

THE EXPLANATION OF CONFLICT IN HOBBS'S *LEVIATHAN*

Pärtel Piirimäe

St. John's College, Cambridge

Abstract. Thomas Hobbes's thesis of the necessity of an absolute sovereign, put forward in *Leviathan* (1651), rests upon the argument that the condition of anarchy is a condition of violent conflict. It is therefore crucial for Hobbes to demonstrate that men, despite being predominantly rational creatures, are unable to arrange and keep cooperative agreements without enforcement by the state. In recent decades it has been fashionable to explain Hobbes's account of conflict with game-theoretical tools borrowed from modern economics. This article accepts the application of game theory as a legitimate and useful way of studying Hobbes, but argues that the commentators have often strayed too far from Hobbes's own text, misrepresenting his fundamental psychological and ethical premises. The article is an attempt to rectify that. After an outline of Hobbes's account of conflict and a critical survey of its current game-theoretical interpretations, it suggests a novel game-theoretical explanation, which, the author hopes, is a more precise representation of what Hobbes actually says in *Leviathan*.

Keywords: Thomas Hobbes, conflict, war, glory, state of nature, game theory

1. Introduction

Thomas Hobbes famously argues in *Leviathan* (1651) that the state of nature is a state of “warre, as is of every man, against every man”. In such a condition, man not only lives in “continuall feare, and danger of violent death” but even his potentially short life is utterly miserable because without security there is no industry, agriculture, commerce, science or arts. In sum, the life of man is “solitary, poore, nasty, brutish and short” (Hobbes 1991: ch. 13, 88–89).¹

Hobbes uses the state of nature as a device for demonstrating the necessity of political society. Furthermore, by showing that the pre-political condition is an

¹ Henceforth I refer to *Leviathan* only with the chapter number, followed by the page number. I shall use “he” and “man” (following Hobbes's usage) in gender-neutral meaning, with the exception of 2-person games where the row player will be “he” and the column player “she”.

intolerable state of permanent conflict, he hopes to demonstrate the necessity of a specific kind of political society, namely that which is governed by an undivided and absolute sovereign. He argues that the worst that can happen to man is a reversal to the state of nature, which is essentially what happens when society gets torn apart by civil war – a situation that Hobbes himself witnessed in his lifetime. A sovereign with absolute power, he claims, is the best security against this ever happening. Whatever evils such unlimited power itself might bring, it is a necessary risk because the evils arising from the lack of such power are far greater.

It is clear that Hobbes's argument for absolutism depends on the success of his argument that the state of nature is indeed a state of war. Some contractarian political philosophers have presented a far more optimistic picture of the pre-political society, so as to be able to refute the necessity of a strong ruler (cf. Kersting 1994). John Locke, for instance, argued in *Two Treatises of Government* (1690) that the state of nature would be a relatively peaceful condition with property, industry and some sort of law enforcement, which renders the prospect of subjection to one person's absolute power distinctly unattractive (Locke 1994: bk.II, ch.2). But Hobbes must not only demonstrate that conflict in the state of nature is inevitable, he must also do it without depicting men as inherently war-loving or madly irrational because his philosophical project as a whole requires a psychological description of man who is sufficiently rational to be able to follow the prescripts of reason to establish and obey a sovereign for the sake of peace. Hobbes's task is thus to show that despite the fact that men rationally prefer peace to war, the condition of the state of nature is such that aggressive conduct advances one's aims better than peaceful behaviour.

Owing to such constraints, Hobbes's explanation of conflict in the state of nature is a fairly complex one, giving rise to considerable disputes in literature as to how it should exactly be interpreted. In the last decades it has been popular to model Hobbes's account of conflict with the help of game-theoretical tools worked out by rational-choice theorists. This approach has been widely criticized by Hobbes-scholars, as it tends to go far beyond what is provided by Hobbes's texts (cf. Hampsher-Monk 1992:25). While I agree with the critics that this has often been the case, I still believe that game theory can offer a valuable insight into Hobbes's theory, as it highlights his less explicit ideas that might otherwise remain unnoticed, and thus helps to reconstruct his account of conflict in a more systematic manner. I also believe that Hobbes fundamentally sees men as instrumentally rational utility-maximizers, which is an important precondition for the application of game theory.²

Therefore, rather than rejecting the game-theoretical devices on the whole as anachronistic, I would advocate a more careful application of them. This means avoiding, firstly, an ambitious agenda of making Hobbes's theory unassailable by standards of modern political philosophy which has led commentators such as

² Hobbes would have no problem endorsing the other important requirements of game theory, namely that people have common knowledge of their rationality and that they know the rules of the game (Hargreaves Heap and Varoufakis 1995:4–31). Cf. Slomp 2000:123.

Gregory Kavka and Jean Hampton to stray too far from Hobbes's text. Kavka does so explicitly, calling his account a "Hobbesian theory" rather than "Hobbes's theory" (Kavka 1986:xii). Hampton, on the other hand, sets out in the introduction to reconstruct Hobbes's own theory, yet argues later that since Hobbes's account "fails us", it needs to be "fixed" by "philosophizing with him", so that she ends up discussing what Hobbes "meant to say" rather than what he actually said in his text (Hampton 1986:2, 69, 86). Secondly, while assuming that Hobbes's picture of human nature permits the application of game theory, we have to be careful not to carry with it the entire set of assumptions of modern economics such as the assumption of fundamental similarity of human motivation. For an early modern thinker it was more natural to assume that people of different classes, such as nobles and commoners, have fundamentally different desires and preferences. I will pay a lot of attention to Hobbes's account of glory-seeking which distinctly reflects his belief in the variety of human motivation.

