

Daniel Eggers, Hobbes and game theory revisited: Zero-sum games in the state of nature (final manuscript)

This is the peer reviewed version of the following article: D. Eggers (2011): „Hobbes and game theory revisited: Zero-sum games in the state of nature“, *Southern Journal of Philosophy*. 49. pp. 193–226., which has been published in final form at <https://onlinelibrary.wiley.com/doi/10.1111/j.2041-6962.2011.00071.x>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.

1. Introduction

One of the trends which have shaped Hobbes scholarship in the past decades is the trend to analyse Hobbes's theory of the state of nature with the means of game theory. Since David Gauthier's pioneering study in 1969, more and more commentators have attempted to game-theoretically describe the natural state and the 'war of every man against every man', so many commentators in fact, that the game-theoretic studies published up to the present day are almost impossible to grasp.ⁱ Most if not all of these studies focused on a question which already occupied Gauthier: whether Hobbes's natural state represents an example of the famous *prisoner's dilemma* (hereafter: PD), a game generally attributed to A. W. Tucker. Although the similarities and dissimilarities between the PD and the state of nature have often been described,ⁱⁱ and although diverging game-theoretic interpretations have been sketched and argued for,ⁱⁱⁱ it still remains an open question to what extent Hobbes's state of nature is equivalent to a PD situation. In the framework of my paper, I shall try to answer this question by reviewing and assessing the previous discussion in the light of the Hobbesian statements. I shall subscribe to the view, advanced by Graham Dodds and David Shoemaker, that Hobbes's state of nature contains differently structured situations of choice for whose game-theoretic

representation the PD *and* its most discussed rival, the *assurance game* (hereafter: AG), *and* the so-called *assurance dilemma* (hereafter: AD) are required. However, I shall demonstrate that even the three games taken together are insufficient to appropriately account for the Hobbesian ‘war of every man against every man’. A full game-theoretic description, I shall argue, needs to pay attention to the situations of competition which turn the state of nature in a state of diffidence and fear because it is only in virtue of these situations that the three games mentioned above, and the AG in particular, can be legitimately applied. As will be shown, however, these situations themselves cannot be described by the three games mentioned because they include situations that represent zero-sum games in which the outcome of mutual cooperation is not available.

While my main purpose in this paper is to identify the appropriate game-theoretic representation of Hobbes’s state of nature, I have also the somewhat broader aim of describing some of the problems and shortcomings of the previous discussion. Before turning to the particulars of Hobbes’s argument, I will, therefore, briefly address the general legitimacy of the game-theoretic approach and outline what I consider to be its proper object.

2. Hobbes and game theory

The increasing enthusiasm for game-theoretic analyses of Hobbes’s argument has received some criticism in the past. While some commentators only express their doubts as to whether the use of game-theoretic methods can provide any fundamental new insights into Hobbes’s theory,^{iv} others consider the application of game theory to Hobbes to be straightforwardly illegitimate.^v Yet, while it is true that the sensible and legitimate application of game theory to Hobbes’s theory is dependent upon some central presuppositions, there do not seem to be compelling reasons to doubt that these presuppositions hold. It is correct that the application

of game-theory to Hobbes would be problematic and perhaps misleading if the behaviour of Hobbesian individuals were restricted by moral obligations in the strict sense of the word, that is, by unconditional or categorical duties. However, the view that there are such obligations or duties in Hobbes is still only a minority view in Hobbes scholarship and one that faces serious problems.^{vi} Likewise, given that game theory presupposes perfect rationality on behalf of all the agents involved, applying game theory to Hobbes would surely be more complicated, but perhaps not impossible, if the predicament of the state of nature were crucially dependent upon human irrationality. The mere fact that a person acts on strong passions, however, is not sufficient for her behaviour to be irrational in the game-theoretic sense. Since the concept of rationality presupposed by game theory is that of instrumental rationality only, a person driven by aggression or hatred can be acting just as rationally as a person acting on the conscious idea of her own good, as long as she acts consistently and tries to maximize what her passions make her desire. The rationality requirement of game theory would, therefore, only be violated if Hobbes's derivation of the 'state of war' presupposed that the individuals would act instrumentally irrational, that is, misjudge their own desires, the means for the satisfaction of those desires or the general situation they are acting in. As will be shown in chapter 3, however, there is no such crucial connection between instrumentally irrational behaviour and Hobbes's 'state of war', at least not in *Leviathan*, where both irrationality and human passions play a far less important role in Hobbes's argument than is often assumed.

Since I consider other challenges of the game-theoretic approach – like the charge of anachronism – not to be any more compelling than the two just discussed,^{vii} I hold applying game theory to Hobbes to be generally legitimate. This is not to deny, however, that the previous game-theoretic debate on Hobbes has been deficient. The most serious problem is that most commentators attempt to game-theoretically grasp the natural state as a whole and

do not distinguish sufficiently between the different situations of choice that may arise within this state. So far, the discussion has brought into focus three major problems,

- (a) the problem of pre-emptive violence (or violence in general),
- (b) the problem of covenant-keeping, and
- (c) the problem of leaving the state of nature and instituting a sovereign power by means of a social contract.

While some contributions to the discussion have been confined to single problems, most commentators tend to run together their analyses of the three problems, treating them as three sides to one and the same story. Yet, the possible value of game theory consists in the thorough description and analysis of individual decisions in particular situations, with particular strategies and particular outcomes. Any sensible application of game theory to Hobbes's argument, therefore, ought to make particular state of nature situations its starting point rather than to suggest that the individuals were involved in one and the same large game all the time.^{viii}

This does not mean, however, that all three problems previously discussed ask for a detailed game-theoretic analysis at all. As has already been pointed out by some commentators,^{ix} the third problem, that is, the problem of instituting government by means of a social contract, hardly deserves the attention it has been given in the past. Even if, as has been claimed by Jean Hampton, it were irrational to keep the social contract because the making and keeping of it represents a PD situation, this would not vitally affect Hobbes's argument. Hobbes's purpose is to provide a justification for civil society and civil obedience with the help of a counterfactual scenario. His aim is to show that no rational individual can

desire to live outside civil society, not to show how a state of nature, should it ever fully obtain, can be left again.

It is important to note that this latter question does not become any more relevant if we take seriously Hobbes's remark that there are historical instantiations of the state of nature.^x First, even if Hobbes holds that there have been stages in human history which shared the state of nature's constitutive features, one may still view the Hobbesian 'war of every man against every man' as some sort of 'model war' which is not absolutely identical to those state of nature-like historical situations Hobbes mentions.^{xi} Secondly, even if there were no rational way out of such a 'state of war', this would not be incompatible with the fact that some historical manifestations of the state of nature have, in fact, been overcome. The success in leaving the state of nature may simply be due to the fact that the individuals living in that state were not acting fully rationally. This concession is by no means as problematic as some commentators seem to think. The assumption that individuals possess perfect rationality may be one of the central presuppositions of game theory. However, it is not a presupposition of Hobbes's argument and surely not part of Hobbes's view of human nature.^{xii} Thirdly and lastly, Hobbes explicitly concedes in the English *Leviathan* that most civil societies did not come into being by way of a social contract, but by way of conquest. As well as the state of nature, the social contract is, therefore, to some extent a logical hypothesis which is meant to provide a theoretical underpinning for the duties of subjects, not to trace the actual emergence of political power and civil society.

Further evidence for this reading is provided by the fact that in *De Cive*, Hobbes explicitly describes the natural *bellum omnium in omnes* as perpetual in its own nature, without, however, viewing this feature as an obstacle to his subsequent description of the institution of sovereign power.^{xiii} Given the overall purpose of Hobbes's state of nature argument, there is indeed no reason for worries of this kind. If the purpose of Hobbes's

argument is to convince citizens of existing societies of the necessity of absolute obedience by painting the blackest possible picture of the relapse into a non-political condition, then this argument can only be strengthened by the assumption that this terrible condition, if it ever fully obtains, can never be left again.^{xiv}

A game-theoretic analysis of Hobbes's state of nature argument, therefore, need not include any discussion of the problem of leaving the state of nature or any game-theoretic analysis of the social contract. In addition to this, I think that the second of the problems listed above, i.e. the question of whether it is generally rational for first or second parties to keep covenants in the state of nature, has also been given too much attention in the past. The reason is that, if compared with the problem of pre-emptive violence, the question only seems to be of minor importance for the reconstruction and assessment of Hobbes's overall argument. Hobbes's describes the natural state as a 'war of every man against every man', and – as will become clear in the next chapter – he does not draw upon the specific problems of making and keeping contractual agreements in doing so, but on the competition for goods and the necessity of anticipatory violence. Since the dangers associated with this 'state of war' provide rational individuals with sufficient reasons for leaving the state of nature, it seems that in order to assess Hobbes's argument for the necessity of civil society, one can dispense with analysing the rationality of covenant-keeping – even if one admits that there can and will be valid contractual agreements in the state of nature^{xv} and that the situation of covenanting parties can sensibly be analysed with the means of game theory.

Such a game-theoretic analysis may very well demonstrate why there will only be few successful covenants in the state of nature and thereby help to support the Hobbesian picture of the state of nature and his pessimistic view of the force of 'covenants without the sword'. One may think, however, that a detailed game-theoretic analysis is not really needed for this. In view of the general insecurity of the 'state of war' and the predominance of 'diffidence and

fear', one can safely assume that covenants, i.e. contracts that involve mutual *trust*, will face crucial problems. Thus the majority of Hobbes's state of nature individuals can be expected to principally doubt the integrity of a possible second party. This means that they will either dispense with entering into covenants altogether or will, if they are the ones to perform first, often refuse to do so when the time has come and thereby void the agreement. Moreover, it can be assumed that first parties who have willingly performed their part of the bargain will in many cases be betrayed by second parties – which experience should, in the long run, lead them to change their strategy. The small number of contractual agreements in Hobbes's state of nature, then, will inevitably decrease further and further. Given this, the exact game-theoretic details of the rationality of covenant-keeping can hardly be expected to bear a particular relevance for the question of what the Hobbesian state of nature is like and whether it is rational to stay in it.

