follows from his premises in that rationality and physical strength are our means to satisfy our desires, and if humans are equally skilled with regard to them, it is quite reasonable that those in the state of nature would share an expectation similar in magnitude of satisfying their desires. We should note that this first short argument of his first argument is strong because the conclusion follows from true premises. Also, three sentences ago, in contending that his conclusion followed from his premises, I made the claim that reason and physical strength are our means to satisfy our desires. In order to fortify my contention, I will explain, albeit briefly, how physical strength is used to satisfy our desires. I will not explain how rationality is used because we considered the three-step process by which its component parts are used in §2.2. Physical strength would be used to undertake those actions selected by rationality.

The second short argument of his first argument concerns his first reason for why anarchic interactions would be violent: competition. It has four premises, two of which he states explicitly. The first stated premise is the conclusion of his first short argument, viz. that dwellers would have an equal hope of attaining their respective ends. We, of course, found that this conclusion is very likely true. The second stated premise is competition. Competition, for Hobbes, is a situation where two or more persons have as their end the same good(s) or piece of land. Certainly, situations of this type occur so this premise is true. From these two stated premises, he concludes that the people would "become enemies; and in the way to their End... endeavour to destroy, or subdue one another"

(*ibid.*). For this outcome to obtain, it would have to be assumed that two characteristics regarding humans are true. The first characteristic is that humans lack an innate moral sense. If they possessed such a sense, they might think twice about using force to obtain the object they desire. The second characteristic is that humans are primarily motivated by, and act on, desires that concern themselves. If they were instead motivated by, and acted on, desires that mainly concern others, they would likely allow the person that they are in conflict with to keep the object of contention. Because these two characteristics regarding humans would have to be assumed true for the conclusion of his short argument to follow, I think there is a third and a fourth hidden premise in his short argument. The third premise is the fifth assumed characteristic of his theory of human nature – that humans are amoral (§2.5). The fourth premise is the sixth assumed characteristic of his theory of human nature – that humans possess limited altruism (§2.6). Recall that I argued that that both characteristics are true. Given that all four premises in his short argument are true or very likely true, we can conclude that his short argument is strong.

Figure 1

First argument of Hobbes' theory of anarchic interactions

1. Biological equality + agents of desire → equal hope of attaining ends
 2. Equal hope of attaining ends + exclusive good desired by two or more persons + amorality + limited altruism → violent competition
 3. Violent competition → diffidence → pre-emptive attack
 4. Desire for glory + perceived undervaluing + amorality → attack for respect
- C. War of all against all

The third short argument of his first argument concerns his second reason for why anarchic interactions would be violent: diffidence. It is composed of two inferences. The first inference is drawn from the conclusion of his second short argument. That conclusion, of course, is that humans would compete violently over exclusive goods. He contends that violent competition would lead to diffidence (*ibid.*). That is, victims of violent competition would lose any hope they had of satisfying their desires (2002: 41). This first inference seems to follow. Anyone on the wrong side of a violent attack is often low in spirits. Given that we found the conclusion of his second short argument to be very likely true, this first inference is strong. From this first inference, Hobbes concludes a second inference. He argues that, to avoid the despair caused by an attack, people would begin pre-emptively attacking each other. This second inference is reasonable. If you attack others, it makes them less able to attack you and therefore less able to cause you despair. Since his first inference is very likely true, that makes this second inference strong. Given that both inferences are strong, we can conclude that his third short argument is strong.

The fourth short argument of his first argument concerns his third reason for why anarchic interactions would be violent: glory. It is composed of three premises – two stated, one hidden – and a conclusion. The first stated premise is the desire for glory. He describes it thus, “every man looketh that his companion should value him, at the same rate he sets upon himselfe” (2002: 88). That is, each individual wishes for members of her society to perceive her as better than both themselves and others. Hobbes considers this desire irrational (and a source of violence) because not every person can

satiated it – not every person can be better than others. His assumption that each human possesses the desire for glory is true. We all desire status of some kind or another. We should also note that this first stated premise relates to the first characteristic of his theory of human nature, that humans are agents of desire. Glory is one of the desires he has in mind when he states that desires arise in humans and we act to fulfil them. Further, his first stated premise also relates to the sixth assumed characteristics of his theory of human nature, that humans possess limited altruism. As was shown in §2.6, the desire for glory provides evidence that he thinks that humans are primarily motivated by, and act to satisfy, desires that concern their person. The second stated premise is that some people at some point will think that others are either showing them “contempt, or [are] undervaluing [them].” Certainly, situations of this type occur so this second premise is true. From the two stated premises, he concludes that those who feel slighted will “extort a greater value from [their] contempters, by damage; and from others, by the example” (*ibid.*). But this conclusion does not clearly follow. Just because people want status and others do not give it to them, it does not mean that they will lash out violently. For that outcome to be likely, they would need to be without an innate moral compass. As such, I think that this fourth short argument shares a hidden premise with his second short argument, viz. that humans are amoral. Once this hidden premise is recognised, it becomes apparent that the conclusion follows. Further, given that the stated premises are true, and that we earlier defended the hidden premise as true, we can infer that this fourth short argument is strong.

We have now found that all four short arguments of his first argument of his theory of anarchic interactions are strong. As such, we can conclude that his first argument is strong.

Let me now outline the game prisoner’s dilemma. In prisoner’s dilemma, two criminals, Jones and Smith, are arrested by police and charged with armed robbery and possession of cocaine for the purposes of trafficking. The two are interrogated separately. The police only have enough evidence to secure a conviction against each woman on the drug charge. Yet they have a preference for securing conviction against both on the bank robbery charge because it is the more serious crime. To collect the evidence needed to convict the women on the robbery charge, they offer each woman the following deal, which they know will yield confessions from each: “if your partner does not confess to robbing the bank and you do, I’ll let you off on both charges and you’ll serve no time in prison. Meanwhile, your friend will be gone for seven years for the bank robbery and another three years for perjury. If she doesn’t confess and you don’t confess, you’ll both do one year in prison for selling coke. If she confesses and you don’t, you’ll do seven years for the robbery and another three years for perjury. And she’ll get off scot-free. If she confesses and so do you, you’ll both do seven years for the robbery.” Table 1 below shows the possible outcomes facing each addict and the years in prison, or ‘payoffs,’ they can expect to serve if that outcome obtains. Jones’ payoffs are to the left of the comma and Smith’s payoffs are to the right.

The solution of the game, which the police know in advance, is that each player will confess. The reason each will confess is because doing so is what game theorists call the ‘dominant strategy,’ or the best decision rule. Confessing is the best decision rule because each player is better off confessing *no matter what the other player does*. Consider Jones. If Smith confesses, she can get either seven years by confessing or 10 years by not confessing. If Smith does not confess, she can get either no years by confessing or one year by not confessing. The crux of the game – why it is a dilemma – is

that the players would be individually and collectively better off if they both did not confess (one year each in prison) but their self-interest leads them to confess anyway (seven years each in prison).

Table 1
Prisoner's Dilemma

		Smith	
		Confess	Stay Silent
Jones	Confess	7,7	0,10
	Stay Silent	10,0	1,1

I will now adapt the prisoner's dilemma game to state of nature interactions by changing the players, their choices, the outcomes, the payoffs, and the solution to mirror the state of nature. The players, Jones and Smith, are no longer criminals. Instead, they are state of nature dwellers. Their choices are no longer confess and not confess, but instead interfere and not interfere. Therefore, the new set of outcomes for each player are as follows: a) interfere/do not interfere, b) do not interfere/do not interfere, c) interfere/interfere, d) do not interfere/interfere, where each player's move is listed to the left of the slash, their partner's move is listed to the right. Let me describe these new outcomes in greater detail. Assume that each dweller starts out with some material objects she has acquired. In (a), she keeps her objects, despoils the other dweller's objects, and subjugates the other dweller. In (b), both she and the other dweller keep their respective goods and remain uninjured and free. In (c), both she and the other dweller keep most of their objects, but suffer injury from battle. In (d), she sees her objects pillaged and she is subdued by the other dweller. The payoffs are no longer years in prison but (non-interpersonal comparable) cardinal felicity functions. Each dweller would attribute the following felicity functions to each outcome: a: 9, b: 7, c: -5, d: -10. The solution of the game is no longer mutual confession but instead mutual interference. The crux of the new game that both are lead to

interfere by their selfishness even though both would have individually and collectively better off if both played not interfere. All of our further discussions concerning the prisoner's dilemma, including the two major arguments concerning iterated prisoner's dilemma games and the two sets of objections to Hobbes' first argument of his theory of anarchic interactions, will centre on this adapted prisoner's dilemma game (rather than the game discussed above). So, when I refer to a particular outcome or payoff of prisoner's dilemma, I will refer to the outcome or payoff from this game. I will write this way so that our future discussions concerning prisoner's dilemma are kept fully relevant to Hobbes' first argument.

