Bargaining with Mr. Hobbes: Choosing a Sovereign

Brian Barry and Charles Cameron
Columbia University[1]

Draft: April 5, 2004
v. 2.0

1. The Transition Problem

In *Hobbes and the Social Contract*, Jean Hampton comments that (on Hobbes's account) individuals in the state of nature ought to clamor for a Sovereign. From this perspective, the transition from life in the raw to a civilized existence has the flavor of a coordination game. And, rational agents ought to be able to solve such a problem pretty easily. For example, a little "cheap talk" might be enough to do the trick ("Hey everybody – enough with the solitary, nasty, brutish and short. Let's get a Sovereign.")[2]

Hampton criticizes this rosy view, using an ingenious argument. She notes that agents in the state of nature (SON) do *not* face a pure coordination game. Rather, they face an *n*-person battle of the sexes. The reason is simple. Each agent prefers life under a Sovereign to life in the state of nature. But each also prefers being Sovereign to being the Sovereign's subject. Conflict over the selection of the Sovereign may lead to a coordination failure, as the agents squabble over who will occupy the favored place. As a result, the unhappy agents may find themselves trapped in the state of nature.

Hampton suggests using a focal point to select the Sovereign. She also notes – in line with a fascinating strand of research in contemporary Political Science – that certain voting procedures allow easy (or at least easier) coordination on a "serious" candidate (§ 6.6).[3] So, the proper voting procedure might endogenously create a focal Sovereign. Still, Hampton's suggested coordinating devices seem rather *ad hoc*. She also tries to use bargaining theory to analyze the transition problem. Unfortunately, when she was writing in the early 1980s, the game theoretic analysis of bargaining was in a primitive state and she could gain little traction with this analytic maneuver.

---

[1] We thank Scott Ashworth, Nolan McCarty, and Norman Schofield, Randy Calvert, John Huber, Gary Miller and other participants at a conference at the Center in Political Economy, Washington University, for helpful comments on an earlier draft. Cameron gratefully acknowledges support from the Center for the Study of Democracy, Woodrow Wilson School of Public and International Affairs, Princeton University, where he was a Visiting Scholar during the chapter's initial composition.

[2] The laboratory evidence on the effect of pre-play communication in coordination games is admirably reviewed in Colin Cammerer, *Behavioral Game Theory*, § 7.2, especially pp. 356-362..

[3] See Myerson and Weber 1992 *APSR* article, recent NYU working paper by Rebecca Morton.

In this chapter, we use modern bargaining theory to analyze the transition problem. We begin with a tinker-toy version of Hobbes's transition problem, a two-person, two-period bargaining game, then move quickly to an $n-$person, $\tau-$period version of the bargaining game. This setting affords surprising insight into the problem. But finite horizon versions of the problem seem quite artificial. So we turn to an infinite horizon bargaining game, as this provides arguably the most natural setting for the problem. We derive a strong result that suggests disagreement in the SON ought to be short, even if agreement requires unanimity. We then extend Hobbes bargaining to include risk aversion and a veil of ignorance, generating an arguably surprising result. We conclude with some observations on the transition problem.

Our analysis leans heavily on two classics in modern bargaining theory, Ariel Rubenstein's fundamental analysis of alternating offer games, but especially John Ferejohn and David Baron's model of legislative bargaining in "divide the dollar" games.[4] We note the formal relationships, below.

2. Preliminaries

The games in this chapter have the form shown in Figure 1. First, a random device picks an agent (the "proposer"), who nominates someone (the "nominee") as Sovereign.[5] If all the agents agree, the nominee becomes Sovereign and all the other agents become Peasants, forever.[6] If the agents do not unanimously accede to the proposed allocation of roles, they remain in the state of nature for that period, and the random device again picks a proposer next period. In the absence of agreement, bargaining continues for a finite number of rounds (in the early versions of the model) or perpetually (in the latter one).

The stage or per-period payoffs take the following form: If the players remain in the state of nature, each receives a per-period payoff of c. If a player becomes a Peasant, he receives a per-period payoff of b, forever.[7] If a player becomes Sovereign, he receives a

---

[4] Ariel Rubenstein, "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50:97-109, David Baron and John Ferejohn, "Bargaining in Legislatures," *American Political Science Review* 83(4): 1181-1206 (1989). An explication of the former is now standard in game theory textbooks, for example, Robert Gibbons, *Game Theory for Applied Economists*, pp. x-y. The Baron-Ferejohn game is a workhorse of contemporary institutional political science, but has yet to receive a textbook treatment (at least, to the best of our knowledge). Readers who follow the arguments in this chapter ought to be able to read Baron and Ferejohn's classic paper with profit. Colin Camerer summarizes experimental evidence on bargaining in his *Behavioral Game Theory*, Chapter 4. Depending on the bargaining protocol, game theoretic models perform very well or quite poorly.

[5] This device can be biased in some fashion, though we do not explore this point here.

[6] The unanimity rule biases the analysis as much toward a coordination failure as possible. But, as discussed below, other thresholds are analyzed easily. Hobbes seemed to believe something like majority rule would prevail, see Hobbes pp. ---.

[7] Obviously, the agents do not live forever, which they know full well. But neither do they face certain death at a specific time. Thus, we treat them as if they imagine themselves living forever, with a present or future orientation captured by delta. For an illuminating discussion of the merits of this (standard) modeling procedure, see Osbourne and Rubenstein pp. --.

per-period payoff of *a*, forever. Thus, *c* indicates how bad the SON is, relative to being a Peasant or Sovereign, while *a* indicates how good being Sovereign is, relative to being a Peasant or remaining in the SON. Following Hobbes, we assume $c \le b \le a$, with $c \ge 0$, unless specifically noted.

What are these payoffs meant to convey? They efficiently represent the choice set typically assumed in the social contract tradition of political thought. The payoff "*c*" represents the return from implementing a decentralized norm in a stateless society. Axiomatic for Hobbes is the unattractiveness of this payoff (but see "States of Nature").[8] The payoff "*b*" represents the (net) payoff to individuals from a set of formal state institutions, which – to some extent – allow the protection of property rights, enforcement of contracts, maintenance of civil order, and so on. In other words, "*b*" is the value of a civil society, given a set of governmental institutions. The payoff "*a*" is the return to the state itself, the reward to state agents for supplying order. Thus, the three payoffs are a shorthand, a "black box," for many possible models of anarchy, the state, and civil society.

There is an important fact about allowable proposals in a specifically Hobbesian state of nature, one worth making explicit. A player can propose only an allocation of roles, saying, "Choose me [or someone else] as Sovereign." He cannot credibly make the following proposal: "Choose me as Sovereign and agree to be a Peasant, and each period I will give each of you a portion *x* of my payoff, so your per-period payoff becomes $b + x$ rather than *b*, and mine $a-(n-1)x$ rather than *a*." The reason why this proposal isn't credible is, of course, the classic Hobbesian dilemma: a Sovereign is bound by naught but her own will and the laws of Nature. Thus, there is no guarantee she will actually deliver *x*, and everyone knows it. The only credible proposals concern *roles*, corresponding to per period payoffs of *b* (for all the Peasants) and *a* (for the single Sovereign). However, if a proposer could make offers of this kind, the game would become an *n*-player Rubenstein game, with a random recognition rule. If a proposer could make offers of this kind to a portion α of the players (say, a majority), whose acceptance would bring about the proposed arrangement, then the bargaining game becomes a Baron-Ferejohn game.

