Learning to Play Bayesian Games

Eddie Dekel

Drew Fudenberg

David K. Levine

Goals

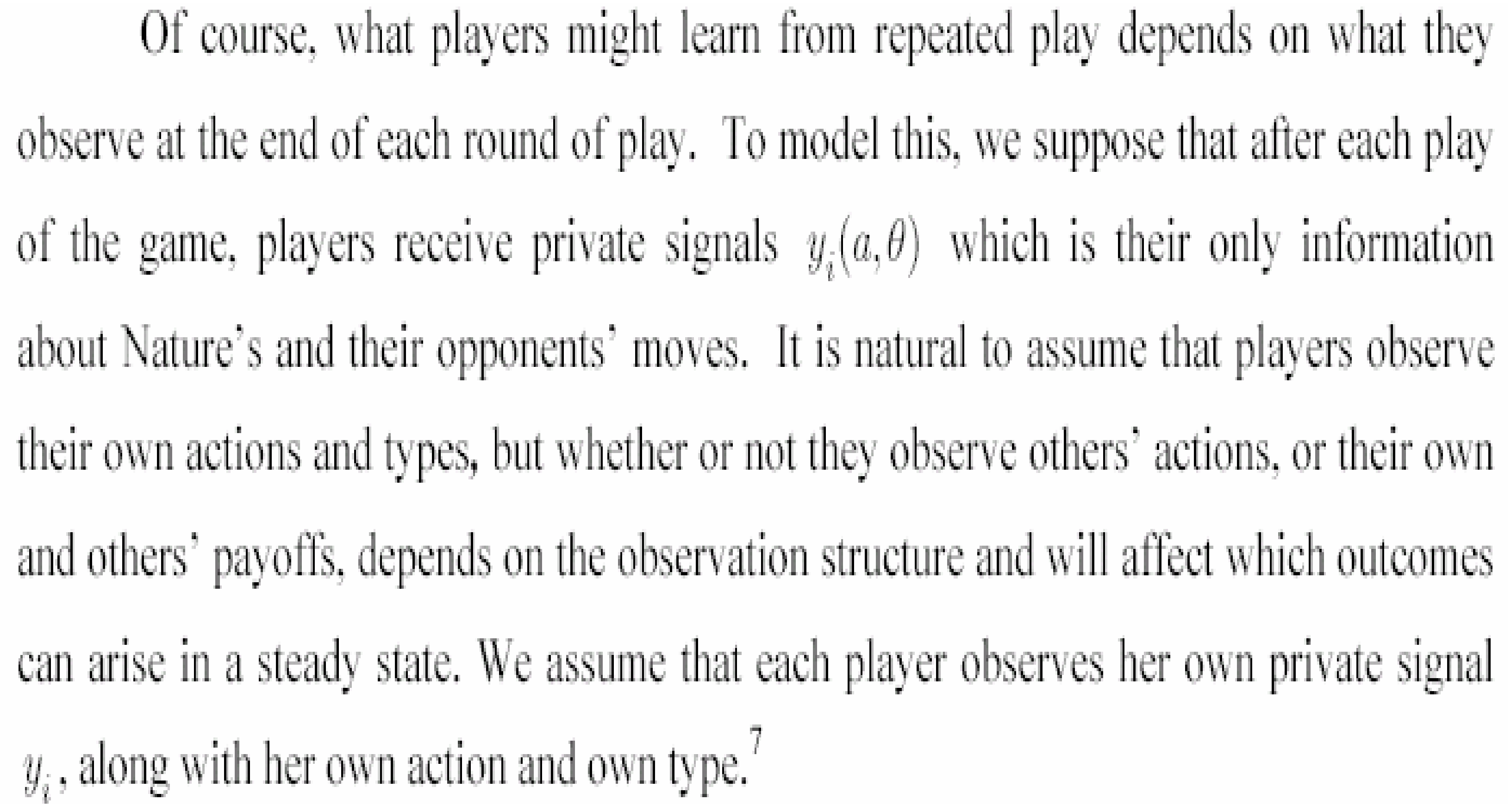
- Our premise is that equilibrium in games arises as the result of learning, and that just what people will learn depends both on the true distribution of Nature's move and on what they observe when the game is played.

- In order for repeated observations to lead players to learn the distribution of opponents’ strategies, the signals observed at the end of each round of play must be sufficiently informative. Such information will tend to lead players to also have correct and hence identical beliefs about the distribution of Nature’s moves.

We consider a static simultaneous-move game with I player roles.[2] (All parameters of the game, including the number of players, and their possible actions and types, are assumed to be finite.) In the static game, Nature moves first, determining players' types, which we denote $\theta_i \in \Theta_i$. To model cases where the types alone do not determine the realized payoffs, we also allow Nature to pick $\theta_0 \in \Theta_0$; we call this "Nature's type." Players observe their types, and then simultaneously choose actions $a_i \in A_i$ as a function of their type, so that a strategy σ_i for player i is a map from her types to mixed actions. Player i's utility $u_i(a,\theta)$ depends on the profile $a=(a_1,...,a_I) \in A$ of realized actions, and on the realization $\theta=(\theta_0,\theta_1,...,\theta_I) \in \Theta$ of Nature's move. When $u_i(a,\theta) = u_i(a,\theta_i)$ we refer to the game as having *private values*.

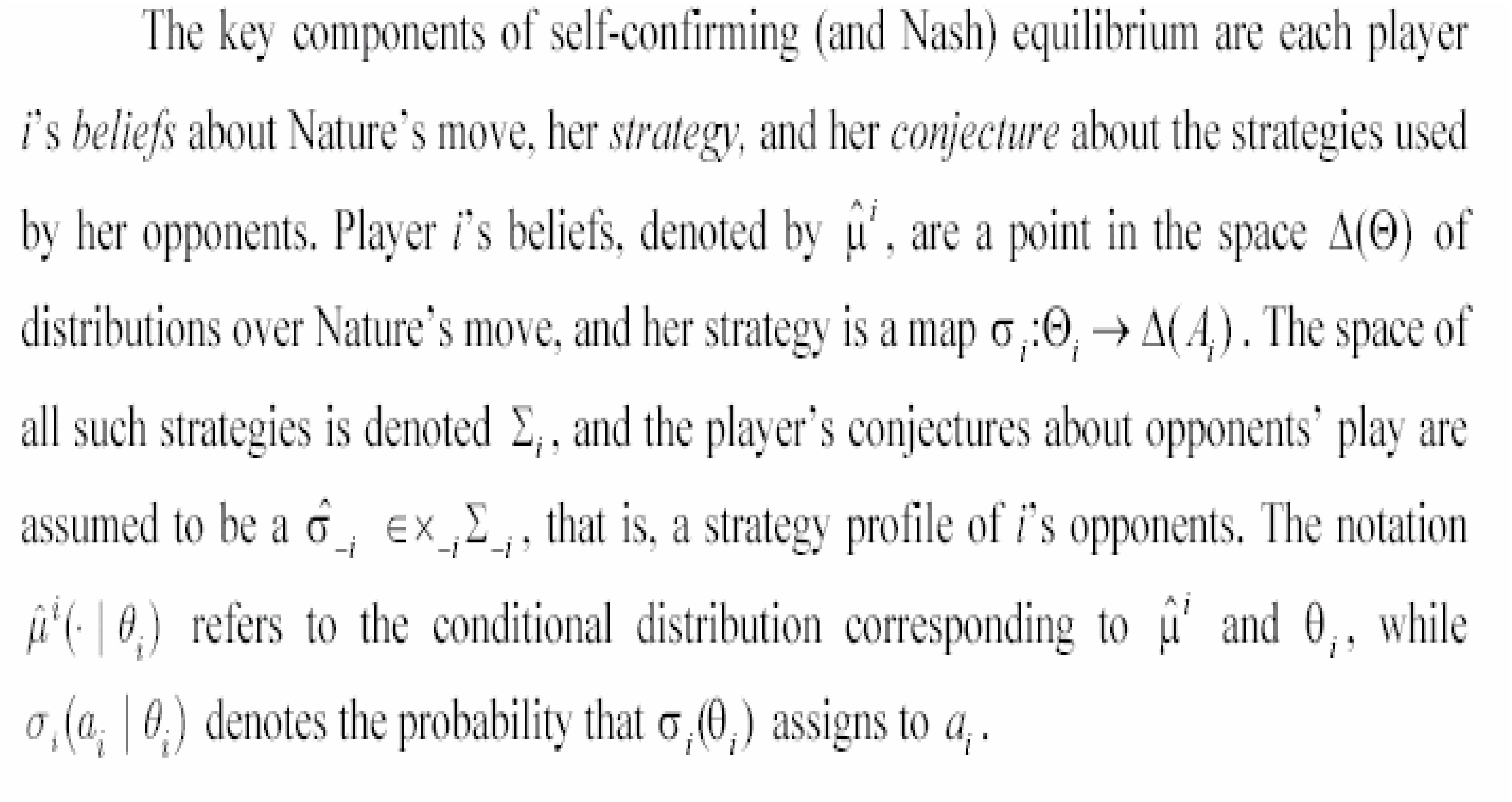
Our solution concept is motivated by thinking about a learning environment in which the game given above is played repeatedly. We suppose that players know their own payoff functions and the sets of possible moves by Nature (Θ) and players (A); but they know neither the strategies used by other players nor the distribution of Nature's move; the players learn about these latter variables from their observations after each period of play. We also suppose that each period the types are drawn independently over time from a fixed distribution p. Thus p corresponds to the true distribution of Nature's move in the stage game, and when $\mu^i = p$ for all players i we say that *the priors are correct*.[6] For the time being, we also assume that there is a single agent in each player role. Section 4 discusses the case where there is a large population of agents in each role who are matched together to play the game; we also discuss there the possibility that types are generated by a more general stochastic process.