I will now outline the two major arguments regarding iterated prisoner's dilemma games, tit-for-tat and backwards induction, and explain their significance for Hobbes' first argument of his anarchic theory of interactions. Let me begin with the tit-for-tat argument. The tit-for-tat argument is that we can expect iterated games of prisoner's dilemma to be characterised by non-interference because tit-for-tat, which incentivises non-interference, is the dominant strategy in iterated games of prisoner's dilemma. The tit-for-tat strategy is play non-interfere on the first move and then, for all subsequent moves, do whatever the other player did on the preceding move (Axelrod 1984: 13). The reason tit-for-tat incentivises non-interference in iterated games of prisoner's dilemma because it makes it playing non-interference in the present game the most lucrative move long-term for that person playing the person employing it. To appreciate how, imagine an iterated series of prisoner's dilemma games between Employer, the person using tit-for-tat, and Opponent, their opponent. After a few games are played, Opponent will become aware of the pattern in Employer's behaviour. For simplicity's sake, assume that Opponent is concerned with maximising his payoffs over the present game and the next game. Further assume that he played non-interfere in the previous game (guaranteeing that Employer will play non-interference this round). If he plays non-interference this round, he will receive a payoff of 7. Next game (because Employer is certain to play non-interference) he will receive either a payoff of 7 if he plays non-interference or a payoff of 9 if he plays interference. Therefore, his overall payoff for this game and the next will be either 14 or 16 if he plays non-interfere this game. By contrast, if he plays interference this game, he will receive a payoff of 9. Next game (because Employer is certain to play interference) he will earn either a payoff of -10 if he plays non-interfere or -5 if he plays interference. Therefore, his overall payoff for this game and the next will be either -1 or 4 if he plays interfere in the present round. So, his payoffs over the present game and the next game will be either 15 higher or 12 higher if plays non-interference in the present game instead of interfere. The tit-for-tat strategy is considered to be dominant in iterated games of prisoner's dilemma is because of the research of Robert Axelrod.³ Axelrod organised two round robin computer tournaments between different strategies for prisoner's dilemma. The strategies were submitted by individuals doing research in game theory. In each of the round robin tournaments, each strategy played every other strategy for one round of two hundred games of prisoner's dilemma. In both tournaments, the tit-for-tat strategy emerged as victorious by earning the highest combined payoff from each round. The significance of the tit-for-tat argument for Hobbes' first argument of his theory of anarchic interactions is that it suggests that we can expect state of nature interactions to be characterised by non-interference, not periodic violence as he maintains.

³ See Axelrod (1984), especially chapters 1 and 2 and appendix A.

I will now consider the backwards induction argument. The backwards induction argument is that if the total number of prisoner's dilemma games in the series to be played is known, always interfere is the dominant strategy. The reasoning behind why always interfere is the dominant strategy goes like this: the final game in the series will play like a single game of prisoner's dilemma. That is, both parties will play interfere. The reason they will both play interfere is that since there is no next game to punish one's opponent for playing interfere in this game by playing interfere the next game and thus guaranteeing that her payoffs over the two games are lower than they otherwise would have been, there is no future benefit to playing non-interfere in the present game. And since there is no future benefit to playing non-interfere in the present game, each player is better off playing interfere no matter what their opponent does. In terms of the second last game, both parties will again play interfere. The reason is that there is no future benefit to playing non-interfere in the present game as interfere cannot be used as a punishment in the next round as one is already guaranteed to play it. As such, both parties are better off playing interfere no matter what the other party does. In the terms of the third last game, dwellers will again play interfere. Like in the second last game, there is no future benefit to playing non-interfere in the present game because interfere cannot be used as punishment in the next game as one is already guaranteed to play it. This reasoning continues 'backwards' until the first game. As such, players end up adopting always defect as their best strategy. The significance of the backward induction argument for Hobbes' first argument of his theory of anarchic interactions is that it predicts that state of nature interactions will develop violently as he contends.

Having considered the tit-for-tat argument and the backwards induction argument, I will now consider the first set of objections to Hobbes' first argument of his theory of anarchic interactions. The first set of objections includes Hampton's knowledge problem and de Jasay's reputation argument. Both objections are arguments for why state of nature interactions would play out as the tit-for-tat argument predicts rather than as how the backwards induction argument predicts. Further, both objections turn on the degree to which the backwards induction argument is applicable to state of nature interactions. Hampton's knowledge problem is that the backwards induction argument does not apply to state of nature interactions because each dweller would not know the number of times they will interact with any other dweller. She contends, "we do not know in this society how many times we will be in a position to cooperate with one another. So how could people in the state of nature be able to determine this fact? Indeed, how could it become *common* knowledge?" (Hampton 1995: 77). Her argument has appeal. It does seem that for the backwards induction argument to apply to state of nature interactions, one would have to make implausible assumptions regarding the information each dweller possessed. De Jasay's reputation argument is that the backwards induction argument does not apply to state of nature interactions because those individuals playing an opponent for a know number of games would have an incentive to play non-interfere in the last game. As such, the last game would not play as a single game of prisoner's dilemma and individuals could not employ the remaining logic of the backwards induction argument backwards to the first game thus guaranteeing that they would play interfere every round. The reason that individuals would have an incentive to play non-interfere in the last game with their current opponent is that their next opponent would be aware of their strategy against their current opponent and their next opponent would use this information in developing their strategy against them (de Jasay 1997: 32). In particular, their next opponent would employ the tit-for-tat strategy to (at least) their last game against their current opponent. The reason

that their next opponent would do so is that it would incentivise them to play non-interfere in the first game against their next opponent thus guaranteeing their next opponent a higher payoff in that game.

I will now consider a successful reply Hobbes could make to the first set of objections. He could argue that the considerations identified – a) lack of knowledge of total games and b) awareness other player's strategies – could engender non-interference were the stakes in state of nature interactions lower than they actually are. Were only some of each individual's property at risk from thievery, one might expect non-interference to develop by the use of tit-for-tat. But for those state of nature interactions that fit the model of the prisoner's dilemma, the stakes involved are so high that there is *never* a decent probability of a next game. The stakes involved are all your property, including your home and your land, your liberty, and your life (Hobbes 2002: 88). With stakes this great, tit-for-tat could not work. If you play non-interference and your opponent plays interfere, you will end up either dead or a slave. There will be no next round. You will either be in a makeshift grave or in shackles. It's true that the slave will have the opportunity to interact with her attacker again, but she won't do so as an equal and therefore she will not be able to punish him for his initial act of interference.

I will now put forward the second set of objections, my argument that individuals would assume a disposition to peace at the outset of the state of nature and that interactions would play out mostly peacefully. My argument is based on the distinction that Hobbes makes between two dispositions: (a) the disposition to violence and (b) the disposition to peace. For my argument to be appreciated, I must first outline his distinction and then argue that dispositions are detectable and therefore the disposition to peace is meaningful way to establish peaceful interactions. Hobbes discusses the disposition to violence in the definition of war considered at the beginning of the section "the Will to contend by Battell." By this, I understand him to mean a willingness to play interfere in state of nature prisoner's dilemma interactions. This disposition makes for a high stakes prisoner's dilemma interaction that makes each player play as if it were the last game. Hobbes discusses the disposition to peace in the second and third arguments of his theory of anarchic interactions below. He argues that low long-term felicity caused by short life and low productivity "encline men to Peace," or the adoption of a peaceful disposition. By this, I understand him to mean a willingness to play non-interfere in state of nature prisoner's dilemma interactions when others do so as well. The disposition to peace lowers the stakes of interactions and makes it so that these interactions are no longer instances of prisoner's dilemma. The only outcome is non-interference/non-interference. I will now argue that dispositions are detectable, so that those disposed to peace can avoid high stakes prisoner's dilemma interactions and enjoy mostly peaceful interactions. We enjoy a disposition to peace all the time in everyday interactions. We look for certain behaviours to in other persons. Are they nervous or calm? The violent are generally uneasy. Do they have a passive or aggressive stance? People who attach often look like they will. How do they react to gestures of goodwill that convey your disposition to peace and measure theirs, such as making eye contact and smiling, nodding, or attempting to shake hands? These tools are not perfect but they do generally allow us to find many peaceful interactions with a high degree of success.

Having outlined Hobbes' distinction between the two dispositions and shown that they are detectable so that those disposed to peace can establish peaceful interactions, we can now appreciate my argument. This is the contention that mostly peaceful, rather than periodically violent, interactions would develop from the outset of the state of nature because individuals would develop a disposition

to peace since they would know the long-term felicity of doing so from the outset. The evidence that individuals would know the long-term felicity to be derived from peace comes from a passage related to the first cause of violence in the state of nature: competition. Recall from my analysis of the second short argument of his first argument above that Hobbes thinks that dwellers would become violent if they desired the same thing. In the passage, Hobbes identifies activities that dwellers would engage in to foster conflicting desires and bring on violence from others. He writes that “if one [dweller] plant, sow, build, or possess a convenient Seat, others may probably be expected to come prepared with forces united” (2002: 87). If dwellers are planting and sowing, it means that they are aware of agriculture. If they are aware of agriculture, they are aware of the long-term felicity derived from doing so – prudential, felicity maximising dwellers would not sweat and toil in the present for any other reason. We can also assume that dwellers know that violence is antithetical to experiencing the long-term felicity of agriculture – it is manifest that you cannot enjoy the corn if Tom over yonder comes and beats you and takes it. So, because dwellers would recognize the long-term felicity to be derived from agriculture and that violence would be antithetical to it, they would develop a disposition to peace from the outset of the state of nature and interactions would develop mostly peacefully. The reason that this objection is problematic for Hobbes is that, as I will show in my analysis of his second and third arguments below, he thinks that the low long-term utility caused by a short life and no productive activity resulting from non-violent interactions would motivate dwellers to adopt a disposition to peace and establish a moral order based on peace and property. However, if dwellers know the long-term felicity offered by agriculture, a productive activity, and that violence would prevent the attainment of that felicity, why would they ever become violent? Why wouldn't they adopt a disposition to peace from the outset of the state of nature?