By avoiding these pitfalls, I hope that I can provide a reconstruction of Hobbes's account of conflict that is closer to his text than those previously offered. As there is no space in this article to trace the development of Hobbes's account throughout his intellectual career, I will focus on *Leviathan* only, which I take as the most mature and systematic presentation of his views. I will first present an analysis of chapter 13 of *Leviathan* where Hobbes explains the causes of conflict in the state of nature (part 2), then offer a critical survey of its most important game-theoretical interpretations (part 3) and finally propose my own game-theoretical model which, I hope, does more justice to Hobbes's account (part 4).

2. Competition, diffidence and glory

In chapter 13 of *Leviathan* Hobbes summarizes his explanation of conflict in the state of nature as follows: "So that in the nature of man, we find three principal causes of quarrell. First, Competition; Secondly, Diffidence; Thirdly, Glory. The first, maketh men invade for Gain; the second, for Safety; and the Third, for Reputation" (ch. 13, 88).

To explain how the *competition* arises in the state of nature, we need first to go back to Hobbes's account of human motivation. The very title of chapter 13 – "Of the Naturall Condition of Mankind, as concerning their Felicity, and Misery" – points to the centrality of the concept of *felicity* as the utmost aim of men, which they apparently fail to achieve in the state of nature. But for Hobbes, a vehement anti-Aristotelian, there is no objective definition of felicity. In Hobbes's mechanistic worldview, every man is in "motion" towards whatever he desires and away from whatever he is averse to (ch. 6, 37–38). Felicity is, accordingly, "continuall successe in obtaining those things which a man from time to time desireth" (ch. 6, 46).

However, there is never a moment when an individual has obtained all that he desires, for "there is no such thing as perpetuall Tranquillity of mind, while we live here; because Life it selfe is but Motion, and can never be without Desire, nor

without Feare” (ibid.). Hobbes’s point is that human activity is future-oriented: men are concerned not only with the satisfaction of their present desires but also with their ability to satisfy their desires in the future. Therefore men are concerned with *power* which Hobbes defines as man’s “present means to obtain some future apparent Good” (ch. 10, 62). Whatever one may lack in “Naturall Power”, that is “the eminence of Faculties of Body, or Mind”, one may compensate by an increase in “Instrumentall Powers” such as wealth, reputation and friends (ibid.; cf. Sorell 1986:100–101).

The problem with power is that it is inflationary, because, in Hobbes’s definition, it is relative to other people’s power (cf. Gauthier 1969:10ff.). My power to obtain and hold on to certain things is sufficient only when it is superior to the power of the others who desire the same things. Therefore, even if people do not seek power for power’s sake, they must necessarily acquire more power in order to safeguard for the future “the power and means to live well” that they have presently obtained (ch. 11, 70). This is the reason for “a generall Inclination of all mankind, a perpetuall and restlesse desire of Power after power, that ceaseth only in Death” (ibid.). The permanently unsatisfied desire for power is the key to understanding why men compete for resources. The competition does not arise just from the general scarcity of natural resources, as some commentators have assumed (Malnes 1993). Hobbes does not imply that the resources are insufficient for the sustenance of the population, were they divided equally, but they are necessarily insufficient for the satisfaction of everyone’s unceasing drive to increase one’s instrumental power.

Hobbes’s second cause of conflict is *diffidence* or lack of trust in others. People in the state of nature realise that their life is conditioned by two fundamental premises: firstly, that they are natural competitors for power (and for resources needed to increase it), and secondly, that no-one can assume a natural superiority in this competition, as they are roughly equal in their ability to kill one another. “For as to the strength of body, the weakest has strength enough to kill the strongest, either by secret machination, or by confederacy with others, that are in the same danger with himselfe” (ch. 13, 87). These factors create a climate of mutual fear which, as Alan Ryan has put it, “drives people to attack one another by the logic of the situation, no matter what their motives” (Ryan 1996:220). If someone fears that he might be attacked, then striking first is a safer option than standing on defence, as it is difficult to be constantly on the alert against machinations or confederacies. And one has reasonable grounds to fear an attack, because one understands that others reason similarly and might want to anticipate one’s possible attack, and so *ad infinitum*.

The situation is further destabilised by the fact that there are “some, that tak[e] pleasure in contemplating their own power in the acts of conquest, which they pursue farther than their security requires”. The existence of such natural aggressors and the uncertainty as to who belongs to this group makes it even more compelling for moderates, “that otherwise would be glad to be at ease within modest bounds”, to strike first. Thus “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” (ch. 13, 87–88; cf. ch. 11, 71).

The problem of natural aggressors leads us to the desire of *glory*, the third cause of conflict in the nature of man. In literature glory-seekers are usually identified with natural aggressors, “warmongers” (e.g. Martinich 2005:69), which is a misrepresentation of Hobbes’s more complex account of glory. One has to note, first, that natural aggressors are introduced in the paragraph which explains how diffidence leads via anticipation to war, whereas the issue of glory-seeking is explained in the next paragraph. This order is not accidental because aggressors are not typical glory-seekers. Hobbes says clearly that only *some* people revel in the acts of conquest, whereas glory-seeking as such is a characteristic of *every man*:

every man looketh that his companion should value him, at the same rate he sets upon himself: And upon all signes of contempt, or undervaluing, naturally endeavours, as far as he dares (which amongst them that have no common power to keep them all in quiet, is far enough to make them destroy each other,) to extort a greater value from his contemners, by damage; and from others, by the example (ch. 13, 88; cf. ch. 17, 119, ch. 18, 126).

Thus for Hobbes, glory-seeking is not an irrational passion of some people but a “natural endeavour” of “every man”. As, according to his mechanistic worldview, all bodies naturally strive towards the enhancement of their vital motion, every natural endeavour of man must be conducive to his self-preservation. How glory-seeking contributes to this becomes clear when we look at previous chapters where Hobbes presents his account of power, value, reputation, glory and honour. Hobbes says that the value of a man is the price that others would pay for the use of his power (ch. 10, 63; cf. Gauthier 1969:16). Reputation means that people set a high value on someone’s power. And this is not something that people desire out of vanity, but they are concerned with reputation because “reputation of Power, is power” (ch. 10, 62). To be undervalued is dangerous for one’s security, because other people are more inclined to attack someone whom they think less able or willing to defend himself (cf. Hampsher-Monk 1992:25). A man can build up reputation by attacking those who contemn him, because they learn from their own experience that his power is actually not inferior to theirs, at least with regard to his capacity to inflict damage on them. And this also sets an example to people not involved in the conflict, as it makes clear that he is not someone who can be subdued without resistance.

