Will they automatically work together? Cooperation Among Non-Fools in Hobbes´s Leviathan

Karim Pluma

Will they *automatically* work together? Cooperation Among Non-Fools in Hobbes´s *Leviathan*

Karim Pluma[1]
Universidad Tecmilenio

Abstract: Thomas Hobbes´s State of War is commonly imagined as a harrowing condition where hostile interactions are the rule and non-hostile encounters are the rare exception. However, while it is generally true that Hobbes purposely outlined his famed condition of anarchy as a condition of perennial conflict, it is also equally true that cooperative behavior was not uncommon. In fact, by only taking into account cooperative behavior, as presented in *Leviathan*, the anarchic humans leave the absolute uncertainties of the State of War and create the Commonwealth for their safety and well-being. Over the past fifty years or so, several exits from the State of War via game theory have been proposed in the literature, with mixed results. Most (if not all) solutions consider the likelihood of betrayal, usually through the figure of the Fool. This is a valid approach since the condition of anarchy can be imagined as being rife with dishonesty. However, the issue of non-Fools – as far as the players willful cooperation and the ultimate responsibility for the creation of government is concerned, has not been addressed yet.
Key words: Thomas Hobbes, State of War, anarchy, cooperation, game theory.

1 Introduction

Written while Hobbes was exiled in Paris during the last stages of the English Civil War, *Leviathan*[2] forwards the general idea that absolutism is the best form of government. During its publication in 1651, Hobbes earned both praise and problems with different audiences at home and abroad. While the scientific community and enlightened readers openly recognized Hobbes´s uncanny rhetorical abilities, ample knowledge of the classical and biblical sources, and profound philosophical insights – all a product of the late Renaissance "first and foremost…literary" [3] curricula – his detractors fervently condemned the "Monster of

[1] karim.pluma@tecmilenio.mx
[2] I quote directly from Tuck´s Cambridge (2015) Edition, which is based on the original English *Leviathan* (1651). I have preserved the spelling and syntaxis as they appear in this edition. After the English Civil War, in 1688, Hobbes wrote the Latin version of *Leviathan*. Pasquino (2001) strongly suggests that the 1651 *Leviathan* was written for a well-educated, albeit not erudite, audience - "the one able to read English" - while the Latin edition was intended for the *communitas doctorum*.
[3] Tuck, 2015.

Malmsbury" for his "…atheism, immorality, and a wide range of unacceptable political views"[4]. Indeed, "…*Leviathan* has always aroused strong feelings in its readers"[5]. One such feeling was about curiosity for the way Hobbes worded many of his most important philosophical formulations. Readers of *Leviathan*, even those with modest mathematical sensibilities, may be quick to notice Hobbes´s mechanistic language[6] when presenting his theories of man and state[7]. Although Hobbes remained a "true humanist"[8] throughout his life, his inclinations shifted to a more "scientific" realm by 1628[9]. He grew interested in geometry by 1629[10]. Though Hobbes never fully mastered the mathematics of his time[11], the regular use of the geometric language[12] allowed him to produce a system "…with statecraft being deduced in an unbroken chain from the principles of logic and first philosophy"[13]. In the context of an axiomatic-deductive system, the most important moment in Leviathan is the exit from the State of War, which could only be achieved using rational decision-making tools. Gauthier was the first commentator to entertain the idea that cooperative behavior in the State of War could be modelled using the game theory. Accordingly, the underlying logic

---

[4] Parkin, 2015. Other unflattering epithets for England´s most important philosopher of the 17th century included: "the Devil´s Secretary", "Angel of Hell", "Nature´s Pest" and "unhappy England´s shame".
[5] Tuck, 2015.
[6] Ryan (1970) observed that Hobbes believed "firmly as one could" that behavior, be it human or animal, animate or inanimate, "was ultimately to be explained, in terms of particulate motion".
[7] See, for example, Hobbes´s description of Motion: "When a Body is once in motion, it moveth (unless something els hinder it) eternally; and whatsoever hindreth it, cannot in an instant, but in time, and by degrees, quite extinguish it: And as wee see in the water, though the wind ceases, the waves give not over rowling for a long time after; so also happeneth in that motion, which is made in the internall parts of man, then, when he Sees, Dreams, &c." (Hobbes, 2015: 15).; Opposition and Liberty: "(by Opposition, I mean externall Impediments of motion;) and may be applyed no lesse to Irrationall, and Inanimate creatures, than to Rationall. For whatsoever is so tyed, or environed, as it cannot move, but within a certain space, which space is determined by opposition of some externall body, we say it hath not Liberty to go further" (Hobbes, 2015: 145); Deliberation: "the whole summe of Desires, Aversions, Hopes, and Fears, continued till the thing either be done, or thought impossible, is that we call Deliberation" (Hobbes, 2015: 44). Other instances, uncited here for brevity´s sake, are the "Theorems of Morall Doctrine" (Hobbes, 2015: 254) and Reason as, essentially, "Adding and Subtracting" (Hobbes, 2015: 32).
[8] Pasquino (2001); the authority on Hobbes´s intellectual background is Strauss (2011).
[9] Sánchez Sarto (2012).
[10] Valasco Gómez (2006); Sánchez Sarto (2012). Euclid´s *Elements* were rediscovered that year. There is an entertaining tale of Hobbes´s almost-divinely inspired attraction to Euclid´s text, but Biener (2016) calls the whole thing "apocryphal and self-promulgated".
[11] Sabine (2012). As Biener (2016) points out, "mathematics" in the 17th century also included astronomy, optics, harmonics, and even geography, not just arithmetic and geometry.
[12] Biener (2016) believes Hobbes´s debt to geometry is that of "system construction", the organized method in which premises and basic argumentative ideas support posterior conclusions.
[13] Biener (2016).

employed by anarchic humans wanting to engage in cooperative behavior would be best understood as a Prisoner´s Dilemma[14] situation. This has inspired generations of political theorists and political economists to propose their own reconstructions, with the number of works published being "almost impossible to grasp"[15]. The level of applicability and relevance of these models – far too many to number, let alone discuss, here – greatly varies, as some follow Hobbes´s text rather closely[16], while other reconstructions can only be thought of as "Hobbesian"[17] game-theoretic exercises. This class of reconstructions has prompted the scientific community to dismiss the game theory as a legitimate tool of analysis for modeling Hobbes´s ideas. They argue that economically rational agents, which are the actors of game theoretic modeling, are similar, yet not identical to, the more psychologically complex humans of *Leviathan*[18]. Although I agree, I also hold that, while game theory is not the most precise tool of analysis to fully model the behavior of anarchic humans, it is a valuable resource to employ situationally. Indeed, game theoretic reconstructions have provided valuable insights into the problem of exiting the State of War by exploring its premises and logical conclusions. Most (if not all) reconstructions have considered the great possibility of betrayal, usually through the figure of the Fool. This is a valid approach since the condition of anarchy is imagined as being rife with dishonesty. However, the non-Fools – the players who willfully cooperate and who are ultimately responsible for the creation of the Commonwealth, has not been addressed in the literature. Along these lines, in this work, a solidly established "truths" in game theory analyses of Hobbes was proved wrong. More specifically, it was demonstrated that two (or more) willing players will *automatically* and

---

[14] Gauthier, 1969. The most careful description applied to the State of War I know of is that of Piirimäe (2006), followed closely by that of Eggers (2011).