Thus the central and most important problem to be analysed with the aid of game theory remains the problem of violence, a problem mostly discussed with regard to the question of why the use of anticipatory violence is rational. Even if one confines the game-theoretic analysis to this problem, it can be demonstrated that the previous game-theoretic interpretations of Hobbes fail to do justice to the complexity of Hobbes's argument and to the complexity of state of nature behaviour. In the remainder of this paper, I will try to provide such a demonstration and try to develop a sufficiently complex game-theoretic representation of the problem of anticipation. In order to be able to do this, I will first summarize Hobbes's discussion of anticipatory violence and his preceding discussion of the competition for vital goods as these are found in the English version of *Leviathan*.^{xvi}

3. The derivation of the 'state of war' in the English *Leviathan*

In the English *Leviathan*, Hobbes's derivation of the 'state of war' begins with the claim that human beings are roughly equal with regard to their physical and mental abilities and the claim that this natural equality of ability results in an equality of hope of human beings to attain their ends. These two claims lay the basis for Hobbes's description of the first source of conflict.

From this equality of ability, ariseth equality of hope in the ataining of our Ends. And therefore if any two men desire the same thing, which neverthelesse they cannot both enjoy, they become enemies; and in the way to their End, (which is principally their owne conservation, and sometimes their delectation only,) endeavour to destroy, or subdue one an other.^{xvii}

It is important to note, especially in comparison with the earlier versions of Hobbes's argument in *The Elements of Law* and *De Cive*, that in the passage above, Hobbes presents the violent competition for goods neither as a result of human irrationality and self-conceit nor as a result of a natural desire to dominate other individuals. Hobbes surely acknowledges the general possibility and probability of human conceitedness. However, at no point of his argument does Hobbes explicitly present the "equality of hopes" as an example of such vanity and conceitedness or criticise it in any other way.^{xviii} On the contrary, the "equality of hopes" appears as a natural and legitimate consequence of the general equality of ability: According to Hobbes, the general equality of hopes "ariseth" from the equality of ability, and this must mean that it arises from the *insight* into the equality of ability. The "equality of hopes", therefore, is due to a realistic assessment of one's own abilities, to the modest and realistic view that one can, in principle, conquer a possible opponent and survive.

What is even more important is that with regard to the situations in question, Hobbes does not seem to recognize any peaceful alternative course of action. The casualness with

which Hobbes moves from the clash of desires to the goal of destroying or subduing one's opponent suggests that there is no way to prevent the competition for goods from turning into violent conflict. The reason for this, however, is not to be sought in any natural tendency to violent behaviour or any natural desire to gain precedence over others, but in the features of the situation. The first indicator that this is Hobbes's view is that the goods in question cannot be divided or used in common ("which neverthelesse they cannot both enjoy"). The second indicator is that, according to the insertion "which is principally their owne conservation, and sometimes their delectation only", the individuals are led into such violent conflicts – at least sometimes – by the mere goal of self-preservation.

Even more disastrous than the violent competition for goods itself is what follows from it: the permanent threat of violent encounters and the resulting necessity of pre-emptive violence which Hobbes describes as the second source of conflict.

And from this diffidence of one another, there is no way for any man to secure himselfe, so reasonable, as Anticipation; that is, by force, or wiles, to master the persons of all men he can, so long, till he see no other power great enough to endanger him: And this is no more than his own conservation requireth, and is generally allowed. Also because there be some, that taking pleasure in contemplating their own power in the acts of conquest, which they pursue farther than their security requires; if others, that otherwise would be glad to be at ease within modest bounds, should not by invasion increase their power, they would not be able, long time, by standing only on their defence, to subsist. And by consequence, such augmentation of dominion over men, being necessary to a mans conservation, it ought to be allowed him.^{xix}

What is important to note about Hobbes's discussion of anticipation is the way in which it is connected to the distinction between 'moderate' and 'vainglorious' men or, to use terms established by Gregory Kavka, between 'moderates' and 'dominators'. While some

individuals, i.e. ‘dominators’, take pleasure in feeling superior to others and hence pursue their attacks further than their security requires, other individuals only strive for the necessary means of survival. The distinction made here provides a further clarification of what is meant by the statement that what leads individuals into violent competition is “principally their own conservation, and sometimes their delectation only”: While the goal that leads moderate individuals into violent conflict is self-preservation, ‘dominators’ are sometimes also led into violent conflict by the goal of enjoying their superiority over others.

Again, it is crucial to see that it is not the ‘dominator’'s striving for precedence that makes anticipation necessary. The sentence “Also because there be some, that taking pleasure in contemplating their own power in the acts of conquest, which they pursue farther than their security requires” shows clearly that the immoderateness of the ‘dominators’ represents an *additional* reason for the application of pre-emptive violence. This is indicated by the use of the initial “Also” and by the fact that the application of violence is implicitly presented once more as a necessary means of security: What distinguishes ‘dominators’ from moderate men is not that the former carry out acts of violence at all, but that they carry those necessary “acts of conquest” *farther* than is actually warranted by the aim of self-preservation. The suggestion that modest or sensible individuals could refrain from settling the competition for goods with the help of force altogether and simply dispense with the goods in question is, therefore, irreconcilable with Hobbes’s account: The rational pursuit of self-preservation not only inevitably leads individuals in the state of nature into competition for goods that cannot be divided; it also leads them into deciding this competition with the help of force and it even leads them into using anticipatory violence.^{xx}

One may ask, of course, why Hobbes makes this rather strong claim and whether it withstands scrutiny. Hobbes’s argument seems to be dependent on the fact that goods are scarce in the state of nature, although Hobbes does not explicitly say so.^{xxi} That Hobbes’s

state of nature argument indeed presupposes a scarcity of goods has been claimed by a series of commentators in the past, and the only thing to be criticized about these claims is that they have not been confined to the version of Hobbes's argument contained in the English *Leviathan*.^{xxii} However, commentators differ on whether this presupposition creates any problems for Hobbes. In order to assess the hidden assumption of Hobbes's derivation of the 'state of war', it is crucial to go back to Hobbes's actual statements concerning the competition for goods. Hobbes does not claim that the striving for self-preservation will always and permanently lead individuals into conflicts that can only be decided by force. Rather, the sentence "And therefore *if* any two men desire the same thing, which nevertheless they cannot both enjoy"^{xxiii} suggests that the described competition for goods can occur and will occur with some frequency, but that there are also alternative settings. The insertion "which is principally their owne conservation, and sometimes their delectation only" seems to confirm this, by suggesting that there could also be violent competitions for goods in the state of nature which are *not* required by the goal of self-preservation. Similarly, the statement according to which 'dominators' are carrying the "acts of conquest" *farther* than is required by their conservation, suggests that in some cases, self-preservation does not force the individuals to engage in "acts of conquest" but could very well be attained without such acts. Even if this may not follow logically from the statements themselves, it is possible to read them in this way and, what is more, the assumption that there will also be situations in the state of nature in which individuals do not directly compete for goods or at least not for vital goods seem much more modest and realistic than the assumption that with each and every encounter, Hobbesian individuals are engaged in competition that directly concerns their survival.

If this is correct, then there should be three types of situations in which Hobbes's state of nature individuals may find themselves:

- a) situations in which striving for self-preservation does not lead the agents into competition with others at all (type 1),
- b) situations in which they are in fact led into such a competition but where the competition can be resolved without the help of force since the goods in question are not actually necessary for one's survival (type 2), and
- c) situations in which individuals are competing for goods that are, with regard to all the individuals involved, indispensable means of survival.

The claim which is the cornerstone of Hobbes's derivation of the 'state of war' in the English *Leviathan* is, then, not the claim that in the state of nature, there are only type 3-situations. It is the far more modest claim that there *also* are such situations, that is, that type 3-situations cannot altogether be avoided, even if all individuals confine themselves to those goods that they need in order to survive. However, whether this claim presupposes a general scarcity of goods seems doubtful. There are good reasons to hold that a partial scarcity, that is, one restricted to particular areas or certain periods of time, would provide a sufficient basis for Hobbes's claim, and in this form, the Hobbesian presupposition seems defensible.

The conclusion that Hobbes only needs to presuppose a partial scarcity of goods for his argument to work is also important for a second reason. Were Hobbes committed to the assumption that goods are generally scarce in the state of nature, this might raise the question of how the establishment of a sovereign power can help to solve the problems of the state of nature at all. Now in a sense, this question is just a variant of the question I have dismissed earlier, namely the question of whether the state of nature, if it ever fully obtains, can ever be left again. However, it seems important to point out that the proposed analysis works even if one is not willing to subscribe to my earlier arguments. The concession that there is a partial

scarcity of goods in the state of nature does not force us to conclude that the establishment of a sovereign power must be futile, because it does by no means imply that the overall number of available goods cannot be increased. First, a partial scarcity of goods, being restricted to particular areas and periods of time, may – among other things – be due to an inefficient distribution of goods. The distribution of goods, however, may be greatly improved once a sovereign power and other political structures are established which allow the relevant efforts to be coordinated in a better way. Second, the partial scarcity of goods may be due to the fact that, in virtue of potential conflicts and the insecurity resulting from it, individuals in the state of nature refrain from trying to produce goods by cultivating land, raising cattle etc. As a matter of fact, Hobbes's example of a man who plants, sows and builds and is then deprived of his possessions seems to suggest just that.^{xxiv} Now it is true that Hobbes presents this problem of depossession as a consequence of the competition for vital goods (“And hence it comes to passe”), which suggests that, according to the above analysis, it must result from the scarcity of goods rather than causing it. However, instead of drawing a hard and fast line between causes and effects, it seems important to recognize that the respective roles of scarcity and violence can vary: Partial scarcity in some area or at some point in time may lead to violent competition which increases insecurity and lowers the incentive to produce goods which, in turn, increases the chance of partial scarcity etc.; the unwillingness to produce goods, however, may also by itself lead to partial scarcity which again may become a cause of violence. Now if this is a plausible view, the development may also be turned around once the sovereign is established and provides security: Successful cultivation of land in some area or at some point in time may temporarily make scarcity less probable, which makes one's possessions less insecure and increases the incentive to produce, which lowers the risk of scarcity even more etc. Of course, an individual who has cultivated land or raised cattle may still get into competition with others who, in order to survive, depend on the produced goods

just as much as he does. However, the existence of a sovereign power still makes a difference because the sovereign can now defend the owner against his attackers who have no title to the goods they desire. Once there is a sovereign power, therefore, individuals who cultivate their land and engage in other productive enterprises need not fear that they lose the fruits of their labour as much as they need to fear it in the state of nature, which means that there is more of an incentive to make the relevant efforts. Again, in fact, this seems to be just what Hobbes tries to suggest with his example of dispossession, because he explicitly ties the risk of dispossession to the condition that a possible attacker has nothing more to fear than the power of the owner himself (“where an Invader hath no more to feare, than an other mans single power”), which may be true of the state of nature but is not true not of a state with an effective sovereign power.