The natural passion of glory-seeking has been often wrongly identified with the passion of vain-glory (e.g. Slomp 2000:29). Vain-glorying is irrational because it does not help to attain the end of glorying which is safety. And all types of glorying “tending to no end” are vain.³ Hobbes defines glorying as “joy, arising

³ Ch. 15, 106: “Besides, revenge without respect to the Example, and profit to come, is a triumph, or glorying in the hurt of another, tending to no end; (for the End is alwayes somewhat to Come;) and glorying to no end, is vain-glory, and contrary to reason.”

from imagination of a mans own power and ability” (ch. 6, 42). Clearly a good measure of such imagination (i.e. confidence) is needed before one starts assuring others of one’s power. However, there are some people who are “vain-glorious” because their imagination of their power is grounded not on their former actions but “on the flattery of others, or only supposed by (themselves)” (ibid.). When it comes to setting examples of their power, they do not have confidence that “begetteth attempt” (ibid.) but are inclined “in the approach of danger, or difficulty, to retire if they can: because, not seeing the way of safety, they will rather hazard their honour, which may be salved with an excuse; than their lives, for which no salve is sufficient” (ch. 11, 72). Obviously, glorying which is not backed by glorious deeds is vain, as it does not help to achieve safety which is the purpose of true glory.

Another type of glorying inconsistent with self-preservation is that of natural aggressors. True, their concept of their power is based on their actual “acts of conquest”, and these acts set an example to the others, who as a result are less likely to attack them. But they also have lost the end of safety from sight, because they pursue power beyond what their security requires. They enjoy glorying to such an extent that they are prepared to risk their lives for its sake. Their glorying has become an end in itself, and we saw that glorying “tending to no end” is vain. Therefore we could plausibly speak of two kinds of vain-glorious people: “cowardly glory-seekers” who hazard their reputations and thus lives by being overly concerned with avoiding death and too little with preserving honour, and “reckless glory-seekers” who hazard their lives by being too concerned with honour and too careless about death.

Nothing in *Leviathan* suggests that these excessive types constitute a majority of the population. But if most people seek glory only in its weak or defensive sense, i.e. they do not allow undervaluing but neither do they relish undervaluing others, why does Hobbes still count glory as one of three “principal causes of quarrell”? The reason is that this type of glorying does not allow the “natural hierarchies” to emerge where natural aggressors would subdue the others without resistance. Thus defensive glory-seeking guarantees a certain measure of equality which is a crucial precondition for war in the state of nature. How this mechanism works becomes clearer when we look at various game-theoretical models which have been proposed as formalized explanations for Hobbes’s account of conflict.

3. Game-theoretical interpretations

a) *Prisoner’s Dilemma*

Game theory is used for modelling strategic interactions of people who have similar, opposed or mixed interests. The strategic aspect means that the outcome of your actions depends on what the other people decide to do; and they in turn adapt their actions to what they expect you to do. Some commentators (e.g. Taylor 1987) have argued that Hobbes’s state of nature is structurally similar to a famous

game known as the Prisoner's Dilemma (henceforth PD) which explains why in certain situations people fail to cooperate for self-interested reasons with the paradoxical result that everyone loses out.

The formal structure of a PD game is as follows. Suppose there is a community of n individuals and each of them must do one and only one of two alternatives: C and D. The situation would be a PD if the preference ordering of each individual satisfies two conditions: (1) given the choice between everyone doing D and everyone doing C, each individual prefers the latter; and (2) no matter what the actions of the others are, the individual is better off doing D rather than C (Sen 1967:112–3).

For the sake of simplicity let us imagine a community where $n = 2$. The preferences of players in a PD game are depicted in the following matrix:⁴

Table 1. Prisoner's Dilemma game

		Column player	
		C	D
Row player	C	3, 3	1, 4
	D	4, 1	2, 2

The immediate result of such preferences is the conflict of individual rationality with collective rationality. For both players strategy D is strictly dominant over strategy C – irrespective of what they expect the other player to do, they prefer to do D. The solution of the game is therefore DD (both doing D) which is the dominant strategy equilibrium. It is also the only Nash-equilibrium of the game, which signifies the outcome where neither player is interested in changing his or her strategy given that the other player sticks to his or hers. Yet this outcome is Pareto-inferior, as it will be regarded as worse by both players than the alternative outcome CC. That outcome, however, cannot be reached without enforcement because even if an agreement to do C is arrived at, it will be in the interest of both players to break it.

Can Hobbes's account of conflict be fitted into this matrix? Let us look at an example of a state of nature consisting of two people who have to choose between peaceful (C) and aggressive (D) strategies. Hobbes argues forcefully that they would both prefer mutual peace to mutual aggression, in order to escape life that is "solitary, poore, nasty, brutish and short". This fulfils the condition (1) which requires that $CC > DD$.

Hobbes also makes clear that despite this general preference for peace, people for some reason actually choose strategy D. Now, in order to qualify as a PD

⁴ The numbers in the matrix show the order of preferences, not the cardinal value of utility, 4 being the most and 1 the least preferred outcome. E.g. the outcome (4,1) means that it is the best outcome for Row player and the worst for Column player. Thus for both players $DC > CC > DD > CD$, the first letter standing for one's own strategy and the second letter for the strategy of the other player.

game, strategy D must be the dominant strategy as required by the condition (2). This condition requires two preference orderings: (2a) $DC > CC$ (unilateral aggression is preferred to mutual peace) and (2b) $DD > CD$ (mutual aggression is preferred to unilateral peace). Thus the entire preference ranking of Hobbesian men in the state of nature should be as follows: unilateral aggression > mutual peace > mutual aggression > unilateral peace.

