The Chairman paradox and proper equilibrium

Kirill Zernikov

August 2024

Abstract

The Chairman Paradox is a classical observation in voting games showing that a Chairman endowed with tie-breaking power might end up with her worst outcome. The analysis posits three players whose preferences build a Condorcet cycle and invokes Iterated Elimination of Weakly Dominated Strategies (IEWDS) to select a unique equilibrium. We generalize this game to the one with m voters of each type and prove that set of all proper equilibria contains a profile with the worst outcome for the Chairmen. If we assume that Chairmen are always voting for their most preferred alternative, which has a support from real life observations, then set of proper equilibria contains only profiles in which the worst alternative for Chaimen wins.

1 Introduction to Chairman Game and Paradox

The classical Chairman Paradox involves a voting game with three alternatives and three voters. Suppose we have three alternatives: x, y, z, three voters: $1, 2, 3$ and their preferences form a Condorcet cycle, i.e.

1) Preferences of voter 1: $x \succ y \succ z$

2) Preferences of voter 2: $z \succ x \succ y$

3) Preferences of voter 3: $y \succ z \succ x$

However, in contrast to problem of regular Condorcet cycle, the ties are broken in favor of whatever alternative the voter 1 voted for. So, for example, in case of sincere voting the alternative x wins. This setting defines a $3 \times 3 \times 3$ normal-form game with strategy sets $S_1 = S_2 = S_3 = \{x, y, z\}$.

This game has 5 Nash equilibria in pure strategies: three unanimous ones – $X = (x, x, x)$; $Y = (y, y, y)$; $Z = (z, z, z)$, one in which the favourite alternative of Chairman (voter 1) wins – $G = (x, x, y)$, and one rather paradoxical one, in which the worst alternative for Chairman wins – $B = (x, z, z)$.

The original argument uses the process of Iterative Elimination of Weakly Dominated Strategies (IEWDS for short) in the following way. Strategy x weakly dominates y and z for player 1, strategy z weakly dominates y for player 2 and strategy y weakly dominates x for player 3. Thus, if we eliminates all these weakly dominated strategies, we will be left with 2×2 game with reduced strategy sets $\tilde{S}_1 = \{x\}$, $\tilde{S}_2 = \{z, x\}$, $\tilde{S}_3 = \{y, z\}$. Now all unanimous equilibria have been eliminated, but both G and B survived. However, strategy z weakly dominates x for player 2 and y for player 3. Hence, via IEWDS only the paradoxical equilibrium B survived and the paradox has been achieved.

2 Further refinements

Even though IEWDS allows us to reach paradox, this solution concept is heavily criticized, mainly because in general case it depends on the order of elimination of weakly dominated strategies. But the main idea of IEWDS remains valuable, since it stems from the notion of *admissibility*, i.e. the argument that weakly dominated strategies should not be played in an equilibrium. This argument seems to be intuitive, because if the opponents were to tremble even slightly from non-admissible equilibrium, the payoff of the player that plays weakly dominated strategy would be strictly worse, than if he would play strategy that weakly dominates.

So the natural thing to do is to conduct a research on tremble-based refinements in Chairman paradox, which was done by Alos-Ferrer in [1]. We will quickly present the results of this work, but before let us introduce some notation.

Consider normal-form game with player set I, strategy sets S_i, mixed strategy sets Σ_i and denote the payoffs of player i when playing $\sigma_i \in \Sigma_i$ by $\pi_i(\sigma_i|\sigma_{-i})$, where $\sigma_{-i} = (\sigma_j)_{j\neq i}$ is the vector of strategies of i's opponents. Let $\Sigma = \prod_{i\in I} \Sigma_i$ be the set of mixed strategy profiles. Call a mixed strategy σ_i completely mixed if $\sigma_i(s_i) > 0$ for any $s_i \in S_i$ and call σ completely mixed if σ_i is completely mixed for every i.

Now we are ready to give formal definition of *trembling-hand perfect equilibrium*, following Selten [2].