Before we turn to those games which have been forwarded as the appropriate representation of the problem of anticipation in the past, it needs to be mentioned that, besides competition and anticipation, Hobbes names a third source of conflict which he refers to as “glory”. Hobbes assumes a general tendency of human beings to compare themselves with others and a desire to be valued by those others in the appropriate measure. According to Hobbes, any signs of contempt may lead individuals “to extort a greater value from his contemners, by damage”^{xxv}. For our purpose, however, the third source of conflict is of no special importance because there is no intimate connection between this source of conflict and the problem of anticipation. This is true even if one is unwilling to attribute as little relevance to the existence of ‘dominators’ as I have claimed to be appropriate. At first glance, the desire of ‘dominators’ to gain superiority over others may appear to be an expression of the striving for “glory”. However, the rather reactive disposition described by Hobbes under this heading, namely the disposition not to accept a possible undervaluation of others, is clearly different from the active desire to gain precedence over others that drives the ‘dominators’. Moreover,

there is no need to square the game-theoretic description of the problem of anticipation with a game-theoretic description of the third source of conflict as is suggested, for instance, by Hampton. That the need of anticipatory violence leads individuals into violent conflict is easily compatible with the view that further conflicts can derive from other sources. The only relevant question is whether the necessity of anticipation is in any way dependent on what Hobbes describes as “glory”, and both the formal structure of Hobbes’s argument and his actual description of the need for anticipation clearly suggest otherwise. Moreover, there can hardly be any doubt that the competition for goods and the necessity of pre-emptive violence are in themselves sufficient to turn the state of nature into a ‘state of war’ – in Hobbes’s sense of the term – which suggests that the third source aggravates conflict rather than initiates it. Neither, therefore, appears a game-theoretic analysis of the third source of conflict necessary in order to appropriately describe the problem of anticipation, nor does it appear wholly indispensable in order to game-theoretically represent Hobbes’s derivation of the ‘state of war’.

4. The prisoner’s dilemma, the assurance game and the assurance dilemma

In trying to critically review the previous game-theoretic discussion of Hobbes’s state of nature, I will focus on three games: the PD, the AG and the AD. While other games have been forwarded as the appropriate representation of the problem of anticipation in the past, the PD, AG and AD have received by far the most attention, and they all in fact promise to capture important aspects of Hobbes’s argument. Moreover, the discussion of the three games and its problems can serve to demonstrate where and why the previous game-theoretic discussion of the problem of anticipation has been insufficient. In light of the analysis of the three games, however, I will also try to briefly sketch what I consider to be wrong about other game-

theoretic interpretations, such as the ‘coordination game’ interpretation or the ‘chicken game’ interpretation.

The crucial aspect of the problem that the PD, AG and AD likewise try to capture is that, according to Hobbes, anticipation is the rational strategy in the state of nature, but leads, if universally applied, to a general ‘state of war’, that is, to a state which a rational person would try to avoid at all costs. This seemingly paradoxical situation is perhaps most clearly exemplified by the PD in which ‘non-cooperation’ is a strict dominant strategy but leads to a pareto-inoptimal outcome.^{xxvi} However, as can be seen by contrasting the three interpretations, the two other games are capable of capturing this crucial point of Hobbes’s argument as well.

In order to elaborate on the similarities and differences between the three interpretations, it seems helpful to include graphic representations of the three interpretations. If we restrict ourselves to a situation in which two individuals (individual 1 and individual 2) are faced with the decision of whether to use anticipatory violence, we can describe two strategies (‘attack’ and ‘not-attack’) and four possible outcomes:

- a) both individuals attack,
- b) neither individual attacks,
- c) individual 1 attacks, but individual 2 does not,
- d) individual 2 attacks, but individual 1 does not.

If these outcomes are represented by ordinal numbers^{xxvii} – 1 representing the best possible outcome and 4 representing the worst –, the graphic representations of the PD, AG and AD interpretations assume the following form:

Matrix 1: The prisoner's dilemma

		individual 2	
		attack	not-attack
individual 1	attack	3,3	1,4
	not-attack	4,1	2,2

Matrix 2: The assurance game

		individual 2	
		attack	not-attack
individual 1	attack	3,3	2,4
	not-attack	4,2	1,1

Matrix 3: The assurance dilemma

		individual 2	
		attack	not-attack
individual 1	attack	3,3	2,4
	not-attack	4,1	1,2

As can be seen from Matrix 1, the PD interpretation takes the 'attack' strategy to be a strict dominant strategy. Viewed from individual 1's perspective: If individual 2 decides to attack, individual 1 will be better off if he also attacks (3 instead of 4); if individual 2 does not attack, it would still be better for individual 1 to attack (1 instead of 2). However, if both individuals indeed act rationally and choose to attack, they are left with an outcome (3,3) which is pareto-inferior to the outcome of mutual non-attacking (2,2).

Although the PD interpretation easily captures the paradox of rationality inherent in the problem of anticipation, however, there are good reasons to doubt whether the PD really provides the appropriate representation of Hobbes's argument. The crucial problem is not, as has sometimes been claimed, that the classical PD cannot account for the fact that more than two individuals are involved in the problem of anticipation,^{xxviii} or the fact that the state of nature is not a 'one-shot' situation, but a dynamic situation in which the individuals have to decide whether to attack or not to attack over and over again.^{xxix} The problem is that, by making the 'attack' strategy a strict dominant strategy, the PD interpretation presupposes that Hobbesian individuals would attack others even if they knew that those others would behave peacefully towards them. As some commentators have pointed out, it is not clear whether this claim is warranted by Hobbes's actual statements. Rather, Hobbes's view seems to be that the individuals should attack only because they cannot be sure that others will indeed behave peacefully.

As can be seen from Matrix 2, the way to incorporate this point into the game-theoretic representation is to describe the problem of anticipation as an AG. Viewed again from the perspective of individual 1: If individual 2 attacks, it is rational for individual 1 to also attack (3 instead of 4); however, if individual 2 does not attack, it is rational for individual 1 not to attack, either (1 instead of 2). Although to attack is hence not a dominant strategy, one important aspect of the problem of anticipation is retained: If both individuals choose to use anticipatory violence, they are left with an outcome (3,3) that is pareto-inferior to the outcome of mutual non-attacking (1,1). The question arises, however, whether to attack is still the rational thing to do and whether the paradox of rationality inherent in the problem of anticipation is saved in the AG interpretation. This question arises because in the AG, the 'dilemma' of the PD no longer exists. Unlike the PD, the AG allows rational individuals to realise the pareto-optimal outcome (1,1): If individual 1 chooses not to attack, individual 2

should also refrain from attacking, and if individual 1 anticipates this, it is indeed rational for him not to attack in the first place. It seems, therefore, that the use of anticipatory violence is not the rational thing to do and that the problem of anticipation, if there still is such a problem at all, would not turn the state of nature into a ‘state of war’.

Yet, I think that the AG interpretation of the problem of anticipation need not have this result. As is sometimes overlooked, one of the possible outcomes for Hobbes’s individuals is their immediate death or the subjection to their opponent. In view of this severe threat, and in view of Hobbes’s claim that human beings necessarily desire “to avoid that which is hurtful; but most of all that terrible enemy of nature, death”^{xxx}, it seems plausible that Hobbesian individuals will apply a maximin-strategy, that is, choose a strategy which allows them to definitely avoid the worst possible outcome. This seems even more plausible given that, according to Hobbes, the state of nature is a state of diffidence and fear in which nobody can be sure about what to expect from other individuals. This, however, would mean that in spite of the general reasonableness of peaceful behaviour, Hobbesian individuals would strike out in order to avoid the outcome of unilateral non-attacking. The AG interpretation, then, has no problem in accounting for the fact that Hobbes considers anticipation the rational thing to do and for his claim that the state of nature will turn into a general ‘state of war’.

The true problem of the AG interpretation is that, just like the PD, it does not pay attention to Hobbes’s distinction between ‘moderates’ and ‘dominators’. Hobbes’s description of moderate individuals may support the AG interpretation by showing that not all Hobbesian individuals have the preferences that the PD interpretation attributes to them: While it seems safe enough to assume that ‘dominators’ place unilateral attacking over mutual non-attacking, it seems just as clear that ‘moderates’, who content themselves with the necessary means of survival, will instead prefer a state of mutual peace to unilateral attacking, even if unilateral attacking may promise them possible benefits. As much as it refutes the PD interpretation,

however, the distinction between ‘moderates’ and ‘dominators’ refutes the AG interpretation: There may be ‘moderate’ individuals who place mutual non-attacking over unilateral attacking, but there are also those others whose striving for superiority will lead them to attack other individuals even where they do not have to expect violent behaviour from them.

It seems, therefore, that an appropriate game-theoretic representation of the problem of anticipation ought to distinguish between two groups of players with diverging orders of preference, and it is this idea that lies at the heart of the AD interpretation.^{xxxii} A closer look reveals how the AD combines the features of the PD and the AG. As in these games, both players have the same two strategies to choose from. However, there is now a certain assymetry in the four outcomes which is the result of the fact that the orders of preference of the two players differ, individual 1 (the moderate individual) owning the preferences of the players of the AG, individual 2 (the ‘dominator’) owning the preferences of the players of the PD. Thus while the outcome of unilateral attacking yields the best possible outcome for the ‘dominator’ (1), it only yields the second best outcome for the ‘moderate’ (2). Accordingly, whereas moderate individuals would benefit most from the outcome of mutual non-attacking (1), ‘dominators’ would in this case attain their second best outcome only (2), which means that they will try to avoid this outcome by choosing the ‘attack’-strategy. For ‘dominators’, therefore, to attack remains a strict dominant strategy even in the AD, and for exactly this reason, it is rational for ‘moderates’ to also refrain from applying the peaceful strategy. The outcome of mutual attacking (3,3), therefore, appears to be inevitable, although – if compared with the outcome of mutual non-attacking (1,2) – it provides a pareto-inferior result, just as it does in the two other games.