Signal content



Of course, what players might learn from repeated play depends on what they observe at the end of each round of play. To model this, we suppose that after each play of the game, players receive private signals $y_i(a,\theta)$ which is their only information about Nature's and their opponents' moves. It is natural to assume that players observe their own actions and types, but whether or not they observe others' actions, or their own and others' payoffs, depends on the observation structure and will affect which outcomes can arise in a steady state. We assume that each player observes her own private signal y_i, along with her own action and own type.[7]

Strategy and Conjecture



The key components of self-confirming (and Nash) equilibrium are each player *i*'s *beliefs* about Nature's move, her *strategy*, and her *conjecture* about the strategies used by her opponents. Player *i*'s beliefs, denoted by $\hat{\mu}^i$, are a point in the space $\Delta(\Theta)$ of distributions over Nature's move, and her strategy is a map $\sigma_i : \Theta_i \to \Delta(A_i)$. The space of all such strategies is denoted Σ_i, and the player's conjectures about opponents' play are assumed to be a $\hat{\sigma}_{-i} \in \times_{-i} \Sigma_{-i}$, that is, a strategy profile of *i*'s opponents. The notation $\hat{\mu}^i(\cdot \mid \theta_i)$ refers to the conditional distribution corresponding to $\hat{\mu}^i$ and θ_i, while $\sigma_i(a_i \mid \theta_i)$ denotes the probability that $\sigma_i(\theta_i)$ assigns to a_i.

Definition: *A strategy profile* σ *is a* self-confirming equilibrium with conjecture $\hat{\sigma}_{-i}$ *and* beliefs $\hat{\mu}_i$ *if for each player* i,

(i) $p(\theta_i) > 0$ *implies* $\hat{\mu}_i(\theta_i) > 0$,

and for any pair $\theta_i, \hat{a}_i$ *such that* $\hat{\mu}^i(\theta_i) \cdot \sigma_i(\hat{a}_i | \theta_i) > 0$ *both the following conditions are satisfied*

(ii) $\hat{a}_i \in \arg\max_{a_i} \sum_{a_{-i},\theta_{-i}} u_i(a_i, a_{-i}, \theta_i, \theta_{-i}) \hat{\mu}^i(\theta_{-i} \mid \theta_i) \hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$,

and

(iii) $$\sum_{\{a_{-i},\theta_{-i}:y_i(\hat{a}_i,a_{-i},\theta_i,\theta_{-i})=\overline{y}_i\}} \hat{\mu}^i(\theta_{-i} \mid \theta_i)\hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$$

$$= \sum_{\{a_{-i},\theta_{-i}:y_i(\hat{a}_i,a_{-i},\theta_i,\theta_{-i})=\overline{y}_i\}} p(\theta_{-i} \mid \theta_i)\sigma_{-i}(a_{-i} \mid \theta_{-i}).$$

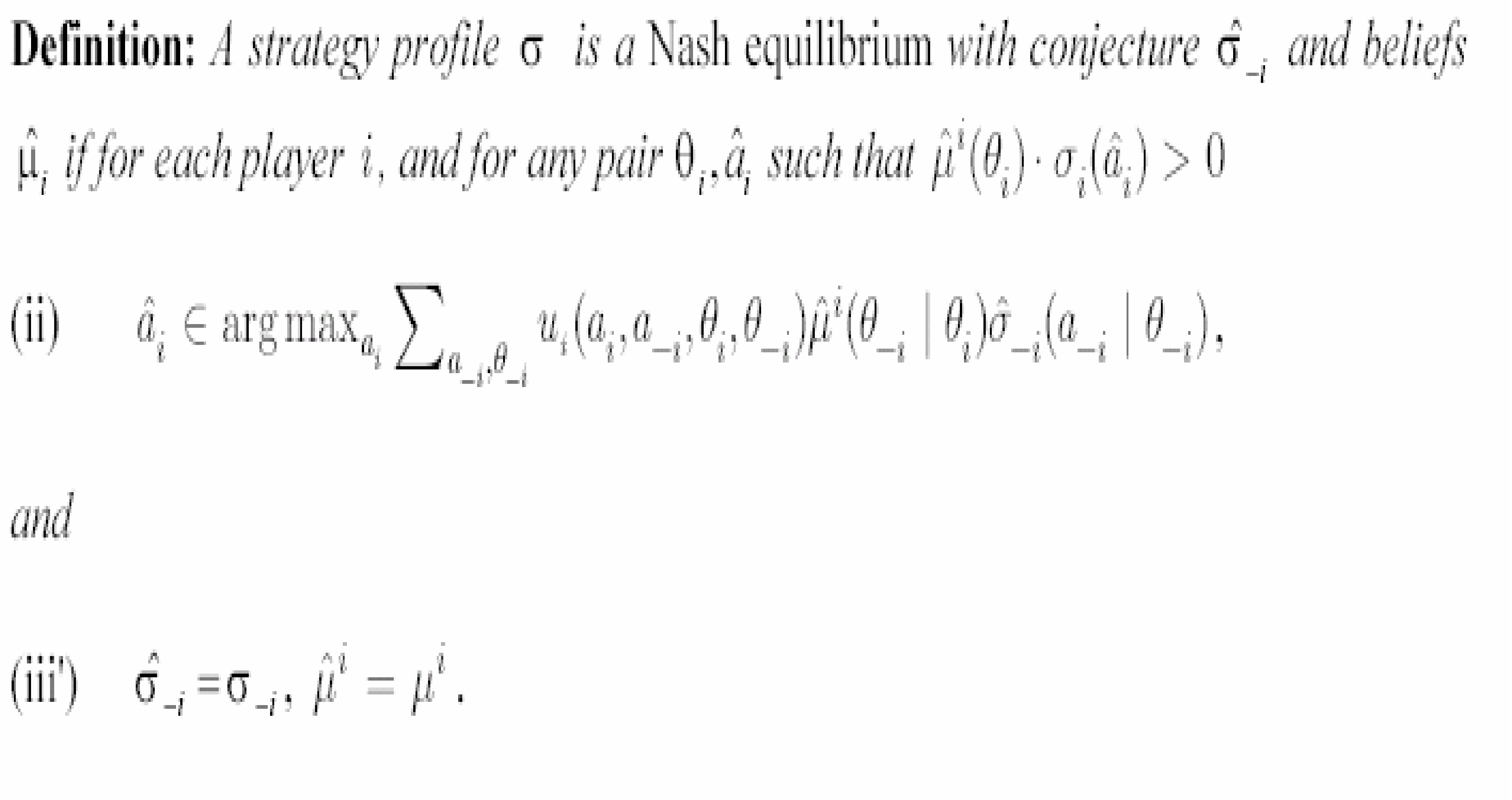
Explanation

- Condition (i) is a consequence of the assumption that players observe their own types.

- Condition (ii) says that any action played by a type of player *i* that has positive probability is a best response to her conjecture about opponents' play and beliefs about Nature's move.

- Condition (iii) says that the distribution of signals (conditional on type) that the player expects to see equals the actual distribution. This captures the least amount of information that we would expect to arise as the steady state of a learning process.