The critics of the application of PD game to Hobbes's account of conflict have, accordingly, questioned whether (2a) or (2b) or both accurately reflect his description of men's motivation in the state of nature. In what follows, I shall have a look at some alternative games which result from disagreements on such ranking of preferences.

b) Chicken game

An explanation put forward by Gabriella Slomp (Slomp 2000: 134-9; cf. Slomp, La Manna 1996) accepts the conditions (1) and (2a) but rejects the condition (2b) of the PD game. Slomp argues that most plausible description of Hobbes's state of nature is the payoff ranking where unilateral aggression (in her apposite terminology "dominion" as this would be the eventual outcome) is preferred to "peaceful independence" ($DC > CC$), but the unilateral avoidance of fight (which entails "subjection") is preferred to war ($CD > DD$). The game that would ensue from such preference structure is called the "Chicken game" (in its evolutionary version known as the "Hawks and Doves game"), which is depicted in Table 2:

Table 2. Chicken game

		Column player			
		C		D	
Row player	C	3	3	2	4
	D	4	2	1	1

Chicken game does not have a unique solution. It has two pure-strategy Nash equilibria (CD, DC) and (DC, CD) which means that if a Row player expects the Column player to avoid fighting (C), he is better off fighting (D), but if he expects her to fight, he prefers to avoid fighting. As the players would not know which equilibrium is the one that will be played out, a mixed-strategy equilibrium where players choose their strategy at random with appropriately weighted probabilities, has been suggested as the rational solution to the game (Rasmusen 1994:72; Slomp, La Manna 1996).

I believe, however, that if we accept the preference ranking of the Chicken game, then Hobbes's theory would offer a more plausible solution than tossing a coin. Slomp and La Manna admit that the mixed-strategy solution is open to the following objection: "Given that each player's action depends on their beliefs regarding the

probability of their opponent being of a certain “type” (for example, more or less intensely glory-seeking), the opponent has the incentive to manipulate information about this probability” (59). They dismiss it – too easily, I believe – on the grounds that individuals in Hobbes’s state of nature, because of the lack of effective communication (no shared language) and common standards, would have no means to manipulate information. However, Hobbes’s account of glory-seeking clearly points to the contrary. We saw in part 2 that glory, in general, “makes men invade for reputation” and, specifically, that reputation could be increased (“extortion of greater value”) both in the eyes of one’s contemners “by damage” and in the eyes of others “by example”. Hobbes leaves no doubt that it is possible to send the message of one’s character across to a wider circle of people than those one directly interacts with. This message is delivered by actions rather than by language.

Thus the disposition to fight can be seen as a long-term strategy towards achieving the more attractive pure-strategy Nash equilibrium (for Row player DC, CD), as it prompts other people to cut their losses by avoiding the fight. To establish the reputation of being a “hawk” means creating an impression that one’s preferences are those of a PD game, i.e. one prefers to fight even when one meets another “hawk”. And surely the people who are genuinely less death-averse (so that they actually hold PD preferences) have better chances than simulators to establish such a reputation. The vainglorious people who do not back their claim to hawkishness by death-defying actions, will be quickly unmasked and end up being dominated by genuine hawks. This solution points to the function of glory-seeking as means to self-preservation – as long as the group of glory-seekers is small enough to keep the fights for reputation at a tolerable level. When the group becomes too large, some genuinely more death-averse people find it more useful to switch to the submissive strategy, and vice-versa. This model is consistent with the sociobiological theory of evolutionarily stable strategy, which explains the genesis of the stable polymorphism of aggressive and submissive individuals in a population (see Dawkins 1976:66–87).

The Hawks and Doves game offers a rather attractive explanation for the restraint of conflict through natural dominance hierarchies in pre-state societies. Surely this is a challenge which Hobbes’s account of conflict cannot overcome? In fact, I believe that the need to avoid such a solution is the very reason why Hobbes introduced the desire of glory as one of the basic characteristics of man. The Chicken game interpretations are therefore based on a fundamental misrepresentation of the preference structure of Hobbesian men.

Let us assume for a moment that condition (2a) holds, as asserted both by PD game and Chicken game interpretations (I will discuss the plausibility of this assumption in part (d) of the chapter). With regard to condition (2b), it is impossible to agree with Slomp’s assertion that it does not hold. Slomp contends – without further explanation – that Hobbesian individuals prefer subjection to war (Slomp 2000:134). However, as we saw previously, Hobbes argues forcefully for the rationality of anticipatory aggression precisely on the grounds that it is the only way to *avoid* being subjected by others. Why does he think that being

subjected to another individual is worse than risking death by war? Indeed, one could maintain against Hobbes that dominators can augment their instrumental power only by having subjects at their disposal and are thus interested in keeping them alive. Yet Hobbes's theory offers several arguments against this charge. First, Hobbes would say that being an instrument for other man's ends necessarily damages your chances of achieving your own ends, considering that in many (if not in most) respects the ends of different men do not coincide. To take the example of security, your submission to a more powerful person does not save you from war, as your small coalition is still liable to attacks by other individuals or coalitions. But the additional problem for you is that your master is likely to force you to take even greater risks for his defence than you would have taken when acting alone or in a more equal coalition. Secondly, Hobbes introduces reckless glory-seekers who do not seek power for security but delight in conquest as such. These men may find additional pleasure in killing the men they have overcome, as this makes their power particularly manifest.

Thirdly, and arguably most importantly, Hobbesian men prefer war to subjection because they are all glory-seekers at least in the weak or defensive sense, i.e. they do not tolerate ostensible undervaluing by others. Hobbes says that men "use Violence [...] for trifles, as a word, a smile, a different opinion, and any other signe of undervalue, either direct in their Persons, or by reflexion in their Kindred, their Friends, their Nation, their Profession, or their Name" (ch. 13, 88). Voluntary submission would be the clearest possible sign of inferiority, which the Hobbesian men, who consider themselves naturally equal to each other, strongly prefer not to give.⁵ It may not be the most plausible account of human nature, but it is Hobbes's account nonetheless and we have to take it seriously when reconstructing his explanation of conflict.