I will now consider a possible successful reply by Hobbes, that an information problem regarding equality of prudence would cause interactions to develop violently. I will first outline his information problem regarding prudence and argue it is true. I will then show why he could argue that it would cause interactions to develop violently. The information problem regarding prudence is that all dwellers are equal with regard to it, but most think that their faculty is better than average. Hobbes writes, “almost all men think they have [prudence] in a greater degree than... all men but themselves, and a few others” (2002:87). The reason that dwellers have this false information regarding the quality of their prudential faculty is that “they see their own [faculty] at hand, and other mens at a distance” (ibid.). That is, dwellers experience their own prudential faculty intimately, so they have full information regarding it, whereas they experience the prudential faculty of others from afar and so initially collect imperfect information regarding it. This information problem that Hobbes assumes with respect to prudence is true. It is an instance of the so-called Lake Wobegon effect – most people tend to believe that they are above average not only with respect to their faculties, like prudence, but also to their achievements. Let me now explain why Hobbes could argue that the information problem regarding prudence would cause interactions to develop violently. He could ask us to return to my analysis of his first and second short arguments of his first argument above. He could argue that I have misread them. In my analysis, I suggest that he argues that dwellers' biological equality leads to an expectation equal in magnitude of satisfying their respective desires which then leads them to be violent in competitive situations. He could argue that what he instead means is that each dweller is biologically equal but thinks that they are more skilled at prudence and so each develops the expectation that they will be able to satisfy their desires *in spite of the efforts of others* to satisfy their

own desires. It is this equal expectation of being able to satisfy their own desires in spite of the efforts of other dwellers to satisfy their own that leads dwellers to attack in competitive situations. They think that they can outwit other dwellers in battle and have their desires satisfied. They do not realize that adopting a disposition to violence makes state of nature interactions high stakes prisoner's dilemmas. Rather, they think that attacking another dweller, whether or not that dweller responds in kind, has a higher payoff than mutual cooperation. Further, Hobbes could argue that once dwellers adopt the disposition to violence and are violent in competitive situations, violence would be sustained for the other two reasons that he offers for violence in the state of nature; diffidence and glory. He could first mention that my analysis of diffidence is off the mark. In my analysis of Hobbes' third short argument above, I argue that he thinks that competitive violence causes diffidence, or lack of hope of satisfying desires in dwellers which then leads them to attack pre-emptively in order to avoid additional despair. He could maintain that what he actually means in this argument is that the initial lack of success in using violence in competitive situations would result in diffidence in dwellers but that they would respond by pre-emptively attacking other dwellers, thinking that they made a mistake in the initial competitive violence but that they are still more prudent and can outwit their competitor in any future battle. He could then argue that my analysis of glory is apt. In my analysis of his third short above, I contend that dwellers would attack other dwellers when those other dwellers failed to recognize their relative superiority. Hobbes would conclude that dwellers would only develop the disposition to peace after the long term had passed. That is, after seeing that their disposition to violence resulted in a short life and no productivity, dwellers would recognize that they were equal to other dwellers in prudence and that they cannot satisfy their desires through battle. They would then adopt a disposition to peace and socially contract to establish a moral order based on peace and property. Although we do not consider Hobbes argument that a state is needed to enforce the social contract, it should be noted that this reply commits us to the view that the social contract can be enforced without a state. If the disposition to peace can bring about peaceful interactions in order to negotiate a social contract to maintain peaceful interactions, then a sovereign is not needed to ensure peaceful interactions – they are already peaceful. This concludes my consideration of the two sets of game theoretic objections to the first argument of Hobbes theory of anarchic interactions.

Having completed everything relevant to my analysis and evaluation of Hobbes' first argument of his theory of anarchic interactions, I will now analyse and evaluate the second and third arguments of his theory. I will begin with the second argument. The second argument concerns the long-term effects of periodically violent state of nature interactions and dwellers' reaction to those effects. It is composed of two short arguments. The first short argument has two stages. The first stage is an inference. The premise is the common conclusion of the four short arguments of the first argument of his theory of anarchic interactions – violent interactions. From this premise, he concludes that each dweller's life would be "solitary, poor, nasty, brutish, and short" (Hobbes 2002: 89). His inference follows. Where lethal violence is spasmodic, lifespans are generally short. The second stage is also an inference. The premise is the conclusion of the first stage. From it, he infers that dwellers would develop a fear of death. The inference follows in that the threat of having one's life cut short generally brings out our innate fear death. Also, by fear of death I understand him to mean low long-term felicity. The reason I do is that fear of death is caused by the prospect of a short life and where one faces such a prospect the forecast for one's future desire satisfaction is at best bleak. The second short argument, like the first, has two stages. The first stage is an inference. The premise is the

common conclusion of the four short arguments of his first argument, or violent interactions. From this premise, he concludes that in the state of nature, “there is no place for Industry; because the fruit thereof is uncertain: and consequently no Culture of the Earth; no Navigation, nor use of the commodities that may be imported by Sea; no commodious Building; no Instruments of moving, and removing such things as require much force...” (*ibid.*). That is, dwellers would not be able to successfully engage in productive activities, such as agriculture, industry, mining, and trade. The inference follows in that if one attempted to engage in the type of activities that produce goods that satisfy desires, such as farming or industry, others could seize the products of one’s efforts by force. The second stage is also an inference. The premise is the conclusion of the first stage. From it, Hobbes concludes that dwellers would develop a “Desire of such things as are necessary to commodious living; and a Hope by their Industry to obtain them” (2002: 90). That is, dwellers would experience a desire both for those material objects that satisfy desire that are a product of productive activities and an ability to procure them by successfully engaging in productive activities. Notice that I am interpreting the phrase “things as are necessary to commodious living” as material objects that satisfy desire. My rationale is that those material objects that make life commodious, or convenient, are those material objects that satisfy desire. Hobbes inference follows in that if the primary means to satisfying desires – i.e. productive activities – is impaired, it reasonable to expect that a desire to successfully engage in them would occur. Also, by “Desire of such things as are necessary to commodious living; and a Hope by their Industry to obtain them” I understand him to mean low long-term felicity. My reason is that this desire is caused by an inability to engage in productive activities and where one cannot engage in the primary means of producing those material objects that satisfy one’s desires one’s expected level of future desire satisfaction is poor.

Figure 2

Second argument of Hobbes’ theory of anarchic interactions

1. Violent interactions → “solitary, nasty, brutish, short” life → fear of death
2. Violent interactions → inability to successfully engage in productive activities → desire both for those goods that satisfy desire and an ability to procure them by successfully engaging in productive activities

The third argument of Hobbes’ theory of anarchic interactions is that fear of death and the desire both for those goods that satisfy desire and an ability to procure them by successfully engaging in productive activities would motivate dwellers to adopt a disposition to peace which would in turn prompt them discover the laws of nature. The argument involves two steps. The first step is an inference drawn from two premises. The first premise is the conclusion of the second stage of the first short argument of his second argument, viz. fear of death. Recall that I interpreted the fear of death as low long-term felicity. The second premise is the conclusion of the second stage of the second short argument of his second argument, viz. the desire both for those goods that satisfy desire and an ability to procure them by successfully engaging in productive activities. Recall that I interpreted this desire as low long-term felicity. From these two premises – or from this one premise since I interpret the content of both premises as low long-term felicity – Hobbes concludes that dwellers would be inclined to peace (2002: 90). That is, they would be motivated adopt a disposition to peace. This

inference follows in that prudential felicity maximisers would no doubt adopt a disposition to peace once they realised doing so was in their long-term interest. The second step is an inference. The premise is the conclusion of the first step. From it, Hobbes concludes that dwellers would discover the laws of nature using reason (*ibid.*). By laws of nature, he means a moral order based on peace and property. That he does will be established in the next chapter. His inference appears to follow in that its reasonable to think that a willingness to engage in peaceful interactions would move felicity maximising dwellers to search for the best way of guaranteeing them.

Figure 3

Third argument of Hobbes' theory of anarchic interactions

Fear of death + desire both for those goods that increase desire satisfaction and an ability to procure them by successfully engaging in productive activities → adopt disposition to peace → discover laws of nature

In this section, I analysed the first, second, and third arguments of Hobbes' theory of anarchic interactions and contended that each argument is strong so his theory is accurate. This concludes the second part of my defence of his foundation for peace and property, namely, analysing the first step of his state of nature thought experiment (i.e. his definition of the state of nature) and analysing, and making arguments in favour of, the second step of his state of nature thought experiment (i.e. his theory of anarchic interactions). I will now take up the third part of my defence of his foundation, analysing, and making arguments in favour of, his foundation.