Compare to a NE



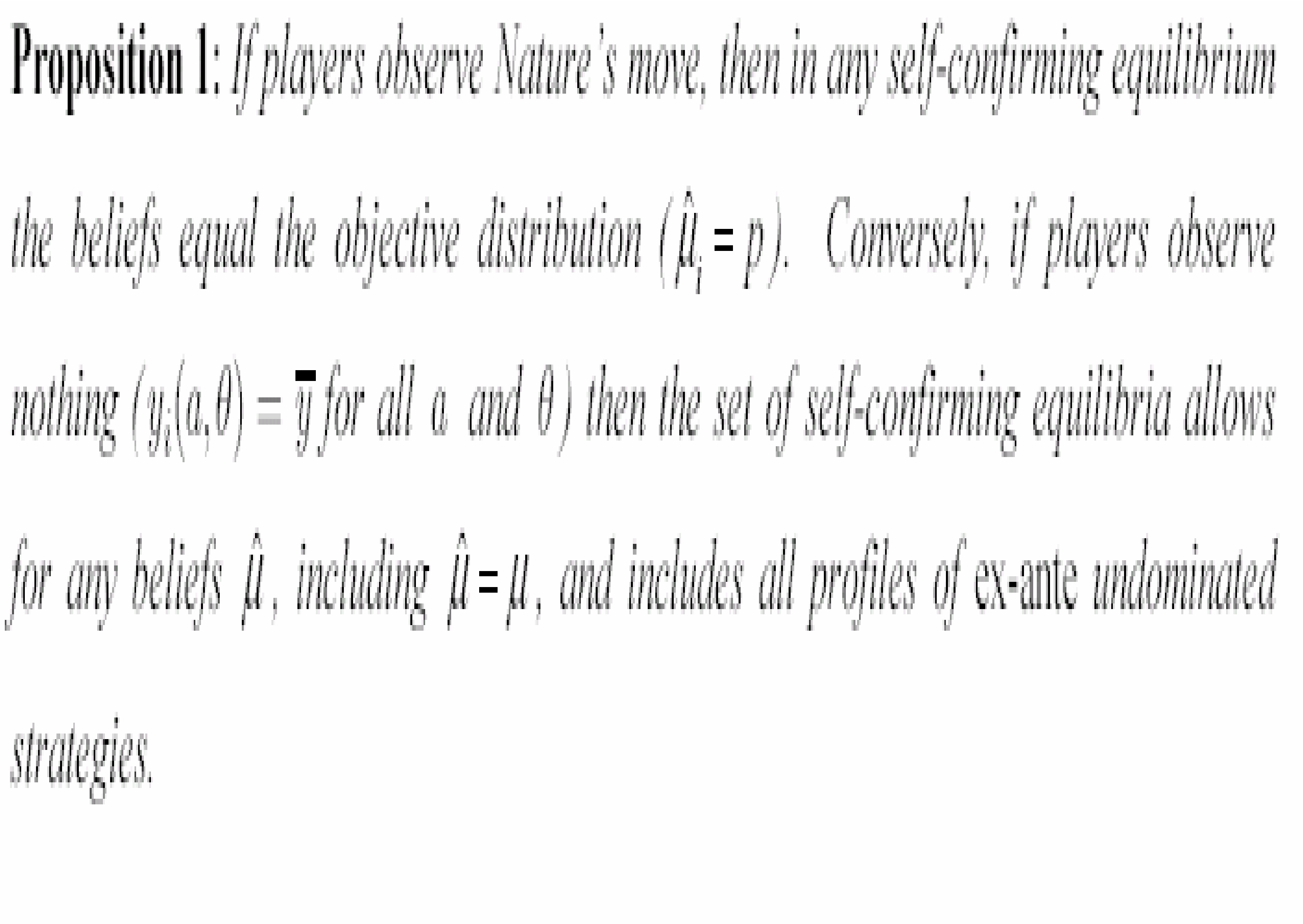
Definition: *A strategy profile* σ *is a* Nash equilibrium *with conjecture* $\hat{\sigma}_{-i}$ *and beliefs* $\hat{\mu}_i$ *if for each player* i, *and for any pair* $\theta_i, \hat{a}_i$ *such that* $\hat{\mu}^i(\theta_i) \cdot \sigma_i(\hat{a}_i) > 0$

(ii) $\hat{a}_i \in \arg\max_{a_i} \sum_{a_{-i},\theta_{-i}} u_i(a_i, a_{-i}, \theta_i, \theta_{-i}) \hat{\mu}^i(\theta_{-i} \mid \theta_i) \hat{\sigma}_{-i}(a_{-i} \mid \theta_{-i})$,

and

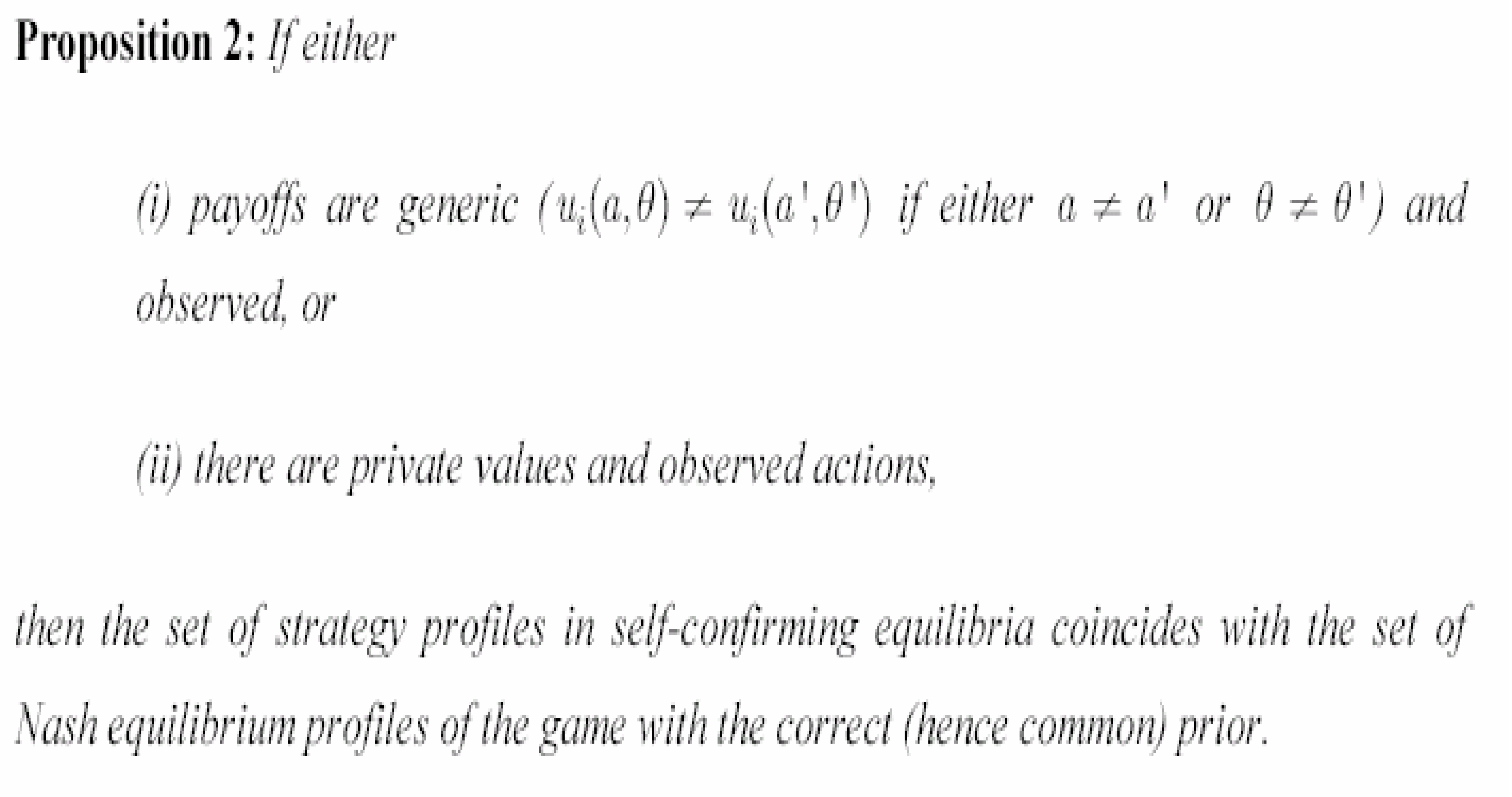
(iii') $\hat{\sigma}_{-i} = \sigma_{-i}$, $\hat{\mu}^i = \mu^i$.

As mentioned above, if players cannot observe or deduce their opponents' actions at the end of each period, then in general there can be self-confirming equilibria that are not Nash equilibria. So we begin by considering the case in which players either directly observe, or indirectly deduce from other observations, the realized actions of their opponents after each play of the game.



Proposition 1: *If players observe Nature's move, then in any self-confirming equilibrium the beliefs equal the objective distribution* ($\hat{\mu}_i = p$). *Conversely, if players observe nothing* ($y_i(a,\theta) = \bar{y}$ *for all* a *and* θ) *then the set of self-confirming equilibria allows for any beliefs* $\hat{\mu}$, *including* $\hat{\mu} = \mu$, *and includes all profiles of* ex-ante *undominated strategies.*