The only support that Slomp and La Manna give to their description of Hobbes's preference rankings is their argument that Hobbesian men attribute "incommensurably negative value" to violent death, which, as they argue, entails the "incommensurably bad outcome of war" and therefore prompts men to choose subjection over war (Slomp 2000:130, 136; Slomp, La Manna 1996:60). This argument is fallacious in several ways. First, even when one agrees with Watkins's interpretation that for Hobbesian men death is the greatest of all evils (Watkins 1965:80–83),⁶ Hobbes is quite clear that men are prepared to *risk* death in order to avoid some other *certain* evil. An evil that for most men overrides the fear of death is the loss of honour: "And because all signes of hatred, or contempt, provoke to fight; insomuch as most men choose rather to hazard their life, than not to be revenged" (ch. 15, 107). Only the vain-glorious are excessively death-averse, so that "they will rather hazard their honour [...] than their lives" (ch. 11, 72), but Hobbes neither considers it as the behaviour of the majority nor commends it as

⁵ Subjection to the sovereign is different, because he is an artificial person.

⁶ Although challenged by Kavka (81) and Curley (1989–90:174) on the basis of Hobbes's other works, it can be seen as a plausible interpretation of *Leviathan*.

rational. Secondly, Slomp and La Manna's assumption of "self-preservation in a strong sense" (which is universal to every man) is not consistent with their other assumption that "some men seek glory, namely the pleasure of superiority and honour" (Slomp, La Manna 1996:61) because the quest for eminence and defence of honour necessarily entail a greater risk of death. Thirdly, the assumption that war brings about certain death and by subjection one can escape death is flawed because Hobbes argues, as we saw, that in the state of nature every course of action entails the risk of death, and that the non-aggressive strategy tends to be more risky than the aggressive one. The problem here is not choosing between voluntary (but glorious) death and living with shame but assessing the probability of violent death in case of each strategy, whereby the considerations of reputation can for most individuals to a certain extent (depending on how glory-prone they are) balance the increased probability of death.

c) Coordination game

Only because of their inconsistent premises, Slomp and La Manna end up constructing a curious game of "chicken with spices" that lacks any rational solution (Slomp 2000:135–138). If they had been consistent in their insistence on "self-preservation in a strong sense", they should have accepted Patrick Neal's Coordination-game solution which carries the assumption of extravagant death-aversion to its logical conclusion. Neal argues that men consider all possible outcomes that can result from the aggressive strategy of one or both parties as unacceptably risky, and therefore cooperation is the dominant strategy in the state of nature (Neal 1988:642). Indeed, if men had no other ends and values than the avoidance of death, they would have no reason to be diffident of each other, as no-one would attack the others for gain or glory, and consequently the state of nature would be a state of peace. This argument, of course, rather than proving that Hobbes failed to establish conflict in the state of nature, demonstrates emphatically why Hobbes could not accept the psychological premises attributed to him by Neal, Slomp and La Manna.

d) Assurance game

I hope to have proved conclusively that Hobbesian men in the state of nature prefer war to subjection and therefore the condition (2b) of the PD game holds. The question whether (2a) holds (i.e. whether the unilateral aggression (dominion) is preferred to peace) is a more complex one. The problem is that chapter 13 does not give the full account of men's motivation in the state of nature. The strategic situation presented in that chapter is simplified in the sense that at any point in time men are entirely free to decide which action is the best means to the satisfaction of their present and future ends. The chapter is written in terms of power alone. In chapters 14 and 15, however, Hobbes introduces moral concepts such as right, law, obligation, covenant (contract) and keeping of faith. Although men in the state of nature are free to do whatever they judge to be in their best interest, reason tells them that the universal right to everything creates war and is therefore

destructive of their ends, including the most fundamental end of self-preservation. Hobbes thus formulates a number of “laws of nature” that a rational man should follow in order to escape war. These laws of nature are true rules of morality, because an action is morally virtuous because of and to the extent that it is conducive to peace: “Vertues and Vices [...] come to be praised as the means of peacable, sociable, and comfortable living” (ch. 15, 111).

The fundamental law of nature is to seek peace. From this Hobbes derives the second law of nature: “That a man be willing, when others are so too, as far-forth, as for Peace, and defence of himselfe he shall think it necessary, to lay down this right to all things; and be contented with so much liberty against other men, as he would allow other men against himselfe” (ch. 14, 92). According to this precept, men should make a covenant to lay down the right to all things. The third law of nature obliges them to “perform their Covenants made, without which, Covenants are in vain, and but empty words; and the Right of all men to all things remaining, wee are still in the condition of Warre” (ch. 15, 100).

Now, the question is whether the precepts of reason to conclude and abide by covenants have any impact on the strategies of players in the state of nature. The PD game explanation said that they have not, because even when someone would adhere to an agreement to lay down weapons, the others quite rationally take advantage of it by breaking the agreement and attacking him. However, in a famous passage in chapter 15 known as “the answer to the Foole”, Hobbes seems to argue for a very different view:

The Foole hath sayd in his heart, there is no such thing as Justice; and sometimes also with his tongue; seriously alleaging, that every mans conservation, and contentment, being committed to his own care, there could be no reason, why every man might not do what he thought conduced thereunto: and therefore also to make, or not make; keep, or not keep Covenants, was not against Reason, when it conduces to ones benefit (ch. 15, 101).