To help readers compare the Hobbes game with the Baron-Ferejohn (BF) and Rubenstein games, let us be a bit clearer about notation, using the notation employed by Baron and Ferejohn. In all three games (BF, *n*-person Rubenstein, and Hobbes), if player *i* gets to be proposer, she makes proposal $x^i = (x_1^i, x_2^i, \ldots x_n^i)$, a vector of individual payoffs to the players. (In this notation, the superscript indexes the proposer, and the subscripts index the individual payoffs for each player.) In the BF and Rubenstein games, each proposed individual payoff must be non-negative and the sum of the proposed payoffs must be less

---

[8] Hobbes is not especially convincing why the State of Nature should be so unpleasant. Hampton asserts that agents in the state of nature are too present-oriented to sustain cooperative norms. Her approach provides a logical basis for Hobbes's assertion but seems *ad hoc*. A modern analysis would stress problems of communication and information, and the difficulty of implementing subgame perfect punishments. See Dixit 2004 Chapter 3. //WE SHOULD DO THIS IN ANOTHER CHAPTER//

than or equal to the dollar under division, so $\sum_{j=1}^{n} x_j^i \leq 1$. In the Hobbes game, each proposed individual per period payoff must be either b or a, and one (and only one) payoff must be a. So proposals always have the form $x^i = (b, b, ..., a, ..., b)$.

We assume the players discount future payoffs, using a common per period discount rate δ, $0 \leq \delta < 1$. That is, one dollar in period 2 is worth only δ in period 1. So if δ is near 1, the players are future oriented, far-sighted, or patient. If δ is near 0, the players are present oriented, near-sighted, or impatient.[9] Note the following. Suppose you receive a dollar two periods in the future. The value of this dollar one period in the future is δ, because it is discounted one period. However, the value of the dollar today is only δ^2, because it must be discounted twice.

We will have frequent recourse to a convenient (if non-obvious) mathematical fact, concerning the present value of a dollar, received each period, forever. It is:

$$1 + \delta + \delta^2 + \delta^3 + ... = \frac{1}{1-\delta} \qquad \text{(CMF)}$$

We need to introduce one more "technical" concept, and a handy equilibrium refinement (which requires us to introduce the concept of weakly dominated strategies).

The technical concept, the *continuation value* of a game, is extremely helpful in finding solutions to bargaining games like the Hobbes game. The basic idea is the following. At each round of bargaining (that is, given an offer x^i by proposer *i*) a player *j* must ask herself whether she prefers the "bird in the hand" or the "bird in the bush": "Do I prefer accepting this offer today and receiving x_j^i forever, or do I prefer to reject this offer, get the SON payoff today, and then receive my expected value of playing the game in the future?" For *j*, the expected value of the game in the future is the game's continuation value.

We define the continuation value as follows (the notation follows that in BF). For any particular sub-game perfect equilibrium, the *value* $v(t, g)$ of sub-game g after t offers (bargaining rounds) is defined as the vector of values $v_i(t, g)$, *i* = 1, …, *n*, to the players that results from the play of that sub-game perfect equilibrium strategy configuration. The continuation value $\delta v_i(t, g)$ is the value to player *i* if the bargaining moves to sub-game g.

Just to be clear, consider the following example. In this game, Player 1 chooses a number, either 0 or 1. Then Player 2 chooses a number, either 0 or 1, and the game ends. As payoff, each receives the product of the two chosen numbers. The game is one of

---

[9] Hampton consistently asserts that Hobbesian players have discount rates near 0, but we make no such assumption.

complete and perfect information. An obvious sub-game perfect equilibrium is: Player 1 chooses 1; Player 2 chooses 1 regardless of what Player 1 chose. Given Player 2's strategy in this equilibrium, the value to Player 1 of the sub-game that results after he plays 0 is 0. The value of the sub-game that follows his play of 1 is 1.

The equilibrium refinement addresses the following problem. Consider the following equilibrium. Whenever chosen as proposer, each agent chooses a nominee not himself, chosen at random. And, every player rejects every proposal. Note two things. First, since all players reject all proposals, no individual player has an incentive to deviate to "accept," since doing so will not bring about a different result. This is true even for the player nominated as Sovereign. Second, since all proposals are rejected, no proposer has an incentive to deviate to "I nominate myself." So this is an equilibrium.

This absurd situation arises because of the coarseness of the choice rule: if more than one player rejects a proposal, it doesn't matter what the others players do. This leaves the players free to use weakly dominated strategies, strategies a player would never employ if there were even the smallest chance she were pivotal.[10] One can "refine" away such equilibria, for example, by assuming that there is an ε chance that choosers "tremble" to their non-strategy choice of accepting a proposal. In that case, there would be an ε^{n-1} chance player *i* is pivotal, in which case his choice would determine whether the proposal is chosen or not. Accordingly, player *i* must deviate from "reject" to "accept," if she actually prefers the proposal to the value of rejection today plus the continuation value of the game (which readers should see is $c+\delta v_i=\frac{c}{1-\delta}$). There is also an ε^{n} chance a proposal will be accepted, so the proposer must deviate from proposing someone else as Sovereign to proposing herself. Thus, tiny trembles eliminate the absurd equilibrium, even as the probability of a tremble goes to zero in the limit.

Rather than check to see if every equilibrium is "trembling hand perfect," we will employ a refinement that is often used in voting games: simply eliminate weakly dominated strategies. Thus, we focus on sub-game perfect equilibria in which players do not use weakly dominated strategies. We will not allude to elimination of weakly dominated strategies again.

3. Bargaining with Mr. Hobbes

With the preliminaries out of the way, we can get down to Hobbes bargaining. We begin with the simplest case and gradually build to a much more natural one.

---

[10] One strategy for a player weakly dominates another if against any strategies by the other players the player does as well with the first strategy as with the second, and against some combination of strategies by the other players, the player does strictly better. For instance, voting against yourself as Sovereign is weakly dominated by voting for yourself as Sovereign.

2.1 Two Rounds of Bargaining

To begin, suppose there are only two players and two rounds of bargaining. In period 1, the random device picks Player 1 or Player 2 to make a proposal and both vote (as it were). If the initial proposal suceeds, the players receive the agreed-on payoffs (*a* or *b*) forever after. If it fails, the players receive the SON payoff *c* in period 1. In that case, the random device again selects a player to make a proposal, and both vote again. If this proposal succeeds, the players receive the agreed-on payoffs (*a* or *b*), then and forever after. If it fails, the players receive the SON payoff *c* in period 2 and forever after.

This game is easily solved via backward induction.

> In period 2, suppose the random device picks Player 2 as proposer. Clearly, Player 1 will accept whatever proposal Player 2 makes, since being a peasant is better than remaining in the state of nature. Given this, Player 2 surely proposes himself as Sovereign. Obviously, the same analysis applies if Player 1 is selected as proposer. So, in the last round of bargaining, whoever is selected as proposer will propose him- or herself as Sovereign, and the other player will accede.
>
> Now consider what must happen in period 1, in any sub-game perfect equilibrium. Focus in particular on the player who is *not* the proposer. What is the continuation value of the game for this player, if she rejects the proposer's offer? First, let us calculate the expected value of the appropriate sub-game. Suppose the chance the random device selects her in the next round is just $\frac{1}{n}$, in this case, ½. So with probability one-half, the player will become the proposer, propose herself successfully, and receive *a* forever. And, with probability ½, she will not be the proposer, accede to the other player's nomination, and thus receive *b* forever (as the Peasant). Therefore, the expected value of any sub-game following a rejected period 1 offer is $\frac{1}{2}\frac{a}{1-\delta}+\frac{1}{2}\frac{b}{1-\delta}=\frac{a+b}{2(1-\delta)}$. Hence, the continuation value of the game is $\frac{\delta(a+b)}{2(1-\delta)}$. So if a player rejects an offer in Period 1, she receives $c+\frac{\delta(a+b)}{2(1-\delta)}$ (that is, the current period payoff from the SON, plus the continuation value of the game). If she is to accept the proposer's offer in period 1, the offer must be at least as good as this. Clearly, if offered the role of Sovereign, she will accept, since $\frac{a}{1-\delta}>c+\frac{\delta(a+b)}{2(1-\delta)}$, for all allowable values of δ. The real issue is: will she accept the role of Peasant? If she does, she receives the payoff of *b*, forever. So she will accept the role of peasant if and only if
>
> $$\frac{b}{1-\delta}\geq c+\frac{\delta(a+b)}{2(1-\delta)}$$
>
> which (via a little algebra) will be true if and only if
>
> $$\delta\leq\frac{2(b-c)}{a-b+2(b-c)} \qquad (1)$$

> Now consider the proposer in period 1. If equation (1) holds, the proposer offers himself as Sovereign (since the proposal will be accepted). Suppose the condition fails, however (so the players are quite patient). In that case, the non-proposer will reject the offer, unless he himself is nominated as Sovereign. What should the proposer do? If the proposer offers sovereignty to the non-proposer, the non-proposer will accept, so the proposer receives $\frac{b}{1-\delta}$ (the discounted value of peasant-hood). If she proposes to retain sovereignty for herself, the non-proposer will reject, so the proposer receives exactly what we calculated above, $c+\frac{\delta(a+b)}{2(1-\delta)}$. This will be greater than the discounted value of peasant-hood if and only if $\delta > \frac{2(b-c)}{a-b+2(b-c)}$. But this is the condition we assumed in this scenario. So in this case, the proposer should nominate himself even though he knows the proposal will be rejected.