[15] Eggers, 2011 and Parietti, 2017 for an illustrative, though not exhaustive, bibliography of studies that have used game theory to explain some decision-making phenomena in Hobbes.

[16] I.e., Moss (2010), proposed a single-shot Prisoners´ Dilemma scenario, where two belligerent players encounter each other at a lakeshore but face a nerve-wrecking lockdown. Should one of the players put her guard down, the other will kill her immediately.

[17] Piirimäe (2006) notes the cases of Kavka and Hampton: both "stray too far from Hobbes´s text" to the point where the former scholar calls his reconstruction "a Hobbesian theory" instead of "Hobbes´s theory", while the latter commentator outright tells her readers that Hobbes´s original need to be "fixed" by "philosophizing with him" and thus sets to fill in her perceived gaps.

[18] Eggers 2011 and Piirimäe, 2006, offer a more detailed account of select scholars who have questioned or altogether rejected the use of game theoretic tools of analysis for Hobbes´s political thought. Tuck (2015), for example, rejects the idea of recasting "Hobbes´s arguments into choice-theoretic terms" because Hobbesian men are rationally compelled to trade in their absolute freedoms for the more restraining safety of the Commonwealth out of self-preservation, which he sees differently from utility maximization.

seamlessly agree to cooperate in the face of life-threatening danger. To prove this argument, first (i) the chief characteristics of the State of War were outlined to fully contextualize the setting in which agreements and disagreements would occur in the absence of a governing authority. Then, (ii) cooperation in the State of War was identified and analyzed and a few basic premises were defined for building the two-game theory models that follow: first, (iii) an agent-based simulation with which it was proved that cooperation is never guaranteed, even under the most auspicious circumstances surrounding economically rational agents, and, second, (iv) a group-based simulation that predicts group stability, success, and failure based on the Battle of the Sexes game. Finally, the Conclusions (v) were presented.

2 The State of War

Hobbes´s State of War is an anarchic condition where human life is described as "…solitary, poore, nasty, brutish, and short"[19]. This perception is attributed to the fact that this construct[20] is characterized by a material primitiveness of pre-government humans in an uncertain world. There is "…no culture of the Earth; no Navigation…no commodious Buildings; no Instruments of moving,..no Knowledge of the face of the Earth; no account of Time; no Arts; no Letters; no Society;" other than the "…continuall feare, and danger of a violent death"[21]. Furthermore, Hobbes pictures anarchic humans as being perennially ambitious, since "Felicity is a continuall progresse of the desire, from one object to another…"[22]. However, material destituteness implies generalized chronic scarcity, and this makes interactions agonistic sum-zero games[23]:

---

[19] Hobbes, [1651] 2015: 89

[20] Pasquino (2001) argues that the State of War was meant to be "an *exemplum,* in the medieval sense of the word", purposely used by Hobbes "with a rhetorical function *ad deterrendum*" to persuade his war-torn audience that even the harshest of governments is better than no government at all. Hume (1961) furthered the idea that Hobbes found partial inspiration for this juxtaposition in Tyndale´s maxim "It is better to have somewhat than to be clean stripped out of all together". Hobbes himself admitted that the State of War was "…peradventure by thought, that there was never such a time, nor a condition of warre as this; and I believe it was never generally so, over all the world: but there are many places, where they live so now. For the savage people in many places of America, except the government of small Families, the concord whereof dependeth on naturall lust, have no government at all; and live at this day in that brutish manner, as I said before" (Hobbes, 2015: 89).

[21] Hobbes, 2015: 89.

[22] Hobbes, 2015: 69-70.

[23] A situation where a player´s gains are other players´ losses.

"And therefore, if any two men desire the same thing, which nevertheless 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"[24].

Hobbes further identifies three reasons for State of War humans to inflict violence on one another:

"The first, maketh men invade for Gain; the second, for Safety; and the third, for Reputation. The first use of Violence, to make themselves Masters of other mens persons, wives, children, and cattell"[25].

While the above-mentioned reasons may seem rational under such extreme circumstances, the fact that anarchic humans may severely injure or kill one another sometimes for "their delectation only" leads to a rather disturbing conclusion: humans in the State of War can oftentimes be notoriously vicious[26]. Cursory readers of Leviathan may thus conclude that the State of War is a "might makes right" world. Nevertheless, brute physical force wouldn´t be a guarantee for success. In a David v. Goliath scenario, well-employed guile and guts could certainly determine a favorable outcome for the weaker player:

"For as to the strength of body, the weakest has strength enough to kill the strongest, either by secret machination, or by the confederacy with others, that are in the same danger with himselfe"[27].

---

[24] Hobbes, 2015: 87.
[25] Hobbes, 2015: 88.
[26] Kavka (1986) believed Hobbes did not intend to picture everyone living in the State of War as a raging psychopath. Accordingly, Kavka distinguished between players in the condition of anarchy: there are the "moderates", or those who only seek to survive, and the "dominators", who actively look to oppress others for non-essential purposes.
[27] Hobbes, 2015: 87.

Since nowhere in Leviathan does Hobbes suggest that players employ a "secret machination" to help themselves in a life-or-death situation, like facing a far stronger player, and given the fact that anarchic humans are hardwired for survival, it is safe to assume that dishonesty would be rife in any transactional activity. However, constantly dealing with potential cheats, would train anarchic humans to detect the "Signes by Inference"[28], a sort of "qualified introspection" [29] that would allow them to avoid falling victim to opportunistic players:

"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 certain ruine…"[30]

Cooperation, it appears, is ultimately irrational, and yet it is not just necessary, but desirable. How can this be? Cooperation can be hugely consequential, as it is the only way to contractually introduce a "Common Power" that successfully "…may be able to defend them from the invasion of Forraigners, and the injuries of one another, and thereby to secure them in such sort…"[31] Furthermore, anarchic humans need a "restraint upon themselves" whose "finall Cause, End, or Designe" is "the foresight of their own preservation, and a more contented life thereby"[32]. The latter clause alludes to secondary and tertiary human needs, such as the establishment of regulations for communal property,[33] lest trouble, and tragedy

---

[28] "Signes by Inference, are sometimes the consequences of Words; sometimes the consequences of Silence; sometimes the consequences of Actions; sometimes the consequence of Forbearing an Action: and generally a sign by Inference, of any Contract, is whatsoever sufficiently argues the will of the Contractor" (Hobbes, 2015: 94).

[29] Missner 1977 uses this term to refer to a rather obscure passage in Hobbes´s Introduction regarding his "method…used to arrive at the theory of human nature": "And though by mens actions wee do discover their designe sometimes; yet to do it without comparing them to our own, and distinguishing all circumstances, by which the case may come to be altered, is to decypher without a key, and be for the most part deceived, by too much trust, or by too much diffidence; as he that reads, is himself a good or evil man" (Hobbes, 2015: 10).

[30] Hobbes, 2015: 110.

[31] Hobbes, 2015: 120.

[32] Hobbes, 2015: 117.