Thus someone who holds the preferences like those of players in a PD game is a paradigmatic “fool”. The Foole is particularly eager to violate the agreements of peace because these instances of covenant-breaking which “shall put man in a condition, to neglect not onely the dispraise, and revilings, but also the power of other men” (ch. 15, 101) seem to him particularly advantageous and thus rational. But Hobbes rejects this kind of reasoning:

This specious reasoning is nevertheless false. For the question is not of promises mutuall, where there is no security of performance on either side; as when there is no Civill Power erected over the parties promising; for such promises are no Covenants: But either where one of the parties has performed already; or where there is a Power to make him performe; there is the question whether it be against reason, that is, against the benefit of the other to performe. And I say it is not against reason (ch. 15, 101).

Hobbes argues that when one party has already performed its part of the covenant, it is rational for the other party to perform as well. Hobbes explains the

rationality of second performance with long-term profitable effects of adopting covenant-keeping as a maxim of behaviour (cf. Darwall 1995:69–79). Although in some cases defection might offer short-term advantages, one cannot expect it as a rule and therefore it is wiser to desist from behaviour which “tendeth to his own destruction” (ch. 15, 102; cf. Gauthier 1969:84). Consistent covenant-keeping is particularly important for building up reputation as someone who is a suitable member of defensive confederations which are the only means to survive in the state of nature:

[...] in a condition of Warre, wherein every man to every man, for want of a common Power to keep them all in awe, is an Enemy, there is no man can hope by his own strength, or wit, to defend himselfe from destruction, without the help of Confederates; where every one expects the same defence by the Confederation, that any one else does: and therefore he which declares he thinks it reason to deceive those that help him, can in reason expect no other means of safety, than what can be had from his own single Power (ch. 15, 102).

Thus a rational man in the state of nature prefers mutual peace (cooperation) to unilateral aggression (defection). The matrix that ensues when conditions (1) and (2b) hold but (2a) does not, is known as the Assurance game, which is depicted in Table 3:⁷

Table 3. Assurance game I

		Column player	
		C	D
Row player	C	4, 4	1, 3
	D	3, 1	2, 2

What is the potential for universal cooperation given these preferences? The Assurance game is more complicated than the PD game. There is no dominant strategy available for players – the optimal strategy depends on the strategy chosen by the other(s) (cf. Sen 1967:114). In a two-person example, if a Row player expects the Column player to do C (keep the contract), he prefers to do C as well; if he expects her to do D (defect), he will do the same. There are two Nash-equilibria in this game: CC (peace) and DD (war). As the outcome CC is Pareto-superior, it seems to be both the individually and collectively rational solution to the game. As Gauthier puts it, “if second performance is rational, then so is first, since by performing first one elicits the performance of the other and so obtains the benefit for which one entered into agreement” (Gauthier 1988:129; cf. Hampton 1986: 65). If the Assurance game were the accurate representation of the strategic situation in the state of nature, then it would be impossible to maintain that it is a state of war and Hobbes’s argument for absolute sovereignty would fail.

⁷ The numeric values represent, again, the order of preferences.

Hobbes, however, never says that the rationality of the second performance entails the rationality of the first performance. He emphasizes that the covenants are only valid when there is no fear that the other party will not perform (ch. 15, 100). In the answer to the Foole he makes clear that in the state of nature (where the sovereign power that compels men to perform does not exist) there can be such security only in the case the other party has performed already. If a man laid down his right to all things, and “the other men will not lay down their Right”, he would “expose himselfe to Prey, (which no man is bound to) rather than [...] dispose himselfe to Peace” (ch. 14, 92). Thus the laws of nature, including the obligation to perform the covenants, are valid only on condition that everyone else abides by the same laws (cf. Barry 1965:253). As Hobbes puts it, they oblige “in *foro interno*; that is to say, they bind to a desire that they should take place: but in *foro externo*; that is, to the putting them in act, not alwayes. For he that should be modest, and tractable, and performe all he promises, in such time, and place, where no man els should do so, should but make himselfe a prey to others, and procure his own certaine ruine” (ch. 15, 110).

So, why would men be afraid of other men not performing if performance was both individually and collectively rational? Gauthier (1988:131) and Hampton (1986:63ff) have explained this with an account of conflict which they call, respectively, “modified rationality account” and “passions account” of conflict. They point out that the cooperation (outcome CC in our model) is unstable because not all men are sufficiently rational to adhere voluntarily to their agreements. As Hobbes puts it, “For he that performeth first, has no assurance the other will performe after, because the bonds of words are too weak to bridle mens ambition, avarice, anger and other Passions, without the fear of some coërcive Power” (ch. 14, 96).

4. Assurance game with glory

The problem is how to represent this situation in game-theoretical terms. David Gauthier points towards the correct game-theoretical explanation of Hobbes’s conflict when he writes that “the state of nature is not a true prisoner’s dilemma but it presents itself as such a dilemma, because of the subversion by the passions of what would otherwise be rational agreement of peace” (Gauthier 1988: 129). He seems to imply that the introduction of passions transforms an Assurance game into a PD but he does not explain how this occurs. I believe that it is possible to demonstrate such a transformation by inserting the desire of glory into the game-theoretical matrix.

We saw earlier that glory-seeking aims at increasing one’s reputation and reputation means that people set a high value on someone’s power. We also saw that power is relative and hence reputation depends not so much on the absolute amount of power but on the relative value of someone’s power compared to the power of other people. Consequently, the utility of a certain outcome for a glory-

seeking man is influenced not only by what he stands to gain but also by how much his payoff is superior to the payoff of the others.

The game that ensues from such preferences has been analyzed by Michael Taylor who calls it the Game of Difference (Taylor 1987:128). Taylor introduces the concept of “eminence” which signifies the difference of payoffs between the individuals in a certain outcome.⁸ In the two person case, if the payoff to the i^{th} individual is P_i , the eminence of the i^{th} over the j^{th} individual is $P_i - P_j$. The Hobbesian glory-seeking individual is not a pure egoist: his utility is a convex combination of his own payoff and his eminence. His preferences contain a mixture of egoism and negative altruism: the smaller the payoff of the partner, the bigger is one’s own utility. Thus Hobbes’s suggestion that “value can be extorted by damage” refers not only to the example he sets of his power but also to a direct increase of one’s eminence over the people damaged.

