**Appendix 1: A Formal Description of Our Model**

Let *N* = {1, 2} be the set of two players in Hobbes’s state of nature. Call them player 1 and

player 2. Let *T* = {m, v} be the set of two different types of the two players, where *m* stands for ‘modest’, and *v* stands for ‘vainglorious’. In other words, each player can respectively

be a modest type or a vainglorious type in Hobbes’s state of nature. Let *A* = {*C, A*} be the set of actions available to the two players. *C* stands for ‘“cooperate’ and *A* stands for ‘initiate a preemptive attack’. Let *S* = {*VV, VM, MV, MM*} be the set of the four possible states of affairs that could be realized in Hobbes’s state of nature. Here, *VV* denotes a state where both player 1 and player 2 are vainglorious; *VM* denotes a state where player 1 is vainglorious and player 2 is modest, and so on.

There is an additional player called, NATURE, who determines the proportion of vainglorious people living in the state of nature and makes the first move in the game by assigning a probability distribution over the four states of affairs in *S* based on the proportion *q*. So, (i.e., the probability that the state in which both player 1 and player 2 are vainglorious is realized is: ), , , .

Let *H* be the set of all possible histories of the game. A history is a sequence of moves played by the players in the game in the order of NATURE, player 1, and player 2 up to that point. So, the sequence (*VM, A, C*) denotes a history of the game in which NATURE realizes a state where player 1 is vainglorious and player 2 is modest, and then vain-glorious 1 initiates a preemptive attack, while modest 2 cooperates. The sequence (*MM, C*) denotes a history in which NATURE realizes a state where both player 1 and player 2 are modest, and then modest 1 cooperates. Set *H* contains *f,* which denotes the empty history. Let’s assume that the game concludes after player 2’s move, and let *E* be the set of all end histories of the game. So, *E* = {(*VV, C, C), (VV, C, A), (VV, A, C), (VV, A, A), (VM, C, C), (VM, C, A), (VM, A, C), (VM, A, A), (MV, C, C), (MV,C, A), (MV, A, C), (MV, A, A),*

*(MM, C, C), (MM, C, A), (MM, A, C), (MM, A, A)*}

Let denote the ‘player-type function’, which assigns a 2-tuple consisting of a player and his/her type to history indicating whose turn it is after the history has been played. So, , and so on.

Let *I* be a partition of such that for all that are in the same partition, . In other words, *I* partitions into separate ‘information sets’ where the type-specified player in the information set knows that such an information set has been reached without knowing which specific history contained in the information set has been previously played. Specifically, let *I* = {{(*VV), (VM)}, {(MV), (MM)}, {(VV, C), (VV, A), (MV, C), (MV, A)}, {(VM, C), (VM, A), (MM, C), (MM, A*)}}, which means that each type of player 1 and player 2 knows his/her own type while being unaware of his/her opponent’s type as well as, in case of player 2, the action performed by player 1.

Let *O* be the set of possible outcomes: *O* = {Death, War, Peace, Power} Let and be outcome-generating functions for player 1 and player 2 respectively. Each player’s outcome-generating function generates an outcome in *O* for all end histories in *E* for each player. Assume that each player (regardless of his/her type) experiences: Power when he/she initiates a preemptive attack while his/her opponent cooperates; Peace when both cooperate; War when both initiate a preemptive attack; and Death when he/she cooperates while his/her opponent initiates a preemptive attack. So, , and so on.

Now, let us make sure that our model meets the five conditions of the state of nature.

**C1 (Equality):** People’s physical and mental capabilities are roughly equal.

Remember that the most important implication of this condition for Hobbes’s state of nature is that the weakest human being has enough power to kill the strongest human being. This is reflected in our model by the fact that both players, 1 and 2, regardless of their type, experience Death whenever they are attacked unprepared—that is, whenever he/she unilaterally seeks peace (i.e., plays *C*) while his/her opponent initiates a preemptive attack (i.e., plays *A*.) The condition of equality (i.e., C1) will be further incorporated into our model by assuming that whenever both players are in a state of War, each player has an equal chance of defeating his/her opponent and retaining his/her life.

**C2 (Competition Due To Scarce Resources):** In the state of nature, resources are scarce in such a way that there will inevitably arise situations where two people would want to obtain the same object.

Such condition implies that our game will be played in Hobbes’s state of nature.

**C3 (Two Types of Men):** In the state of nature, there exist two types of people: the modest type and the vainglorious type. Furthermore, it is common knowledge that there is a certain proportion of the entire human population that is vainglorious—people who enjoy having superior power over others and pursue power, not as a means to secure their self-preservation, but for its own sake.