Generic=very informative



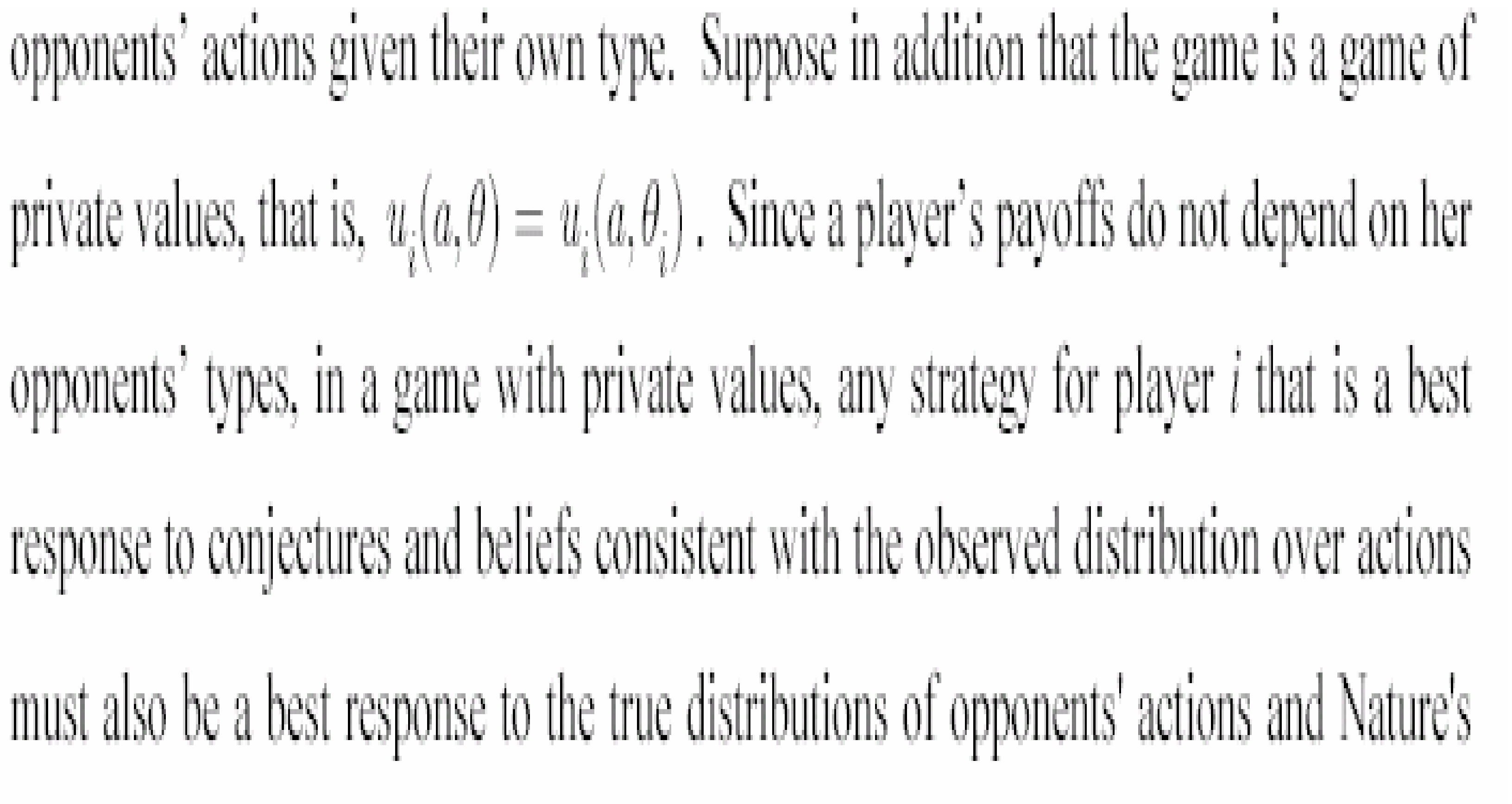
Proposition 2: *If either*

(i) payoffs are generic ($u_i(a,\theta) \neq u_i(a',\theta')$ *if either* $a \neq a'$ *or* $\theta \neq \theta'$) *and observed, or*

(ii) there are private values and observed actions,

then the set of strategy profiles in self-confirming equilibria coincides with the set of Nash equilibrium profiles of the game with the correct (hence common) prior.

Private values



opponents' actions given their own type. Suppose in addition that the game is a game of private values, that is, $u_i(a,\theta) \equiv u_i(a,\theta_i)$. Since a player's payoffs do not depend on her opponents' types, in a game with private values, any strategy for player *i* that is a best response to conjectures and beliefs consistent with the observed distribution over actions must also be a best response to the true distributions of opponents' actions and Nature's

Proposition 4 : *(i) Without private values ($u_i(a,\theta) \neq u_i(a,\theta_i)$ for some (a,θ)), if neither types nor payoffs are observed, but actions are ($y_i(a,\theta) = a$), there can be self-confirming equilibria with correct beliefs about Nature ($\hat{\mu}^i = p$) that are not Nash even with correct priors (Example 3).*

(ii) Even if the set of strategy profiles in self-confirming equilibria with beliefs $\mu = \hat{\mu}$ coincides with the set of Nash equilibria, conjectures about opponents' play may fail to be correct ($\hat{\sigma}_{-i} \neq \sigma_{-i}$). Consequently the profile can fail to be self confirming once actions are added to the available information (Example 4).

Example 4: A game where when payoffs are observed, Nash equilibrium and self-confirming equilibrium are equivalent iff actions are not observed.

Consider a two-player game in which Nature chooses the left or right matrix. Neither player has private information. Proposition 2 (*i*) does not apply because the payoffs include ties as shown below.

	A	*B*
A	1, 1	0, 0
B	0, 0	0, 0

	A	*B*
A	0, 0	1, 1
B	1, 1	0, 0

To analyze Nash equilibria, suppose that the stage game prior is that both players think the left matrix is chosen with probability $1-\varepsilon$. The strategic form for this game given the common beliefs μ is

	A	*B*
A	$1-\varepsilon$, $1-\varepsilon$	ε , ε
B	ε , ε	0, 0

The unique Nash equilibrium for the specified beliefs is (*A*, *A*).

Now suppose that in the learning environment, the true probability of the left matrix is ε. If players observe only their payoffs, then (*A*, *A*) is a self-confirming equilibrium with beliefs $(1-\varepsilon, \varepsilon)$ and conjecture that the opponent is playing *B*: in this case each player believes that playing *A* yields 1 with probability ε, and *B* yields 0. However, if players were to also observe actions, then the Nash equilibrium (*A*, *A*) would no longer be self confirming.

Example 5: A game where Nash equilibrium and self-confirming equilibrium coincide for a specific diverse belief about Nature's move

	L	*R*
U	1, 1	0, 0
D	0, 0	-1, -1

	L	*R*
U	-1, -1	0,0
D	0, 0	1, 1

This is a two-player game in which Nature chooses the left (*l*) or right (*r*) payoffs, and neither player observes Nature's move. The row player believes the left payoffs are chosen, the column player believes the opposite: $\mu^1(l) = \mu^2(r) = 1$. So the unique Nash equilibrium is for the row player to play *U* and the column player *R*, with payoffs (0, 0). Whether or not players observe their opponent's actions or their own utility, this profile is self-confirming with beliefs equal to the given stage-game priors. However, the subset of self-confirming equilibria with beliefs in which $\hat{\mu}^1 = \hat{\mu}^2$ is either (*U*, *L*), (*D*, *R*), or the entire strategy space.