Unfortunately Taylor derives no profit from his inventive argument because he believes that men in the state of nature find themselves in a PD game anyway (ibid.: 132). The addition of the desire for eminence does not change the fundamental character of the PD game, it just makes the game “more severe” (ibid.: 121). But Taylor supposes that there could exist some games which would be transformed into a PD if eminence was taken into account as a part of individual’s utility: “It would be of some interest to discover which sort of games, not themselves Prisoner’s Dilemmas, become Prisoner’s Dilemmas when transformed to Games of Difference” (ibid.: 143).

It can be shown that the Assurance game is just a sort of game which, given certain numerical payoffs and a certain rate of eminence, transforms into a PD game. Let us have a look at the utilities (U) in a modified Assurance game where both players consider eminence (α) half as important as their own payoff: $\alpha = 0.5$ (if $\alpha = 0$, then the player is interested only in his own payoff; if $\alpha = 1$, then he is only interested in the difference of payoffs). The following formula will be used: $U_i = \alpha_i(P_i - P_j) + (1 - \alpha_i)P_i$ (ibid.: 116). If we take the payoffs of the Assurance game in their simplest possible form, adopting the ordinal utilities of the Assurance game I (Table 2) as the cardinal payoffs, then we get the transformed game presented in Table 4.

Table 4. Modified Assurance game I ($\alpha = 0.5$)

		Column player			
		C		D	
Row player	C	2	2	-0.5	2.5
	D	2.5	-0.5	1	1

⁸ It has to be noted that Taylor’s Game of Difference goes beyond the standard game theory as axiomized by von Neumann and Morgenstern (1947), which does not allow interpersonal comparison of utility. While I believe that the psychology of Hobbesian man with its peculiar mixture of reason and passions can only be represented by the game-theoretical toolbox extended in such way, it also points to the limits of the application of rational-choice theory to Hobbes.

The inclusion of the consideration of eminence reverses the preference $CC > DC$, so that the condition (2a) is satisfied, with the result that the Assurance game transforms into a PD game and the conflict becomes inevitable. Interpersonal comparison (the desire of glory) prompts a non-Foole to act like a Foole. It can be easily figured out that the preference rankings of an Assurance game will always transform into the preference rankings of a PD, if the rate of eminence

$$\alpha > P_{CC} - P_{DC} / (P_{CC} - P_{CD})^9 \quad (\text{Formula 1})$$

There are several possible counterarguments to this model. First, the attribution of numerical values to the outcomes of strategic interactions in the state of nature is highly conjectural, and it can be argued that at some other values the transformation requires an implausibly high rate of eminence (in case of Assurance game I the required value is $\alpha > 1/3$). It is a fair criticism but what this model certainly demonstrates, is that the introduction of the desire of glory *can* transform the game. I believe it is also possible to argue that a Hobbesian account would support a greater difference in the original Assurance game between the payoffs of mutual cooperation and unilateral cooperation (CC-CD), compared to the payoff difference between mutual cooperation and unilateral defection (CC-DC), which lowers the threshold for the rate of eminence required to transform it into a PD game. We must remember that mutual cooperation was preferred over unilateral defection only because of the long-term advantage of trustworthiness which supposedly outweighs any possible short-term advantages that can be gained from defection. But the difference between mutual cooperation and unilateral cooperation is far greater, because one entails peace and the other “betraying oneself to the enemy”. Although it would be futile to try to determine the exact payoffs of a Hobbesian Assurance game, the payoffs of the Assurance game II (Table 5) are probably more in line with Hobbes’s argument. The rate of eminence required to transform it into the PD is only $\alpha > 1/7$.

Table 5. Assurance game II

		Column player			
		C		D	
Row player	C	4	4	0.5	3.5
	D	3.5	0.5	2	2

The second objection could be developed from Jean Hampton’s argument that in order to generate sufficient conflict in the state of nature, one must assume an implausibly high proportion of men who are dominated by passions. She claims

⁹ $(1-\alpha)CC = (1-\alpha)DC + \alpha(DC-CD) \rightarrow$
 $CC - \alpha CC = DC - \alpha CD \rightarrow$
 $CC - DC = \alpha(CC - CD) \rightarrow$
 $\alpha = CC - DC / CC - CD$

that since this is not consistent with Hobbesian psychology and jeopardizes the “applicability of his hypothetical imperative to institute a sovereign”, Hobbes’s account of conflict “fails us” (Hampton 1986:69–79). This is a valid objection because until now we have only described the strategic interaction of two people whom reckless glory-seeking has turned into Fooles. Our task is only accomplished if we manage to show that non-Fooles would choose non-cooperative strategy as well. One way to do it is to argue that the probability that the other player happens to be a Foole is sufficiently high so that the expected utility of defection is higher than that of cooperation. When, for example, the payoffs of the Row player are equal to those depicted in the Assurance game II (Table 5), he is indifferent between his strategies if he assigns the probability 25% that the Column player has the preferences of a Foole. Such a solution is, again, liable to the charge that we can fiddle with the numerical values so as to achieve the results we need. *Pace* Hampton, I would nevertheless argue that the potential harm when the other player defects is so great that only a small probability of meeting a Foole is needed to make defection advisable. Another plausible way to argue here is that the potential harm is even so great as to warrant the application of the maximin rule instead of the expected utility calculation (so Elster 1979: 20ff). The maximin rule would advise against cooperation. The lesson is that if someone fears that he is dealing with a reckless glory-seeker – and such fear in the uncertainty of the state of nature is natural and pervasive – he should not trust whatever promises he has been given, but rather prepare himself for defence and anticipation.

In addition to these arguments it should be remembered that most people are moderately glory-seeking in the sense that they do not tolerate undervaluing by others. These people might not be interested in dominion but are keen to avoid the risk of subjection, which can only be achieved by defection.