**C4 (Non-Universal Egoism):** Not everybody seeks to maximize his or her own self-interest (i.e., power). The modest types, who compose the majority of the entire population, would strictly prefer to cooperate with other people given that these other people cooperate in return. By contrast, the vainglorious types are the type of people for whom maximizing self-interest is the primary aim; hence, they would gladly enjoy taking advantage of other people’s good intentions whenever it is to their advantage and increase their power.

The fact that there exist two types of people in Hobbes’s state of nature is reflected in our model by the fact that both players, 1 and 2, can be randomly assigned to be any of the two types—modest or vainglorious–by NATURE, who performs the first move of the game. The fact that it is common knowledge that there exist two distinct types of people in Hobbes’s state of nature is reflected in our model by the fact that each player-type knows the portion of vainglorious people in Hobbes’s state of nature that NATURE initially determines.

In order to distinguish the modest type from the vainglorious type and to incorporate the fact that the two different types have distinct psychologies and motivations, we would need to further specify the preferences of the two types regarding the outcomes in *O*. Assume that the preferences of each type of individual regarding the sure outcomes that are relevant to each type are as follows:

**The preferences of the modest type—i.e., (\_, modest) (\_= 1 or 2)**

: (‘’ means ‘is strictly preferred to’)

**The preferences of the vainglorious type—i.e., (\_, vainglorious) (\_ = 1 or 2)**

:

In other words, in Hobbes’s state of nature, the modest types give utmost priority to peace, while the vainglorious types give utmost priority to power, which is exactly what we would expect. Now, let denote the value of Power (or Glory), let denote the value of Peace, let denote the value of Life, and let denote the value of Death. Now, let be the vNM utility function for (player *i*, type *t*) and assign the vNM utilities for each type of player 1 and player 2 as follows:

For

Here, we are assuming that the modest type does not put additional value on Power while the vainglorious type does not put additional value on Peace. Furthermore, we are assuming, based on the condition of equality, that each player has an equal chance of surviving when both players engage in War. With these sets of vNM utilities we have incorporated conditions C3 and C4 into our model. Now for our last condition of Hobbes’s state of nature:

**C5 (Uncertainty):** In the state of nature, people cannot reliably know other people’s types—that is, whether the person with whom he/she is interacting is modest or vainglorious.

Such condition is incorporated into our model by partition *I* = *{{(VV), (VM)}, {(MV), (MM)},*

*{(VV, C), (VV, A), (MV, C), (MV, A)}, {(VM, C), (VM, A), (MM, C), (MM, A*)}} which divides into distinct ‘information sets’ for each type of player.

Everything that has been described about our model is summarized in figure 1.

**Appendix 2: Main Results of the Model and Their Proofs**

Then, let us observe what will happen in our current model that incorporates all major characteristics of Hobbes’s state of nature. To do this, we would first need to understand what is called an assessment. An assessment is a pair (b,) consisting of two components: (i) a profile of behavioral strategies b for each type of player, and (ii) a profile called a system of beliefs.

Basically, a behavioral strategy is a probability distribution over the available actions in a given information set; it represents the probability with which the given player in the information set plays each action available there. In our model, there are four information sets—each belonging to (1,*v*), (1,*m*), (2,*v*), (2,*m*)—and, for each information set, there are two actions available to each type of player: namely, to cooperate (C) or to initiate a preemptive attack (A).

Let be the probability that (1,v) plays C in his/her information set [note that this implies that (1,v) will play A with probability in his/her information set]; let be the probability that (1,m) plays C in his/her information set; let be the probability that (2,v) plays C in his/her information set; and let be the probability that (2,m) plays C in his/her information set. Then, ) represents a profile of behavioral strategies played by each type of player in his/her respective information set.

A system of beliefs is a profile of probability distributions on each information set, where is a probability distribution on the information set {(VV), (VM)}, is a probability distribution on the information set {(MV), (MM)}, is a probability distribution on the information set {(VV,C), (VV,A), (MV,C),(MV,A)}, and is a probability distribution on the information set {(MM,C), (MM,A), (VM,C),(VM,A)}. A system of beliefs represents each player’s beliefs concerning the likelihood that a given history within his/her information set has been reached. For instance, the information set {(VV), (VM)}, which belongs to (1,v), has two histories; (VV) and (VM). Here, (which is the first component of the system of beliefs ) assigns probabilities to each history (VV) and (VM) in the information set {(VV), (VM)}; the probabilities represent (1,v)’s beliefs about the likelihoods of each of these histories being reached given that the information set {(VV), (VM)} has been reached.