One important difference between the AG interpretation and the other two interpretations is that the former does not, and cannot, claim exclusiveness. If it is claimed that any appropriate game-theoretic description of the problem of anticipation needs to pay

attention to the distinction between ‘dominators’ and ‘moderates’, then it follows that the AD matrix *alone* cannot appropriately represent the state of nature because it only describes one type of encounter in Hobbes’s state of nature, namely the encounter of ‘dominators’ and ‘moderates’. In order to account for the two other possible cases, namely the encounter of two ‘moderates’ and the encounter of two ‘dominators’, the AD needs to be supplemented by the AG and the PD. Hobbes’s distinction between ‘moderates’ and ‘dominators’, therefore, renders the enterprise of describing the problem of anticipation by the use of one single game matrix altogether impossible: The problem of anticipation has the form of an PD whenever two ‘dominators’ face each other, the form of an AG whenever two ‘moderates’ face each other, and the form of an AD whenever a ‘moderate’ and a ‘dominator’ face each other.

One may think that this preliminary conclusion creates a new problem for legitimately applying game theory to Hobbes because it seems to make Hobbes’s argument incompatible with one of the central presuppositions of game theory: the presupposition that each player knows about the preferences of all the other players involved and can hence anticipate their probable choices.^{xxxii} Given that the state of nature is a state of diffidence and fear, it can hardly be assumed that Hobbesian individuals will always know which kind of individual they are up against. If, however, Hobbesian individuals do not know which ‘game’ they are actually playing, it seems that they will not be able to make a thorough rational calculation in the game-theoretic sense which seems to reduce any game-theoretic description of their decision to absurdity. In our case, however, there is no reason for worries of this kind. As has already been indicated, to attack is a strict dominant strategy for ‘dominators’ both in PD and in AD situations. Any knowledge about the other player’s preferences is hence irrelevant to their decision. Moreover, the possible lack of knowledge does not crucially affect the decision of ‘moderates’, either. As has been shown, even in AG situations it is rational for ‘moderates’ to attack because of the general insecurity of the state of nature which makes it reasonable to

apply a maximin-strategy and not to trust the rationality of others. If, in addition to this, ‘moderates’ may also face individuals for whom to attack is a strict dominant strategy – as is the case in AD situations in which they face a ‘dominator’ – they have even more reason to refrain from behaving peacefully. In the end, therefore, even ‘moderates’ will use the ‘attack’-strategy as if it were a strict dominant strategy. Thus even if there may be many cases in which Hobbesian individuals will not be able to fully grasp the situation they are in, the matrices of the PD, AG and AD are still suited to cast light upon their choices because they show why it is, after all, rational for the individuals to choose the uncooperative strategy and to use anticipatory violence.^{xxxiii}

However, our analysis still raises an important point because it strongly suggests that the game-theoretic analysis of the problem of anticipation cannot wholly be separated from the competition for vital goods which turns the state of nature into a state of diffidence and mutual fear in the first place. That it is rational for Hobbes’s moderate individuals to use pre-emptive violence cannot be explained without referring to the general insecurity of the natural state, and this insecurity, in turn, cannot be explained without referring to the initial problem of competition. Therefore, an appropriate game-theoretic representation of the problem of anticipation cannot dispense with a game-theoretic representation of the competition for vital goods. As far as I can see, such a representation has not been provided so far. As a matter of fact, it has not even been recognized that and why such a representation is required. The subsequent attempt to provide this representation will show that the Hobbesian ‘war of every man against every man’ starts neither from a PD, nor from an AG, nor from an AD, but from zero-sum games in which the outcome of mutual cooperation does not exist.

Before we proceed with the game-theoretic analysis of the competition for goods, it seems important to give further support to the previous analysis by briefly pointing out why game-theoretic interpretations that do not draw upon the PD, AG and AD, but on other games

– such as the ‘coordination game’ or the ‘chicken game’ – do not provide an appropriate or even a better representation of the problem of anticipation. That Hobbes’s state of nature ought to be represented by a ‘coordination game’ has been argued by Patrick Neal.^{xxxiv} Neal’s interpretation is based on the fact that, according to Hobbes, the only way for state of nature individuals to permanently avoid violent death is to cooperate with each other and leave the state of nature. Neal infers from this that we should attribute one and the same utility to the outcomes of unilateral non-cooperation, unilateral cooperation and mutual non-cooperation because they will all, in the long run, lead to the death of the agent. The resulting coordination game would have the following structure:

Matrix 4: The coordination game

		individual 2	
		not-cooperate	cooperate
individual 1	not-cooperate	2,2	2,2
	cooperate	2,2	1,1

Now the first thing to be noted about Neal’s interpretation is that his attempt to describe the state of nature and its overall predicament with one game runs counter to the aim of distinguishing more clearly between the different situations of choice within the Hobbesian state of nature on which our discussion of the problem of anticipation rests. In itself, however, this may not seem a sufficient reason to reject Neal’s interpretation: If we restrict ourselves to the problem of anticipation, the ‘coordination game’ may still provide the appropriate representation of this problem – or at least of one of its variants. However, a closer look at the features of the game shows that this is not the case. As indicated above, the ‘coordination game’ interpretation assumes that three of the four possible outcomes of the situation in question are in effect equivalent. This means that the individuals should be indifferent as to

which of these three outcomes will obtain. However, neither Hobbes's remarks on the reasonableness of anticipation nor our previous analysis provides any basis for this view. It seems clear that moderate individuals will prefer the outcome of mutual non-attacking both to the outcome of unilateral non-attacking and to the outcome of mutual attacking, and it seems just as clear that they will prefer unilateral attacking and mutual attacking to unilateral non-attacking – otherwise Hobbes's claim that it is reasonable to strike out pre-emptively would not make sense. The case is similar for 'dominators': Given the reasonableness of anticipation, 'dominators' can safely be assumed to prefer the outcome of unilateral attacking to mutual attacking and to unilateral non-attacking; and they can just as safely be assumed to prefer the outcome of mutual attacking to unilateral non-attacking. If this much is admitted, however, the 'coordination game' interpretation already breaks down because it cannot account for the fact that 'moderates' and 'dominators' attribute different utilities to at least three of the four possible outcomes.

The 'chicken game' interpretation, which has been forwarded by Gabriella Slomp and Manfredi La Manna, runs into similar problems. What distinguishes the 'chicken game' from the PD, AG and AD, is that the players of the 'chicken game' prefer the outcome of unilateral cooperation to mutual non-cooperation, whereas in the three other games, mutual non-cooperation yields a higher utility than unilateral cooperation. Now, if applied to the standard example – two car drivers approaching each other in a game meant to display their courage, trying not to 'chicken' out before the other one does – the 'chicken game' order of preferences seems plausible: Neither of the individuals wants to 'chicken' out, but both will probably prefer being considered a coward to dying in an actual crash. If applied to the problem of anticipation, however, the order of preferences is clearly inappropriate. If Hobbesian individuals would prefer unilateral non-attacking to mutual attacking, it would not be reasonable for them to use pre-emptive violence if they knew that another individual was

about to attack them. For Hobbes, however, this seems to be the paradigm case for when anticipation is reasonable, and we have good reasons to follow him here. A ‘chicken game’ interpretation of the problem of anticipation is only plausible if mutual non-attacking provides a lower chance of survival than does being attacked by a unilateral aggressor. This, however, just doesn’t seem right: First, the individual thus attacked runs the risk of being killed before even getting the chance to strike back. Second, even if he gets the chance to respond to the violence of the attacker, it must be assumed that the opportunity of striking out first has put his adversary in a somewhat superior position, which means that the prospects of the attacked person are worse than they would be if he had willingly entered into the fight from the beginning and on more equal terms.

That Slomp/La Manna still favor the ‘chicken game’ reading is due to their presupposition that unilateral non-attacking will result in the peaceful submission of the attacked individual and not in a fight.^{xxxv} In view of what is at stake, however, it seems much more realistic that the attacked individual will try to defend himself. This seems to be Hobbes view as well, who generally allows for two outcomes of violent encounters – the killing and the submission of the losing party –, but, with regard to the latter, does not talk so much about peaceful submission, but rather about the loser being ‘subdued’.^{xxxvi} If, however, submission results from losing a fight rather than from having avoided it in the first place, Hobbesian individuals always face the risk of being killed before even having the chance to submit themselves to their opponent. This provides sufficient reason to expect that they will, after all, prefer a fight on more equal terms – which, as a matter of fact, would just as well leave the possibility of submitting themselves once they do no longer believe in their own victory.

A last game-theoretic interpretation which deserves to be mentioned is Pärtel Piirimäe’s view that the Hobbesian state of nature constitutes an “assurance game with glory”. One problem of Piirimäe’s interpretation is, once more, that it does not distinguish

sufficiently between the different problems that plague the state of nature. While generally addressing the problem of war and peace, Piirimäe argues for the AG interpretation by drawing on the problem of covenant-keeping and Hobbes's 'reply to the Foole'.^{xxxvii} He then buttresses the AG reading against possible objections by reading back the problem of "glory" into the problem of anticipation.^{xxxviii} The main problem with this, however, is that, as has been emphasized in chapter 3, there is no intimate connection between the problem of "glory" and the necessity of pre-emptive violence, which, according to Hobbes's argument, follows solely from the competition for vital goods. As demonstrated above, it is true that the AG interpretation captures at least part of the problem of anticipation. It is also true that it needs to be backed up by further considerations, namely by an explanation of why the state of nature turns into a state of diffidence and fear in which mutual cooperation will not even be achieved by moderate individuals. However, Piirimäe's proposal – to bring in the problem of "glory" – should only be accepted if there were no other interpretation at hand that is more in line with the structure and content of Hobbes's argument. In the remainder of this paper, I shall try to show that there is, in fact, such an interpretation which manages to both save the AG interpretation and to stay close to Hobbes's statements.