Definition 1. *A trembling-hand perfect equilibrium (THPE) is a strategy profile $\sigma^* \in \Sigma$ such that there exists a sequence $\sigma^n \to \sigma^*$ with σ^n completely mixed for all n such that, for each n and each $i \in I$,*

$$\sigma_i \in \arg\max_{\sigma_i \in \Sigma_i} \pi_i(\sigma_i|\sigma^n_{-i})$$

To apply trembling-hand perfection, the cardinal-payoff representation of the preferences of the players needs to be specified. However, all results below are independent of the cardinal representation of preferences and remain unaffected if the specific payoffs are perturbed, as long as the induced ordinal preferences are not changed. Since there are only three options, preferences only depend on the selected option, and no voter is indifferent among two different options, the utilities of players can be normalized such that the worst option of a player yields payoff 0, the best option yields payoff 1, and the intermediate option yields payoff $a \in (0, 1)$ (whether one uses the same a for all three players or three different constants is inconsequential, as there will be no interpersonal comparisons in the analysis). This notation will be fixed for the remainder of the paper.

Using this cardinal-payoff presentation, Alos-Ferrer in his paper proves the following theorem:

Theorem 1. *In the Chairman Game, for any representation of the voters' preferences, there are two and only two THPE, $G = (x, x, y)$ and $B = (x, z, z)$. None of the unanimous Nash equilibria (x, x, x), (y, y, y), and (z, z, z), and none of the non-degenerate mixed Nash equilibria, is a THPE.*

Thus in some sense seemingly resolving the paradox. However, he also shows that B is also a *truly perfect equilibrium*, meaning that the property in Definition 1 holds for any sequence $\sigma_n = (\sigma^n_1, \sigma^n_2, \sigma^n_3)$ of completely mixed strategies converging B, $\sigma_n \to B$ as $n \to \infty$. This also implies that it is a *strictly perfect equilibrium*, which hints that

there might be an equilibrium that eliminates G. And indeed there is one, none other the popular concept known as the *proper equilibrium*.

We now turn to proper equilibria as introduced by Myerson [3]

Definition 2. *A proper equilibrium is a strategy profile $\sigma^* \in \Sigma$ such that there exists a sequence $\sigma^n \to \sigma^*$ with σ^n completely mixed for all n and a real-valued sequence $\varepsilon_n \to 0$ with $0 < \varepsilon_n < 1$ such that, for each n, $i \in I$ and for all $s_i, s_i' \in S_i$,*

$$\text{if } \pi_i(s_i|\sigma_{-i}^n) < \pi(s_i'|\sigma_{-i}^n), \text{ then } \sigma_i^n(s_i) \leq \varepsilon_n \cdot \sigma_i^n(s_i')$$

Alos-Ferrer also proved the second theorem:

Theorem 2. *In the Chairman Game, for any representation of the voters' preferences, the only proper equilibrium is $B = (x, z, z)$.*

Thus reinstating the Chairman Paradox.

3 General Case

In our work we would like to generalize the Chairman paradox in the following way: suppose than now there are m people with preferences $x \succ y \succ z$, m people with preferences $z \succ x \succ y$ and m people with preferences $y \succ z \succ x$.

Our main result is the following theorem:

Theorem 3. *$G = (x, \dots, x; x, \dots, x; y, \dots, y)$ and $B = (x, \dots, x; z, \dots z; z, \dots z)$ are both THPE. The equilibrium B is also the proper equilibrium.*

Our proof will be divided intro two parts: in the first part we will prove that B is a proper equilibrium and thus a THPE equilibrium. In the second part we will prove that G is a THPE.

Proof. Part 1:

Let's begin by proving that B is a proper equilibrium. We will do it by definition, using the following system of equilibria, approaching B:

$$\sigma_1^n = \dots = \sigma_m^n = (1 - \frac{1}{n^{100m!}} - \frac{1}{n^{200m!}}; \frac{1}{n^{100m!}}; \frac{1}{n^{200m!}})$$
$$\sigma_{m+1}^n = \dots = \sigma_{3m}^n = (\frac{1}{n^{3m}}; \frac{1}{n^2}; 1 - \frac{1}{n^{3m}} - \frac{1}{n^2})$$

Now we need to prove the statement from Definition 2.