Chapter 4

Hobbes' foundation for peace and property: mutual benefit

Hobbes' foundation for peace and property is the principle of mutual benefit, or so I will argue in this chapter. Specifically, I will maintain that his foundation is that it is rational, or felicity maximising, for each dweller to socially contract to lay down their liberty to interfere in order to establish a moral order based on peace and property. My interpretation of his foundation is based on a) an argument that I infer from my interpretation of his definition of law of nature and my interpretation of his first law of nature and b) my interpretation of his second law of nature. In the next few sentences, I will i) present my interpretation of his definition of law of nature, my interpretation of his first law of nature, and the argument that I infer from them, ii) prove that the argument that I infer from them can, in fact, be inferred from them, iii) present my interpretation of his second law of nature, and iv) prove that my interpretation of his foundation can be inferred from my interpretation of his second law of nature and from the argument that I infer from my interpretations of his definition law of nature and his first law of nature. My interpretation of his definition of law of nature is that laws of nature are, or at least concern, felicity maximising moral rules, or rules of interaction. My interpretation of his first law of nature is that individuals should pursue a moral order based on the principle of peace and that such an order would include the principle of private property. The argument that I infer from these two interpretations is that a moral order based on the principles of peace and property is felicity maximising. If laws of nature are, or at least concern, felicity maximising rules of interaction, and if the first law of nature commands that dwellers establish a moral order based on peace and property, we can conclude that he thinks that a moral order based on peace and property is felicity maximising. My interpretation of his second law is that dwellers should socially contract to lay down their liberty to interfere because doing so will establish a moral order based on peace and property. I will now demonstrate how my interpretation of his foundation can be inferred from my interpretation of his second law of nature and from the argument that I infer my interpretations of his definition of law of nature and his first law of nature. If it is true that a moral order based on peace and property is felicity maximising, and if it is true that socially contracting to lay down one's liberty to interfere establishes a moral order based on peace and property, then it is true that socially contracting to lay down one's liberty to interfere to establish a moral order based on peace and property is felicity maximising. Now, the two proofs that I have undertaken in the previous few sentences – showing that my interpretation of his foundation can be inferred from my interpretation of his second law of nature and from the argument that I infer from my interpretations of his definition of law of nature and his first law of nature and that the argument that I infer from my interpretations of his definition of law of nature and his first law of nature can, in fact, be inferred from them – show that my interpretation of his foundation is ultimately derivable from my interpretations of his definition of law of nature, his first law of nature, and his second law of nature. As such, I will undertake my defence of my interpretation of his foundation in this chapter by presenting my interpretations of his definition of law of nature, his first law of nature, and his second law of nature (in that order). Also, defending my interpretation of his foundation is only my first task in this chapter. My second task is arguing in favour of his foundation. Once I have considered my interpretations of his definition of law of nature,

his first law of nature, and his second law of nature, I will argue that it is felicity maximising for state of nature dwellers to socially contract to establish a moral order based on peace and property because they could not better improve upon their state of nature felicity by socially contracting to establish another type of moral order. I will make my argument by constructing a hypothetical negotiation between dwellers wherein they attempt to come to an agreement on one of Gauthier's three provisos: the weak proviso, the proviso, and the strong proviso. The reason that I use Gauthier's three provisos is because each is a rule, or pair of rules, that accurately represent one of Hobbes' state of nature, his peace and property moral order, and the sort of moral orders that most contemporary political theorists would propose as an alternative to his order. I will show that dwellers will select the proviso (Hobbes' moral order based on peace and property) over the strong proviso (the sort of moral order that most political philosophers would recommend) as their improvement over the weak proviso (state of nature) because it is the only proviso that is purely mutually beneficial.

4.1 Definition of law of nature

Let me take up my first task, defending my interpretation of his foundation by presenting my interpretations of his definition of law of nature, his first law of nature, and his second law of nature. I will begin by advancing my interpretation of his definition of law of nature. His definition reads:

A LAW OF NATURE, (*Lex Naturalis*,) is a Precept, or generall Rule, found out by Reason, by which a man is forbidden to do, that, which is destructive of his life, or taketh away the means of preserving the same; and to omit, that, by which he thinketh it may be best preserved. (2002: 91)

I will argue that he means that laws of nature are, or at least concern, moral rules, or rules of interaction, that are felicity maximising. I will make my argument by arguing for specific readings of three phrases in the above definition. The first phrase is "a man is forbidden to do, that, which is destructive of his life." I will argue that it means that laws of nature prohibit interactions that result in the first cause of each dweller's low state of nature felicity, viz. a short life. The second phrase is "a man is forbidden to do, that, which... taketh away the means of preserving [his life]." I will argue that it means that laws of nature prohibit interactions that result in the second cause of each dweller's low state of nature felicity, viz. no productivity. The third phrase is "omit, that, by which he thinketh [his life] may be best preserved." I will argue that it means that laws of nature do not prohibit any interactions not already prohibited by the first and second phrases that any dweller thinks will maximise her felicity. So, with these specific readings of the three phrases, I am arguing that laws of nature are, or at least concern, moral rules, or rules of interaction, that are felicity maximising because they remove the two causes of each dweller's low state of nature felicity and leave them otherwise free to engage in interactions they think will maximise felicity.

I will begin by defending my reading of the first phrase, "a man is forbidden to do, that, which is destructive of his life", that it means that laws of nature proscribe interactions that result in the first cause of each dwellers low state of nature felicity, viz. a short life. My defence turns on my understanding of the sub-phrase "that, which is destructive of his life." I understand it to mean 'that

which ends life.’ This is important because if people are permitted to engage in interactions that end life and they do periodically (as is the case in Hobbes’ state of nature), the probability of anyone living only a short amount of time will be quite high. And if the laws of nature prohibit actions that end life, it means that the probability of anyone living a short life will be quite low (or, inversely, anyone’s chances of living long life will be quite high).

I will now defend my reading of the second phrase, “a man is forbidden to do, that, which... taketh away the means of preserving [his life],” that it means that laws of nature prohibit interactions whose outcome is the second cause of each dweller’s low state of nature felicity, viz. no productivity. My defence will consist of arguing that the sub-phrase “means of preserving [his life]” means material objects that satisfy one’s desires over time and then showing why this proves that the second phrase means that laws of nature prohibit interactions that obstruct production. I will now argue that the sub-phrase means material objects that satisfy one’s desires over. Recall from §3.1 that I assume that self-preservation means maximising one’s desire satisfaction, or felicity. As such, the means of preserving one’s life must be the means of maximising one’s felicity. Recall from §2.3 that the means by which one maximises felicity is by acting now to obtain material objects that will satisfy one’s desires over time. If the means of preserving one’s life is the means to maximising felicity, and if the means to maximising felicity are material objects that satisfy one’s desires over time, then the means to preserving one’s life are material objects that satisfy one’s desires over time. Now that I have established that the sub-phrase “means of preserving [his life]” means material objects that satisfy one’s desires over time, I can prove that the second phrase means that laws of nature prohibit those interactions that prevent production. If material objects that satisfy one’s desires over time are the means of preserving oneself, and if laws of nature prevent those interactions that obstruct obtaining the means of preserving oneself, then the laws of nature prevent those interactions that prevent one from obtaining material objects that satisfy one’s desires over time. This is significant with regards to production because productive actions are the most efficient means for obtaining material objects. So, because laws of nature prohibit interactions that prevent one from obtaining material objects that satisfy one’s desires over time, they also prevent interactions that obstruct production.

Finally, I will defend my reading of the third phrase, “omit, that, by which he thinketh [his life] may be best preserved,” that it means that laws of nature do not prohibit any interactions not already prohibited by the first and second phrases that any dweller thinks will maximise her felicity. To make my argument, I will divide the phrase into three sub-phrases and explain the meaning of each. The sub-phrases are “omit,” “that, by which,” and “he thinketh [his life] may be best preserved.” By “omit,” he means not proscribe. By “that, by which,” he means any interaction not already prohibited by the first and second phrases. It might instead be thought that he means any interaction at all. But this reading, if accurate, would be problematic in that it would make the laws of nature indistinguishable from the right of nature. And what would be the purpose of the laws of nature if they allowed dwellers to engage in the very interactions that make Hobbes’ state of nature such a nasty place? By “he thinketh [his life] may be best preserved,” he means whatever any dweller thinks will maximise her felicity. As I asked the reader to recall in my defence of the second phrase, I assume that self-preservation means maximising felicity.

4.2 First law of nature

I will now put forward my interpretation of Hobbes' first law of nature. His first law reads, "*to seek Peace, and follow it*" (2002: 92). I will argue that he means that dwellers should attempt to establish a moral order based on the principle of peace, or non-interference, since they lack one. I will also argue that he thinks that the moral based on peace that dwellers should attempt to establish would include the principle of private property, or do not interfere with other's use of things in their possession. I will make the former argument first followed by the latter. I will make the former argument by dividing the law into two phrases and establishing the meaning of each. The two phrases are a) "to seek peace" and b) "follow it." By "to seek peace," he means 'lacking a moral order based on peace.' The evidence that he means as much is that an established, though archaic, definition of the phrase 'to seek.' That definition is lacking or not yet found (Dictionary 2005). By "follow it," he means 'pursue it.' The evidence that he means 'pursue it' is an archaic definition of 'follow.' That definition is strive after or aim at (*ibid.*). So, because he means 'lacking a moral order based on peace' by "to seek peace" and 'pursue it' by "follow it," we can conclude that the first law means lacking a moral order based on peace, dwellers should attempt to establish one.