5. The competition for vital goods: Zero-Sum Games in Hobbes's state of nature

In our discussion of the competition for goods in chapter 3, we distinguished three types of situations into which the inhabitants of the state of nature may be led by their striving for self-preservation: situations in which they can take a desired good into possession without even getting into any competition with other individuals at all (type 1); situations in which they get into such competition, but do not necessarily depend on the good in question in order to preserve themselves (type 2); and situations in which they compete with others for a good

vital to their preservation and are forced to decide this competition with the help of force (type 3). Obviously, if we want to game-theoretically describe the competition for goods, only type 2- and type 3-situations are of any interest. Now type 2-situations and type 3-situations require a different game-theoretic approach because Hobbes's distinction of 'dominators' and 'moderates' is relevant to the former, but not to the latter. If a competition for goods is characterised by the fact that both individuals desperately need the goods in order to preserve themselves, the fact whether the individuals are only striving for self-preservation or also for something else is not pertinent to their decision. However, if the competing individuals do not really need the desired goods in order to survive, it can be assumed that the said fact should be much more relevant. In trying to develop a game-theoretic representation of type 2-situations, I will, therefore, start again from Hobbes's distinction between 'moderates' and 'dominators'.

If two individuals are in a type 2-situation, they have two strategies to choose from: They can try to take the desired good into possession or renounce it. There are, therefore, four possible outcomes:

- a) individual 1 and individual 2 try to acquire the good in question and fight for it,
- b) individual 1 does not try to acquire the good in question and individual 2 can acquire it without a fight,
- c) individual 2 does not try to acquire the good in question and individual 1 can acquire it without a fight, and
- d) both individuals refrain from trying to acquire the good in question.

If we try to order the preferences of 'moderates' and 'dominators' with regard to the four possible outcomes, we can assume, first, that both groups of individuals will prefer mutual renunciation of the good in question to unilateral renunciation. The reason is that mutual

renunciation leaves the possibility to still take hold of the good at a later time, a possibility which does not present itself in the case of unilateral renunciation. Secondly, it should be obvious that the unilateral attempt to acquire the good represents the best possible outcome for both ‘moderates’ and ‘dominators’ because it allows the individual to take possession of the desired good without any opposition. The outcomes (b) and (c), therefore, are the best possible results for individual 2 and individual 1 respectively. What remains to be answered, therefore, is only whether the individuals will prefer a fight for the desired good to unilateral renunciation, and there are good reasons to assume that these preferences will differ with regard to ‘moderates’ and ‘dominators’. Since the striving for superiority leads ‘dominators’ to enter into “acts of conquest” even where this is not required by self-preservation, it seems that they prefer a fight for the desired good to unilateral renunciation. Given the undeniable risks involved in a fight, however, it seems that even ‘dominators’ should prefer mutual renunciation to such fighting, not least because mutual renunciation does not necessarily mean to let go of the good in question forever. Given this, we should attribute to ‘dominators’ the following order of preferences:

1. unilateral attempt to acquire the good
2. mutual renunciation
3. mutual attempt to acquire the good
4. unilateral renunciation

In contrast, moderate individuals are generally willing to content themselves with what is needed for survival and will enter into violent conflict only where there is no alternative route to be taken. In type 2-situations, in which such alternative routes *are* available, moderate

individuals should hence prefer both mutual renunciation and unilateral renunciation to an actual fight. Their order of preference would, therefore, be the following:

1. unilateral attempt to acquire the good
2. mutual renunciation
3. unilateral renunciation
4. mutual attempt to acquire the good

If we try to describe the three possible type 2-situations – ‘dominator’ vs. ‘dominator’, ‘dominator’ vs. ‘moderate’ and ‘moderate’ vs. ‘moderate’ – by matrices, we get the following three representations:

Matrix 5

		dominator	
		acquire	renounce
dominator	acquire	3,3	1,4
	renounce	4,1	2,2

Matrix 6

		dominator	
		acquire	renounce
moderate	acquire	4,3	1,4
	renounce	3,1	2,2

Matrix 7

		moderate	
		acquire	renounce
moderate	acquire	4,4	1,3
	renounce	3,1	2,2

The matrices show that ‘dominators’ who compete with others in type 2-situations find themselves in a straightforward PD. The two other cases, however, in which a ‘dominator’ competes with a ‘moderate’ or two ‘moderates’ compete with each other, are not equivalent to the AD and AG. The reason for this is that the preferences that ‘moderates’ have with regard to the competition for goods are different from their preferences concerning the use of preemptive violence. That ‘moderates’ are unwilling to attack other individuals unless they need to do so in order to survive, does not mean that they are unwilling to take possession of desired goods if they can do so without a fight. Even if they do not need these goods in order to survive, the mere fact that they desire them should provide them with a sufficient reason for taking action where they can do so safely.

The chance to thus take the desired good into possession, however, will only actually present itself when a ‘moderate’ faces another ‘moderate’. As can be seen from matrices 6 and 7, to try to acquire a desired good is always a strict dominant strategy for a ‘dominator’. Again, therefore, ‘dominators’ do not require any knowledge about their possible opponent in order to make their decision. As soon as ‘moderates’ are up against such a ‘dominator’, they will have to renounce the good in question in order not to be drawn into a fight. If two moderate individuals face each other, which situation is equivalent to the ‘chicken game’ discussed earlier, the case is not so obvious. Both individuals will certainly attempt to avoid a fight, which suggests that the outcome of mutual renunciation (2,2) will obtain. However, this outcome is not an equilibrium which means that both individuals would alter their strategy if they found out about the decision of their opponent. Since it is not impossible that the individuals will in fact acquire this knowledge and, the desired good still being available, will get the chance to renew their decision, the outcome of mutual renunciation is not stable. Even

if there are repetitions of the game, however, it can be assumed that a fight will be avoided at any costs, even if this means that both individuals have to finally give up on the desired good.

For exactly this reason, the knowledge of whether they face a 'dominator' or a 'moderate' is, again, not of vital importance for the decision of 'moderates', either. Their initial move will always be to refrain from trying to take the desired good into possession in order to avoid a fight. And if, due to the actual behaviour of the other individual, the opportunity arises to acquire the good at some later point, moderate individuals will still be most anxious to avoid a violent confrontation. Type 2-situations will hence only result in actual fighting if two 'dominators' face each other. Thus the game-theoretic analysis of type 2-situations demonstrates once more that those individuals who are merely interested in their own preservation will enter into violent competition for goods only if there are type 3-situations, that is, situations in which the desired good is of direct and vital importance for their own survival.

In type 3-situations, we have principally the same two strategies as in type 2-situations: to enter into a fight or to give up the good in question. This would mean that there are the same four possible outcomes. However, the first important difference between type 2-situations and type 3-situations is that in the latter situation, there are, effectively, only three possible outcomes. Since in type 3-situations, the desired goods are necessary means of self-preservation, we can say in advance that the outcome of mutual renunciation would result in the certain death of both individuals involved. We can also say that an individual who unilaterally refrains from trying to acquire the desired good will surely die, too, while his opponent – who can take the good into possession without a fight – will survive. Therefore, individuals who find themselves in a type 3-situation can be said to face only three possible outcomes, namely certain death, certain survival and the chance of survival associated with

entering into a fight. It should be obvious that all individuals have the following order of preferences:

1. certain survival (CS)
2. chance of survival (COS)
3. certain death (CD)

If we include these specified outcomes into our game-theoretic representation of type 3-situations, we get the following matrix:

		individual 2	
		acquire	renounce
individual 1	acquire	COS,COS	CS,CD
	renounce	CD,CS	CD,CD

Now given Hobbes's anthropological and psychological assumptions, one of the two strategies, namely to dispense with the good in question, is practically not available to Hobbesian individuals. Since this strategy would directly result in the certain death of the agent, the choice of the strategy is not only wrong or irrational, but practically impossible, given that the fundamental and strongest desire of Hobbesian individuals is to avoid death. What follows from this is that type 3-situations ought to be represented by a diagram which includes only a single strategy – namely 'acquire' –, and, accordingly, only one outcome – namely 'COS,COS'. While the exact result of this inevitable violent conflict cannot be predicted, it is obvious that only one of the individuals involved will get the benefit of the good fought for and that only this individual will survive. Given this, the outcome 'COS,COS' can be further specified, because what will obtain in the end, is either the

outcome ‘CS,CD’ or the outcome ‘CD,CS’. Since it cannot be indicated who will survive in the course of the actual conflict, it seems appropriate to describe type 3-situations as a game against nature. A graphic description could then assume the following form:

		Nature			
individual 1	acquire	1 wins CS,CD	2 wins CD,CS	acquire	individual 2

The diagram shows that type 3-situations represent zero-sum games in which no mutually cooperative strategy exists, but in which one player necessarily loses what the other gains. Since Hobbes’s derivation of the ‘state of war’ starts from such situations and since the necessity of anticipatory violence is dependent on their occurrence, it can be concluded that a full game-theoretic representation of Hobbes’s ‘war of every man against every man’ needs to draw upon the zero-sum game ‘competition for vital goods’.

6. Hobbes and game theory revisited

Our game-theoretic analysis of Hobbes’s derivation of the state of war has shown that the previous attempts to describe Hobbes’s natural state with the means of game theory fail to do justice to the complexity of Hobbes’s argument. Thus it has not thoroughly been distinguished between the distinct problems of anticipation and covenant-keeping, and in dealing with the former problem, commentators have not given due attention to Hobbes’s distinction between ‘moderates’ and ‘dominators’. If this distinction is allowed its proper place, it becomes obvious that neither the PD, nor the AG, nor the AD, if taken on their own, provide an appropriate representation of the problem of anticipatory violence. Instead, all the games are

required in order to describe the underlying situations of choice, a fact that has properly been acknowledged only by Graham Dodds and David Shoemaker.

Yet, even these commentators fail to identify a second crucial omission of the previous game-theoretic debate, namely the general neglect of the problem of competition from which the necessity of anticipatory violence arises in the first place. The game-theoretic analysis of this particular problem reveals that, in addition to the occurrence of PD, AG and AD situations, Hobbes's natural state is plagued with other games as well, most importantly with zero-sum games in which mutual cooperation is straightforwardly impossible. To give attention to these games is all the more important given that their influence on the problem of anticipation is a permanent one. The function assigned to the competition for goods in the framework of Hobbes's argument may only be to explain the initial occurrence of violent conflicts which turns the state of nature into a state of diffidence and creates the need for anticipation. It should be obvious, however, that within the general 'state of war', such situations will remain to occur. In fact, there are good reasons to assume that situations of competition for necessary but scarce goods will strongly increase in number, given the overall increase in violent confrontations and its negative effect on human culture and human industry.