■

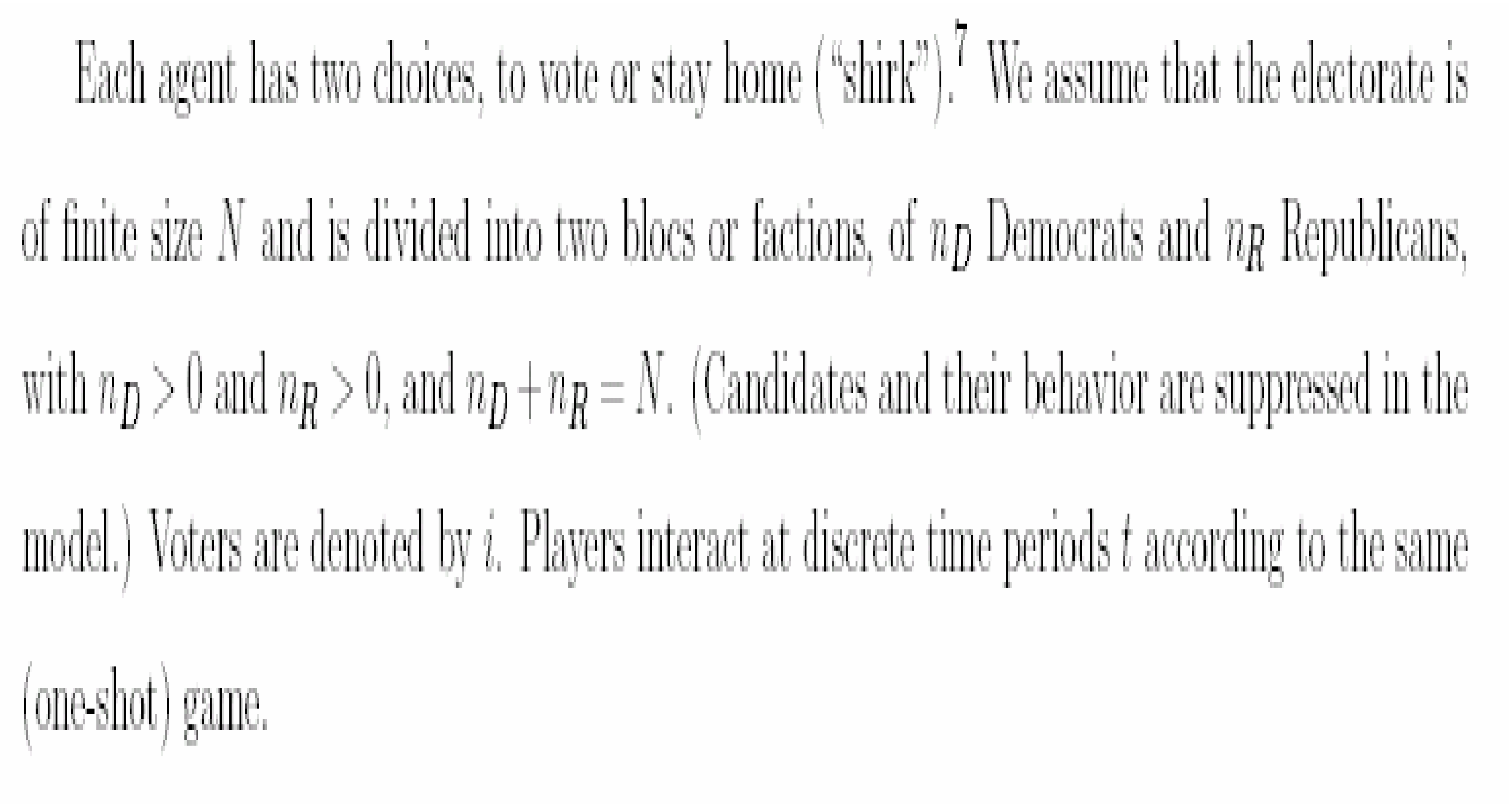
A Behavioral Model of Turnout

Bendor, Diermeier and Ting

BDT 2003

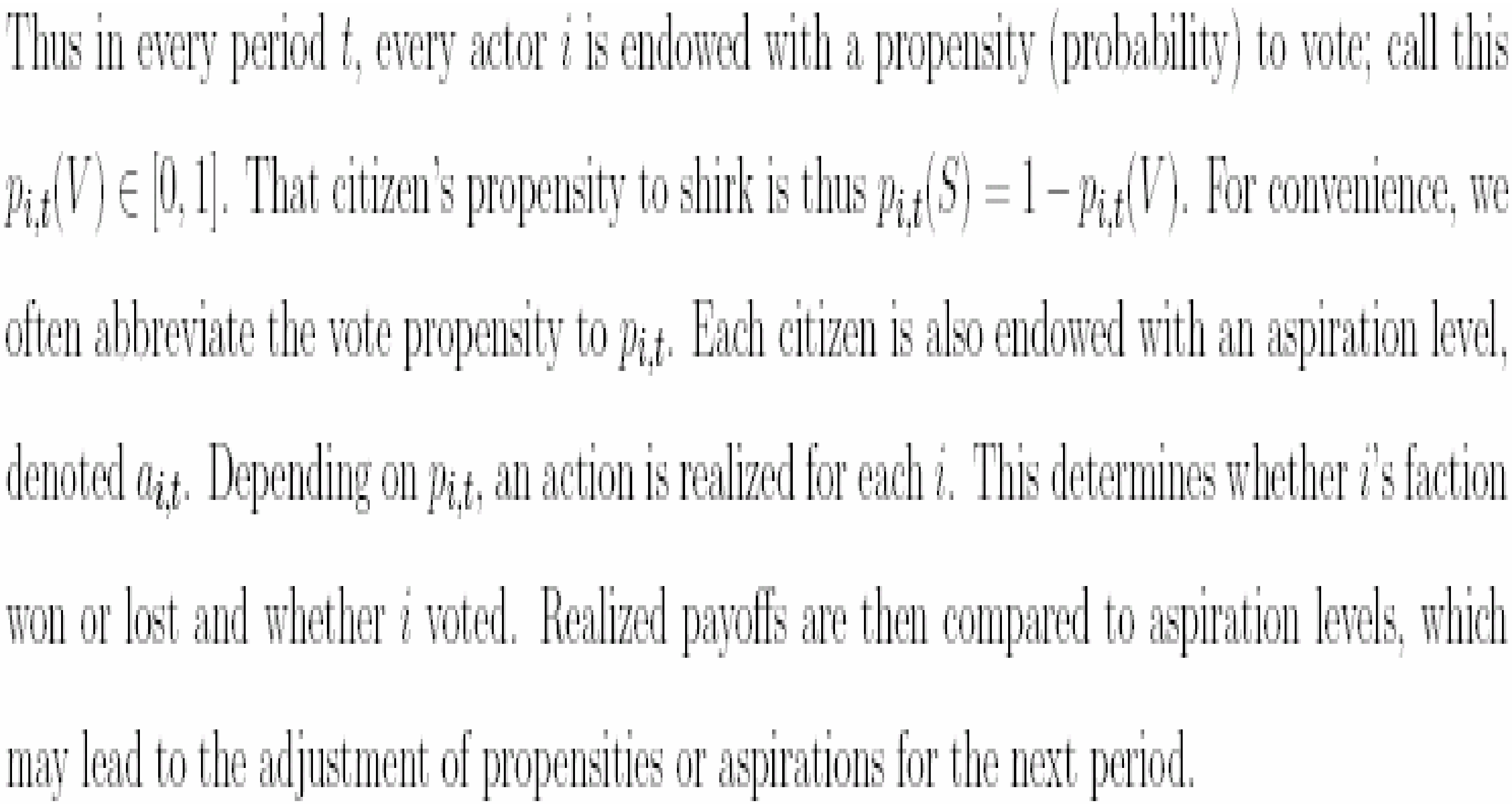
- We construct a model of adaptive rationality: citizens learn by simple trial-and-error, repeating satisfactory actions and avoiding unsatisfactory ones.
- Their aspiration levels, which code current
- payoffs as satisfactory or unsatisfactory, are also endogenous, themselves adjusting to experience.
- Our main result is that agents who adapt in this manner turn out in substantial numbers even in large electorates and even if voting is costly for everyone.

Choices



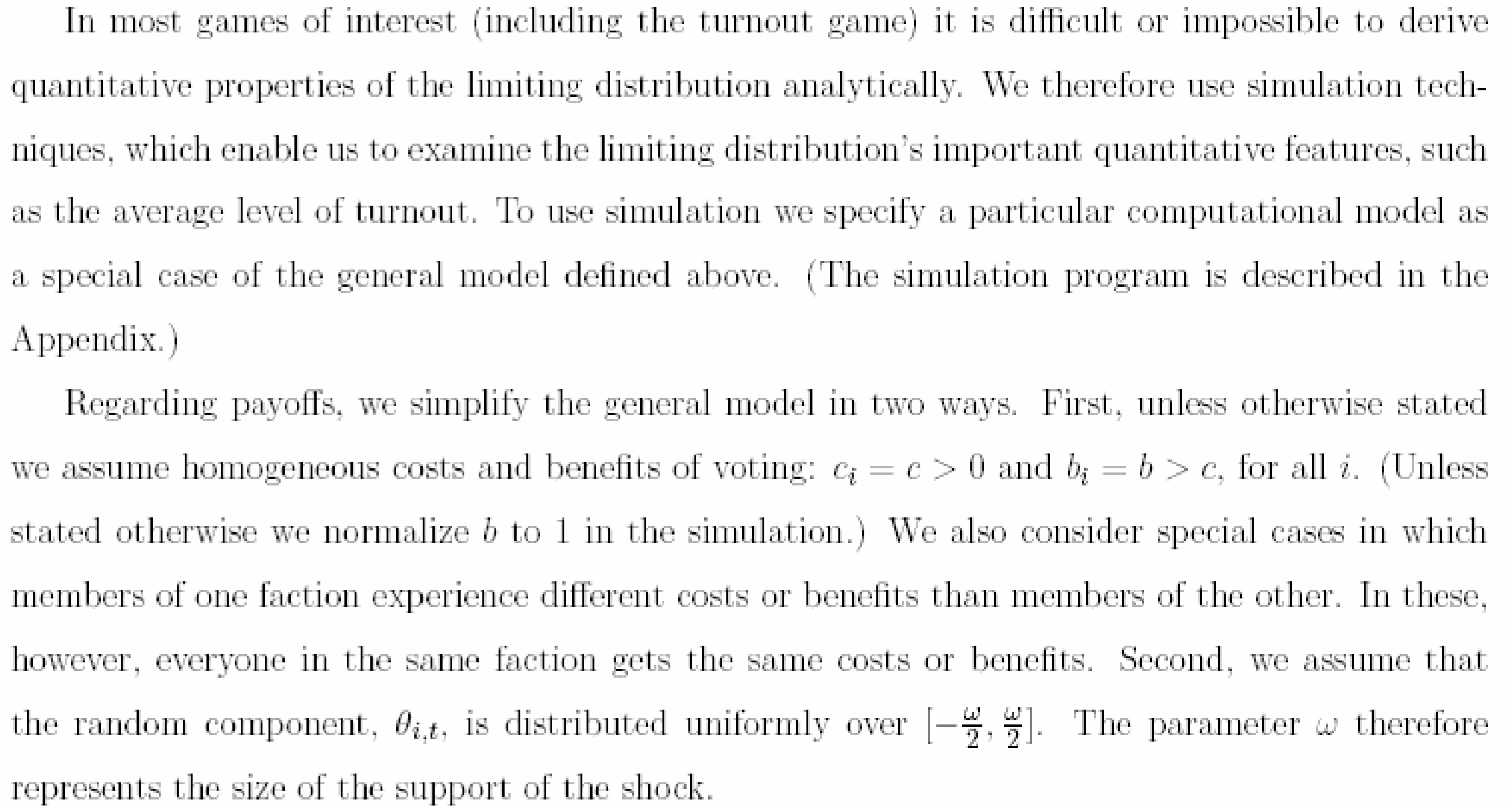
Each agent has two choices, to vote or stay home ("shirk").[7] We assume that the electorate is of finite size N and is divided into two blocs or factions, of n_D Democrats and n_R Republicans, with $n_D > 0$ and $n_R > 0$, and $n_D + n_R = N$. (Candidates and their behavior are suppressed in the model.) Voters are denoted by i. Players interact at discrete time periods t according to the same (one-shot) game.