[33] See, for example, the Twelfth Law of Nature: "That such things as cannot be divided, be enjoyed in Common, if it can be; and if the quantity of the thing permit, without Stint; otherwise Proportionably to the number of them that have Right" (Hobbes, 2015: 108). In the 16th century England, "villages were designed in such a fashion that in the center of the village there would be a piece of green land that everyone could use"

ensue[34]. Hobbes believed that most anarchic humans would ultimately appreciate the long-term benefits of cooperation for various reasons. In fact, he even suggested that deciding *against* cooperating would not just be undesirable, but outright *foolish*. Only a

"Foole hath sayd in his heart, there is no such thing as Justice; and sometimes also with his tongue; seriously alleging, 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 conducted thereunto: and therefore also to make, or not make; keep, or not keep Covenants, *was not against Reason*, when it conduced to ones benefit"[35]

Despite their bruteness, anarchic humans are naturally able to exercise reason, which Hobbes identifies as

"… nothing but *Reckoning* (that is, Adding and Subtracting) of the consequences of generall names agreed upon, for the *marking* and *signifying* of our thoughts; I say, *marking* them, when we reckon by our selves; and *signifying*, when we demonstrate, or approve our reckonings to other men"[36]

"Adding and subtracting" "consequences" heavily implies that humans, in Hobbes´s view, are naturally capable of some degree of foresight, formally identified in the text as Anticipation:

"And from this difference 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

(Dutta 2001). The community kept these lands, known as *commons*, for various purposes, including public celebrations and pastureland.

34 "The tragedy of the commons develops in this way. Picture a pasture open to all. It is to be expected that each herdsman will try to keep as many cattle as possible on the commons. Such an arrangement may work satisfactorily for centuries because tribal wars, poaching, and disease kept the numbers of both man and beast well below the carrying capacity of the land. Finally, however, comes the day of reckoning, that is, the day when the long-desired goal of social stability becomes a reality. At this point, the inherent logic of the commons remorselessly generates tragedy" (Hardin, 1968).

35 Hobbes, 2015: 72. Italics my own.

36 Hobbes, 2015: 32.

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."[37]

Hence, fully aware that

"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, or not".

Hobbes ultimately concludes that

"… it is not against reason. For the manifestation whereof we are to consider; First, that when a man doth a thing, which notwithstanding any thing can be foreseen, and reckoned on, tendeth to his own destruction, howsoever some accident which he could not expect, arriving, may turne it to his benefit; yet such events do not make it reasonably or wisely done"[38].

Hobbes also recognizes his objectors´ concerns voiced by the Fool that cooperation

"…may not sometimes stand with that Reason,…and particularly then, when it conduceth to such a benefit, as shall put a man in a condition, to neglect not onlely the dispraise, and revilings, but also the power of other men"[39]

---

[37] Hobbes, 2015: 87-88.

[38] Hobbes, 2015: 102. This passage, known as "Hobbes´s Reply to the Fool", has been extensively analyzed and debated by scholars. As Hapmton (1986) famously pointed it out: "Hobbes´s answer to the fool is remarkable, because it directly contradicts the position taken in the chapters we have previously discussed in which Hobbes appears to adopt the fool´s position to explain the failure of contracts in the state of nature". Referring to the apparent contradiction described above, LeBuffe (2006) holds that "Hobbes seems too unlikely to commit such a serious blunder, especially because the two discussions in question occur so closely in the text" regarding the apparent contradiction in Hobbes´s response to the Fool. Tuck (*apud* Pasquino [2001]) is right in rendering the passage "notoriously obscure", but as Pasquino (2001) pointed out, the vagueness disappears in the 1668 edition; he then provides a translation of Hobbes´s reply from the Latin, which I omit here in the interest of brevity.

[39] Hobbes, 2015: 101.

But argues that, whoever acts in bad faith while everyone is partaking in the "Covenant", or agreement, then

"He therefore that breaketh his Covenant, and subsequently declareth that he thinks that he may with reason do so, cannot be received into any Society, that unites themselves for Peace and Defence,…and therefore if he be left, or cast out of Society, he perisheth[40].

There are, of course, other elements of note regarding the State of War. Nevertheless, in the current analysis, the textual evidence presented above allows for the following set of assumptions.

3 Cooperation in the State of War

Cooperation[41] is a fact in *Leviathan*. There are at least two identifiable scenarios in the text where anarchic humans would most likely engage in cooperative behavior. Simply put, collective human action will, in turn, produce two different effects.

Scenario A. *Inconsequential (or negligible) effects*. Fending off *a single (or few)* hostile individual(s) with the aid of others, as seen above[42].

Scenario B. *Consequential effects*. Fending off *a multitude* of hostile individuals with the aid of others.

Here, however, individual and group security would be harder to achieve and maintain. In Scenario A, numerical superiority would win the day

[40] Hobbes, 2015: 102-103.

[41] By this I strictly mean, following Hobbes, a concerted effort at creating an institution whose end "is the peace and defence" (Hobbes, 2015) of the group.

[42] See note 27.

"because in small numbers, small additions on the one side or the other make the advantage of strength so great as is sufficient to carry the victory, and therefore gives encouragement to an invasion."[43]

Presumably, once the threat is gone or eliminated, it is almost certain that the "confederacy with others" would dissolve. There would be no need to maintain a standing security force to guard against the occasional bully or bandit. But this would not be the case in Scenario B. Hobbes warns that

"The Multitude sufficient to confide in for our Security, is not determined by any certain number, but by comparison with the Enemy we feare; and is then sufficient, when the odds of the Enemy is not of so visible and conspicuous moment, to determine the event of warre, as to move him to attempt."[44]

Hobbes did not explain in *Leviathan* what compelled anarchic humans to form larger groups. However, it was made sufficiently clear that the formation of Multitudes would create a significant disruption in their immediate surroundings. "Comparison with the Enemy we feare" is but the natural response to such an event where smaller player units (individuals, families, Confederacies) would be driven to evaluate a notoriously larger and perceivably dangerous presence. In an anarchic world, a Multitude would be immediately recognized as an existential problem by smaller player units. Nevertheless, a knee jerk reaction like rushing to create a competing Multitude would not be a good solution because the very nature of this unit severely limits the development of any real competitive advantages over other Multitudes. Like families and Confederacies, Multitudes are, foremost, *voluntary* associations. This means that their material resources – men, arms, horses – are in a state of constant volatility. Furthermore, whatever keeps the Multitude together is, in all probability, a string of fragile agreements, in the absence of a central authority. It is thus sufficiently clear that to eliminate or at least keep enemy Multitudes comfortably at bay, smaller player units need to go *beyond* the Multitude not in *size*, but in *structure*. Since no individual player can

---

[43] Hobbes, 2015: 118.
[44] Ibid.

force others to join her, association would have to remain consensual, though not in the same way as when joining other groups. Hobbes differentiates this "yes" to post-Multitude membership as

"….more than consent, or concord; it is a real unity of them all in one and the same person, made by covenant of every man with every man, in such manner as if every man should say to every man: I authorise and give up my right of governing myself to this man, or to this assembly of men, on this condition; that thou give up, thy right to him, and authorize all his actions in like manner."[45]

Once the Multitude is "…so united in one person" - effectively introducing an enforcer of agreements - it morphs into a far more sophisticated unit "…called a Commonwealth; in Latin, *Civitas.*"[46]

The task at hand now, as was mentioned above, is to propose some model with the twin intention to: (a) describe mathematically the way the humans of *Leviathan* may cooperate in dangerous situations and (b) pave the way out of the State of War using game theory. Before diving headfirst into the problem, however, it is important to consider the following.

a. *The State of War is a chaotic system*.