There will, therefore, be a dynamic in the state of nature towards these situations, and it seems plausible to assume that this dynamic will directly affect the behaviour of the individuals. As soon as the state of nature individuals discover that there will be situations of competition for vital goods and that these situations tend to increase, they shall generally be anxious to avoid the occurrence of such situations as well as possible. They will do so by trying to accumulate as much power and as many goods as possible in order to prevent themselves from coming to depend on any particular goods which are not yet in their possession. This means that they will attack other individuals and attempt to kill or subdue

them even where they could, in principle, refrain from doing so, like, for example, in situations in which they deal with individuals who seem peaceful and willing to cooperate. Thus even ‘moderates’ will in many if not in most cases tend to treat the ‘attack’ strategy as a dominant strategy and attack peaceful individuals in order to avoid a personal scarcity of goods which forces them into a fight with another, possibly just as desperate individual. These considerations help to confirm once more that and why constant cooperation will not come about in the natural state. Moreover, one may wonder whether this unnoticed impact of zero-sum games could be one of the reasons why the false view that Hobbes’s state of nature constitutes a PD in which the use of force represents a strict dominant strategy for all individuals has been so persistent.

In order to conclude our analysis, we should try to assess the value of applying game theory to Hobbes’s state of nature argument and ask whether commentators such as Sorell are right in questioning its benefit. It seems to me that analysing Hobbes’s argument with the means of game theory can make a fruitful contribution to interpreting Hobbes’s theory. The application of game theory not only requires a particularly precise and thorough analysis of Hobbes’s statements, but also furthers such an analysis by sharpening the view for the implications of those statements and for possible inconsistencies in Hobbes’s argument. This view seems to be confirmed by the influence the previous game-theoretic discussion has exacted on Hobbes scholarship in the past. The attempts to game-theoretically analyse Hobbes’s theory have resulted in a better understanding of Hobbes’s state of nature argument and have caused a greater sensitivity among Hobbes scholars for the nature and implications of Hobbes’s derivation of the ‘state of war’. This positive impact of the game-theoretic discussion of Hobbes clearly outweighs the fact that the debate sometimes moved into the wrong direction because too much emphasis was given to problems of minor importance.

However, although the application of game theory to Hobbes has proven useful, and although our analysis has not provided any new reasons for denying its general legitimacy, there remain good reasons not to become too enthusiastic about it. It seems that the sensible application of game theory to Hobbes's theory is, after all, a very limited one. It seems to be restricted to the relatively general enterprise of representing the crucial situations of choice within Hobbes's natural state by means of more or less general matrices. A more detailed use of game theory, one that would include more detailed mathematical descriptions and evaluations of the different strategies and outcomes, is prevented by the fact that Hobbes neither spells out these strategies and outcomes nor the actual reasons which lead the individuals to prefer one strategy to another with the accuracy required for such an analysis. It should be obvious that the attack strategy can *de facto* be realized by very distinct actions, and that, due to this fact, the outcome of mutual attacking can bear many different results which need to be assessed quite differently.^{xxxix} A detailed game-theoretical analysis of the relevant state of nature situations of choice would need to take into account all these cases and develop sophisticated patterns of preference for the players with regard to all the possible outcomes. There can hardly be any doubt, however, that Hobbes's quite concise statements do not provide a sufficient basis for such an enterprise.

What adds to this problem is that this detailed game-theoretic description cannot dispense completely with attributing probabilities to the possible outcomes. As has been indicated above, as long as the predicament of the state of nature is taken to be that of a PD situation, which means that individuals have a strict dominant strategy, it seems sufficient to just order the possible outcomes in view of the general preferences of the agents, and something to the same effect applies to zero-sum games. However, if there are also situations in which no truly dominant strategy is available, like in the AG and the AD, the question of how the different outcomes relate to each other in terms of probability and actual utility

becomes much more pressing. Whether the decision of ‘moderates’ to attack other individuals is really, all things considered, rational in a given situation, can only be evaluated if the probability of each possible outcome and the utility these outcomes provide are taken into account. Yet, relations such as these can simply not be determined on the basis of Hobbes’s statements. A more detailed reconstruction of Hobbes’s state of nature argument, therefore, depends upon informations which Hobbes does not provide.

It can be concluded, therefore, that it is hardly possible to provide a full game-theoretic evaluation of Hobbes’s state of nature argument and remain, at the same time, close to Hobbes’s actual text. The game-theoretic analysis of Hobbes’s statements can help to show that and how individually rational behaviour can lead the inhabitants of the state of nature into a state which is contrary to reason. It can thereby help to dispel the impression, widespread among Hobbes’s contemporaries, that Hobbes’s argument gets caught up in inconsistencies at this point. However, the game-theoretic analysis does not suffice to really evaluate Hobbes’s state of nature argument in his own terms. It may generally be possible to develop more complex and sophisticated game-theoretic descriptions of state of nature situations and to fill out some of the gaps in Hobbes’s description. In doing so, however, one will inevitably veer away from Hobbes’s actual theory and move from the enterprise of interpreting and explicating Hobbes’s theory to the enterprise of providing a new Hobbesian – or anti-Hobbesian – theory of the dangers of life outside civil society.

Bibliography:

Alexandra, Andrew. 1992. “Should Hobbes’s State of Nature be represented as a Prisoner’s Dilemma?” *Southern Journal of Philosophy* 30:1–16.

Barry, Brian. 1965. *Political Argument*. New York: Humanities Press.

- Berkowitz, Peter. 1999. *Virtue and the Making of Modern Liberalism*. Princeton: Princeton University Press.
- Boonin-Vail, David. 1994. *Thomas Hobbes and the science of moral virtue*. Cambridge: Cambridge University Press.
- Boulting, Noel. 2005. "Ought Hobbes's Natural Condition of Mankind be represented as a Prisoner's Dilemma?" *Hobbes Studies* 18:27–49.
- Buchanan, James M. 1965. "Ethical Rules, Expected Values, and Large Numbers." *Ethics* 76:1–13.
- Buchanan, James M. 1975. *The Limits of Liberty. Between Anarchy and Leviathan*. Chicago: The University of Chicago Press.
- Curley, Edwin. 1989. "Reflections on Hobbes: Recent Work on his Moral and Political Philosophy." *Journal of Philosophical Research* 15:169–250.
- Curley, Edwin. 1994. Introduction and Notes to *Leviathan*, by Thomas Hobbes. With selected variants from the Latin edition of 1668. Edited by Edwin Curley. VIII–LXXVII and 1–548. Indianapolis: Hackett.
- Dodds, Graham G., and David W. Shoemaker. 2002. "Why we can't all just get along: Human variety and game theory in Hobbes's state of nature." *Southern Journal of Philosophy* 60:345–374.
- Eggers, Daniel. 2008. *Die Naturzustandstheorie des Thomas Hobbes. Eine vergleichende Analyse von The Elements of Law, De Cive und den englischen und lateinischen Fassungen des Leviathan*. Berlin; New York: de Gruyter.
- Eggers, Daniel. 2009. "Liberty and contractual obligation in Hobbes" *Hobbes Studies* 22:70–103.
- Esfeld, Michael. 1995. *Mechanismus und Subjektivität in der Philosophie von Thomas Hobbes*. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

- Ewin, R. E. 1991. *Virtues and rights. The moral philosophy of Thomas Hobbes*. Boulder; San Francisco; Oxford: Westview Press.
- Gauthier, David. 1969. *The Logic of Leviathan. The Moral and Political Theory of Thomas Hobbes*. Oxford: Clarendon Press.
- Gauthier, David. 1988. "Hobbes's Social Contract." In *Perspectives on Thomas Hobbes*, edited by G. A. J. Rogers, and Alan Ryan, 125–152. Oxford: Oxford University Press.
- Haji, Ishtiyaque. 1990. "The Symmetry Enigma in Hobbes." *Dialogue* 29:189–204.
- Haji, Ishtiyaque. 1991. "Hampton on Hobbes on State-of-Nature Cooperation." *Philosophy and Philosophical Research* 51:589–601.
- Hampton, Jean. 1986. *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.
- Hampton, Jean. 1991. "Cooperating and contracting. A Reply to I. Haji's „Hampton on Hobbes on State-of-Nature Cooperation”." *Philosophy and Phenomenological Research* 51: 603–609.
- Hardin, Russell. 1991. "Hobbesian Political Order." *Political Theory* 19:156–180.
- Harrison, Ross. 2003. *Hobbes, Locke, and Confusion's Masterpiece. An Examination of Seventeenth-Century Political Philosophy*. Cambridge: Cambridge University Press.
- Harvey, Martin. 2004. "Teasing a limited deontological theory of morals out of Hobbes." *The Philosophical Forum* 35:35–50.
- Hobbes, Thomas. 1651. *Leviathan, or The Matter, Forme, & Power of a Commonwealth Ecclesiasticall and Civill*. London: Andrew Crooke.
- Hobbes, Thomas. 1668. *Leviathan, sive De Materia, Forma, & Potestate Civitatis Ecclesiasticae et Civilis*. Amsterdam: Blaeu.
- Hobbes, Thomas. 1969. *The Elements of Law Natural and Politic*. Edited by Ferdinand Tönnies. Second Edition by M. M. Goldsmith. London: Frank Cass.