4.2.1 Peace includes property

I will now make the latter argument, that he thinks that the moral order based on peace that dwellers should attempt to establish would include the principle of private property. I will make the argument by presenting evidence that confirms it. The evidence that confirms it is one of the two reasons he offers at the beginning of chapter XV for why a sovereign is needed to enforce the social contract to abandon the liberty to interfere. He argues that a sovereign is needed to "make good that Propriety, which by mutuall Contract men acquire, in recompence of the universal Right [i.e. the liberty to interfere] they abandon" (2002: 101). To understand why his argument confirms the latter argument, we must make sense of it. To make sense of it, we need to appreciate the purpose of socially contracting to abandon the liberty to interfere. As I will show in §4.3.3 below, Hobbes thinks dwellers should socially contract to abandon the liberty to interfere because doing so will establish the moral order based on the principle of peace. So, the purpose of socially contracting to abandon the liberty to interfere is establishing the moral order based on the principle of peace. Understanding the purpose of socially contracting to abandon the liberty to interfere, we can make sense of his argument. He is contending that establishing the moral order based on peace would establish property rights in those material objects individuals acquire by contracting with each other. Put another way, establishing the moral order based on peace would establish the principle of property. Having made sense of his argument, I can now show how his argument verifies the latter argument. If Hobbes states that dwellers should attempt to establish a moral based on the principle of peace, and if he later states that establishing such an order will establish the principle of property, then he thinks that the moral order based on the principle of peace dwellers should attempt to establish would include the principle of property.

It should be noted that my confirmation of the latter argument in the preceding paragraph leaves two important questions unanswered. First, how does socially contracting to abandon the liberty to interfere establish peace? Second, how does establishing a moral order based on the principle peace

establish the principle of property? The first question will be answered in §4.3.4 below, so I will forgo answering it here. In terms of the second question, Hobbes' argument considered in preceding paragraph suggests the answer, viz. peace entails property. If individuals cannot interfere with each other, then they cannot obstruct each other's use of the material objects they acquire by trade. If Jones trades for a field to plant corn, Smith cannot successfully plant cotton on that same field without physically preventing Jones from approaching to plant corn as she had planned. Problematically, Hobbes' argument suggests that he does not seem to capture the full extent to which peace entails property. His argument suggests that he thinks that peace necessitates property only over material objects acquired by trade. This suggests that he does not appreciate that peace also establishes *initial* ownership rights over material objects. In fact, we know he does not appreciate this point because he writes elsewhere that the initial distribution of land is to be carried out by the sovereign (2002: 171). Peace establishes initial ownership rights over material objects if the material objects are acquired in accordance with the principle of first use. The principle of first use, according to Narveson, is "She who gets there first *and* commences to use [the material object], in ways that require ongoing access to it" (Narveson 2002: 119). Under this principle, ongoing use must be socially verifiable and not pie in the sky (2002: 119-20). Putting up rock formations at the corners of a field that you are hunting wild boar on is one thing, claiming ownership of the Rocky Mountains because you saw them first is another. The reason that peace establishes property rights over material objects acquired in accordance with the principle of first use is because it is not possible to use something someone else is already using in a continual way without hindering their regular activities. If Harris clears a thicket to plant hemp and Thompson comes the next day and plants canola, Thompson is impeding Harris' recurrent use.

Hobbes' failure to recognise that peace establishes initial ownership rights over material objects acquired in accordance with the principle of first use means that the moral order based on peace he thinks dwellers should attempt to establish would include a principle of property that is wanting. I will strengthen his conception of the moral order dwellers should attempt to pursue by assuming that he thinks that peace implies initial property rights over those objects acquired in accordance with the principle of first use.

4.3 Second law of nature

I will now advance my interpretation of Hobbes' second law of nature. His second law reads:

That a man be willing, when others are so too, as farre-forth, as for Peace, and defence of himselfe 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 for other men against himself.
(2002: 92)

My interpretation is that each dweller should socially contract to forgo their absolute liberty to use all material objects to the extent that it interferes with other dwellers because doing so will establish a moral order based on the principles of peace and property. Put more briefly, he thinks that dwellers should socially contract to forgo their liberty to interfere because doing so will establish a moral order

based on the principles of peace and property. I will show that my interpretation is accurate by undertaking a proof. The proof will involve me carrying out four tasks: i) I will divide his second law into four phrases, ii) I will explain how the four phrases fit together, iii) I will argue for the meaning of each phrase, and iv) I will put the interpretations of each phrase argued for in task (iii) together according to the criteria laid out in task (ii). The end result of the proof will be my interpretation as stated in the first sentence of this paragraph. In the remainder of this paragraph, I will undertake tasks (i) and (ii). I will also outline how I will simultaneously perform tasks (iii) and (iv) in the remainder of §4.3. I divide his law into these four phrase: 1) “*That a man be willing, when others are so too,*” 2) “*as farre-forth, as for Peace, and defence of himselfe he shall think it necessary,*” 3) “*to lay down this right to all things,*” and 4) “*and be contented with so much liberty against other men, as he would allow for other men against himself.*” The phrases fit together in the following way: phrase (3) outlines the main action dwellers are to perform. Phrases (4) and (1) stipulate the two conditions under which they are to perform that action. Phrase (2) identifies the purpose for undertaking the action mandated by phrase (3) when the conditions specified by phrases (4) and (1) are met. Let me now explain how I will simultaneously perform tasks (iii) and (iv) in the remainder of the section. I will begin by putting forward my interpretations of phrases (3), (4), and (1). I will argue that phrase (3) means that dwellers should forgo their absolute liberty to use all material objects. I will argue that phrase (4) effectively means ‘to the extent that their liberty interferes with others.’ I will next argue that phrase (1) means the principle of social contract. Having put forward my interpretations of phrases (3), (4), (1), I will conclude that since phrase (4) and (1) are conditions on the action mandated in phrase (3), the combined meaning of phrases (3), (4), and (1) is that dwellers should socially contract to forgo their liberty to use all material objects to the extent that their liberty interferes with others. I will next advance my interpretation of phrase (2). I will argue that that we can read him as meaning that the purpose of performing the action mandated by the combined meaning of phrases (3), (4), and (1) is to establish a moral order based on the principles of peace and property. Given that I will have already shown that the combined meaning of phrases (3), (4), and (1) is that dwellers should socially contract to forgo their absolute liberty to use all material objects to the extent that their liberty interferes with others, I will conclude that the combined meaning of all four phrases is that each dweller should socially contract to forgo their absolute liberty to use all material objects to the extent that it interferes with other dwellers because doing so will establish a moral order based on the principles of peace and property.

4.3.1 Forgoing the absolute liberty to use all material objects

I will now put forward my interpretations of phrases (3), (4), and (1), beginning with phrase (3). My interpretation of phrase (3), “*to lay down this right to all things,*” is that it means that each dweller should forgo their absolute liberty to use all material objects. I will show my interpretation is correct by establishing the meaning of the two terms in the phrase: a) lay down and b) right to all things. I will start with the term (b) and argue that by it Hobbes means the absolute liberty to use (or not use) all material objects. Further, because term (b) is heretofore unmentioned, I will explain its place in his larger political theory. I will then argue that by term (a) he means permanently forgo the absolute liberty to interfere.

Let me begin with term (b), right to all things. I will establish its meaning by determining the meaning of the two words that compose it, right and things. He defines right as "liberty to do, or to forbear" (2002: 91). That is, as I argued in §3.1, a "sphere of action wherein the right holder enjoys the unrestricted choice to do or not do those actions included in the sphere." By things, he means material objects. He appears to use the word in the sense in which he uses it in his definition of felicity in §2.3. The right to all things, then, is the freedom to act on or not act on any physical object capable of satisfying desire, "even one another's body" (2002: 91). Put another way, it is a sphere of action wherein the right holder enjoys the unrestricted choice to do or not do all those actions included in the sphere and the sphere is composed of all possible actions on all extended objects.

I will now explain the place of the right to all things in his political theory. I do so because I have not yet discussed it, and, as such, it leaves the impression that Hobbes is introducing the concept at an important point in his theory without prior mention. This picture, however, is inaccurate. The place of the right to all things in his political theory is that it is another term for the right of nature. That is, he thinks of the right to all things and the right of nature as being one and the same. We know this because he attributes to the right to all things the same definition as the right of nature in the sentence directly following his statement of the second law. In that sentence, he refers to the right of all things as "this Right, of doing any thing he liketh..." (2002: 92). That is, that liberty to do or not do any action that satisfies desire. Recall from §3.1 that he defines the right of nature as doing whatever one thinks best satisfies desire. Also, that he equates the two concepts offers us a further and important insight into his thought. That insight is that he assumes that all actions are performed on material objects. If liberty to do whatever one desires is the same as the liberty to use whatever material objects one desires, then all actions must be performed on material objects. We should note that his reasoning is not entirely accurate. We do perform actions, albeit not many, that have no material object as their focus. For example, we can fart or wave our arms in the air or scream at nothing in particular simply because we feel like it. But these cases are few so that Hobbes' assumption is reasonable.