Recall that Hobbes understood humans as bodies in perpetual motion[47]. This means that, among other things, the fabric of reality is made up of random events. As was seen above, humans can exist as (i) individuals, or as members of (ii) families, (iii) Confederacies, (iv) Multitudes, and (v) Commonwealths. However, Hobbes never intended for these units to follow a natural progression of some sort. There is no guarantee that, say, families will

[45] Hobbes, 2015: 120-121.
[46] Ibid.
[47] See notes 6 and 7.

eventually "evolve" into a Multitude or Commonwealth[48] in the same way that, if kept alive, a child will reach old age over time.

b. *The rise and fall of groups in the State of War are best described by punctuated equilibrium*.

In the field of evolutionary biology, "Darwinian evolution is a force of continuous change – a slow and unceasing accumulation of the fittest traits over vast periods of time" (Siebel, 2019: 1). Conversely, "punctuated equilibrium suggests that evolution occurs as a series of bursts of evolutionary change" which happen "…in response to an environmental trigger and are separated by periods of evolutionary equilibrium" (ibid.). Viewed this way, groups in the State of War do not follow a "Darwinian" (i.e., continuous) evolutionary path, as depicted in (a). Instead, group formations are punctuated, which means they are spontaneous, largely unpredictable, and with the capacity to change the landscape permanently and eventually reach a period of equilibrium, or *stasis*, with little to no change, until the next major disruption.

c. The Commonwealth is an artificial man

With the rise of a single Commonwealth, there is every reason to believe that other player units of lesser power will erect other Commonwealths to counter the threat. That is, this would be the weaker player units´ *best response*[49]. Commonwealths now become the norm.

---

[48] As it happens, according to Hobbes (2015: 89), in "…many places of America…" where there is nothing beyond the "…government of small Families," which, despite existing for millennia, "…have no government at all."

[49] A common type of response in game theory. It is, basically, the best possible answer other players have to another player´s actions. Group evolution as the best response strategy in Leviathan is akin to "keeping up with the Joneses". Imagine the Joneses moving into an affluent neighborhood. After a few weeks of studying his surroundings, Mr. Jones notices that no one has a sports car. He wishes to be the first one to impress his new neighbors. Mr. Jones then becomes the proud owner of a brand-new Ferrari 488. Not to be outclassed by the newcomers, other neighbors go out and buy similar luxury vehicles. Before long, every family in the neighborhood owns one. Disillusioned, the attention-seeking Mr. Jones realizes that, whatever he does, his neighbors will equal or top it. The "moral of the story" here is that it takes very little time before a major disruption (Ferrari 488) gets answered (other luxury cars) and creates a new balance (every family owns one now).

In addition, since groups tend "…to appoint one Man, or Assembly of men, to beare their Person" [50], i.e., to represent them in decision-making, Commonwealths become "artificial men" [51]. This means that a Commonwealth will share many of the characteristics of individuals when dealing with other Commonwealths.

Unable to conquer, permanently subdue or destroy one another, Commonwealths inadvertently reach a stalemate,

"But withal, they live in the condition of a perpetual war, and upon the confines of battle, with their frontiers armed, and cannons planted against their neighbours round about"[52].

This is consistent with the idea of punctuated equilibrium in (b), as a Commonwealth may rise suddenly (spontaneity), force the creation of other Commonwealths as the best response (disruptive change), and partake in a balance of power with other Commonwealths (stasis) until something disrupts it.

d. Every player is assumed to be an imperfect economically rational agent.

The humans of *Leviathan* are strongly influenced by their Passions[53]. They can, however, behave like perfect *economically rational agents*[54] from time to time.

e. Fear is the most powerful driving force in humans.

---

50 Hobbes, 2015: 120.

51 "For by art is created that great Leviathan called a Commonwealth, or State…which is but an artificial man, though of greater stature and strength than the natural, for whose protection and defence it was intended" (Hobbes, 2015, 9). In fact, "Hobbes was the first major philosopher to organise a theory of government around the person of the state" (Skinner, 1999).

52 Hobbes, 2015: 149. This is a clear reference to Hobbes´s contemporary international relations.

53 "And therefore the voluntary actions, and inclinations of all men, tend, not onely to the procuring, but also to the assuring of a contended life; and differ only in one way: which ariseth partly from the diversity of passions, in divers men" (Hobbes, 2015: 69-70).

54 In classical economic theory, *economically rational agents* refer to players who "…always act to maximize their own expected utility" (Kampik, Nieves, and Lindgren, 2019).

For Hobbes, fear – and specifically fear of a violent death – constitutes "…the wellspring of human behavior."[55] Fear tops the list of compelling reasons anarchic humans had to create the Commonwealth[56]. Hobbes mentioned no other Passion is with as much frequency in *Leviathan*: the string of letters f-e-a-r (which would also include the 17th century variant, "feare", and words like "fearsome", "fearful", etc.) occurs some 177 times throughout the text[57]. Cooperative behavior, then, would have to be a learned behavior (as opposed to fear, which occurs naturally in the brain). Furthermore, cooperating with others would require conviction, a function of reason.

f. Fools and non-Fools exist within a single spectrum.

Anyone who denies the existence of justice and thus approves of breaking Covenants for self-serving reasons counts, according to Hobbes, is a Fool[58]. Selfishness, not stupidity, is what mainly defines the condition of Foolishness. It is hardwired in the human mind: happiness itself is a function of selfishness[59]. According to Hobbes, personal gain is the chief motivation for violence. The desire to possess is even stronger than the desire to act against others for self-preservation![60] There are ample pieces of evidence in Hobbes´s theory of man to safely assume that humans are *inherently* Foolish but may certainly overcome, under certain circumstances, certain aspects of their primitive nature, and go on to create complex cooperative structures like the Commonwealth. This last affirmation not only preserves Hobbes´s original Fool – non-Fool dichotomy, but also takes it a step further by understanding it within the larger context of a single continuous spectrum:

Figure 1. Cooperation Spectrum

---

55 Gillespie, 2008.

56 "The passions that incline men to peace are: fear of death; desire of such things as are necessary to commodious living; and a hope by their industry to obtain them" (Hobbes, 2015: 90).

57 This count was made possible using a PDF version of the Tuck Edition and the word search function.

58 See note 35.

59 See note 22.

60 See note 23.

Perfect Fool (0) α Perfect non-Fool (1)

The Cooperation Spectrum (CS) represents a player´s probability of cooperation. Intuitively, the two ends of this probability line are defined by a certainty of noncooperation ($\rho = 0$) and cooperation ($\rho = 1$) in the figures of the perfect Fool and perfect non-Fool, respectively. There is a third probability value, namely α, which is the *cooperation threshold*, and it represents the minimum probability value at which players may think it's rational to cooperate. The cooperation threshold splits the CS into two hemispheres: if $0 < \rho < \alpha$, then that player is simply a Fool; conversely, $\rho > \alpha > 1$, a non-Fool. Players who satisfy $\rho = \alpha$ may simply be assumed to be undecided.