- Hobbes, Thomas. 1983. *De Cive. The latin version entitled in the first edition Elementorum Philosophiae Sectio Tertia De Cive and in later editions Elementa Philosophica De Cive*. Edited by Howard Warrender. Oxford: Oxford University Press.
- Höffe, Otfried. 1982. "Wissenschaft im Dienste freier Selbsterhaltung. Zum Theorie-Praxis-Verhältnis in Thomas Hobbes' Staatsphilosophie." In *Furcht und Freiheit. Leviathan-Diskussion 300 Jahre nach Hobbes*, edited by Udo Bernbach, and Klaus-Michael Kodalle, 30–64. Opladen: Westdeutscher Verlag.
- Hoekstra, Kinch. 2007. "Hobbes on the Natural Condition of Mankind." In *The Cambridge Companion to Hobbes's Leviathan*, edited by Patricia Springborg, 109–127. Cambridge: Cambridge University Press.
- Hüning, Dieter. 1995. "Kant auf den Spuren von Thomas Hobbes?" In *Proceedings of the Eighth International Kant Congress Memphis 1995. Vol.II. Part 2. Sections 10-18*, edited by Hoke Robinson, 761–771. Milwaukee/Wisconsin: Marquette University Press.
- Hüttemann, Andreas. 2004. "Naturzustand und Staatsvertrag bei Hobbes." *Zeitschrift für philosophische Forschung* 58:29–53.
- Kavka, Gregory S. 1983. "Hobbes's war of all against all." *Ethics* 93:291–310.
- Kavka, Gregory S. 1986. *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.
- Kersting, Wolfgang. 1992. *Thomas Hobbes zur Einführung*. Hamburg: Junius.
- Kisner, Matthew. 2004. "Lions, Foxes, and Polecats: Is it rational for Hobbesian subjects to covenant?" *History of Philosophy Quarterly* 21:81–100.
- Kleemeier, Ulrike. 2002. *Grundfragen einer philosophischen Theorie des Krieges. Platon – Hobbes – Clausewitz*. Berlin: Akademie Verlag.
- Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions. Introduction and Critical Survey*. New York; London; Sydney: Wiley.

- Ludwig, Bernd. 1998. *Die Wiederentdeckung des Epikureischen Naturrechtes. Zu Thomas Hobbes' philosophischer Entwicklung von De Cive zum Leviathan im Pariser Exil 1640-1651*. Frankfurt a.M.: Klostermann.
- McLean, Iain. 1981. "The Social Contract in Leviathan and the Prisoner's Dilemma Supergame." *Political Studies* 29:339–351.
- McNeilly, F. S. 1968. *The Anatomy of Leviathan*. London; Melbourne; Toronto; New York: Macmillan.
- Mueller, Dennis C. 2003. *Public Choice III*. Cambridge: Cambridge University Press.
- Neal, Patrick. 1988. "Hobbes and Rational Choice Theory." *Western Political Quarterly* 41:635–652.
- Nida-Rümelin, Julian. 1996. "Bellum omnium contra omnes. Konflikttheorie und Naturzustandskonzeption im 13. Kapitel des Leviathan." In *Thomas Hobbes. Leviathan oder Stoff, Form und Gewalt eines bürgerlichen und kirchlichen Staates*, edited by Wolfgang Kersting, 109–130. Berlin: Akademie Verlag.
- Pasquino, Pasquale. 2001. "Hobbes, Religion, and Rational Choice: Hobbes's Two Leviathans and the Fool." *Pacific Philosophical Quarterly* 82:406–419.
- Piirimäe, Pärtel. 2006. "The Explanation of Conflict in Hobbes's *Leviathan*." *Trames. A Journal of the Humanities and Social Sciences* 10:3–21.
- Raphael, D. D. 1977. *Hobbes. Morals and Politics*. London: Allen & Unwin.
- Rawls, John. 1971. *A theory of justice*. Cambridge, MA: Harvard University Press.
- Ryan, Alan. 1996. "Hobbes's political philosophy." In *The Cambridge companion to Hobbes*, edited by Tom Sorell, 208–245. Cambridge: Cambridge University Press.
- Sen, Amartya K. 1967. "Isolation, assurance and the social rate of discount." *The Quarterly Journal of Economics* 81:112–124.
- Sen, Amartya K. 1969. "A game-theoretic analysis of theories of collectivism in allocation."

- In *Growth and Choice*, edited by T. Majumdar, 1–17. Oxford: Oxford University Press.
- Shaver, Robert. 1990. “Leviathan, King of the Proud.” *Hobbes Studies* 3:54–74.
- Siep, Ludwig. 1974. “Der Kampf um Anerkennung. Zu Hegels Auseinandersetzung mit Hobbes in den Jenaer Schriften.” *Hegel-Studien* 9:155–207.
- Slomp, Gabriella, and Mafredi La Manna. 1996. “Hobbes, Harsanyi and the Edge of Abyss.” *Canadian Journal of Political Science* 29:47–70.
- Strauss, Leo. 1952. *The political philosophy of Hobbes. Its basis and its genesis*. Translated by Elsa M. Sinclair. Chicago: University of Chicago Press.
- Taylor, A.E. 1993. “The ethical doctrine of Hobbes.” In *Thomas Hobbes. Critical Assessments. 4 Volumes. Volume 2: Ethics*, edited by Preston King, 22–39. London; New York: Routledge & Kegan.
- Taylor, Michael. 1976. *Anarchy and Cooperation*. London; New York; Sydney: Wiley.
- Tricaud, Francois. 1988. “Hobbes’s conception of the state of nature from 1640 to 1651: Evolution and ambiguities.” In *Perspectives on Hobbes*, edited by G. A. J. Rogers, and Alan Ryan, 107–123. Oxford: Oxford University Press.
- Tuck, Richard. 1989. *Hobbes*. Oxford: Oxford University Press.
- Tuck, Richard. 1996. “Hobbes’s moral philosophy.” *The Cambridge companion to Hobbes*, edited by Tom Sorell, 175–207. Cambridge: Cambridge University Press.
- Ullmann-Margalit, Edna. 1977. *The Emergence of Norms*. Oxford: Clarendon Press.
- Warrender, Howard. 1957. *The political philosophy of Hobbes. His theory of obligation*. Oxford: Oxford University Press.
- Wolf, Friedrich O. 1969. *Die neue Wissenschaft des Thomas Hobbes. Zu den Grundlagen der politischen Philosophie der Neuzeit*. Stuttgart-Bad Cannstatt: Frommann.

* For comments on earlier versions of this paper, I am grateful to Kurt Bayertz, Annabel Brett, Wilfried Hinsch, Dirk Brantl and Anna Maj Blundell. I am also indebted to two anonymous referees for the *Southern Journal of Philosophy* for their helpful criticisms and suggestions.

ⁱ See, for example, Taylor 1976; Ullmann-Margalit 1977; McLean 1981; Kavka 1983; Kavka 1986; Hampton 1986; Neal 1988; Gauthier 1988; Haji 1990; Shaver 1990; Hampton 1991; Haji 1991; Hardin 1991; Boonin-Vail 1994; Nida-Rümelin 1996; Slomp/La Manna 1996; Alexander 2001; Pasquino 2001; Dodds/Shoemaker 2002; Kisner 2004; Hüttemann 2004; Boulting 2005; and Piirimäe 2006. See also Barry 1965; Rawls 1971; Curley 1989; Curley 1994; Ryan 1996; Harrison 2003; and Hoekstra 2007.

ⁱⁱ That Hobbes's state of nature shows all the essential features of the PD has, for example, been argued by Michael Taylor, Edna Ullmann-Margalit and Ian McLean (see Taylor 1976: 7; Ullmann-Margalit 1977: 62f.; and McLean 1981: 339f.). See also Rawls who even describes Hobbes's state of nature as the "classic example" of the PD (see Rawls 1971: 269). Jean Hampton concedes that at least the 'rationality account of conflict' developed by Hobbes in chapter 13 of *Leviathan* is appropriately represented by a PD matrix (see Hampton 1986: 61f.). More recently, PD interpretations have been advocated by Boonin-Vail, Nida-Rümelin, Pasquino and Hüttemann, and Boulting has tried to defend the PD interpretation against some of the most popular challenges.

ⁱⁱⁱ See, for example, Alexandra who holds that the state of nature is more appropriately represented by the assurance game, a game attributed to Amartya Sen (see Alexandra 1992: 12f.; for a general description of the assurance game, see Sen 1967; and Sen 1969). This view is also suggested by Shaver (see Shaver 1990: 58). Gauthier seems to be favouring the AG reading as well, although he refrains from using the term (see Gauthier 1988: 132). A variant of the AG interpretation has been recently been forwarded by Pärtel Piirimäe who describes

the state of nature as an “Assurance game with glory” (Piirimäe 2006: 16). Graham Dodds and David Shoemaker have argued that in game-theoretically describing the state of nature, one needs to also draw upon a game they call the “assurance dilemma” (Dodds/Shoemaker 2002: 355; for similar considerations see Taylor 1976: 110; and Kavka 1986: 113f.). Further readings have been forwarded by Patrick Neal who describes the state of nature as a ‘coordination game’ (see Neal 1988: 642), and by Gabriella Slomp and Manfredi La Manna who favor a ‘chicken game’ interpretation (see Slomp/La Manna 1996: 57).

^{iv} See, for example, Sorell 1986: 152.

^v See, for instance, Ewin who rejects all game-theoretic analyses of Hobbes as anachronistic (see Ewin 1991: 46), and Hüning and Ludwig who deny that the theoretical conditions for legitimately applying game theory to Hobbes’s argument are satisfied (see Hüning 1995: 762; and Ludwig 1998: 300f.). For a similar point, see Tuck 1989: 106ff. See also Neal and Slomp/La Manna who apply game theory to Hobbes, but only in order to point to the insufficiency of the orthodox rational choice-approach (see Neal 1988: 635f.; and Slomp/La Manna 1996: 65f.).

^{vi} For deontological readings of Hobbes’s theory of obligation, see Taylor 1993. See also Raphael 1977; Ludwig 1998; and Harvey 2004. For a detailed refutation, see Eggers 2009.

^{vii} For a discussion of the charge of anachronism, see Curley 1994: XXVI.

^{viii} To clearly distinguish the different state of nature problems and the particular situations of choice associated with them, is all the more required because, according to some commentators, these problems ask for different game-theoretic representations. Thus Robert Shaver and David Boonin-Vail criticise Hampton’s treatment of the problems of violence and covenant-keeping and hold that, contrary to what Hampton claims, Hobbes can, without any contradiction, emphasize the rationality of anticipatory violence and consider a second-party’s nonperformance of a valid covenant irrational (see Shaver 1990: 58f.; and Boonin-Vail 1994:

129. See also Dodds/Shoemaker 2002: 346). That the different situations of choice that can arise in the state of nature may constitute opposite games is also suggested by the striking fact that most authors who view the state of nature as a PD tend to base their interpretations on Hobbes's derivation of the 'state of war' and the necessity of anticipation while those authors who view the state of nature as an AG tend to appeal to Hobbes's remarks concerning the performance of contracts and covenants. It would appear, then, that the situations of choice associated with these two problems might in a relevant sense be different.

^{ix} See Shaver 1990: 64. See also Neal 1988: 643ff.; Hardin 1991: 159 and 172; Nida-Rümelin 1996: 123; Pasquino 2001: 407f. and 411; and Hüttemann 2004: 45.