What have we established? We have identified two quite distinct equilibria. In the first, the player selected as proposer always nominates herself as Sovereign, and the other player accepts any proposal. So the players reach immediate agreement. This equilibrium – call it the "no delay equilibrium" – requires somewhat impatient players. In the second equilibrium, the player selected as proposer also always nominates herself as Sovereign. But the non-proposer rejects the first period offer if he himself has not been nominated Sovereign. Then, the non-proposer accepts any proposal in the second round. Call this the "delay" equilibrium, since the first period offer is rejected. We might also call this the "brinkmanship" equilibrium, since the players delay in the hope of making the last offer.

It should be obvious that the logic of the previous paragraphs extends immediately to any two-period, *n*-player game.[11] A little algebra confirms that the critical condition separating the two equilibria is

$$\delta^* = \frac{n(b-c)}{a-b+n(b-c)} \tag{2}$$

2.2 Many Rounds of Bargaining

Now let's extend the *n*-person game to any finite number of rounds, τ.

Again, we need to identify the critical value of δ, δ^*, such that all non-proposers with values of $\delta > \delta^*$ reject an offer of Peasant, in favor of delaying in the hope of nominating himself as Sovereign in the last round, τ. The following question is critical: in which round does a non-proposer have the *least* incentive to delay? The answer, of course, is the

---

[11] One minor caveat: with more than two players, the strategy "I nominate myself as Peasant, someone else as Sovereign" is not well defined because it does not identify a specific person as nominee Sovereign. So, one might imagine such a proposer using some simple random device to pick a Sovereign to nominate. For example, if the players are indexed 1,…, *n*, when player *i* is selected as proposer, he nominates player $i+1$ (player *n* nominates player 1). Such proposals are not used in equilibrium, of course.

first round, since τ is then as far away as it can be. If the non-proposers are not patient enough to delay in the first round, they will accede to the initial proposal, ending the bargaining – while if they are patient enough to delay then, they will be patient enough to delay in all the later rounds. This insight allows us to calculate the critical value δ^*.

In the Appendix, we show that

$$\delta^* = \left(\frac{n(b-c)}{a-b+n(b-c)} \right)^{\frac{1}{\tau-1}} \tag{3}$$

When $\tau = 2$, this result reduces to that identified in Equation (2). But for larger τ, the critical value of δ becomes *larger* (since the quantity in the parentheses must lie between zero and one, and the $\tau - 1^{\text{th}}$ root of such a number increases in τ).[12]

We have now established the following proposition.

> **Proposition 1**. In the τ-period, *n*-player Hobbes game, if $\delta \leq \delta^*$, the player initially selected as proposer always nominates herself as Sovereign, and all the players accede to the proposal. If $\delta > \delta^*$, the player selected as proposer nominates herself as Sovereign, and all the non-proposers reject the offer of peasant-hood in every period but the last. In the last period, the players accept any proposal.

The Proposition offers some interesting insights. First, as we noted, there are two distinct equilibria, a "no-delay" equilibrium, and a "delay" equilibrium. The first occurs when the players are rather impatient, the second when they are sufficiently patient.

Second, one might expect the proposer in Period 1 to have a first mover advantage, implying something like a race to offer oneself as Sovereign. No doubt this would seem plausible to Hobbes himself. And indeed there is a whopping first mover advantage – but only in the "no delay" equilibrium, where the players are rather impatient ($\delta \leq \delta^*$). But there is no first mover advantage at all in the "delay" equilibrium, where the players are more patient. In that case, the *last* mover has a huge advantage, since delay enables her to make the take-it-or-leave-it proposal "Me Sovereign," since the "leave it" is the nasty SON. In that case, one sees something like brinkmanship, each player refusing to accede

---

[12] A slight subtlety is the following. If $\delta < \delta^*$, the non-proposers will accede in the first round. But their value of δ might be high enough so that they would delay if the game reached a round closer to the end. A correct specification of the non-proposers' acceptance strategy will involve a comparison between the actual δ and $\left(\frac{n(b-c)}{a-b+n(b-c)} \right)^{\frac{1}{\rho-1}}$, where ρ is the number of rounds remaining in the bargaining (including the current round). A non-proposer will delay if δ is greater than this value, but not otherwise.

to the others until the very last moment, hoping to be the last player to put an offer on the table. This hints at the possibility of a coordination failure.

Third, note the following about the critical value of delta separating the two equilibria.

> **Corollary 1**. In the τ -period *n*-player game, for any value of $\delta < 1$, the delay equilibrium vanishes when a) the number of players becomes sufficiently large, b) the SON becomes sufficiently unattractive, c) the value of being Sovereign is sufficiently close to that of being a Peasant, or d) the length of the game is sufficiently long.

Proof. Note that $\lim \delta^* = \lim \left(\frac{nc}{a+nc} \right)^{\frac{1}{\tau-1}} = 1$ as a) $n \to \infty$, b) $c \to -\infty$, c) $a \to b$, and d) $\tau \to \infty$. By assumption, $\delta < 1$. Hence, from Proposition 1, only the non-delay equilibrium can hold as the value of those parameters approaches the indicated values. ■

The four results in Corollary 1 ground a Hobbesian analysis of the transition problem. Before discussing the results, however, we will conduct a "robustness" check: how sensitive are the results to the obviously artificial assumption of a finite number of bargaining rounds?

2.3 Bargaining Without an Arbitrary Number of Rounds

Central to the calculations of the players has been the chance of making the proposal in the final round of bargaining. Yet the notion of an arbitrary "final" round of bargaining in the SON is terribly artificial. What happens if we abandon the idea of some definite final round, and allow the players to bargain indefinitely, even perpetually? In some respects, this infinite horizon setting seems the most natural version of the Hobbes game.

First we establish a result that only a game theorist could relish: almost anything short of eternal delay can happen in such a setting. To establish this result, we allow players to condition their strategies on the history of the game. Thus, they can use trigger strategies, as in the Folk Theorem. (In other words, we establish a Folk Theorem for the infinite horizon Hobbes Game).

The structure of these "any finite delay" equilibria seems quite absurd, however. In essence, individuals in the state of nature must have a norm of delay. Within this norm, if a member votes to accept a proposal too early, all the other players punish her by delaying agreement even more – and each must deliver this punishment or face punishment themselves, in the form of more delay. But why on earth would people in the state of nature have such a bizarre norm? There is no reason to believe they would.

Accordingly, we turn to much simpler – and seemingly more sensible – strategies, ones that are invariant to the history of the game. With these *stationary* strategies, anyone selected as proposer makes the same proposal regardless of when she is called upon, and all the non-nominees always accept with the same probability regardless of the prior history of the game. Since nothing intrinsically differentiates one period from another, these strategies seem intuitively appealing. We will seek a so-called *Markov perfect equilibrium*, one in which stationary strategies form a subgame perfect Nash equilibrium in every subgame of the game. As it transpires, the Hobbes bargaining game has a simple Markov perfect equilibrium with appealing properties.