In what follows, I will use the solution concept ‘Perfect Bayesian Equilibrium (PBE)’ to solve the game. We say that an assessment (b,) is a Perfect Bayesian Equilibrium (PBE) if and only if both the profile of behavioral strategies b is sequentially rational and the system of beliefs assigns beliefs according to Bayes’s rule whenever an information set is reached with positive probability. Given any system of beliefs, a profile of behavioral strategies b is sequentially rational if and only if the behavioral strategies contained in such a profile generate an expected payoff greater than or equal to that of any other set of behavioral strategies for each and every player-type given their beliefs.

**Proposition 1:** For any system of beliefs , is the only set of sequentially rational behavioral strategies for (1,v) and (2,v). In other words, the vainglorious types, (1,v) and (2,v), will play A (i.e., they will initiate a preemptive attack) for sure.

**Proof of Proposition 1**

Consider any system of beliefs where , , , , , .

Consider information set {(VV), (VM)}. For all — that is, it is the vainglorious type of player 1’s turn to play. Playing A gives (1,v) an expected payoff of:

,

while playing C gives (1, v) an expected payoff of:

.

Note that

Therefore, for any system of beliefs , playing A for sure (i.e. ) is sequentially rational for (1,v).

Now, consider the information set {(VV,C), (VV,A), (MV,C), (MV,A)}. For all }, —that is, it is the vainglorious type of player 2’s turn to play. Playing A gives an expected payoff of:

while playing C gives (2, v) an expected pay-off of:

Note that

Therefore, for any system of beliefs , playing A for sure (i.e. ) is sequentially rational for (2,v). ∎

**Proposition 2:** Consider the two modest types, (1,m) and (2,m). Let the beliefs of (1,m) and

(2,m) (i.e., and ) obey Bayes’s rule in their respective information sets. Then,

(1,m) will:

cooperate (i.e., play C) if and only if

initiate a preemptive attack (i.e., play A) if and only if

be indifferent between playing C and playing A if and only if

Similarly, (2,m) will:

cooperate (i.e., play C) if and only if

initiate a preemptive attack (i.e., play A) if and only if

be indifferent between playing C and playing A if and only if

**Proof of Proposition 2**

Consider (2,m)’s information set: {(MM,C), (MM,A), (VM,C), (VM,A)}. Let (2,m) assign beliefs to each history in his/her information set according to Bayes’s rule. Then, we have:

(2,m) will play C if and only if the expected payoff of playing C is greater than the expected payoff of playing A; that is, . This is so if and only if

Now, consider (1,m)’s information set: {(MV), (MM)}. Let (1,m) assign beliefs to each history in his/her information set according to Bayes’s rule. Then, we have:

(1,m) will play C if and only if if and only if

∎

**Proposition 3:** As becomes large, the increases and the threshold of attack T evaluated at (i.e., ) decreases.

**Proof of Proposition 3**

Since and are differentiable, we may take the partial derivative of the two values with respect to l, which gives us:

Focus on the signs of these two partial derivatives: as the value of l increases, the -intercept

increases, while the value of T evaluated at decreases. ∎

**Proposition 4:** As , the converges to 1, and converges to 0.

**Proof of Proposition 4**

∎

**Proposition 5:** Let be the proportion of the vainglorious types in the state of nature. For any arbitrarily small , there exists a threshold such that the modest types will attack for sure whenever .

**Proof of Proposition 5**

Pick any . By Proposition 2, (1, m) optimally attacks if and only if . Rearranging this in terms of , we get . Note that the right-hand side of this inequality is weakly increasing in as its partial derivative with respect to is: . As , we have for all . Set . Then, whenever , (1, m) will optimally attack for sure (i.e. ). A similar argument holds for (2, m) as well (i.e. . ∎

**Proposition 6:** Let be the proportion of the vainglorious types in the state of nature. For any arbitrarily small , the assessment where and such that

is the unique PBE of the game whenever , where .

**Proof of Proposition 6**

By Proposition 1, and are the only set of sequentially rational strategies for (1,v) and (2,v). By Proposition 5, given , and given that (1,m) and (2,m)’s beliefs in their respective information set obeys Bayes’s rule, and are the only sequentially rational strategies for (1,m) and (2,m). Therefore, is the only set of behavioral strategies that is sequentially rational.

Now, given , in order for the system of beliefs to assign beliefs according to Bayes’s rule, we must have:

Therefore, the assessment is the unique PBE of the game.