^x See *Leviathan*, 63. For a more detailed discussion of the historicity of Hobbes's state of nature, see Eggers 2008: 34ff.

^{xi} See Kleemeier 2002: 127f.

^{xii} That perfect rationality is presupposed by Hobbes's state of nature theory is suggested by Haji (see Haji 1990: 189ff.). Kavka seems to have a similar view, but is less explicit (see Kavka 1983: 303; and Kavka 1986: 84f.). That Hobbes does not conceive of his state of nature individuals as perfectly rational is emphasised by Curley (see Curley 1989: 181) and Berkowitz: "Far from having a rational actor model of politics, Hobbes may be said to have an irrational actor model." (Berkowitz 1999: 51)

^{xiii} See *De Cive*, 96.

^{xiv} For a similar point, see Neal 1988: 646.

^{xv} This has influentially been argued by Howard Warrender (see Warrender 1957: 30ff.).

^{xvi} One reason for this is that virtually all game-theoretic analyses of Hobbes forwarded have thus been confined to the English *Leviathan*. Another reason is that the argument of the English *Leviathan* is the one most sympathetic to a game-theoretic analysis of Hobbes's argument. As can be shown by thoroughly comparing Hobbes's different works, the

derivation of the ‘state of war’ experiences a significant shift between 1640 and 1668 (for a detailed discussion, see Eggers 2008: 60ff. See also Strauss 1952; McNeilly 1968; Wolf 1969; Siep 1974; Hampton 1986; Tricaud 1988; and Esfeld 1995). In the two earlier works, Hobbes allows human irrationality a larger impact on the emergence of the ‘state of war’ than in the two versions of *Leviathan* in which he tries hard to derive the ‘state of war’ from the rational pursuit of self-preservation alone. Moreover, it is only in *Leviathan* that the necessity of anticipatory violence – which functions as some kind of tacit presupposition in both *The Elements* and *De Cive*, but is explicitly established only when the derivation of the ‘state of war’ is already completed –, is given its proper place. Given this, the *Leviathan* seems to provide the appropriate basis for a game-theoretic analysis of Hobbes’s discussion of anticipatory violence, the English version chosen here mainly for the sake of convenience.

^{xvii} *Leviathan*, 61.

^{xviii} This is claimed by Kleemeier (see Kleemeier 2002: 140f.).

^{xix} *Leviathan*, 61.

^{xx} In the Latin version of *Leviathan*, the passage stating the necessity of anticipation is altered in a way that may seem to deprive this reading of its basis. Due to the substitution of “Quoniam” for “Also” and of “qui animi & gloriae causâ universum terrarum orbem superare vellent” for “that taking pleasure in contemplating their own power in the acts of conquest, which they pursue farther than their security requires”, the striving of ‘dominators’ is no longer explicitly presented as an additional reason for the ‘moderates’’s use of anticipatory violence but rather as the only reason: “Quoniam enim sunt, qui animi & gloriae causâ universum terrarum orbem superare vellent, nisi alii (alioqui modicis contenti finibus) potentiam suam alios invadendo auferent, sed tantummodo se & sua defendere conarentur, subsistere diu non possunt.” (Latin *Leviathan*, 64). The problem with this, however, is that the former passage describing the competition for goods is now even less ambiguous than the

English in claiming that the goal of self-preservation leads individuals into violent conflict with others, no matter whether they are moderate or not: “Quoties ergo duo idem cupiunt, quo frui ambo non possunt, alter acrius hostis fit, & ad finem sibi propositum (*quae est conservatio propria*) alter alterum conatur subjugare vel interficere.” (Latin *Leviathan*, 63; my emphasis). Since Hobbes continues to derive the necessity of anticipation from the competition for goods thus described (“In tanto, & mutuo hominum metu”), the Latin text is not wholly coherent on what constitutes the need for anticipation and on whether this need depends on the existence of ‘dominators’. There do not seem to be good reasons, then, to prefer the Latin to the English version which, in contrast, present a perfectly consistent picture.

^{xxi} As a matter of fact, one may even think that the presupposition of scarcity is already needed in order to make sense of Hobbes’s reference to goods that two people cannot enjoy together. The claim that a good for which two individuals compete cannot be divided or used in common seems plausible if we are allowed to assume that the good in question is needed in its entirety to ensure any of the two individual’s survival for more than the immediate future. It seems even more plausible if we are allowed to assume that the individuals desire the good not only for themselves, but in order to ensure the survival of their families. However, if we are denied these assumptions and are not allowed to view the goods as scarce in this sense, it is hard to see what kinds of goods Hobbes could be thinking of.

^{xxii} For the view that Hobbes’s argument presupposes a scarcity of goods, see McNeilly 1968: 143; Gauthier 1969: 18; Hampton 1986: 60; Kersting 1992: 104f.; Esfeld 1995: 223; Nida-Rümelin 1996: 126; Ryan 1996: 220f.; Kleemeier 2002: 136f.; Dodds/Shoemaker 2002: 347; and Hoekstra 2007: 110. For a diverging view, see Höffe 1982: 43; and Tuck 1996: 185. For a more detailed discussion of the claimed differences between *Leviathan* and the earlier works, see Eggers 2008: 100ff. and 119f.

^{xxiii} My emphasis.

^{xxiv} See *Leviathan*, 61.

^{xxv} *Leviathan*, 61.

^{xxvi} For a more detailed characterisation of the PD, such as for the general features of game theory, see Luce/Raiffa 1957.

^{xxvii} It is quite common in the relevant literature on Hobbes to merely use ordinal numbers and not to attribute actual utilities to the possible outcomes. The reason for this is, probably, that the discussion has focused from the beginning on the similarities between the state of nature and the PD. Since ‘non-cooperation’ is a strict dominant strategy in the PD, both the actual utilities and the probabilities of the outcomes are virtually irrelevant because under no possible scenario, ‘cooperation’ can become the rational thing to do. However, if the state of nature, or the problem of anticipation, cannot be represented by the PD, such issues may become important again. I will try to discuss the implications of this in the final chapter.

^{xxviii} The fact that the state of nature is a state in which more than two individuals interact loses much of its relevance if we do not try to describe the state of nature as one large game, but – as proposed in chapter 2 – focus on single situations of choice which are much more sympathetic to being described as two-person or two-party games. Moreover, an expansion of the PD to a ‘multi-party’ game leaves the essential features of the game unaffected.

^{xxix} Some commentators claim, in view of the dynamics of the state of nature, that the situation of Hobbesian individuals ought to be represented by an ‘iterated’ PD or PD ‘supergame’ (see, for instance, Taylor 1976: 105ff.; McLean 1981: 342ff.; Hampton 1986: 80ff.), or even claim that this situation cannot be grasped with game-theoretical devices at all (see Sorell 1986: 152). However, even if these worries might have some foundation with regard to the problem of covenant-keeping, they do not seem to apply to the problem here discussed, that is, to the problem of anticipation. One of the outcomes that the individuals face

in the relevant situations is their immediate death or at least their enslavement by a superior opponent. This means, however, that when the individuals have to decide whether to attack or not to attack each other, they do not know whether they will survive or retain their freedom at all, that is, whether there will be any further choices for them to make. There seems, therefore, to be no place for those considerations that distinguish the (infinitely-repeated) super game from the one-shot PD, like, for example, the consideration to behave cooperatively in order to acquire a certain reputation that might pay in future encounters and might then prevent oneself from being pre-emptively attacked by others. Moreover, even if an individual might value such a reputation so highly as to be generally willing to risk losing his life by not attacking another individual, it still does not seem rational for him to actually do so. Thus it is doubtful whether it will be possible at all to gain a reputation in this sense or for such a reputation to actually pay. The view that this is, in fact, possible depends upon very strong assumptions that cannot be taken to hold with regard to Hobbes's state of nature as a state with large numbers of individuals and without stable social structures: The assumption that individual A can reasonably assume to encounter individual B again in the future, or the assumption that if he does not encounter individual B again, individual B will at least have spread the news of his earlier behaviour to others which he will encounter in the future and who will then, while still contemplating to strike out pre-emptively, be able to identify him as individual A and take his previous behaviour into consideration (That the size of a group is highly relevant for the rationality of non-cooperative behaviour is emphasized in Buchanan 1965; and Buchanan 1975: 66ff. See also Mueller 2003: 12f. and 41f.).

^{xxx} *The Elements of Law*, 71. See also *De Cive*, 94.

^{xxxi} As indicated in the introduction, only few of the many commentators who applied game-theory to Hobbes have given any attention to the distinction between 'dominators' and 'moderates'. One of those few exceptions is Kavka who, instead of describing the state of

nature as a PD, describes it as a “quasi-prisoner’s dilemma” in which the use of pre-emptive violence is not a strict dominant, but only a “quasi-dominant” strategy (see Kavka 1986: 113). From my point of view, however, the AD interpretation, forwarded by Dodds/Shoemaker, is the more promising suggestion. The AD interpretation can already be found in Michael Taylor’s study (see Taylor 1976: 110). Yet, Taylor rejects this way of representing the ‘state of war’ in the end, with the somewhat perplexing statement that there is not sufficient evidence in the English *Leviathan* for assuming that Hobbes did not attribute to all individuals exactly the same nature.

^{xxxii} See, for example, Luce/Raiffa 1957: 5 and 49.

^{xxxiii} For similar considerations, see Dodds/Shoemaker 2002: 356ff.

^{xxxiv} See Neal 1988: 642.

^{xxxv} See Slomp/La Manna 1996: 50f.

^{xxxvi} See *Leviathan*, 61.

^{xxxvii} See Piirimäe 2006: 13ff.

^{xxxviii} See Piirimäe 2006: 16ff.

^{xxxix} Thus the use of anticipatory violence could, for instance, be realized by attacking someone in an open field, by attacking someone while he is sleeping, by pretending to be peaceful and attacking after one has made friends etc., and the attack itself could be launched by using a knife, by using a piece of wood, by using one’s own bare hands etc. It should be obvious, however, that the exact nature of the attack will have an important impact on the result of the ensuing fight. Generally, it needs to be emphasised that even the best possible outcome of unilateral attack can lead to the death of the aggressor while the unsuspecting individual thus attacked can still emerge victorious.