2.3.1 A Folk Theorem for Hobbes Bargaining

> **Proposition 2.** Any delay short of infinite delay is sustainable as a sub-game perfect equilibrium in the Hobbes Bargaining Game.

// Write out proof. (Introduce idea of sub-game perfect trigger strategies). (This is the equivalent of Proposition 2 in BF).//

2.3.2 Stationary Strategies

/We require strategies to be symmetric: all non-nominees must use the same strategy (even if indifferent between actions), and all individuals chosen as proposer must use the same strategy//

> **Proposition 3**. The following is the unique symmetric Markov perfect equilibrium in the Hobbes game. In each period, the proposer nominates himself as Sovereign and votes for himself. Each of the remaining $n-1$ agents accepts the proposal with probability σ, where $\sigma = 1$ if $\delta \leq \frac{n(b-c)}{a-b+n(b-c)}$ or if the non-proposer has been nominated Sovereign, and $\sigma_i = \left(\frac{n(b-c)}{a-b}\frac{(1-\delta)}{\delta}\right)^{\frac{1}{n-1}}$ otherwise.

We supply a proof in the Appendix.

We now derive the equivalent of the earlier corollary, for the Markov perfect equilibrium.

> **Corollary 2**. Agreement in the Hobbes game is immediate if 1) *c* is sufficiently small, 2) *a* is sufficiently small, and 3) *n* is sufficiently large.

Proof. From inspection of $\sigma_i = \left(\frac{n(b-c)}{a-b}\frac{(1-\delta)}{\delta}\right)^{\frac{1}{n-1}}$.■

2.4 Discussion

The first parts of the Corollary are intuitive but interesting. Hobbes argues strongly that the State of Nature is extremely unpleasant, that is, *c* is very large. In particular, he emphasizes the threat of violent death. //In an earlier chapter, we argued that cooperation in anarchy will breakdown if social relations in a community are too anomic, or if there is a sizeable population of violence-prone boors. If so, c will indeed be large.// The Corollary indicates that, under such conditions, agents in the SON will agree instantly to the institution of a Sovereign.

Hobbes also suggests that a is likely to be fairly small. //Hobbes argument about "stationary bandits" a la Olson… perhaps elaborating at more length?// Raises some interesting questions about limiting the value of Sovereignty … //Note African countries where the value of a is enormous – vicious fights over holding title to the Sovereign's position.//

The third part is somewhat surprising: given a unanimity rule, one might think that increasing the number of agents would lead to a greater chance for stalemate. But the logic should be now be clear: with large n, the chance you will be the proposer in the near future is small. So it is better to agree immediately to a proposal and end the state of anarchy, rather than hold out in the hopes of becoming Sovereign.

In fact, it is rather difficult to get long disagreements in the Hobbes bargaining game. Figure 3. //Discuss//

4. Risk Aversion and the Veil of Ignorance

We have assumed the bargainers are risk neutral and know their fate for certain, if they establish a Sovereign. But suppose they are risk averse and somewhat uncertain about the prospect of life under a Sovereign, as seems quite plausible. In other words, suppose risk averse agents bargain behind a "veil of ignorance." What happens then?

Theorists in the social contract tradition often claim great virtues for risk aversion and a veil of ignorance. Risk averse agents, they claim, will modify the social contract in order to protect themselves from bad events. For example, the bargainers may require pledges from all that if wealthy post-transition they will transfer income to the post-transition poor. If credible, such pledges provide an insurance policy against a life in poverty. Similarly, the bargainers may insist on explicit rights and a written constitution, as supposed insurance against tyranny. In turn, devices that reduce risk, if credible, make approval of a social contract more likely – and subsequently more just to boot.

Needless to say, Hobbes would reject this argument, on the ground that agreements made before the Sovereign arrives are worth nothing afterwards, and therefore cannot be credible. In a later chapter, we will consider Hobbes's point. But without yet committing to whether laws and constitutions can really bind a Sovereign, we can examine the effects of risk aversion and the veil of uncertainty on Hobbesian bargaining. The results are rather surprising.

In order to keep the exposition as simple as possible, we focus on a game with two rounds of bargaining, since Proposition 3 shows this simple model captures most of the interesting features of Hobbes bargaining. In addition, for the sake of simplicity, we will normalize payoffs so that the Sovereign's payoff is $a = 1$ and the worst possible payoff is 0, which will equal *c* unless explicitly indicated.

Risk Aversion

We first consider risk aversion outside a veil of ignorance. Even absent uncertainty about the future magnitude of *b*, risk aversion has consequences for Hobbes bargaining, because it always involves risk.

We have already introduced the idea of risk averse agents //(in the Primer).// Recall that risk aversion implies, and is implied by, a concave utility function, such as the one shown in Figure 4. Hence, to incorporate risk aversion, we need only give the agents a concave utility function, defined over the payoffs. Thus, they will receive *U*(*b*) rather than *b* if they accept a proposal, *U*(*c*) rather than *c* if they remain in the state of nature, and *U*(*a*) rather than a if accepted as Sovereign.

At times we will make the discussion more concrete by using the utility function $U(x;\alpha) = 1-(1-x)^{\alpha}$. When $\alpha = 1$, this function reduces to $U(x;1) = x$, the risk neutral utility function of Section 3. When $\alpha > 1$ the function is concave and hence its possessor is risk averse. When $\alpha = 2$, the function is an appropriately scaled quadratic utility function. In fact, this is the utility function shown in Figure 4. Interested readers can confirm that when $0 < \alpha < 1$, the utility function's possessor is risk seeking.

We can re-write the two period condition explored earlier, that is, accept in period 1 if

$$\frac{U(b;\alpha)}{1-\delta} > U(c) + \delta\left(\frac{1}{n}\frac{U(a;\alpha)}{(1-\delta)} + \frac{(n-1)}{n}\frac{U(b;\alpha)}{(1-\delta)}\right)$$

and reject in period 1, otherwise.[13] The critical condition then becomes (after a little algebra)

$$\delta(\alpha) < \frac{n\big(U(b;\alpha)-U(c;\alpha)\big)}{U(a;\alpha)-U(b;\alpha)+n\big(U(b;\alpha)-U(c;\alpha)\big)} \tag{2a}$$

It should be clear that is simply a slightly more general version of Equation (2).

The following proposition is almost immediate:

> **Proposition 4**. If $U(x;\alpha) = 1-(1-x)^{\alpha}$, $\alpha > 0$, then the critical δ for a risk averse agent is greater than that for a risk neutral agent, for all *b* between 0 and 1.

---

[13] Implicitly we are using Proposition 1, so we know all agents will accept the period 2 proposal.

Proof. Recall we have normalized so c = 0 and a = 1. So $U(a;\alpha)=1$ for all $\alpha \geq 1$ and similarly $U(c;\alpha)=0$ for all $\alpha \geq 1$. Using these facts, re-write (2a) as

$$\delta(\alpha) < \frac{n\big(U(b;\alpha)-U(c)\big)}{U(a)-U(b;\alpha)+n\big(U(b;\alpha)-U(c)\big)}$$

But note that this has the form $\dfrac{k}{U(a)-U(b;\alpha)-k}$, so that $\delta(\alpha)$ increases in $\alpha > 1$ iff $U(b;\alpha)$ increases in $\alpha > 1$. And $\dfrac{\partial U(b;\alpha)}{\partial \alpha} = -(1-b)^{\alpha}\log(1-b) > 0$ for all b between 0 and 1, given the assumed utility function. By an identical construction, the critical delta for a risk seeking agent is everywhere less than that of a risk neutral one, for b between 0 and 1.■

The proposition says: for any b that is better than the SON payoff but worse than the Sovereign's payoff, there are agents who will accept a period 1 offer to be a Peasant if they are risk averse, but will reject it if they are risk neutral.[14] In other words, *absent a veil of ignorance, risk aversion makes immediate agreement more likely.*

Figure 5 illustrates the proposition, comparing the critical deltas with the risk averse quadratic utility function ($\alpha = 2$) and the risk neutral utility function ($\alpha = 1$) (in the figure, n = 2). It is easy to calculate that

$$\delta(2) < \frac{nU(b;2)}{1+(n-1)U(b;2)} = \frac{nb(2-b)}{1+(n-1)b(2-b)} \tag{2b}$$

At n = 2 and b = ½, for example, $\delta(1)=\frac{2}{3}$ (using Equation (2)) while $\delta(2)=\frac{6}{7}$ (using Equation (2b)). Hence, if b = ½, risk neutral agents who have deltas between 2/3 and 6/7 will reject a first period offer to be Peasant, but will accept the same offer if they are risk averse ($\alpha = 2$).