The value of α strongly depends on how dangerous a given situation is. Interestingly, Hobbes claimed anarchic humans are capable of some level of fear-driven foresight called Anticipation[61]. However, the probability of cooperation, ρ, also depends on how players perceive one another. Nonetheless dire the situation, if the Signes by Inference[62] gathered from other players are not reassuring, then the perceiving player may not cooperate. After all, dishonesty would be a part of a larger survival strategy for every player. Because every player in the State of War would be inclined to lie, cheat, and steal from time to time, no one´s word alone can be taken at face value.

4 Agent-Based Simulation

I hold that the Battle of the Sexes aptly models an interaction between two non-Fools willing to cooperate in the face of danger[63]. The premises of this game state that a husband and wife wish to spend time together but cannot agree on how. Clearly, the optimal scenario is hanging

---

[61] "And from this difference 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 sees no other power great enough to endanger him:" (Hobbes, 2015: 87-88)

[62] See notes 26 and 27.

[63] I am in no way saying or assuming that the Battle of the Sexes is the only way to model interactions between players in the State of War. Interactions with Fools, for example, are probably best modeled through a Prisoners´ Dilemma game, as the literature shows. However, for the conditions I set – two willing non-Fools – the Battle of the Sexes is the superior choice given its striking similarity.

out together, while the least desirable one is to go about their separate ways, with the likelihood of being mad at each other for failing to compromise. In the interest of brevity, the basic premises of Battle of the Sexes were outlined right next to the summarized interactions between two non-Fools.

Table 1. Battle of the Sexes formulation compared to non-Fool – non-Fool interaction.

Battle of the sexes	Players in the condition of anarchy
Two players – husband and wife – wish to an event.	Two players - π_1 and π_2- wish to perform a certain non-hostile action.
Both husband and wife want to spend time with each other, but the former wants to attend Event A and the latter, Event B.	Both players want to cooperate with one another, but π_1 has Plan A, while π_2, has Plan B.
If husband and wife do not agree on where to spend time together, they both risk going about their different ways – the least desirable scenario.	If π_1 and π_2 do not agree on where to cooperate, they both risk going about their different ways – the least desirable scenario.
Should husband and wife fail to reach an agreement, they will not be happy with each other.	Should π_1 and π_2 not reach an agreement on how to cooperate, they both risk alienating each other and even interpreting one another´s reluctance as hostile.

Put simply, players strive to win the game. To achieve their goal, players rely exclusively on a solid strategy, "a blueprint for action" where "for every decision node it tells the player how to choose."[64] Game theorists distinguish among several types of strategies. However, for the Battle of the Sexes, there are two strategies that immediately stand out.

[64] Dutta, 2001.

The game, like many others, has *pure strategies*, "where each player chooses to play an action in a deterministic, non-aleatory manner."[65] In the Battle of the Sexes, players have two pure strategies, Plan A and Plan B. There is a third choice, however, and that is leaving the decision to chance by, say, flipping a coin[66]. Pure strategies tend to be conspicuous since players can see or set them from the start. *Mixed strategies*, however, aren´t so obvious. Unlike pure strategies, mixed strategies are player choices that are "…not deterministic and are regulated by probability distributions."[67] Mathematically, a mixed strategy is *lottery*[68], or a probability distribution over a pure strategy. Flipping a coin to determine which plan to follow is an example of mixed strategies. Players randomize their strategies when wishing to bluff and make their moves unpredictable to competing players[69], with the intention of (hopefully) modifying their behavior. Randomization largely takes place when players believe they can obtain a greater payoff than by employing pure strategies.

The classic game theory builds on the foundation that players are economically rational agents with little to no intention to deviate from what they deem the most rational choices. Their preferred strategies will then be those that offer the largest *payoff*, or *utility*[70]. These are the values that players assign to a given situation or event. The arguably simplest way to summarize this information is through a payoff matrix.

---

65 Gottlob, Greco, and Scarcello, 2005.

66 The coin example appears in Dutta, 2001.

67 Gottlob, Greco, and Scarcello, 2005.

68 Lotteries are of great importance in game theory. The simplest way to describe them is as situations where there is some level of uncertainty. For example, Jake, a college student, is going out Saturday night – his "pure strategy", i.e., certain strategy. There are two options (lotteries): going to the campus club, where his payoff is meeting a cute girl on the dancefloor, or going to a Shakespeare reading club, which meets the same day at the same time in the student center with a payoff of possibly making new friends with similar interests. Both have uncertain payoffs: Jake may not meet the cute girl he hopes to in Lottery 1, and he may not make friends in Lottery 2. Jake must decide. Which is the better lottery, i.e., the one with the promise of a higher payoff? This is exactly why the concept of expected utility or payoff is so important: because it reduces the value of a lottery to a single number that can be compared more easily with other from other lotteries.

69 "The canonical example of this is bluffing in poker. If you hold a bad hand, you will sometimes bet heavily on it and sometimes not, choosing (in each instance) randomly between bluffing (betting) and not. The idea is that you don´t want your betting behavior to signal your opponents what cards you hold; you randomize between bluffing and not so that when you bet heavily, your opponent is confused as to whether you hold a good hand or not" (Kreps, 2009).

70 Both terms are used interchangeably in game theory.

Table 2. Payoff matrix for Battle of the Sexes.

		π_2	
		A_2	B_2
π_1	A_1	1, 2	0, 0
	B_1	0, 0	2, 1

Table 2 contains every possible strategy and its corresponding payoffs. Plans A and B are represented on the matrix by the letters A and B, respectively, with the added subscripts to further distinguish strategies in relation to players. Naturally, Players 1 and 2 assign the highest payoff (2) to their proposed plans, B and A, respectively. Opting for the other player´s plan (A for Player 1 and B for Player 2) yields a lower payoff because while one of the players is not executing their preferred plan, they are at least cooperating with the other player. Finally, when players decide, they do not want to work together, they each get a payoff of 0. Evidently, strategies A_1B_2 and B_1A_2 are the least desirable, as noncooperation can potentially translate into a certain death for both players. Strategies A_1A_2 and B_1B_2, on the contrary, are attractive, as both players benefit with a nonzero payoff, *but not equally*. Who will yield? This observation is a basic notion of a *Nash equilibrium*.

A result of paramount importance in economics, a Nash equilibrium occurs when "an array of strategies, one for each player, such that no player has an incentive…to deviate from his part of the strategy array."[71] Both Players 1 and 2 must evaluate if it is rational for them to

[71] Kreps, 2009.

deviate (→), or change, from a strategy based on its payoff. Table 3[72] summarizes the criteria of players.