1) Firstly we will prove it for the first type of voters (the first m voters). Without loss of generality, it is enough to prove that $\pi_1(x|\sigma_{-1}^n) > \pi_1(y|\sigma_{-1}^n) > \pi_1(z|\sigma_{-1}^n)$, i.e. the inequalities in Definition 2 only for player 1. Notice that for each $j \in \{x, y, z\}$:

$$\pi_1(j|\sigma_{-1}^n) = \sum_{\{i_2,\dots,i_{3m}\} \in \{x,y,z\}^{3m-1}} \sigma_2^n(i_2) \dots \sigma_{3m}^n(i_{3m}) \cdot \pi_1(s_1 = j; s_2 = i_2; \dots; s_{3m} = i_{3m}),$$

where $\pi_1(s_1 = j; s_2 = i_2; \dots; s_{3m} = i_{3m})$ is the payoff of player 1 when he plays pure strategy j, 2^{nd} player plays pure strategy i_2, …, the last player plays pure strategy i_{3m}, so it is either 1, a or 0.

One can also notice that this sum is a polynomial of n^{-1} and it has 3^{3m-1} initial summands (before we open the brackets). Each of these 3^{3m-1} summands is itself a polynomial of n^{-1}. Call the minimal power of n^{-1} in this polynomial the *degree of polynomial*. Let's order these polynomials by their degree of n^{-1} from the least one to the biggest one. So the first summand will be the one that has $\{i_2 = x, \dots, i_m = x; i_{m+1} = z, \dots, i_{3m} = z\}$ and the power of n^{-1} will be zero. But notice that in $\pi_1(x|\sigma^n_{-1})$, $\pi_1(y|\sigma^n_{-1})$ and $\pi_1(z|\sigma^n_{-1})$ this first summand will be multiplied by the same number, since it does not matter what player 1 will play — z will win regardless.

Now let's look at the next summands. After the first one, by design of our system of equilibria, we will have $2m$ summands, in each of which $\{i_2 = x, \dots, i_m = x\}$ and one of the $\{i_{m+1}, \dots, i_{3m}\}$ is equal to y, while all the other ones are equal to z. The power of n^{-1} of each of these $2m$ summands is 2. But again, in $\pi_1(x|\sigma^n_{-1})$, $\pi_1(y|\sigma^n_{-1})$ and $\pi_1(z|\sigma^n_{-1})$ all these summands will be multiplied by the same number, since it does not matter what player 1 will play — z will win regardless.

And so on.

But since we are only interested in difference between $\pi_1(x|\sigma^n_{-1})$, $\pi_1(y|\sigma^n_{-1})$ and $\pi_1(z|\sigma^n_{-1})$, we need to look only at the first summand, in which the voice of player 1 actually matters. Again, by design of our system, it will happen only after m players of $\{m+1, \dots, 3m\}$ will play y instead of z. In this case

- $m-1$ players are playing x
- m players are playing y
- m players are playing z

So wins the alternative in which favor 1^{st} player will vote for. If he votes for x he will get 1, if he votes for y he will get a, if he votes for z he will get 0. Thus, formally speaking:

$$\pi_1(x|\sigma^n_{-1}) - \pi_1(y|\sigma^n_{-1}) = (1-a) \cdot \left(\frac{1}{n^2}\right)^m + o\left(\frac{1}{n^{2m}}\right)$$
$$\pi_1(y|\sigma^n_{-1}) - \pi_1(z|\sigma^n_{-1}) = a \cdot \left(\frac{1}{n^2}\right)^m + o\left(\frac{1}{n^{2m}}\right)$$

So $\pi_1(x|\sigma^n_{-1}) > \pi_1(y|\sigma^n_{-1}) > \pi_1(z|\sigma^n_{-1})$ for n large enough.