I will now explain what he means by term (a), 'lay down.' Explaining what he means requires analysing his definition of laying down a right. It reads:

To lay downe a mans Right to any thing, is to divest himselfe of the Liberty, of hindring another of the benefit of his own Right to the same. For he that renounceth, or passeth away his Right, giveth not to any other man a Right which he had not before; because there is nothing to which every man had not Right by Nature: but onely standeth out of his way, that he may enjoy his own originall Right, without hindrance from him; not without hindrance from another.
(*ibid.*)

We should note two things. First, when a dweller lays down a right to a thing it means that she permanently forgoes her liberty to use (or not use) some material object. Second, Hobbes focuses our attention on the particular social effect of doing so in the state of nature. He contends that when a dweller forgoes her liberty use (or not use) a material object in the state of nature she is simply giving

up her ability to interfere with other dwellers' use of that same object. This is because each dweller, enjoying the right to all things, already has the liberty to use (or not use) all material objects. In another state of affairs, by contrast, one where each dweller does not possess the right to all things and one where the principle of peace, or general non-interference, is enforced, the social effect of a dweller laying down her right to a thing might be that another dweller, or some other dwellers, gain the liberty to use some object they previously could not. The reason Hobbes draws our attention to the social effect of laying down a right in the state of nature is to bring to light its value for establishing the principle of peace, or general non-interference. If each dweller forgoes her liberty to use those objects that both she and others both wish to use, then general non-interference will be established between dwellers.

We have so far established what Hobbes means by both right to all things and lay down. As such, we can conclude that by phrase (3), he means that each dweller should forgo her absolute liberty to use all material objects (including others' bodies).

4.3.2 The limit on forgoing the absolute liberty to use all material objects

I will now put forward my interpretation of phrase (4), "*be contented with so much liberty against other men, as he would allow for other men against himself,*" that it effectively means 'to the extent that their liberty interferes with others.' In phrase (4), Hobbes is arguing that each dweller should leave herself as much of her absolute liberty to use all material objects in so far as it interferes with other dwellers as she would allow them to keep against her. This raises the question, how much of other dwellers' absolute liberty to use all material objects in so far as it interferes with them does he think each dweller would allow other dwellers to keep? Since he identifies the cause of each dweller's low long-term felicity in the state of nature as periodic violence from others and since he makes no mention of the benefits of being able to use one's own absolute liberty to use all material objects in so far as it interferes with other dwellers in certain ways in spite of the cost of letting others use their absolute liberty to use all material objects in so far as it interferes with them in those same ways, we can infer that he is arguing that dwellers would allow their contemporaries to retain very little, if any, liberty to use all material objects in so far as it interferes with them. So, in effect, he is arguing that dwellers should forgo their liberty to use all material objects to the extent that it interferes with other dwellers. Since we have to fit this phrase together with phrase (3) somehow – they are part of the same statement – in a way that makes sense, I will shorten the phrase to "to the extent that their liberty interferes with others." That will allow it fit nicely with phrase (3). If we were to simply combine our unshortened interpretation of phrase (3) with our interpretation of phrase (4), the result would be the longwinded statement "dwellers should forgo their absolute liberty to use all material object and they should forgo their absolute liberty to use all material objects to the extend that their liberty interferes with others."

4.3.3 By social contract

I will now present my interpretative argument concerning phrase (1), "*That a man be willing, when others are so too,*" that it means the principle of social contract. My argument is simply that it is quite

clear that he is stating that an individual should agree to do x, if and only if everyone else, or nearly everyone else, agrees to do it too.

We have so far considered my interpretations of phrases (3), (4), and (1). Recall from the introduction that I asserted that phrases (4) and (1) are conditions on the actions mandated by phrase (3). We can conclude, then, that the three phrases together mean that each dweller should socially contract to forgo her absolute liberty to use all material objects to the extent that it interferes with other dwellers.

4.3.4 To establish a moral order based peace and property

Let me now propound my interpretation of phrase (2) “*as farre-forth, as for Peace, and defence of himselfe he shall think it necessary,*” that we can read him as meaning that the purpose of performing the action mandated by the combined meaning of phrases (3), (4), and (1) is to establish a moral order based on the principles of peace and property. In phrase (2), he is clearly pointing out that the end of performing the action mandated by the combined meaning of phrases (3), (4), and (1) is to establish a moral order based on the principle of peace, or the rule of non-interference, and hence preserve one’s self, or maximise one’s felicity. That performing the action mandated by the combined meaning of phrases (3), (4), and (1) would establish a moral order based on the principle of peace is fairly obvious. If each dweller socially contracts to forgo her absolute liberty to use all material objects to the extent that it interferes with others, then it will be recognised that no one should interfere with anyone else.

To show that we can read him stating that the purpose of performing the action mandated by the combined meaning of phrases (3), (4), and (1) is to establish a moral order based on the principles of peace and property, I have to argue that we can read property into, and felicity maximisation out of, phrase (2). We can read property into phrase (2) because, as I established in §4.2.1, he thinks that the moral order based on the principle of peace would include the principle of property. We can read felicity maximisation out of phrase (2) because it is superfluous. The reader can already infer that the moral order based on the principle of peace is felicity maximising based on his definition of law of nature and his first law of nature. As was argued in §4.1, his definition of law of nature shows that laws of nature concern moral orders that maximise felicity by prohibiting those actions that result in the two causes of each dweller’s low state of nature felicity and by otherwise leaving them free to engage in any other interactions they wish. As was argued in §4.2, his first law of nature mandates that dwellers should attempt to establish a moral order based on the principle of peace. From these two aspects of his theory of natural law, the reader already knows that a moral order based on the principle of peace would maximise felicity because it would prohibit those actions that result in the two causes of each dweller’s low state of nature felicity and otherwise leave them free to engage in any other interactions they wish.

My demonstration of my interpretation of phrase (2) is now complete. Given that we can read phrase (2) as meaning that the purpose of performing the action mandated by the combined meaning of phrases (3), (4), and (1) is to establish a moral order based on the principles of peace and property, and that we have shown that the combined meaning of phrases (3), (4), and (1) is that each dweller should socially contract to forgo her absolute liberty to use all material objects to the extent that it

interferes with other dwellers, we can conclude that the combined meaning of all four phrases is each dweller should socially contract to forgo their absolute liberty to use all material objects to the extent that it interferes with other dwellers because doing so will establish a moral order based on the principles of peace and property. Or more tersely, each dweller should socially contract to forgo their liberty to interfere because doing so will establish a moral order based on the principles of peace and property. With the combined meaning of the four phrases established, I conclude that my interpretation of his second law is accurate.

4.4 Arguing in favour of Hobbes' foundation

So far in this chapter, I have completed the first task identified in the opening paragraph, showing that my interpretation of Hobbes' foundation is accurate by presenting my interpretations of his definition of law of nature, his first law of nature, and his second law of nature. In this section, I will take up the second task identified in the opening paragraph, arguing in favour of his foundation (or, at least, his foundation as I interpret it). I will argue that it is felicity maximising for state of nature dwellers to socially contract to establish a moral order based on the principles of peace and property because they could not better improve on their low state of nature felicity by socially contracting to establish another type of moral order. I will make my argument by presenting a hypothetical negotiation between dwellers wherein they choose between two of the provisos Gauthier discusses in chapter VII of his classic text *Morals by Agreement* – the proviso and the strong proviso – as their best improvement over the third proviso he discusses – the weak proviso (1986: 200-08). The reason that I use his three provisos for the hypothetical negotiation is because each is a rule of interaction, or pair of rules of interaction, that neatly captures one of the baseline of Hobbes' foundation (his state of nature), his purported improvement (his moral order based on peace and property), and the sort of moral order most contemporary political philosophers would propose as an alternative to his moral order (viz. one based on general non-interference with some redistributive requirement). I will present the hypothetical negotiation between dwellers in three steps. In the first step, I will show which proviso is equivalent to which of Hobbes' state of nature, his moral order based on peace and property, and the sort of moral order most contemporary political philosophers would propose as alternatives to his moral order. In particular, I will argue that the weak proviso is a rule of interaction that effectively captures his state of nature, that the proviso is a rule of interaction that encapsulates his moral order based on peace and property, and that the strong proviso is a pair of rules of interaction that represents the sort of moral order most contemporary political philosophers would propose as alternatives to his order. In the second step, I will present each dweller's evaluation of each proviso. Their evaluation will consist of determining which type of situation each rule of proviso governs, figuring out which outcome each rule mandates for those interactions that occur in the type of situation it governs, and assigning each rule a felicity per interaction function for the type of situation it governs. In the third step, I will show why dwellers will socially contract to establish the proviso (i.e. Hobbes' moral order based on peace and property) instead the strong proviso (i.e. the sort of moral order most political theorists would present as an alternative to Hobbes' order) as their best improvement over the weak proviso (i.e. the state of nature). The reason is that the proviso is purely mutually beneficial while the strong proviso is only partly mutually beneficial. What I mean by this reason will be explained in greater detail during the third step.