Table 3. Deviations

Player 1

		π_2	
		A_2	B_2
π_1	A_1	1, x	x, x
	B_1	0, x	x, x

A→B? No, A>B

Player 2

		π_2	
		A_2	B_2
π_1	A_1	x, 2	x, 0
	B_1	x, x	x, x

A→B? No, A>B

Player 1

		π_2	
		A_2	B_2
π_1	A_1	x, x	0, x
	B_1	x, x	2, x

B→A? No, B>A

Player 2

		π_2	
		A_2	B_2
π_1	A_1	x, x	x, x
	B_1	x, 0	x, 1

B→A? No, B>A

As predicted, this game has two equilibria, as presented in blue in Table 4. Since these equilibria occur with pure strategies, they are more formally known as *pure strategy* Nash equilibria. As was stated above, however, this game needs a "tiebreaker" strategy.

Table 4. Pure Strategy Nash Equilibria

		π_2	
		A_2	B_2
π_1	A_1	1, 2	0, 0

[72] I have left the relevant values for these analyses and used x´s in place of values that are not, so as to avoid confusion.

B_1	0, 0	2, 1

When there are multiple pure strategy Nash equilibria, players are expected to randomize their strategies. This is when they opt for mixed strategies. Since these rely on probability distributions, let´s assume Player 1 chooses Option A with probability p and Player 2 opts for Option B with probability q. Since both players wish to cooperate, this game is an "either-or" scenario. This means that Player 1 will choose Option B with probability $(1 - p)$ while Player 2 will choose Option A with probability $(1 - q)$.

Recall that mixed strategies are a type of lottery, which is simply a "probability distribution over possible outcomes"[73]. A mixed strategy Nash equilibrium represents a third choice – much like tossing a coin when there are other clear options. These clear options – the pure strategies – are then *in the support* of the mixed strategy, as it is referred to by game theorists, and since the player randomizing, it can only mean, intuitively, that she is indifferent to the mixed strategies.

The payoffs from this type of lottery (mixed strategies) can be calculated by using the following algorithm:

Step 1. Calculate the expected utility for both players.
Step 2. Calculate the probability that each player will be indifferent to either strategy.
Step 3. Calculate the probability of both players choosing each strategy, based on the probabilities found in Step 2.
Step 4. Multiply the probability distribution found in Step 3 by each payoff for each player. The sum of these products is the mixed strategy payoff.

[73] Dutta, 2001.

Step 1 requires players to compute their *expected utility*. Most economics textbooks present some slight variation of the following formula:

$$EU = \sum_{i=1}^{n} p_i U(x_i) = p_1 U(x_1) + p_2 U(x_2) \dots p_n U(x_n)$$

where the expected utility (EU) is the summation of the product between utilities $U(x_n)$ and likelihood or probability p_n they will occur.

Applied to this case, there will be four expected utility equations: 2 for Player 1 and 2 for Player 2, one for each option individually:

$$EU_{Player\ 1,A} = pU(x_1) + (p-1)U(x_2) \qquad EU_{Player\ 2,A} = qU(x_1) + (q-1)U(x_2)$$

$$EU_{Player\ 1,B} = pU(x_1) + (p-1)U(x_2) \qquad EU_{Player\ 2,B} = qU(x_1) + (q-1)U(x_2)$$

Let´s begin with the expected utilities for Player 1. Table 5 contains the selected information for these calculations.

Table 5. Mixed Strategies of Player 2 (π_2).

		$\boldsymbol{\pi_2(q)}$	
		$\boldsymbol{A_2}$	$\boldsymbol{B_2}$
$\boldsymbol{\pi_1(p)}$	$\boldsymbol{A_1}$	1, x	0, x
	$\boldsymbol{B_1}$	0, x	2, x

$$EU_{Player\ 1,A} = pU(x_1) + (p-1)U(x_2)$$
$$= p(1) + (1-p)(0)$$

$$= p$$

$$EU_{Player\ 1,B} = pU(x_1) + (p-1)U(x_2)$$
$$= p(0) + (1-p)(2)$$
$$= 2 - 2p$$

Then, by Step 2,

$$EU_{Player\ 1,A} = EU_{Player\ 1,B}$$
$$p = 2 - 2p$$
$$3p = 2$$
$$p = \frac{2}{3}$$

The probability for Player 1 to randomize his strategies (that is, not care whether to follow Plan A or Plan B) is $\frac{2}{3}$.

The expected utilities for Player 2 are as follows. Table 6 contains the selected information for these calculations.

Table 6. Mixed Strategies of Player 1 (π_1).

		$\boldsymbol{\pi_2(q)}$	
		$\boldsymbol{A_2}$	$\boldsymbol{B_2}$
$\boldsymbol{\pi_1(p)}$	$\boldsymbol{A_1}$	1, x	0, x
	$\boldsymbol{B_1}$	0, x	2, x

$$EU_{Player\ 2,A} = qU(x_1) + (q-1)U(x_2)$$
$$= q(2) + (1-q)(0)$$

$$= 2q$$

$$EU_{Player\ 2,B} = qU(x_1) + (q-1)U(x_2)$$
$$= q(0) + (1-q)(1)$$
$$= 1 - q$$

By Step 2,

$$EU_{Player\ 2,A} = EU_{Player\ 2,B}$$
$$2q = 1 - q$$
$$3q = 1$$
$$q = \frac{1}{3}$$

For Player 2 to randomize, her payoffs need to be equal, and the probability for this $q = \frac{1}{3}$. Ultimately, this means that Player 1 will opt for Plan A with a probability of $\frac{1}{3}$. However, there is a greater chance – of $p = \frac{2}{3}$, to be specific - that he will wind up following Plan B, her preferred course of action. Similarly, Player 2 has a probability of $\frac{2}{3}$ choosing Plan A – her proposed way of avoiding impending doom – and a probability of just $\frac{1}{3}$ to follow Plan B. By Step 3,

$$p_1 = pA_1pA_2 = \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{2}{9} \qquad p_3 = pB_1pA_2 = \left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = \frac{4}{9}$$

$$p_2 = pA_1pB_2 = \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) = \frac{1}{9} \qquad p_4 = pB_1pB_2 = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) = \frac{2}{9}$$

By Step 4, for Player 1,

$$= \left(1 * \frac{2}{9}\right) + \left(0 * \frac{1}{9}\right) + \left(0 * \frac{4}{9}\right) + \left(2 * \frac{2}{9}\right)$$

$$= \frac{1}{3}$$

for Player 2,

$$= \left(2 * \frac{2}{9}\right) + \left(0 * \frac{1}{9}\right) + \left(0 * \frac{4}{9}\right) + \left(1 * \frac{2}{9}\right)$$

$$= \frac{1}{3}$$

Table 7. Mixed Strategy Payoffs of Players 1 and 2.

Mixed Strategy Payoffs of Player 1 (π_1).

$\pi_1(p)$ \ $\pi_2(q)$	A_2	B_2
A_1	$1 * \frac{2}{9}$	$0 * \frac{1}{9}$
B_1	$0 * \frac{4}{9}$	$2 * \frac{2}{9}$

Mixed Strategy Payoffs of Player2 (π_2).