Another possible objection to my model is Hampton’s contention that the state of nature is actually a repeated PD game where cooperation is the rational strategy in the long run, and Hobbesian men fail to cooperate only because their short-sightedness does not enable them to understand the actual nature of the game (Hampton 1986:80–89). However, Hampton’s argument that one can “teach” the defectors the usefulness of cooperation by unilaterally cooperating is fatally undermined by her own admission that such teaching is plausible only in the situations of lesser risk (*ibid.*: 81). As Slomp succinctly puts it, “she accepts that one cannot afford to play iterated prisoner’s dilemmas in high-risk situations, but she does not seem to concede that the state of nature is exactly such a situation” (Slomp 2000:131). The examples Hampton brings of low-risk situations, such as an agreement between Alice and Bill to exchange her horse for his cow (Hampton 1986: 65), do not account for what is actually at stake in the state of nature. Edwin Curley has correctly pointed out that such bargaining can only occur in the condition of security (Curley 1989–90:185). The agreement that Hobbes has in mind is the covenant “to lay down one’s right to all things” which most plausibly can be interpreted as mutual disarmament. And we have repeatedly seen Hobbes emphasizing the extremely high risks involved in the first performance of such an

agreement. Hampton's "Shortsightedness account of conflict" is therefore entirely unnecessary.¹⁰

5. Conclusions

I hope to have shown that the game-theoretical explanations currently circulating in literature are unsatisfactory, because they either misrepresent Hobbes's psychological or ethical premises (Prisoner's Dilemma game, Chicken game and Coordination game) or fail to generate sufficient conflict in the state of nature (Assurance game in its original form). I suggest that the modification of the Assurance game by the addition of the desire of glory overcomes both these problems. It demonstrates why the state of nature is a state of war, without needing an implausible interpretation of Hobbes's description of human psychology, particularly with regard to his views on glory-seeking and death-avoidance. Whatever faults his picture of the human nature may have, I believe he has managed to establish a consistent and powerful account of human conflict.

Acknowledgements

The research for this article has been conducted in the framework of the grant No 4618 of the Estonian Science Foundation. I am particularly grateful to Manfred Holler for the substantial discussion which helped to improve the paper, and to Eva Piirimäe and Richard Samworth for their comments on an earlier draft.

Address:

Pärtel Piirimäe
St. John's College, Cambridge
CB2 1TP
United Kingdom

E-mail: pp228@cam.ac.uk

References

- Barry, Brian (1965) *Political Argument*. London: Routledge and K. Paul.
Curley, Edwin (1989–90) "Reflections on Hobbes: Recent Work on his Moral and Political Philosophy". *Journal of Philosophical Research* 15, 169–250.
Darwall, Stephen (1995) *The British Moralists and the Internal 'Ought': 1640–1740*. Cambridge: Cambridge University Press.
Dawkins, Richard (1976) *The Selfish Gene*. Oxford: Oxford University Press.

¹⁰ It is therefore unsurprising that Hampton finds only as much "textual evidence" for her account as to be able to claim that Hobbes "may even have been confusedly trying to present it himself in both *De Cive* and *Leviathan*" (80).

- Elster, Jon (1979) *Ulysses and the Sirens. Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press.
- Gauthier, David (1969) *The Logic of Leviathan*. Oxford: Oxford University Press.
- Gauthier, David (1988) "Hobbes's Social Contract". In *Perspectives on Thomas Hobbes*. G. A. J. Rogers, Alan Ryan, eds. Oxford: Clarendon Press.
- Hampton, Jean (1986) *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.
- Heap, Shaun P. Hargreaves and Yanis Varoufakis (1995) *Game Theory. A Critical Introduction*. London: Routledge.
- Hobbes, Thomas (1991) *Leviathan* [1651]. Richard Tuck, ed. Cambridge: Cambridge University Press.
- Hampsher-Monk, Ian (1992) *A History of Modern Political Thought*. Oxford: Blackwell.
- Kavka, Gregory S. (1986) *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.
- Kersting, Wolfgang (1994) *Die politische Philosophie des Gesellschaftsvertrags*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Locke, John (1994) *Two Treatises of Government* [1690]. Peter Laslett, ed. Cambridge: Cambridge University Press.
- Luce, Duncan R. and Howard Raiffa (1957) *Games and Decisions*. New York: Wiley, 1957.
- Malnes, Raino (1993) *The Hobbesian Theory of International Conflict*. Oslo: Scandinavian University Press.
- Martinich, A. P. (2005) *Hobbes*. New York and London: Routledge.
- Neal, Patrick (1988) "Hobbes and the Rational Choice Theory". *Western Political Quarterly* 41, 635–52.
- Neumann, John von and Oskar Morgenstern (1947) *Theory of Games and Economic Behaviour*. Princeton: Princeton University Press.
- Rasmusen, Eric (1994) *Games and Information. An Introduction to Game Theory*. Oxford: Blackwell.
- Ryan, Alan (1996) "Hobbes's Political Philosophy". In *The Cambridge Companion to Hobbes*, 208–245. Cambridge: Cambridge University Press.
- Sen, Amartya (1967) "Isolation, Assurance and the Social Rate of Discount". *Quarterly Journal of Economics* 81, 112–124.
- Slomp, Gabriella (2000) *Thomas Hobbes and the Political Philosophy of Glory*. Basingstoke: MacMillan.
- Slomp, Gabriella and Manfredi M. A. La Manna (1996) "Hobbes, Harsanyi and the Edge of the Abyss". *Canadian Journal of Political Science* 29, 1, 47–70.
- Sorell, Tom (1986) *Hobbes*. London: Routledge and Kegan Paul.
- Taylor, Michael (1987) *The Possibility of Cooperation*. Cambridge: Cambridge University Press.
- Watkins, J. W. N. (1965) *Hobbes's System of Ideas*. London: Hutchinson.