The Veil of Ignorance

The essence of a veil of ignorance is that one is unsure of one's exact life prospects after ratification of the social contract. This complication is easily handled in the framework of Hobbes bargaining by allowing b to be a random variable. Conceptually, all the previous analysis goes through, simply replacing $U(b)$ with $EU(b)$.

A slightly different version of the veil of ignorance may appeal to theorists of a psychological bent. That is, suppose you are quite sure what the social contract will bring (that is, you know b, or believe you do). But never having experienced a life of that kind, you cannot be entirely sure how happy it will make you. Hobbes bargaining easily

---

[14] These agents have δ 's that satisfy the condition $\delta(\alpha = 1) < \delta < \delta(\alpha > 1)$.

handles this variant as well, by allowing a random shock to perturb $U(b)$. Again, all the previous analysis goes through, simply replacing $U(b)$ with $EU(b)$. However, we will focus on the first approach as it is the "standard" one.

What happens if $EU(b)$ replaces $U(b)$? We investigate this question in a particularly tractable setting, though the answers are more general.

First, let's begin with a risk neutral agent, so $U(b;\alpha) = U(b;1) = b$, as explained in the previous section. Second, let's be more explicit about the agent's uncertainty concerning b. We have normalized the payoffs so that the best possible payoff is 1 and the worst possible payoff is 0. So any realization of b must lie within this range. (In a minute we will consider what happens if, as Hume suggested, a realization of b may be worse than the State of Nature). One could specify some arbitrary values for b and associate probability values with them. This would provide us with a "discrete" probability distribution. But a less artificial approach allows b to take *any* value between 0 and 1. This approach requires a "continuous" probability distribution.

It may help to illustrate with an example. There is a well-known class of continuous distributions whose range is the interval [0,1], the so-called beta distribution. Beta distributions are extremely flexible, assuming an amazing variety of forms depending on the values of two shape parameters, here called κ and λ to avoid any confusion with our other proliferating Greek letters. Figure 7 shows a beta distribution with shape parameters (5,5). In a beta distribution, the mean of the distribution of the random variable b is just $\bar{b} = \frac{\kappa}{\kappa + \lambda}$, so the mean of this centrist distribution is ½. The variance of a beta distribution is ---. Figure 8 shows a beta distribution with shape parameters (.05,.05) (both parameters must always be greater than zero). Again, the mean of this distribution is ½ but it is now hyper-polarized, with variance ___. These two distributions also illustrate the idea of a *mean-preserving spread* – the means of the two distributions are the same but the spread – the variance – in the second is greater than in the first. In what follows we assume b is a random variable with continuous probability distribution on [0,1] (in other words, we assume it is something like a beta distribution, though it need not be that particular distribution).

It is straightforward to calculate the expected value of $U(b)$, when $U(b) = b$ and b is a random variable with continuous probability distribution $f(b)$.[15] It is simply $EU(b;1) = \bar{b}$.

It will now be seen that the veil of ignorance has no effect on risk neutral agents. That is, suppose a risk neutral agent is beyond the veil of ignorance and faces a particular value of b – call it b' -- if a Sovereign is chosen. This agent has a critical value of delta, indicated by Equation (2). If this same risk neutral agent were placed behind the veil of ignorance so she cannot be sure of the value of b but believes it will average b', her critical value of

---

[15] It is $EU(b;1) = \int_0^1 bf(b)db = \bar{b}$, where $f(b)$ is the probability density function for a beta distribution. This result is standard in textbooks on decision theory.

delta (derived from Equation 2a) would be *exactly the same as the value without the veil of ignorance*. And it will be invariant to mean preserving spreads of the random variable *b*.

Now consider a risk averse agent behind the veil of ignorance. In particular, let $\alpha = 2$, so the agent has the scaled quadratic utility function shown earlier. In this case, a well-known result from decision theory allows us to calculate the expected utility of *b* as $EU(b;2) = 1 - (1 - \bar{b})^2 - \sigma^2$. In other words, it is the value of the utility function evaluated at the average value of *b*, *minus the variance of b*. So, the more uncertain the agent is about life under the Sovereign, the *smaller* is her expected utility of life under the Sovereign. A moment's reflection should show that qualitatively this is indeed a sensible result, if risk aversion means anything.

Thoughtful readers may already anticipate the next result, but let us make it explicit. In Equation (2a), replace *U*(*b*) with *EU*(*b*;2) and recall that *U*(*a*;2) = 1. Let us associate the worst possible outcome with the State of Nature, so *U*(*c*;2) = 0. Then after a little algebra we have:

$$\delta(2) = \frac{n\left(\bar{b}(2-\bar{b}) - \sigma^2\right)}{1 + (n-1)\left(\bar{b}(2-\bar{b}) - \sigma^2\right)} \tag{2c}$$

Equation (2c) allows us to isolate quite precisely the effect of the veil of ignorance: it adds terms involving the variance of the distribution to both the top and bottom of the relevant ratio. (Note that if $\sigma^2 = 0$, Equation (2c) collapses to Equation (2b), derived from a quadratic utility agent outside the veil of ignorance). But what is the net effect of the uncertainty induced by the veil of ignorance? The answer is: increasing the variance of *b* (with a mean-preserving spread, so that the mean value remains constant) *decreases the critical delta*.[16]

In fact, we have now proven the following proposition:

> **Proposition 5**. If $U(x) = 1 - (1 - x)^2$ and *b* is a continuous random variable defined on [0,1], the critical δ is smaller behind the veil of ignorance than outside the veil of ignorance.

Proposition 3 means that if the prospects of life under the Sovereign are at all uncertain, there will be a range of agents who will turn down the Period 1 proposal, who otherwise would accept it. In other words, given risk aversion, *the veil of ignorance makes delay more likely*.

---

[16] In Equation 2c, $\frac{\partial \delta(2)}{\partial \sigma} = \frac{-2n\sigma}{\left(1 + (n-1)(\bar{b}(2-\bar{b}) - \sigma^2\right)^2} < 0$ for all $\sigma > 0$.

The intuition behind the result is worth noting. If the prospects of life under the Sovereign are uncertain, then the State of Nature become relatively less unattractive. Simultaneously, the Sovereign's payoff becomes relatively more attractive. Hence, delay becomes relatively less painful and relatively more attractive.

In some sense, this result violates the intuitions of some social contract theorists: in the world of Hobbesian bargaining, a veil of ignorance makes it harder to come to agreements, relative to no veil of ignorance (at least, assuming mean preserving spreads). Yet the contract theorists' intuition is partially correct. Clearly, the veil of ignorance does provide agents with an incentive to reduce their future uncertainty. If they can find a way to do so, bargaining over the Sovereign becomes easier. And many of the social contract theorists who have argued for a veil of ignorance would evaluate a less risky post-transition world as a juster, better world.

Thus, the analysis in this section clarifies a central normative issue concerning the veil of ignorance. In evaluating its normative properties, one must weigh the benefits (if any) from incentives to create a less risky and thus arguably better post-transition society, against the costs of delay and more time spent in the State of Nature. For Hobbes, this evaluation would be easy: any time spent in the State of Nature far outweighs non-existent gains from non-credible risk-reduction schemes. He would conclude that a veil of ignorance is simply bad – but unfortunately, like many bad things in life, there is probably nothing one can do about it.