$\pi_1(p)$ \ $\pi_2(q)$	A_2	B_2
A_1	$2 * \frac{2}{9}$	$0 * \frac{1}{9}$
B_1	$0 * \frac{4}{9}$	$1 * \frac{2}{9}$

Hence, the mixed strategy payoff for Players 1 and 2 is as follows:

Player 1: 2 > 1> $\frac{1}{3}$

Player 2: 2 > 1> $\frac{1}{3}$

Payoffs 1 and 2 are from the pure strategy Nash equilibria; the payoff of $\frac{1}{3}$ is the mixed strategy payoff. Clearly, the mixed strategy payoff is inferior to the pure strategy payoffs for both players. Alas, randomizing did not solve the game, and with two pure strategy Nash equilibria, there is no one clear choice both players will be compelled to elect. It may seem like this game is a stalemate. From a pure decision-making perspective, this might be a problem. Nevertheless, the above-mentioned result, proves something that, to this day, has been taken for granted in game theoretic approaches to cooperation in *Leviathan*: that *willingness to cooperate* from two (or more players) automatically and smoothly materializes into *cooperation*. Through the Battle of the Sexes model, however, I have proven that

(i) willingness to cooperate, however strong and well-meaning, from two or more players is not enough, as there appear to be two immediately rational and "right" options (pure strategy Nash equilibria) from which to choose, and

(ii) that if players wish to expand their options by randomizing their strategies, there is a legitimate concern for this willingness to cooperate to erode and eventually fade away, as they see that their payoffs are lower than they are first to (pure strategy) possible gains.

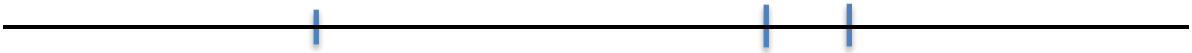
Now what? The solution, however bland, is simple: cooperation can only occur if one of the players caves. This result is typical in a Battle of the Sexes situation. When

"…players agree on the need to collaborate but are in conflict regarding the specific method, one player must always compromise regarding the specific method of coordination (in other words, accept smaller gains)". However, "it has always been believed that if an agreement is reached under such circumstances, the players do not have the incentive to withdraw from the agreement." [74] This prediction, where one player cooperates despite a smaller payoff, is also consistent with the way anarchic humans would behave. In fact, this parallel between the Battle of the Sexes and Hobbes´s theory of the state is of paramount importance for creating the Commonwealth. I will explain this jump from an agent-based simulation to a group-based simulation next.

[74] Sekiyama, 2014.

5 Group-Based Simulation

Let´s first assume two (individual) players in the condition of anarchy realize the presence of an existential threat nearby and determine that, to survive, they must cooperate. Willful though they are, however, they cannot decide on the specifics: Player 1 has Plan B, and Player 2, Plan A. Both players strongly believe their plan is the more rational way to avoid what appears to be a certain death. As was demonstrated above using the Battle of the Sexes game, however, consensus on one or the other plan seems distant since acting as economically rational agents (*d*), they will initially act upon the course of action which promises the largest payoffs. The Cooperation Spectrum (*f*) would look something like this:



Perfect Fool (0) α ρ_2 ρ_1 Perfect non-Fool (1)

However, as the threat draws nearer, Fear, the most powerful of the Passions (*e*), makes Player 1 cave to Player 2, and follow, say, Plan B. The situation is quickly reaching a point where *not* cooperating would be irrational. Graphically, this would make the cooperation threshold (α), a function of time, shift to the right significantly.

The concession from Player 1 has two effects: her Signes by Inference makes Player 2 cast any doubts of cooperation and both players, being dead serious about cooperating, effectively become perfect non-Fools. Their CS has now changed:

Perfect Fool (0) α ρ_2, ρ_1

By the commutative property, it can be obtained $\rho_2 = \rho_1 = 1$. This allows for one of the players to represent the other player and act as a single unit (*c*). This iterative process may go on until, from smaller units, larger groups are created. This means that, even if the original

formulation of Battle of the Sexes involves two *individual* players, by the fact that the Commonwealth is an artificial man (*c*), two Commonwealths can still "play", even if players represent millions of subjects.

Will these two players cooperate after neutralizing the existential threat? According to Hobbes, the second reason why players establish a Commonwealth is to have "commodious living"[75] – the good life – whereof they may profit from countless arts and sciences[76]. This is where Leviathan and the Battle of the Sexes mirror each other: those players who have less to gain from cooperation will still do so *as long* the payoffs of staying together are greater than those of going about their separate ways. It is then "…not against reason"[77], as Hobbes so famously replied to the Fool, to cooperate for less than originally bargained for.

Now, the model. Let´s assume Players 1 and 2 have settled for a pure strategy Nash equilibrium of the Battle of the Sexes game, say, B_1B_2. Their situations immediately follow this decision:

Table 8. Chosen Pure Strategy Nash Equilibrium.

		π_2	
		A_2	B_2
π_1	A_1	1, 2	0, 0
	B_1	0, 0	2, 1

[75] See note 50.

[76] See note 21.

[77] "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, or not. And I say it is not against reason" (Hobbes, 2015: 102).

1. There is only one Nash equilibrium – cooperation – and because this is the most desired behavior because it produces a good payoff, hence it is *valuable.*
2. As the default behavior, players must work towards preserving cooperation.
3. But since, as was mentioned above, cooperation is desired but not the natural and immediate option for anarchic humans, keeping it comes at a *cost*. Likewise, since cooperation was chosen in relation to its opposite (noncooperation, or anarchy) there is a perennial *comparison* between the two.

Let´s identify the payoffs mentioned in (3) as the expected utility of being part of a group, EU_g. Situation (1) informs us there is value to cooperation, v. The third variable, w, represents the proportion of players that a player or players expect to cooperate, and it is linked to a player or player´s sense of Anticipation. Finally, there are the costs of cooperation, $-c$, which include the costs of opportunity. This model appears elsewhere[78] in a slightly different form.

Hence,

$$EU_g = vw - c \qquad (1)$$

The CS α represents the minimum probability value at which players may cooperate. From a player´s point of view, however, this threshold may indicate how rational or irrational cooperating with others is. Since being a member of a group makes players, by extension, perfect non-Fools, each player may estimate the costs of cooperation in relation to their current position (1) and its distance from α. Rewriting c as $1 - \alpha$, then

[78] In *Assholes. A theory*, the philosopher Aaron James (2012) defines the figure of the asshole, as a person who "1. Allows himself to enjoy special advantages and does so systematically; 2. Does this out of an entrenched sense of entitlement; and 3. Is immunized by his sense of entitlement against the complaints of other people". This profound sense of entitlement denies others recognition before the asshole, and, by extension, moral respect. This is what makes, in James´s view, the problem of the asshole a philosophical one. In the appendix of the book, entitled "A Game Theory Model of Asshole Capitalism", James addresses his concern about a disproportionate growth of asshole characters that may ultimately lead to what he terms Asshole Capitalism, a modern might-makes-right dystopia where asshole attitudes become are the norm, not the exception. The link between the Battle of the Sexes and James´ original model, however, is my own work.

$$EU_g = vw - (1 - \alpha) \qquad (2)$$

where the condition for a player to cooperate is $EU_g > 0$ and where cooperation is the Nash equilibrium (i.e., every player´s best choice) if and only if $vw > -(1 - \alpha)$.