Let me begin with the first step of my presentation of the bargain, showing which of the provisos are equivalent to which of Hobbes' state of nature, his moral order based on peace and property, and the sort of moral orders most contemporary political philosophers would propose as alternatives to his moral order. I will begin with my argument that the weak proviso effectively captures Hobbes' state of nature. The weak proviso reads, "in order to better one's own situation, one may worsen that of others" (1986: 206). To show how this proviso effectively captures Hobbes' state of nature, we must make sense of it. To make sense of it, we need to know what Gauthier means by bettering and worsening. He defines bettering as follows: "one situation is better for some person than another, if and only if it affords him a greater expected utility" (1986: 203). As such, we can assume that by worsening he means that one situation is worse for a person than another if and only if it affords him a lower expected utility. Knowing what Gauthier means by bettering and worsening, we can determine the meaning of the weak proviso. It states, in effect, that in acting to improve one's expected utility, one may engage in any interaction, including those that lower the expected utility of the person with whom one interacts. Put another way, in acting to maximise one's felicity, one may partake in any interaction, including those that interfere with others (i.e. lower their long-term felicity). Knowing what the weak proviso means, I can now show how it captures Hobbes' state of nature. The weak proviso effectively captures the state of nature because it, like the right of nature possessed by all dwellers, permits all interactions, including interference interactions, in the pursuit of felicity maximisation. Recall from §3.1 that the right of nature is the liberty to do or not do whatever actions a dweller thinks will best satisfy her desires (i.e. maximise her felicity). And the liberty to do or not do whatever actions a dweller thinks would best maximise her felicity includes the liberty to partake in or not partake in any interactions a dweller thinks would best maximise her felicity, including those interactions that lower another dweller's felicity (i.e. interference interactions).

I will now present my argument that the proviso encapsulates Hobbes' moral order based on peace and property. The proviso "prohibits bettering one's situation through interaction that worsens the situation of another" (1986: 205). That is, you can increase your expected utility by any means you wish save interactions that lower the expected utility of another (i.e. interference interactions). This proviso encapsulates the Hobbesian moral order based on peace and property because the Hobbesian order allows dwellers to exercise their right of nature, or their liberty to do or not any action that maximises their felicity, unless it interferes with others.

I will now argue that the strong proviso is a pair of rules of interaction that represent the sort of moral order most contemporary political philosophers would propose as an alternative to Hobbes' moral order based on peace and property. The reason that the strong proviso represents the sort of moral order most contemporary political philosophers would propose as an alternative is because it represents those moral orders based on general non-interference and some kind of egalitarian redistribution requirement, such as Marx's needs principle – from each according to ability to each according to need (Marx 2006: 215) – or Rawls' difference principle – differences in inequality should benefit the least well-off attached to offices and positions open all under conditions of fair equality of opportunity (Kymlicka 2002: 56) – and most contemporary political philosophers support such an order. I will now demonstrate that the strong proviso represents those moral orders based on general non-interference and some kind of egalitarian redistribution requirement. The key to the demonstration is analysing the strong proviso's stated command. Its stated command reads, "as a

condition of bettering one's own situation, one must better that of others" (1986: 206). That is, if a person acts to increase her expected utility, she must also act to increase the expected utility of some other persons within a reasonable period of time according to some requirement that outlines how, and to what extent, their expected utility is to be improved. Put another way, if an individual acts to maximise her felicity, she must also improve the felicity of some other individuals within an acceptable time period in keeping with some criteria that stipulates the means by which, and the amount by which, their felicity is to be increased. Clearly, the stated command of the strong proviso applies in addition to the stated command of the proviso. If it did not, actions undertaken to increase the felicity of others according to the requirement – whatever it is – would be offset by interference interactions (i.e. interactions that lower expected utility). If Jones betters Thompson at t_1 according to requirement x , Thompson could then be worsened at t_2 by Smith, negating the bettering. Because the strong proviso includes both its stated command and the command of the proviso, we can restate it as follows:

Rule (1): you can maximise your felicity by any means save interference (i.e. lowering the felicity of others through interaction)

Rule (2): if you act to maximise your felicity, you must also increase the felicity of some others according to requirement x

Rule (1) requires general non-interference and rule (2) requires some kind of egalitarian redistribution.⁴ As such, we can conclude that the strong proviso represents those moral orders based on general non-interference and some kind of egalitarian redistribution requirement.

Having demonstrated which proviso is equivalent to which of Hobbes' state of nature, his moral order based on peace and property, and the sort of moral order most contemporary political philosophers would propose as an alternative to his moral order, the first step of my presentation of the hypothetical negotiation between dwellers is complete. I will now undertake the second step of my presentation, putting forward each dweller's evaluation of each proviso. To appreciate each dweller's evaluation of each proviso, we must understand their method of evaluation. To understand their method of evaluation, we must be aware of how each dweller perceives each future interaction. Each dweller will view each future interaction as the state of nature prisoner's dilemma game outlined in §3.2. Recall that this game has the following four possible outcomes for each dweller: a) interference/non-interference, b) non-interference/non-interference, c) interference/interference, and d) non-interference/interference, where each dweller's play is listed to the left of the oblique stroke and their opponent's to the right. Recall as well that each dweller attributes the following (non-interpersonal comparable) felicity functions to each possible outcome: a: 9, b: 7, c: -5, d: -10. Now that we are aware of how each dweller's perceives each future interaction, we can understand their method of evaluation. Each dweller will evaluate each proviso by assessing each of its rules. They will assess each rule in three stages. First, they will determine which type of situation it governs. Second, they will establish which outcome it mandates for each interaction that occurs in that type of

⁴ I should note that rule (2) of the strong proviso captures not just egalitarian redistributive requirements, but also inegalitarian redistribution requirements, e.g. where peasants must maintain a clergy class through a tithe.

situation. The outcome will be one of the four outcomes of the state of nature prisoner's dilemma game, either a) interference/non-interference, b) non-interference/non-interference, c) interference/interference, or d) non-interference/interference. Third, they will determine the 'felicity per interaction function' for each rule. This function represents the felicity associated with the outcome that each rule mandates for each interaction that occurs in the type of situation it governs, i.e. a: 9, b: 7, c: -5, d: -10.

Understanding each dweller's method of evaluation, we can now appreciate each dweller's evaluation of each proviso. Let me begin with each dweller's evaluation of the weak proviso. The weak proviso is each dweller's baseline and thus each dweller has plenty of experience with it. They know that its lone rule governs the type of situation where an individual is maximising her felicity. They know this because it describes what an individual may do in such situations. They are aware that its rule permits all four types of outcomes for those interactions that occur in such situations. However, they also know that those interactions that do occur in such situations far too often result in their third most preferred outcome of interference/interference. The reason they do is because most dwellers have until now overestimated the relative quality of their prudential faculty. They foolishly thought that they were wiser than other dwellers and so they figured they could get their way by using the shrewdest types of interference. This error in judgement led them to the false belief that interference/interference offered them a higher felicity than non-interference/non-interference. Although, they will not make this arrogant mistake again, they consider remaining with the lone rule of the weak proviso an endorsement of their mistake and so they assign it a felicity per interaction function of -5. I will now present each dweller's evaluation of the proviso. Each dweller will recognise that its single rule, like the lone rule of the weak proviso, governs the type of situation where an individual is maximising her felicity. They will recognise that it governs such situations because it describes what individuals may not do in such situations. They will further reason that since it prohibits interference in such situations, any interactions that occur will have the outcome of non-interference/non-interference. As such, they assign it a felicity per interaction function of 7. I will now put forward each dweller's evaluation of the strong proviso, beginning with their assessment of rule (1). Each dweller would determine that rule (1), like both the lone rule of the weak proviso and the single rule of the proviso, governs the type of situation where an individual is maximising her felicity. They would come to this conclusion because rule (1) outlines what an individual may or may not do in such situations. Since rule (1) prohibits interference in such situations, they would infer that it mandates the outcome of non-interference/non-interference for those interactions that occur in them. As such, they would assign it a felicity per interaction function of 7. I will now outline each dweller's assessment of rule (2). Each dweller would recognise that rule (2) governs the type of situation where an individual *has already* acted to maximise her felicity. They would recognise that it governs such situations because it mandates that a redistributive interaction take place in them. For those interactions that occur in such situations (viz. those rule (2) mandates), not all dwellers would have the same finding regarding which outcome rule (2) mandates. Those who are to have their felicity increased will recognise that rule (2) mandates their most preferred outcome of interference/non-interference. As such, they would assign rule (2) a felicity per interaction function of 9. Those who are to increase the felicity of others will determine that rule (2) would mandate their least preferred outcome of non-interference/interference. Based on this finding, they would assign rule (2) a felicity per interaction function of -10.