2) Now lets prove it for the second type type of voters (the second batch of m voters). Without loss of generality we have to prove that $\pi_{m+1}(z|\sigma^n_{-(m+1)}) > \pi_{m+1}(y|\sigma^n_{-(m+1)}) > \pi_{m+1}(x|\sigma^n_{-(m+1)})$.

The idea of the proof is the same as in 1), i.e. we are looking for the first summand in which the vote of player $m+1$ actually matters. Again, by design of our system, it will happen only after $m-1$ players of $\{m+2, \dots, 3m\}$ will play y instead of z. In this case:

- m players are playing x
- $m-1$ players are playing y
- m players are playing z

So if he votes for y, x will win and player $m+1$ will get a. But if he votes for z, z will win and he will get 1. Thus:

$$\pi_{m+1}(z|\sigma^n_{-(m+1)}) - \pi_{m+1}(y|\sigma^n_{-(m+1)}) = (1-a)\cdot\left(\frac{1}{n^2}\right)^{m-1} + o\left(\frac{1}{n^{2m-2}}\right)$$

So $\pi_{m+1}(z|\sigma^n_{-(m+1)}) > \pi_{m+1}(y|\sigma^n_{-(m+1)})$.

Now we just need to prove that $\pi_{m+1}(y|\sigma^n_{-(m+1)}) > \pi_{m+1}(x|\sigma^n_{-(m+1)})$. In the previous case it did not matter if player $m+1$ voted for x or y. Thus the first summand in which the vote of player $m+1$ towards x or y actually matters will be only after m players of $\{m+2,\ldots,3m\}$ will play y instead of z. In this case:

- m players are playing x
- m players are playing y
- $m-1$ players are playing z

So if player $m+1$ plays x, x will win and he will get a; but if he plays y, y will win and he will get zero. Formally speaking:

$$\pi_{m+1}(y|\sigma^n_{-(m+1)}) - \pi_{m+1}(x|\sigma^n_{-(m+1)}) = a\cdot\left(\frac{1}{n^2}\right)^{m} + o\left(\frac{1}{n^{2m}}\right)$$

So $\pi_{m+1}(y|\sigma^n_{-(m+1)}) > \pi_{m+1}(x|\sigma^n_{-(m+1)})$.

3) Now we finally move on to the third type of voters (the last batch of m voters). Without loss of generality we have to prove that $\pi_{2m+1}(z|\sigma^n_{-(2m+1)}) > \pi_{2m+1}(y|\sigma^n_{-(2m+1)}) > \pi_{2m+1}(x|\sigma^n_{-(2m+1)})$.

The idea of the proof is the same as in 1), i.e. we are looking for the first summand in which the vote of player $2m+1$ actually matters. Again, by design of our system, it will happen only after $m-1$ players of $\{m+2,\ldots,2m,2m+2,\ldots 3m\}$ will play y instead of z. In this case:

- m players are playing x
- $m-1$ players are playing y
- m players are playing z

So if he votes for y, x will win and player $m+1$ will get 0. But if he votes for z, z will win and he will get a. Thus:

$$\pi_{2m+1}(z|\sigma^n_{-(m+1)}) - \pi_{2m+1}(y|\sigma^n_{-(m+1)}) = a\cdot\left(\frac{1}{n^2}\right)^{m-1} + o\left(\frac{1}{n^{2m-2}}\right)$$

So $\pi_{2m+1}(z|\sigma^n_{-(m+1)}) > \pi_{2m+1}(y|\sigma^n_{-(m+1)})$. And since y weakly dominates x for voter $2m+1$: $\pi_{2m+1}(y|\sigma^n_{-(m+1)}) > \pi_{2m+1}(x|\sigma^n_{-(m+1)})$ and our proof for equilibrium B is done.