Finally, let us turn to Humes's point: suppose there is a chance that life under the Sovereign might actually turn out worse than life in the State of Nature, though on average expects improvement. (One expects a mediocrity, hopes for a Washington, and fears a Pol Pot). …

7. Discussion and Conclusion

In this chapter, we used contemporary bargaining theory to reformulate a Hobbesian approach to foundational issue in classical political theory, the selection of a Sovereign. We discovered that a modern approach retains much of the old wine in the new bottle, though invariably in somewhat clearer form. For example, Hobbes's intuitions about the prospect of agreement and the number of bargainers, relative nastiness of the State of Nature, and attractiveness of being Sovereign, all receive confirmation and refinement. A modern framework also allowed us to examine the hoary notion of a veil of ignorance, revealing its normative properties as more problematic than many have suggested. In this case, perhaps, the new bottle contains a little new wine (if only sour).

A critical issue in the chapter has been the ability of agents in the state of nature to bind the hands of a Sovereign ex ante, or even to imagine they can do so ex post. In the spirit of Hobbes, we ruled out explicit moves in this direction. But now it is time to see whether there might be allowed back – which would have large implications for the normative evaluation of the veil of ignorance, among other matters.

6. Extensions

Hobbes bargaining offers a tractable framework for analyzing classical problems in political theory. In addition, it suggests some new ones.

Some of the are obvious: suppose only a majority or super-majority of the agents are needed to end the SON; or suppose a specific number *k* is needed. How would this affect the bargaining?

A new question Suppose agents have different discount rates: does this confer an advantage on the patient? If so, how much?

8. Appendix

Critical δ in the Finite Horizon Game

We calculate the value of δ that will lead non-proposers to reject the "Peasant" offer in Round 1 and every subsequent round, until the last one, in which they accept the offer.

In such an equilibrium, in Round 1, a non-proposer who rejects the initial offer receives in expectation

$$\begin{aligned} & c + \delta v_i(0, g) \\ &= c + \delta c + \delta^2 v_i(1, g) \\ &= c + \delta c + \delta^2 c - \ldots - \delta^{\tau-2} c + \delta^{\tau-1} v_i(\tau - 1, g) \end{aligned} \tag{A1}$$

where $v_i(\tau - 1, g) = \dfrac{a + (n-1)b}{n(1-\delta)}$. In other words, the player expects to receive the SON payoff for $\tau - 1$ rounds, plus the continuation value of the game in the last round.

A1 can be decomposed into a stream of payments of c for $\tau - 1$ rounds, plus a one-shot payment worth $\dfrac{\delta^{\tau-1}(a + (n-1)b)}{n(1-\delta)}$. A convenient mathematical fact (derivable from CMF) indicates the present value of a finite stream of one dollar per period:

$$1 + \delta + \delta^2 + \delta^3 + \ldots + \delta^{\tau-1} = \frac{1 - \delta^\tau}{1 - \delta} \tag{CMF2}$$

Hence, we can re-write A1 as

$$c\frac{1 - \delta^{\tau-1}}{1 - \delta} + \frac{\delta^{\tau-1}(a + (n-1)b)}{n(1-\delta)} = \frac{c}{1-\delta} + \frac{\delta^{\tau-1}(a - b + n(b - c))}{n(1-\delta)}$$

In Round 1, a non-proposer will compare this value with the present value of accepting the offer to be Peasant, namely $\frac{b}{1-\delta}$. She will reject the initial Peasant offer if and only if

$$\frac{c}{1-\delta}+\frac{\delta^{\tau-1}(a-b+n(b-c))}{n(1-\delta)} > \frac{b}{1-\delta}$$

which will be true if and only if

$$\delta > \left(\frac{n(b-c)}{a-b+n(b-c)}\right)^{\frac{1}{\tau-1}}$$

Proof of Proposition 3

First we offer two lemmas.

Lemma 1. In any symmetric Markov perfect equilibrium in which weakly dominated strategies have been eliminated, a) $\sigma_i = 1$ if *i* is the nominee, and b) $\sigma_j \neq 0$ for all non-nominees.

Proof. The first part is obvious. To see the second part, suppose all non-nominees employ $\sigma_j = 0$ but there is a very small chance, ε, that they tremble to $\sigma_j = 1$. If so, for a non-nominee *j* employing $\sigma_j = 0$, there is a small chance, ε^{n-2}, that *j* is pivotal. Accordingly, deviating to $\sigma_j = 1$ brings $\varepsilon^{n-2}\frac{b}{1-\delta}+(1-\varepsilon^{n-2})\left(c+\delta v_j(t)\right)$ while playing the prescribed $\sigma_j = 0$ brings $c+\delta v_j(t)$. If $\sigma_j = 0$ for all non-nominees, for small ε, $c+\delta v_i(t) \cong \frac{c}{1-\delta}$ (using CMF2). So the non-nominee should deviate if $\varepsilon^{n-2}\frac{b}{1-\delta}+(1-\varepsilon^{n-2})\left(\frac{c}{1-\delta}\right) > \frac{c}{1-\delta}$. But this will be true if *b* > *c*, which is true by assumption. So the non-nominee will deviate from $\sigma_j = 0$ to $\sigma_j = 1$.■

The intuition is that a chance, however small, of receiving the Peasant payoff beginning today and lasting forever is better than a near certainty of remaining in the state of nature for a very, very long time. Hence, non-nominees cannot use assured rejection.[17]

---

[17] Lemma 1 holds only for sufficiently small ε, in an "all non-nominees surely reject" equilibrium. As shown below, essentially the same construction fails for sufficiently large "trembles" by sufficiently patient players, thereby allowing the construction of a mixed strategy equilibrium.

Lemma 2. In any equilibrium in which a proposal of someone other than the proposer as Sovereign would be accepted with positive probability, the proposer must nominate himself as Sovereign and vote for himself.

Proof. In such an equilibrium, it must be sequentially rational for non-proposers to vote for proposals in which they become peasants. Given this, from symmetry, the single non-proposer nominated Sovereign would also vote for a proposal in which he became a peasant and the proposer became Sovereign. Accordingly, the proposer will nominate himself Sovereign.■

Lemma 2 means that the proposer in the equilibrium in Proposition 3 will not have an incentive to deviate. In addition, it eliminates the following type of equilibria. In the mixed strategy equilibrium, non-nominated individuals are indifferent between voting against a proposal and voting for it. Hence, a single individual could credibly threaten to vote against a proposal in which he himself was not nominated. In this case, the proposer would be indifferent between nominating the person, and not nominating the person, so he could nominate the person. The symmetry requirement eliminates equilibria of this type.

We now present the proof of Proposition 3.