Adaptations



Thus in every period t, every actor i is endowed with a propensity (probability) to vote; call this $p_{i,t}(V) \in [0,1]$. That citizen's propensity to shirk is thus $p_{i,t}(S) = 1 - p_{i,t}(V)$. For convenience, we often abbreviate the vote propensity to $p_{i,t}$. Each citizen is also endowed with an aspiration level, denoted $a_{i,t}$. Depending on $p_{i,t}$, an action is realized for each i. This determines whether i's faction won or lost and whether i voted. Realized payoffs are then compared to aspiration levels, which may lead to the adjustment of propensities or aspirations for the next period.

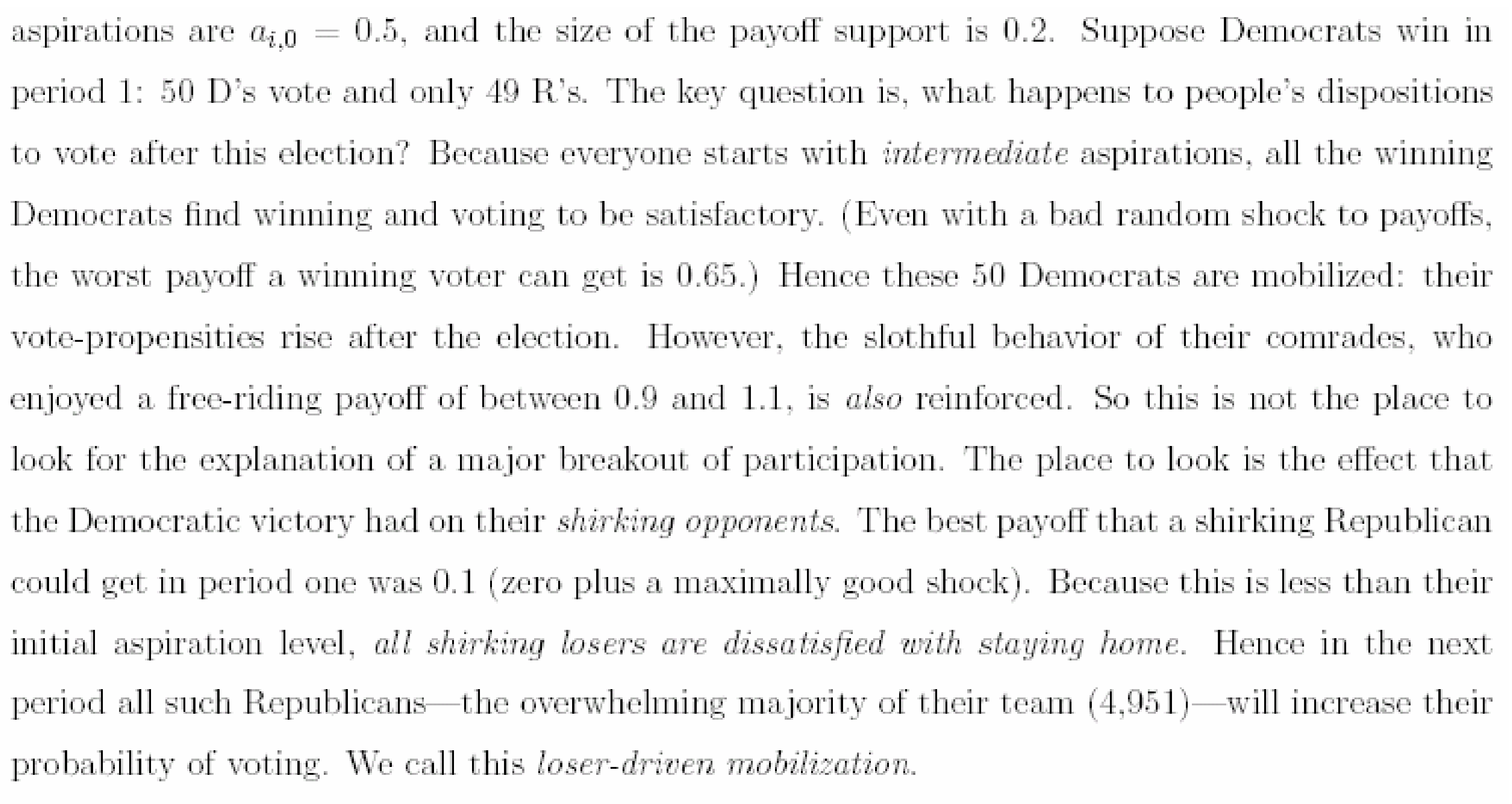
Means of inference



In most games of interest (including the turnout game) it is difficult or impossible to derive quantitative properties of the limiting distribution analytically. We therefore use simulation techniques, which enable us to examine the limiting distribution's important quantitative features, such as the average level of turnout. To use simulation we specify a particular computational model as a special case of the general model defined above. (The simulation program is described in the Appendix.)

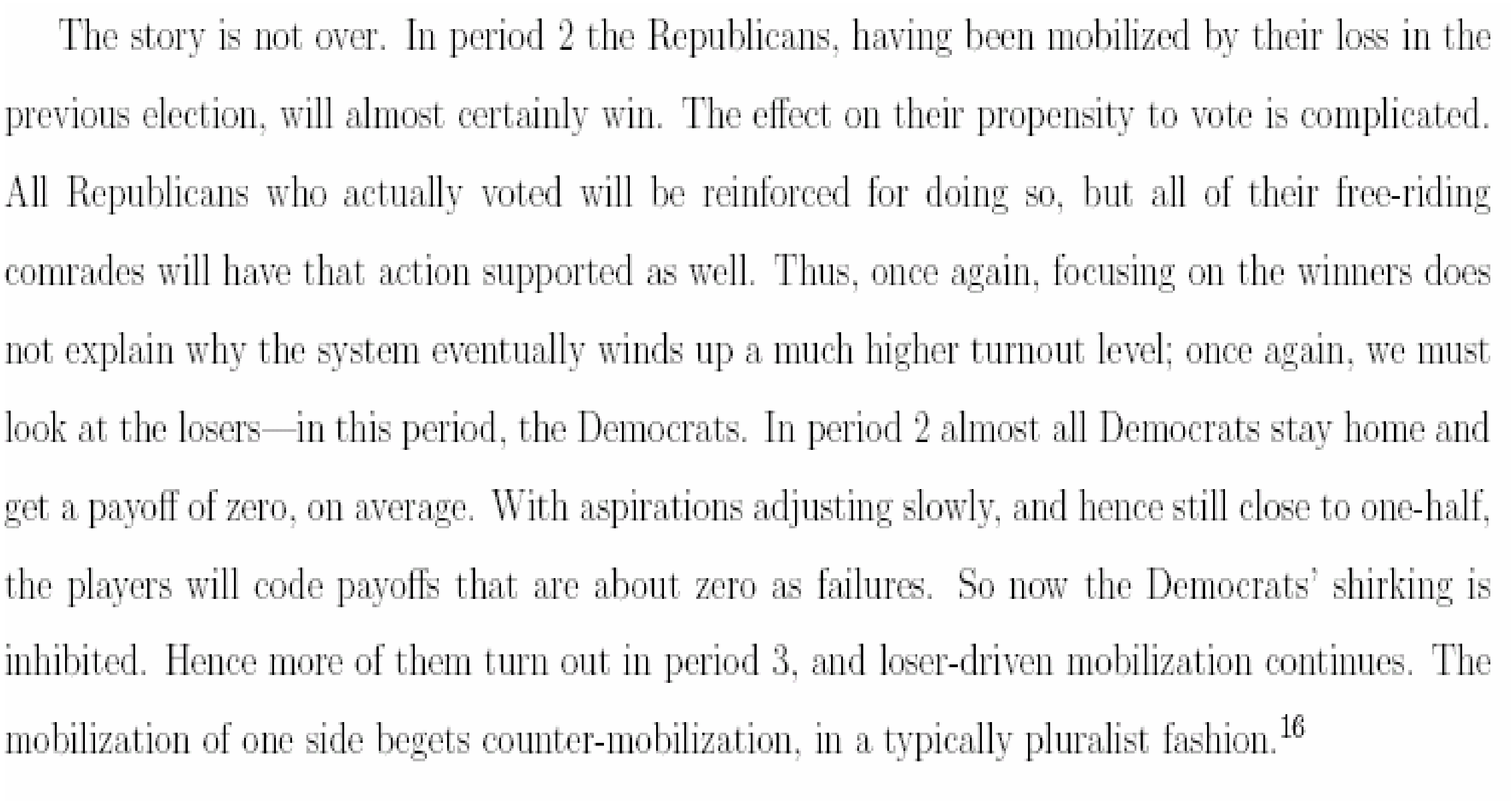
Regarding payoffs, we simplify the general model in two ways. First, unless otherwise stated we assume homogeneous costs and benefits of voting: $c_i = c > 0$ and $b_i = b > c$, for all i. (Unless stated otherwise we normalize b to 1 in the simulation.) We also consider special cases in which members of one faction experience different costs or benefits than members of the other. In these, however, everyone in the same faction gets the same costs or benefits. Second, we assume that the random component, $\theta_{i,t}$, is distributed uniformly over $[-\frac{\omega}{2}, \frac{\omega}{2}]$. The parameter ω therefore represents the size of the support of the shock.