Now that I have presented each dweller's evaluation of each proviso, the second step of my presentation of the hypothetical negotiation between dwellers is complete. I will now perform the third step of my presentation, showing why dwellers will socially contract to establish the proviso instead of the strong proviso as their best improvement over the weak proviso. To understand why dwellers will agree to the proviso instead of the strong proviso, we need to first appreciate both each dweller's method for comparing the three provisos and each dweller's negotiation strategy. I will begin by describing each dweller's method of comparing the provisos, which involves two tasks. In the first task, they will compare the single rule of the proviso and rule (1) of the strong proviso to the lone rule of the weak proviso. They will do so because each of these rules governs the same type of situation, wherein an individual acts to maximise her felicity. Their second task will involve comparing an assumed second rule of the proviso and rule (2) of the strong proviso to an assumed second rule of the weak proviso for the type of situation where an individual has already acted to maximise her felicity. Each dweller will assume a second rule for both the proviso and the weak proviso so that a comparison can take place for such situations. They will reason that if a comparison for such situations does not take place, no agreement could be reached regarding either the strong proviso or the proviso and they will be stuck in the sad state of affairs offered by the weak proviso. The assumed second rule of the proviso will be that inference is prohibited in interactions that occur in such situations. Each dweller will assume this rule because it mirrors the single rule of the proviso. The assumed second rule of the weak proviso will be that all outcomes are permitted in interactions that occur in such situations. Each dweller will assume this rule because it mirrors the lone rule of the weak proviso. Also, dwellers would reason that since every dweller is a prudential felicity maximiser, actions (including interactions) will only take place in situations where an individual acts to maximise her felicity unless there is a rule mandating otherwise. Therefore, they will assign both the assumed second rule of the proviso and the assumed second rule of the weak proviso a felicity per interaction function of 0. I will now present each dweller's negotiation strategy. For both comparisons, each dweller will attempt to agree to that rule of the proviso or strong proviso that offers them the greatest improvement (i.e. greatest gain in felicity per interaction) over the rule of the weak proviso. They will not be willing to agree to any rule of either the proviso or the strong proviso that would make them worse off (i.e. a lower felicity per interaction). Ostensibly, this strategy could lead to agreement on a hybrid proviso with one rule from the proviso for one type of situation and another from the strong proviso for the other type of situation. However, as I show presently, that will not occur.

With an appreciation of both each dweller's method for comparing the three provisos and each dweller's negotiation strategy, we can now understand why dwellers would agree to the proviso instead of the strong proviso as their improvement over the weak proviso. I will show why dwellers would agree to this by presenting the results of both of the component tasks of each dweller's method for comparing the three provisos. In the first task, each dweller would find that both the single rule of the proviso and rule (1) of the strong proviso offered them an increase in their utility per interaction function of 12 over the lone rule of the weak proviso. As such, each dweller would be willing to agree to either the single rule of the proviso or rule (1) of the strong proviso as their improvement over the lone rule of the weak proviso. Further, they would be indifferent between which was agreed to. In the second task, each dweller would find that the assumed second rule of the proviso offered them no change in their utility per interaction function over the assumed second rule of the weak proviso. For

rule (2) of the strong proviso, those who increased the felicity of others under rule (2) would find that it offered them a decrease in their felicity per interaction of 10 as compared to the assumed second rule of the weak proviso. As such, they would be willing to agree to the assumed second rule of the proviso but not rule (2) of the strong proviso. Those who would have their felicity increased under rule (2) of the strong proviso would find that it would offer them an increase in their felicity per interaction of 9 over the assumed second rule of the weak proviso. As such, they would be willing to agree to both the assumed second rule of the proviso and rule (2) of the strong proviso, but would prefer the latter. Therefore, the reason that dwellers would agree to the proviso instead of the strong proviso as their improvement over the weak proviso is that the proviso is purely mutually beneficial and the strong proviso is only partly mutually beneficial. By the proviso is purely mutually beneficial, I mean that each dweller is at least as well off under both rules of the proviso as under both rules of the weak proviso. By the strong proviso is partly mutually beneficial, I mean that while each dweller is at least as well off under the first rule of the strong proviso as under the lone rule of the weak proviso, at least one dweller will be worse off under the second rule of the strong proviso than they would be under the assumed second rule of the weak proviso. And this fact will destroy the unanimous consent which is an essential feature of the social contract.

Let me now consider a strong objection that most contemporary political philosophers could make to the third step of my presentation of the hypothetical negotiation between dwellers, why dwellers would agree to the proviso instead of the strong proviso as their improvement over the weak proviso. They could argue that my presentation of the results of the second task of each dweller's method for comparing the three provisos wrongly assumes that it is not in the interest of those who better others under rule (2) of the strong proviso to agree to that rule. They would argue that although those who would better others must accept a decrease in their felicity by engaging in redistributive interactions in those situations after they act to maximise their felicity, this would more than be made up for over the long-term by the increase in felicity they would experience in those interactions that occur in those situations where they act to maximise their felicity. They would experience an increase in felicity in such interactions because, by helping those whose felicity is increased, they would help them realise their potential and help them become better contributors to society. The effect of this would be an increase in overall production. With an increase in overall production, the felicity that resulted from those interactions that occur in situations in which an individual acts to maximise her felicity would increase significantly.

There are two problems with the possible objection most contemporary political theorists could make. First, and most fundamentally, it ignores the nature of incentives. Generally, if you reward something, you encourage it. If do not reward something, you do not encourage it. This is not to say that providing alms to the less fortunate is never justified because it will ensure they remain in their condition indefinitely. Certainly there are many instances of people turning their lives around as a result of the generosity of others. What it is instead to say is that if you create a guarantee that if someone reaches a given relative position they are certain to receive alms, you encourage those who are in that position to stay there (and quite possibly you encourage some others to join them). Second, the reply ignores what we know about increasing productive output. Generally, those who would be required to increase the felicity of others under rule (2) of the strong proviso have more wealth because they put their resources to more efficient uses, that is, uses that create more wealth (e.g.

savings and investment, successful business ventures, etc.). Similarly, those who would have their felicity increased under rule (2) of the strong proviso generally have less wealth because they put their talents and resources towards less efficient uses (e.g. consumption, unsuccessful business ventures). If we wish to increase the marginal gains that each individual receives from the production process, we should leave who use their resources in the most efficient way alone. If redistribute their resources to those who use their own resources less efficiently, we would expect, contra the objection, for each individual's marginal change in income to be less than it otherwise would be. It might not fall, but it would not increase at the same rate.

Chapter 5

Conclusion

In this thesis, I defended Hobbes' foundation. My defence centred on the idea that a proper justification of his foundation must include a defence of his theory of anarchic interactions and his theory of human nature since his foundation is based upon both. Therefore, I made three considerable arguments. The first such argument concerned his theory of human nature. I argued that each of the five assumptions that compose it are true. The second argument regarded his theory of anarchic interactions. I contended that his theory is accurate because each of its three major arguments are strong. I also considered two sets of game theoretic objections to its contentious first argument. I showed how he could successfully reply to both. The third argument pertained to his foundation itself. I argued he is correct, that it is felicity maximising for dwellers to socially contract to establish a moral order based on peace and property. I showed that this is the case by constructing a hypothetical bargain between state of nature dwellers wherein they chose Hobbes' moral order over the sort of moral order most contemporary political philosophers would propose as an alternative as their improvement over the state of nature. I argued that dwellers would choose in this fashion because the Hobbesian moral order is purely mutually beneficial while the sort of moral order most contemporary political philosophers would propose as an alternative is only partly mutual beneficial.

References

- Axelrod, Robert (1984). *The Evolution of Cooperation*. United States of America: Basic Books.
- Blackburn, Simon (1996). *The Oxford Dictionary of Philosophy*. Oxford, United Kingdom: Oxford University Press.
- De Jasay, Anthony (1997). "Contradictory Contractarianism." *Against Politics*. New York, NY: Routledge.
- Dictionary (2005). Dictionary Version 1.0.2 by Oxford American Dictionaries. Cupertino, CA: Apple Computer, Inc.
- Frost, Samantha (2001). "Faking it: Hobbes's Thinking-bodies and the Ethics of Dissimulation." *Political Theory* 29 (1): 30–57.
- Gauthier, David (1986). *Morals by Agreement*. New York, NY: Oxford University Press.
- Hampton, Jean (1995). *Hobbes and the Social Contract Tradition*. Reprint. New York, NY: Cambridge University Press.
- Hausman, Daniel M. and Michael S. McPherson (2006). *Economic Analysis, Moral Philosophy, and Public Policy*. Second edition. New York, NY: Cambridge University Press.
- Hobbes, Thomas (2002). *Leviathan*. Revised student edition. Ed. by Richard Tuck. 5th edition. New York, NY: Cambridge University Press.
- Kymlicka, Will (2002). *Contemporary Political Philosophy: An Introduction*. Second Edition. New York, NY: Oxford University Press.
- Marx, Karl (2006). "Critique of the Gotha Programme." *Marx: Later Political Writings*. Ed. Terrell Carver. Sixth Printing. Cambridge, United Kingdom: Cambridge University Press.
- Narveson, Jan (2002). "Property Rights: Original Acquisitions and Lockean Provisos," *Respecting Persons in Theory and Practice*. Lanham, Maryland: Roman & Littlefield Publishers, Inc, pp. 111-129.

Works Consulted

Kavka, Gregory (1986). *Hobbesian Moral and Political Theory*. Princeton, NJ: Princeton University Press.

MacPherson, Clive Broughton (1970). *The Political Theory of Possessive Individualism*. Reprint. Oxford, United Kingdom: Oxford University Press

Narveson, Jan (1988). *The Libertarian Idea*. Philadelphia, PA: Temple University Press.