Proof. The proof has three steps. Step 1. Lemma 1 means that a “perpetual disagreement” equilibrium cannot survive elimination of weakly dominated strategies (that is, it cannot be that $\sigma_i = 0$ for all non-nominees). So we focus on symmetric equilibria in which non-nominees affirm with probability $\sigma_i, 0 < \sigma_i \le 1$. From Lemma 2, in all such equilibria, the proposer nominates himself and votes for himself with certainty. Step 2. Assume $\sigma_i = 1$, for all non-nominees. If a non-nominee adheres to the indicated strategy, he receives $\frac{b}{1-\delta}$. If he deviates to any other strategy, σ_i', $0 < \sigma_i' < 1$, he receives $\sigma_i' \frac{b}{1-\delta} + (1-\sigma')\big(c + \delta v_i(t)\big)$, where $v_i(t) = \frac{1}{n}\frac{a}{1-\delta} + \frac{n-1}{n}\frac{b}{1-\delta}$. This deviation will be profitable iff $\frac{b}{1-\delta} < c + \delta \frac{a}{n(1-\delta)}$, that is, iff $\delta > \frac{n(b-c)}{a-b+n(b-c)}$. Thus, an equilibrium in which non-nominees affirm with certainty requires $\delta \le \frac{n(b-c)}{a-b+n(b-c)}$. (Note: we have not yet shown that mixed strategy equilibria cannot hold in this case). Step 3. Assume $0 < \sigma_i < 1$ for all non-nominees. This is a mixed strategy equilibrium, so for non-nominee *i* the expected utility from rejecting with certainty must be the same as affirming with certainty. The expected utility of rejecting with certainty is $c + \delta v_i(t)$. Because the proposer nominates himself and votes for himself with certainty (using Lemma 2), the probability that *i* is pivotal is σ^{n-2}. So the expected utility of affirming with certainty is $\sigma^{n-2}\frac{b}{1-\delta}+(1-\sigma^{n-2})\left(c+\delta v_i(t)\right)$. Thus we have the indifference condition (IC):

$$c+\delta v_i(t)=\left(1-\sigma^{n-2}\right)\left(c+\delta v_i(t)\right)+\sigma^{n-2}\frac{b}{1-\delta}$$

$$v_i(t)=\frac{b-c(1-\delta)}{(1-\delta)\delta} \tag{IC}$$

for $0<\sigma_i<1$. Note that the value to *i* of the game in any sub-game after *t* offers, in the mixed strategy equilibrium is:

$$\begin{aligned}v_i(t)&=\frac{1}{n}\left(\sigma^{n-1}\frac{a}{1-\delta}+(1-\sigma^{n-1})(c+\delta v_i(t+1))\right)+\frac{n-1}{n}\left(\sigma^{n-1}\frac{b}{1-\delta}+(1-\sigma^{n-1})(c+\delta v_i(t+1))\right)\\&=\sigma^{n-1}\frac{a+(n-1)b}{n(1-\delta)}+(1-\sigma^{n-1})(c+\delta v_i(t+1))\end{aligned} \tag{A2}$$

However, because strategies are stationary, the value of every sub-game must be the same irrespective of *t*. Hence, in equation (A2) $v_i(t+1)=v_i(t)$. Using this fact and simplifying results in

$$v_i(t)=\frac{a+(n-1)\sigma^{n-1}b-nc(\sigma^{n-1}-1+\delta(1-\sigma^{n-1}))}{(1-\delta)n(1-\delta(1-\sigma^{n-1}))} \tag{A3}$$

Combining (A3) and the indifference condition (IC) yields $\sigma^{n-1}=\frac{n(b-c)(1-\delta)}{(a-b)\delta}$ and thus $\sigma_i=\left(\frac{n(b-c)(1-\delta)}{(a-b)\delta}\right)^{\frac{1}{n-1}}$, which is the condition in the proposition. Note that as δ approaches $\frac{n(b-c)}{a-b+n(b-c)}$ from above, σ_i approaches 1. For $\delta<\frac{n(b-c)}{a-b+n(b-c)}$, $\sigma_i>1$, indicating that (IC) cannot hold. We have now shown that if $\delta\le\frac{n(b-c)}{a-b+n(b-c)}$, $\sigma_i=1$ and cannot take any other value, while if $\delta>\frac{n(b-c)}{a-b+n(b-c)}$, $\sigma_i=\left(\frac{n(b-c)(1-\delta)}{(a-b)\delta}\right)^{\frac{1}{n-1}}$ and cannot take any other value. In tandem with Lemma 2, this completes the proof.■

References

Baron, David and John Ferejohn. 1989. “Bargaining in Legislatures,” *American Political Science Review* 83(4): 1181-1206.

Dixit, Avinash. 2004. *Lawlessness and Economics: Alternative Modes of Governance*. Princeton: Princeton University Press.

Gibbons, Robert. *Game Theory for Applied Economists*. MIT Press.

Hampton, Jean. *Hobbes and the Social Contract Tradition*.

Hobbes, Thomas. *Leviathan*.

Kavka, Gregory.

Morton, Rebecca. NYU working paper.

Myerson and Weber 1992 *APSR* article.

Osbourne and Rubenstein.

Rubenstein, Ariel. 1982. “Perfect Equilibrium in a Bargaining Model,” *Econometrica* 50: 97-109.

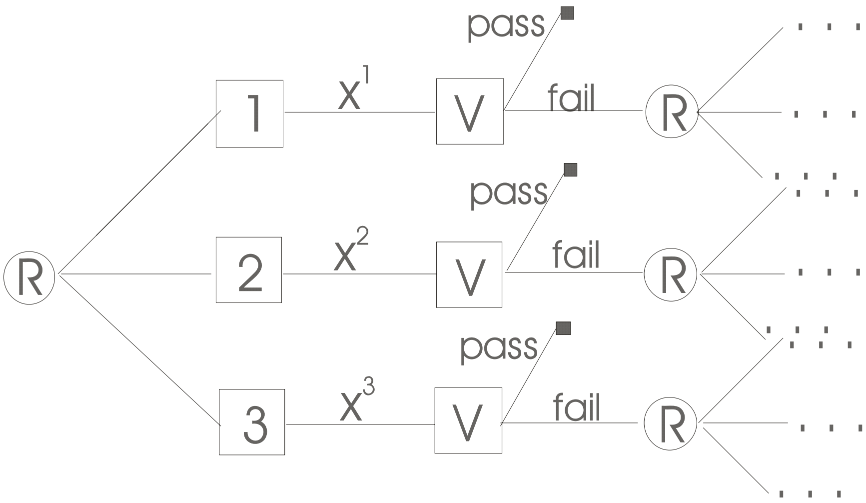

Figure 1. The Hobbes Bargaining Game

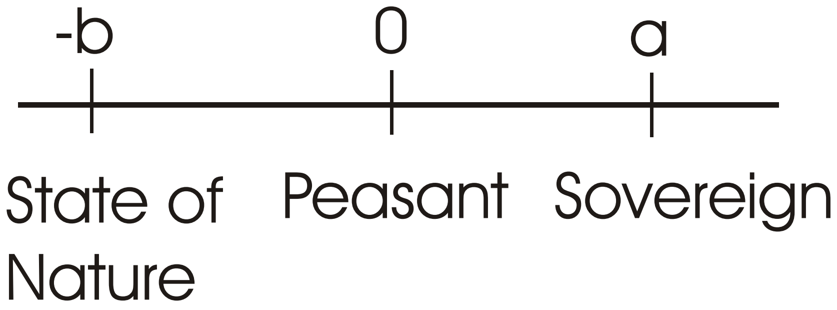

Figure 2. Stage Payoffs in the Hobbes Game

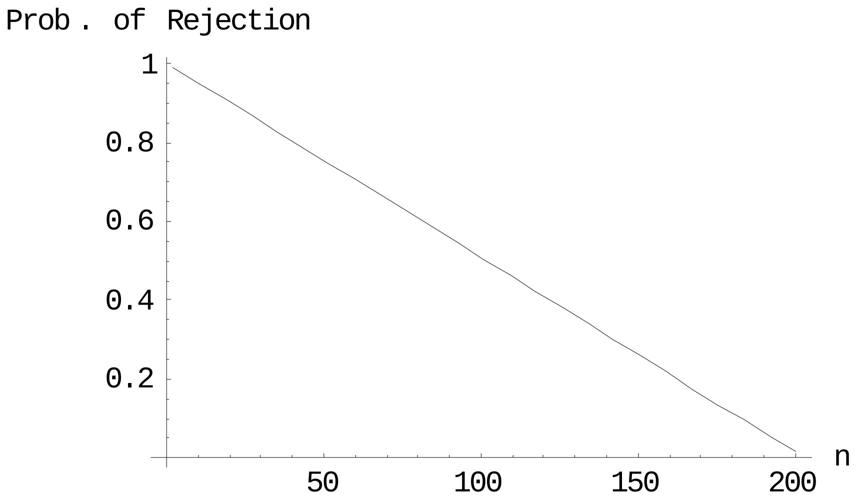

Figure 3. Probability of a Disagreement (a = 100, b = 1, $\delta = 2/3$).