Part 2:

Now let's prove that G is a THPE equilibrium. We again do it by definition, using the following system of equilibria:

$$\sigma_1^n = (1 - \frac{1}{n^2} - \frac{1}{n^{3m}}; \frac{1}{n^2}; \frac{1}{n^{3m}})$$
$$\sigma_2^n = \cdots = \sigma_{2m}^n = (1 - \frac{1}{n^4} - \frac{1}{n^{3m}}; \frac{1}{n^4}; \frac{1}{n^{3m}})$$
$$\sigma_{2m+1}^n = \cdots = \sigma_{3m}^n = (\frac{1}{n^{200m!}}; 1 - \frac{1}{n^{100m!}} - \frac{1}{n^{200m!}}; \frac{1}{n^{100m!}})$$

The proof is a simple check that is very similar to the one in *Part 1* and is omitted.

□

4 Chairman Game in real life

In 2021 Granic and Wagner conducted a study on how people vote in Chairman Game in real life. The results were quite surprising and I genuinely recommend reading it, but will just point out some of the most surprising facts:

1) $B = (x, z, z)$ happens only in 40% of elections
2) (x, x, z) happens in 28% of elections
3) (x, z, y) and $G = (x, x, y)$ got 14% and 5% respectively
4) Overall alternative x won 52% of all elections, while z won 46% of all elections
5) Player 1 votes for x in 94% of cases!

Last point is the inspiration for our final result, which is the following theorem:

Theorem 4. *Suppose that it is common knowledge that first m players are playing x, i.e. we will be looking at a restricted game, in which $\tilde{S}_1 = \cdots = \tilde{S}_m = \{x\}$. In this reduced game IEWDS is independent of the order of elimination of strategies, it selects only one equilibrium and this equilibrium is $B = (x, \ldots, x; z, \ldots, z; z, \ldots, z)$. B is also a proper equilibrium and in any other proper equilibrium alternative z is also winning.*

Proof. Since payoffs of players are based only on the outcome of election, the condition on independence of IEWDS from order of elimination of strategies is trivially fulfilled (check Marx and Swinkels [4, 5] for more details).

Now let's prove that $B = (x, \ldots, x; z, \ldots, z; z, \ldots, z)$ is the only surviving strategy after IEWDS. Notice that for players of 2^{nd} type z weakly dominates y, so they will not be be playing y in an equilibrium that survives IEWDS. But that means that there is no more than m people voting for y in an equilibrium of such type, so y cannot win. Then z weakly dominates y for players of the 3^{rd} type and for them z obviously dominates x, so last m players will be playing z in an equilibrium that survives after IEWDS. Now if the first m players are playing x and the last m players are playing z, z weakly dominates x for players $\{m + 1, \ldots, 2m\}$. Thus B is the only equilibrium, satisfying IEWDS. In the previous section we have already shown that B is a proper equilibrium in the full generalization of the Chairman Game, thus it is a proper equilibrium in this case.

Koriyama and Nunez [6] in their 2015 paper proved that if a normal-form game is solvable through IEWDS and fulfills a condition in Marx and Swinkels [4, 5], the unique IEWDS outcome must coincide with the payoff of a proper equilibrium. Thus, alternative z winning in a proper equilibrium and the Chairman Paradox is restored in general case (keeping in mind the restriction from the statement of the theorem).

□

References

[1] C. Alós-Ferrer, "The trembling chairman paradox," *Games and Economic Behavior*, vol. 131, pp. 51–56, 2022.

[2] R. Selten, "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, vol. 4, pp. 25–55, 1975.

[3] R. B. Myerson, "Refinements of the nash equilibrium concept," *International Journal of Game Theory*, vol. 7, pp. 73–80, 1978.

[4] L. M. Marx and J. M. Swinkels, "Order independence for iterated weak dominance," *Games and Economic Behavior*, vol. 18, no. 2, pp. 219–245, 1997.

[5] L. M. Marx and J. M. Swinkels, "Order independence for iterated weak dominance," *Games and Economic Behavior*, vol. 31, no. 2, pp. 324–329, 2000.

[6] Y. Koriyama and M. Núñez, "How proper is the dominance-solvable outcome?," *PSN: Game Theory (Topic)*, 2015.