Let´s assume there is a Commonwealth with millions of players. This Commonwealth enjoys a peaceful existence, despite a small but somewhat noisy party (of Fools) seeking to overthrow the government. Some of its leaders have even been arrested, tried for treason, made public enemies [79] and banished from the realm. Despite this, there is stasis, or equilibrium (*b*): most players pay little to no attention to this and carry on with their daily lives. Under these conditions, a player may evaluate her payoff to cooperate as follows:

$$EU_g = (7)(.92) - .25 = 8.03$$

The cost of cooperation will almost always be a nonzero value. To support the Commonwealth, players must lose their Liberty[80] and comply with the responsibilities of civil society: paying taxes, abiding by the law, keeping societal expectations, etc. However, the high value, along with the high proportion of players willing to do their part, make cooperative behavior worthwhile.

By assuming that the separatist faction mentioned above grows in power and popularity and eventually drives the Commonwealth into a civil war, the stasis or equilibrium will be suddenly disrupted. The proportion of those willing to cooperate and the value players assigned to cooperative behavior have fallen severely. The Fools promoting secession have consequently made the costs of cooperation rise dramatically, as now fewer players are willing to finance the burdens of war. Hence, from the same player´s perspective, the payoff for cooperating under these conditions is now:

---

[79] "for in denying subjection, he denies such punishment as by the law hath been ordained, and therefore suffers as an enemy of the Commonwealth; that is, according to the will of the representative" (Hobbes, 2015: 216).

[80] See note 7.

$$EU_g = (2)(.30) - 1 = -0.4$$

A negative EU_g ultimately means this player has no incentive to continue cooperating. If other players come to similar conclusions, then the Commonwealth will most likely dissolve, and every player returns to the condition of anarchy.

6 Conclusions

Game theory reconstructions of Hobbes´s Leviathan have always, more indirectly than directly, assumed (among others) the following two premises: (a) two (or more) willing players will *automatically* and seamlessly agree to cooperate in the face of life-threatening danger and (b) there is a way out of anarchy. However, the situation in reality differs. As was demonstrated above, two willing economically rational agents faced with the decision to work together in the face of great danger will almost invariably lead to a Battle of the Sexes scenario where players would still have to choose *how* to cooperate (that is, which plan to follow). Unless one of the players yields or submits, there is a chance they will choose against cooperating and go their separate ways. Lastly, a group-based simulation was presented that predicts group stability or decline stemming from the Battle of the Sexes game.

References

Biener, Z. (2016). Hobbes on the Order of the Sciences: A Partial Defense of the Mathematization Thesis. *The Southern Journal of Philosophy.* 54: 312-332.

Dodds, G. G. and Shoemaker, D. W. (2002). Why We Can´t All Just Get Along: Human Variety and Game Theory in Hobbes´s State of Nature. *The Southern Journal of Philosophy.* 60: 345-374.

Dutta, P. K. (2001). *Strategies and Games. Theory and Practice.* Cambridge, MA: MIT Press.

Eggers, D. (2011). Hobbes and Game Theory Revisited. Zero-Sum Games in the State of Nature. *The Southern Journal of Philosophy.* 49: 193-225.

Gauthier, D. (1969). *The Logic of Levithan. The Moral and Political Theory of Thomas Hobbes.* Oxford: Clarendon Press.

Gillespie, M. A. (2008). "7. Hobbes' Fearful Wisdom". In *The Theological Origins of Modernity*. Chicago: University of Chicago Press, pp. 207-254. https://doi.org/10.7208/9780226293516-009.

Gottlob, G., Greco, G., and Scarcello, F. (2005). Pure Nash Equilibria: Hard and Easy Games. *Journal of Artificial Intelligence Research.* 24: 357-406

Hampton, J. (1986). *Hobbes and the Social Contract Tradition.* Cambridge: Cambridge University Press.

Harding, G. (1968). The Tragedy of the Commons. *Science*.162: 1243-1248.

Hobbes, T. (2015). *Leviathan, or The Matter, Forme, & Power of a Commonwealth Ecclesiasticall and Civill.* Cambridge: Cambridge University Press.

Hume, A. (1961). "A Study of the Writings of the English Protestant Exiles, 1525-35". Unpublished Ph.D. thesis, University of London.

James, A. (2012). *Assholes. A Theory*. New York. Anchor Books.

Kampik, T., Nieves, J.C., Lindgren, H. (2019). Explaining Sympathetic Actions of Rational Agents. In: Calvaresi, D., Najjar, A., Schumacher, M., Främling, K. (eds) Explainable, Transparent Autonomous Agents and Multi-Agent Systems. EXTRAAMAS 2019. Lecture Notes in Computer Science(), vol 11763. Springer, Cham. https://doi.org/10.1007/978-3-030-30391-4_4

Kavka, G. S. (1986). *Hobbesian moral and political theory.* Princeton: Princeton University Press.

Kreps, D. M. (2009). *Game Theory and Economic Modelling.* Milton Keynes: Oxford University Press.

LeBuffe, M. (2007). Hobbes´s Reply to the Fool. *Philosophy Compass.* 2: 31-45.

Missner, M. (1977). Hobbes´s Method in *Leviathan. Journal of the History of Ideas.* 38: 607-621.

Moss, L. S. (2010). Hobbes´s and the Early Uses of Economic Method. *American Journal of Economics and Sociology.* 69: 499-523.

Neal, P. (1988). Hobbes and Rational Choice Theory. *The Western Political Quarterly*. 41: 635-652.

Parietti, G. (2017). Hobbes on Teleology and Reason. *European Journal of Philosophy.* 1-25

Parkin, J. (2015). Hobbes and the Reception of "Leviathan". *Journal of the History of Ideas, 76*, (2).

Pasquino, P. (2001). Hobbes, Religion, and Rational Choice: Hobbes´s Two Leviathans and the Fool. *Pacific Philosophical Quarterly.* 82: 406-419.

Piirimäe, P. (2006). The Explanation of Conflict in Hobbes´s *Leviathan. Trames*. 10: 3-21.

Ryan, A. (1970). *The philosophy of the social sciences.* New York: Pantheon Books.

Sabine, G. H. (2012). *Historia de la teoría política.* Mexico City: Fondo de Cultura Económica.

Sánchez Sarto, M. (2012) Prólogo. In *Leviatán, O de la material, forma y poder de una república eclesiástica y civil.* Mexico City: Fondo de Cultura Económica.

Sekiyama, T. (2014). Coordination, Compromise, and Change: An Implication of the Repeated Games of "the Battle of the Sexes". *Journal of Mathematics and System Science 4*, 557-568.

Siebel, T. (2019). *Digital Transformation: Survive and Thrive in an Era of Mass Extinction.* New York: Rodin Books.

Skinner, Q. (1999). Hobbes and the Purely Artificial Person of the State. *The Journal of Political Philosophy 7*, 1-29.

Strauss, L. (2011). *La filosofía política de Thomas Hobbes.* Mexico City: Fondo de Cultura Económica.

Stauffer, D. (2007). Reopening the Quarrel between the Ancients and the Moderns: Leo Strauss´s Critique of Hobbes´s "New Political Science". *The American Political Science Review.* 101: 223-233.

Tuck, R. (2015). Introduction. In *Leviathan, or The Matter, Forme, & Power of a Commonwealth Ecclesiasticall and Civill.* Cambridge: Cambridge University Press.