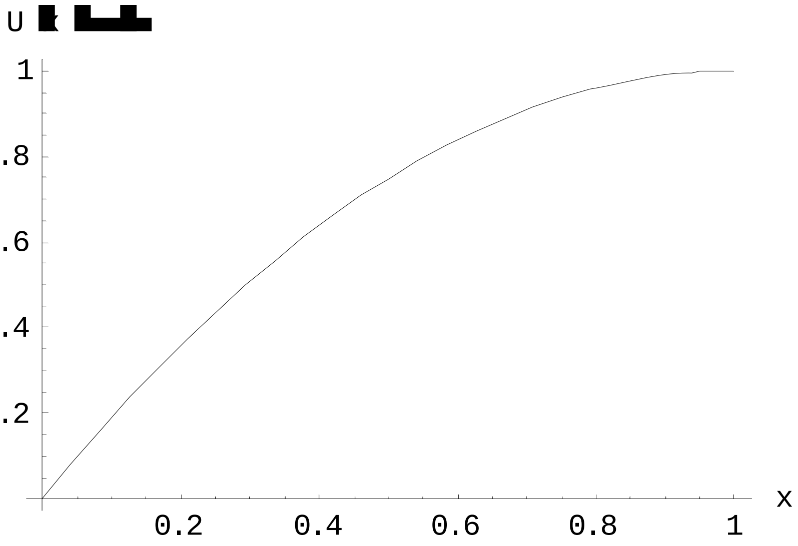

Figure 4. A Risk Averse Utility Function for Hobbes Bargainers

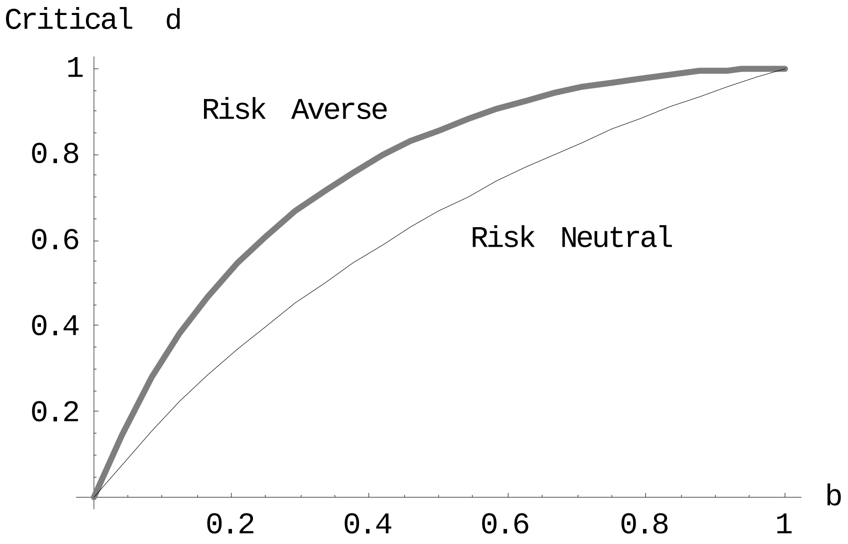

Figure 5. Critical Deltas for Risk Averse vs. Risk Neutral Agents

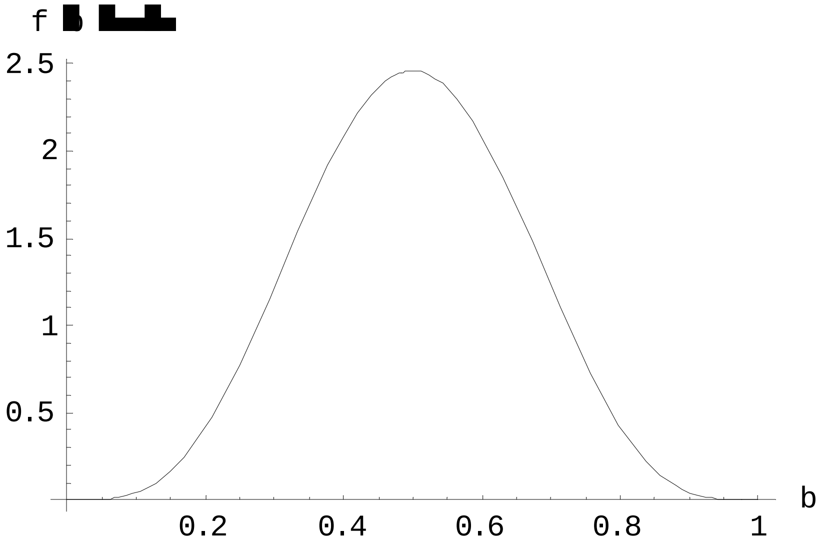

Figure 6. Beta distribution with Shape Parameters (5,5).

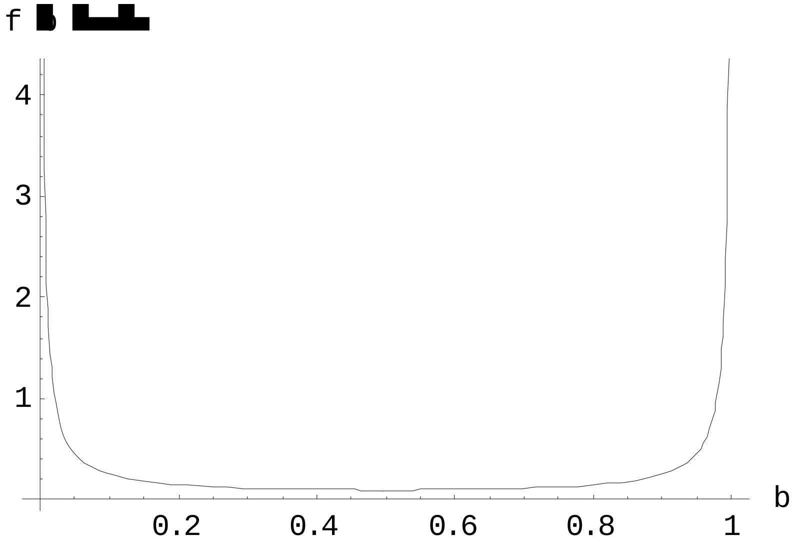

Figure 7. Beta Distribution with Shape Parameters (.05, .05)

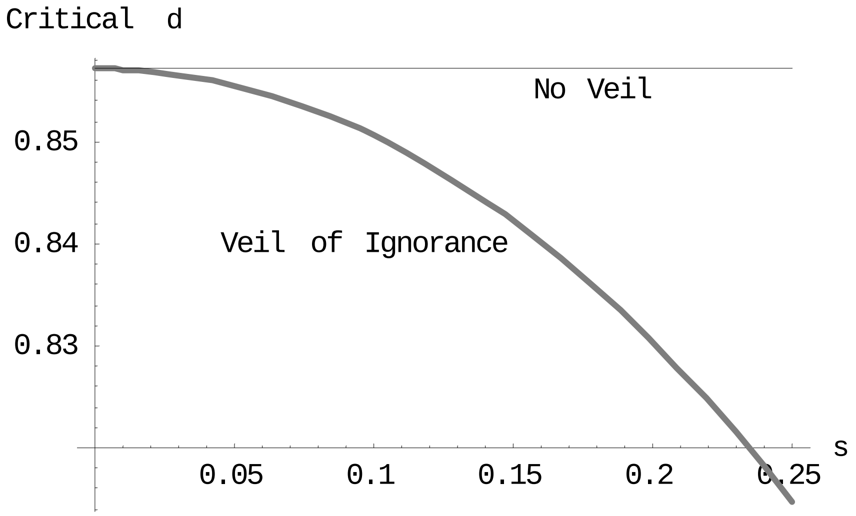

Figure 8. Critical Deltas Behind and Beyond the Veil of Ignorance (quadratic utility, $n = 2$, $\bar{b} = \frac{1}{2}$)