How it works - 1



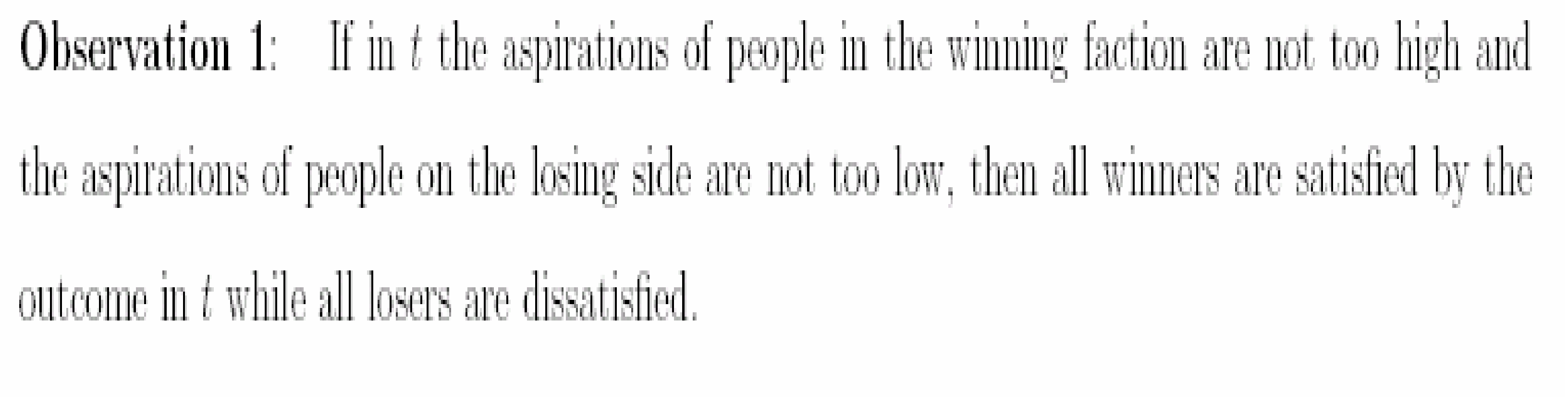
aspirations are $a_{i,0} = 0.5$, and the size of the payoff support is 0.2. Suppose Democrats win in period 1: 50 D's vote and only 49 R's. The key question is, what happens to people's dispositions to vote after this election? Because everyone starts with *intermediate* aspirations, all the winning Democrats find winning and voting to be satisfactory. (Even with a bad random shock to payoffs, the worst payoff a winning voter can get is 0.65.) Hence these 50 Democrats are mobilized: their vote-propensities rise after the election. However, the slothful behavior of their comrades, who enjoyed a free-riding payoff of between 0.9 and 1.1, is *also* reinforced. So this is not the place to look for the explanation of a major breakout of participation. The place to look is the effect that the Democratic victory had on their *shirking opponents*. The best payoff that a shirking Republican could get in period one was 0.1 (zero plus a maximally good shock). Because this is less than their initial aspiration level, *all shirking losers are dissatisfied with staying home*. Hence in the next period all such Republicans—the overwhelming majority of their team (4,951)—will increase their probability of voting. We call this *loser-driven mobilization*.

How it works –2

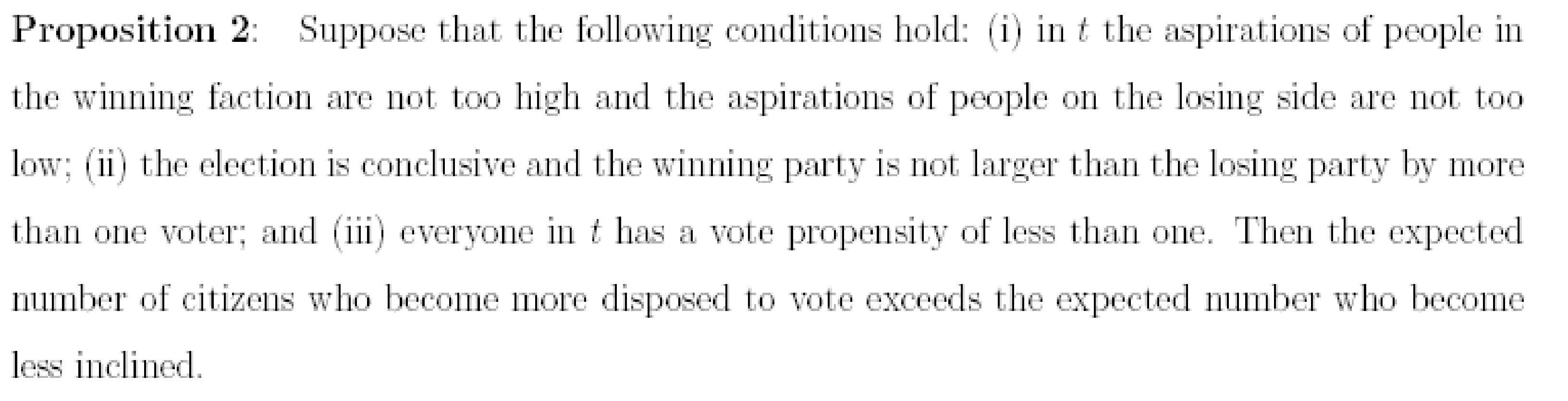


The story is not over. In period 2 the Republicans, having been mobilized by their loss in the previous election, will almost certainly win. The effect on their propensity to vote is complicated. All Republicans who actually voted will be reinforced for doing so, but all of their free-riding comrades will have that action supported as well. Thus, once again, focusing on the winners does not explain why the system eventually winds up a much higher turnout level; once again, we must look at the losers—in this period, the Democrats. In period 2 almost all Democrats stay home and get a payoff of zero, on average. With aspirations adjusting slowly, and hence still close to one-half, the players will code payoffs that are about zero as failures. So now the Democrats' shirking is inhibited. Hence more of them turn out in period 3, and loser-driven mobilization continues. The mobilization of one side begets counter-mobilization, in a typically pluralist fashion.[16]

Further implications



Observation 1: If in t the aspirations of people in the winning faction are not too high and the aspirations of people on the losing side are not too low, then all winners are satisfied by the outcome in t while all losers are dissatisfied.



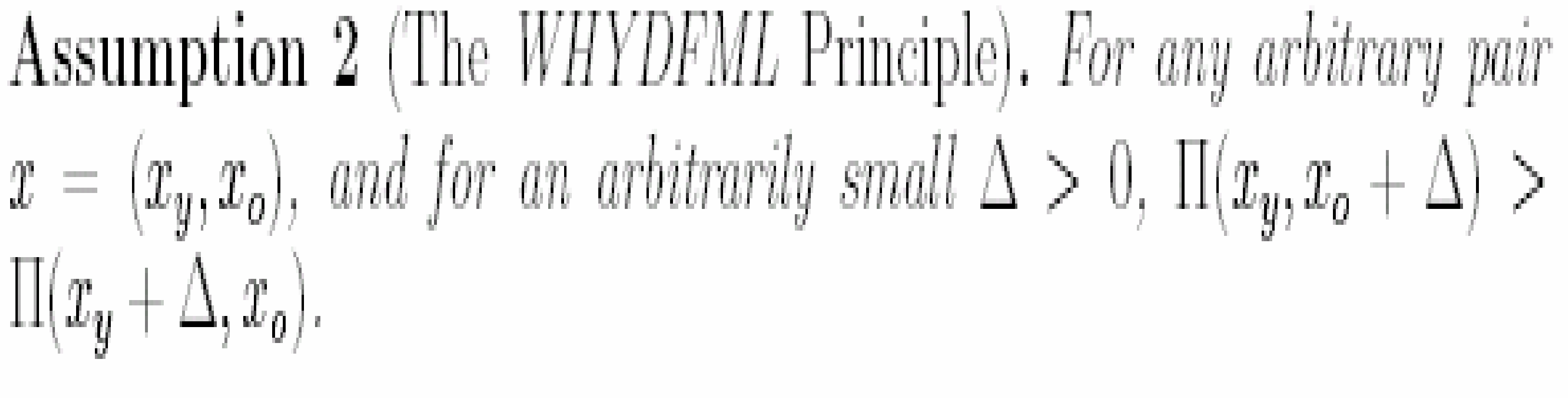
Proposition 2: Suppose that the following conditions hold: (i) in t the aspirations of people in the winning faction are not too high and the aspirations of people on the losing side are not too low; (ii) the election is conclusive and the winning party is not larger than the losing party by more than one voter; and (iii) everyone in t has a vote propensity of less than one. Then the expected number of citizens who become more disposed to vote exceeds the expected number who become less inclined.

Agenda Setting Powers in Organizations with Overlapping Generations of Players

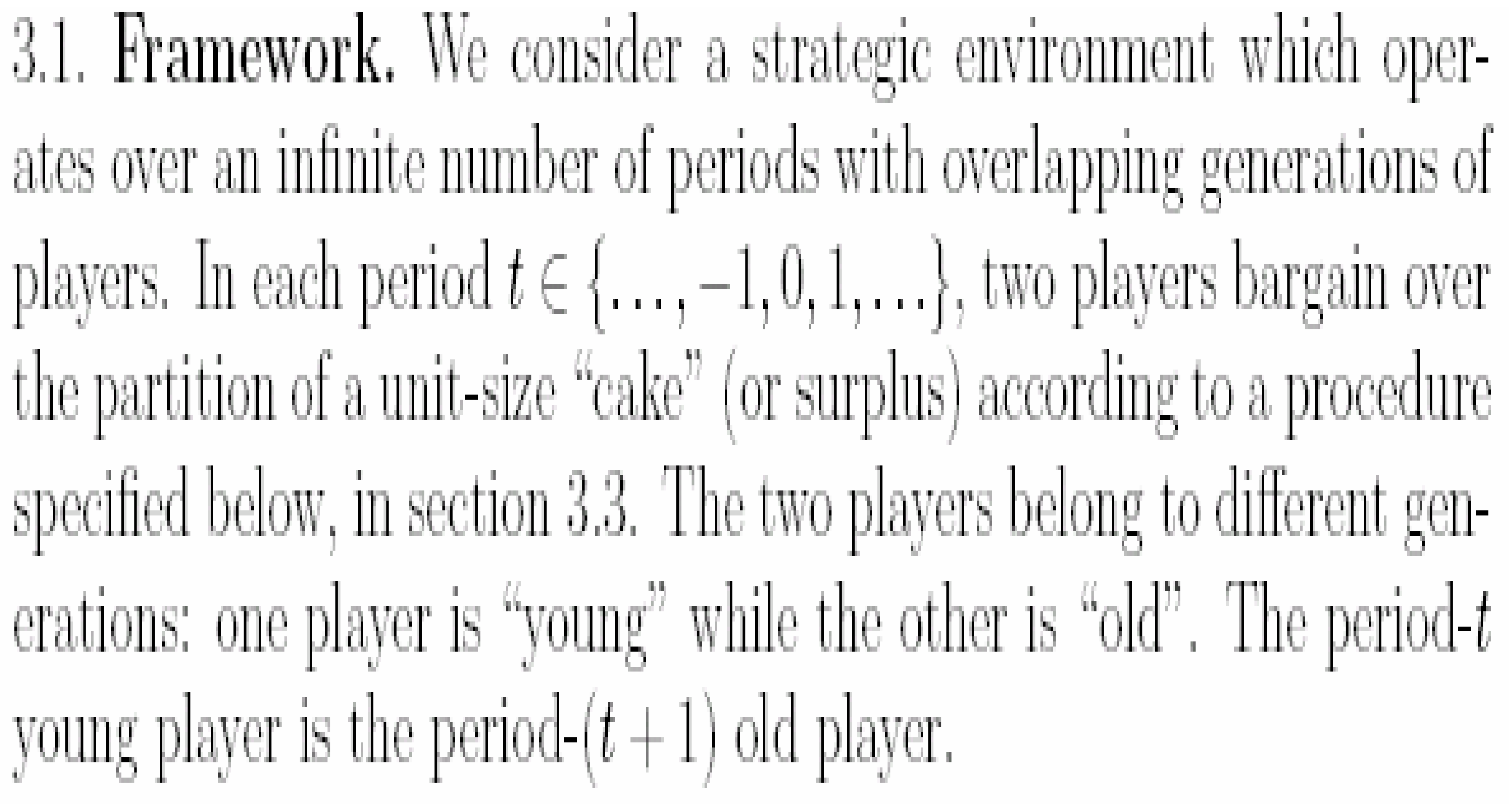
Abinay Muthoo and Kenneth Shepsle

Shepsle/Muthoo premises

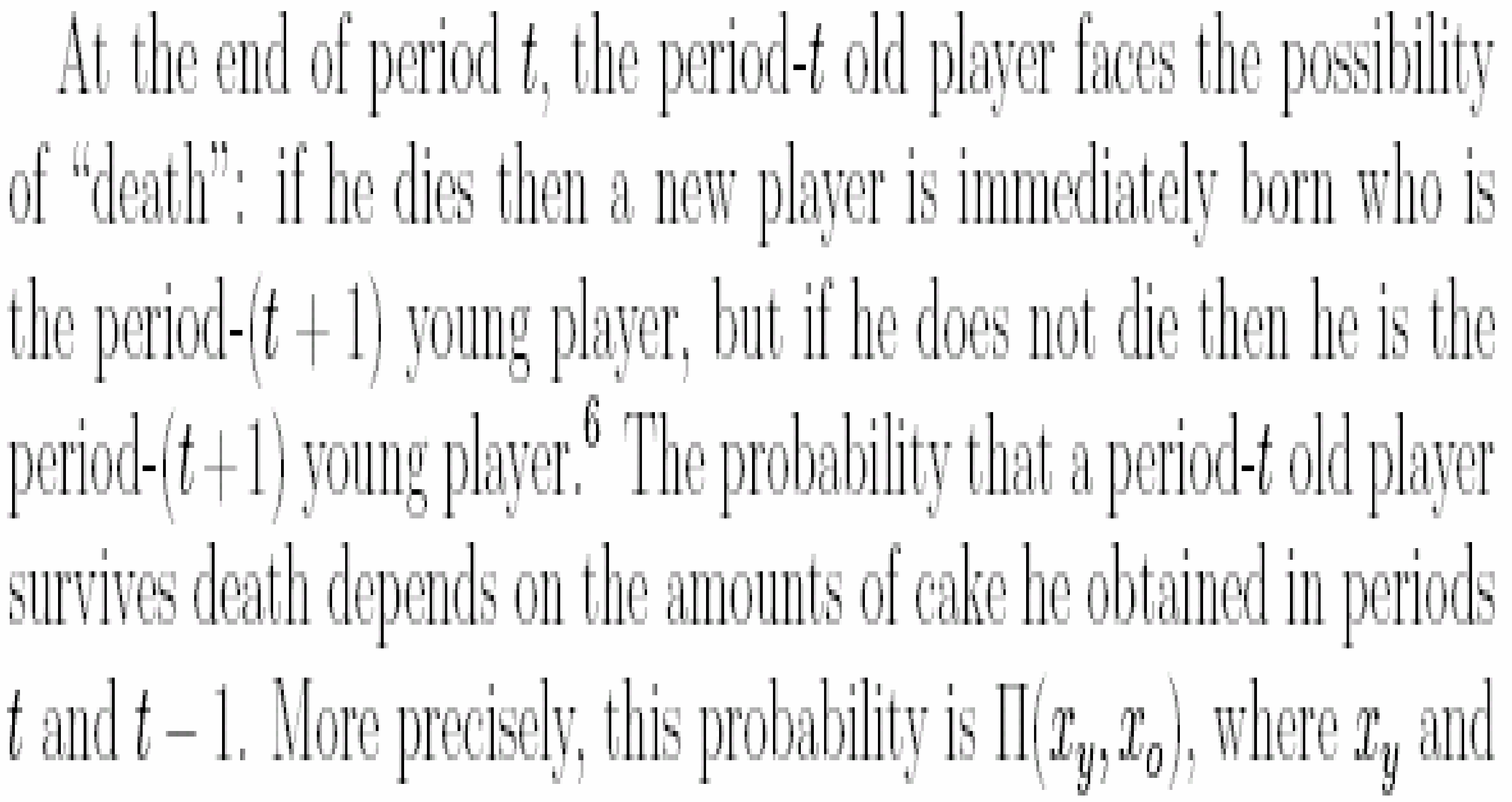
- Only the oldest generation of players face re-election at the end of each period.
- Success in re-elections depends on a retrospective assessment by voters with the WHYDFML Principle lying at the heart of this matter.



Assumption 2 (The *WHYDFML* Principle). *For any arbitrary pair* $x = (x_y, x_o)$, *and for an arbitrarily small* $\Delta > 0$, $\Pi(x_y, x_o + \Delta) > \Pi(x_y + \Delta, x_o)$.



3.1. **Framework.** We consider a strategic environment which operates over an infinite number of periods with overlapping generations of players. In each period $t \in \{\ldots, -1, 0, 1, \ldots\}$, two players bargain over the partition of a unit-size "cake" (or surplus) according to a procedure specified below, in section 3.3. The two players belong to different generations: one player is "young" while the other is "old". The period-t young player is the period-$(t+1)$ old player.



At the end of period t, the period-t old player faces the possibility of "death": if he dies then a new player is immediately born who is the period-$(t+1)$ young player, but if he does not die then he is the period-$(t+1)$ young player.[6] The probability that a period-t old player survives death depends on the amounts of cake he obtained in periods t and $t-1$. More precisely, this probability is $\Pi(x_y, x_o)$, where x_y and

With probability θ, where $\theta \in [0,1]$, the young player makes a "take-it-or-leave-it" offer to the old player, and with probability $1-\theta$ it is the old player who makes a "take-it-or-leave-it" offer to the young player.

Proposition 1. *Define the following inequality:*

$$\frac{1}{2}\Pi(1,1) + \frac{1}{2}\Pi(0,0) > \Pi(0,1). \tag{1}$$

(i) If Π satisfies inequality 1, then there exists a $\theta^* \in (0,1)$ such that $W(.)$ is maximized at $\theta = \theta^*$, where

$$\theta^* = \frac{\Pi(1,1) + \Pi(0,0) - 2\Pi(0,1)}{2[\Pi(1,1) + \Pi(0,0) - \Pi(0,1) - \Pi(1,0)]}.$$

(ii) If Π does not satisfy inequality 1, then $W(.)$ is maximized at $\theta = 0$.

It is easy to verify that θ^* is strictly less than 0.5; and that is the case because of Assumption 2 (the *WHYDFML* Principle). This result indicates that even when it is optimal to allocate some bargaining power to the young player, most of it should reside with the old player.