# האוניברסיטה העברית בירושלים
## THE HEBREW UNIVERSITY OF JERUSALEM


---

# GAMES IN THE NERVOUS SYSTEM: THE GAME MOTONEURONS PLAY

**by**


**IRIT NOWIK, IDAN SEGEV and SHMUEL ZAMIR**



**Discussion Paper # 440 December 2006**



## מרכז לחקר הרציונליות

## CENTER FOR THE STUDY OF RATIONALITY


---


**Feldman Building, Givat-Ram, 91904 Jerusalem, Israel**
**PHONE: [972]-2-6584135 FAX: [972]-2-6513681**
**E-MAIL: ratio@math.huji.ac.il**
**URL: http://www.ratio.huji.ac.il/**

# GAMES IN THE NERVOUS SYSTEM: THE GAME MOTONEURONS PLAY

Irit Nowik[12], Idan Segev[1], Shmuel Zamir[2]

December 24, 2006

Abstract. Game theory is usually applied to biology through evolutionary games. However, many competitive processes in biology may be better understood by analyzing them on a shorter time-scale than the time-course considered in evolutionary dynamics. Instead of the change in the "fitness" of a player, which is the traditional payoff in evolutionary games, we define the payoff function, tailored to the specific questions addressed. In this work we analyze the developmental competition that arises between motoneurons innervating the same muscle. The "*size principle*" — a fundamental principle in the organization of the motor system, stating that motoneurons with successively higher activation-threshold innervate successively larger portions of the muscle — emerges as a result of this competition. We define a game, in which motoneurons compete to innervate a maximal number of muscle-fibers. The strategies of the motoneurons are their activation-thresholds. By using a game theoretical approach we succeed to explain the emergence of the size principle and to reconcile seemingly contradictory experimental data on this issue. The evolutionary advantage of properties as the size principle, emerging as a consequence of competition rather than being genetically hardwired, is that it endows the system with adaptation capabilities, such that the outcome may be fine-tuned to fit the environment. In accordance with this idea the present study provides several experimentally-testable predictions regarding the magnitude of the size principle in different muscles.

## 1. INTRODUCTION

The human brain contains about $10^{11}$ neurons, and about $10^{14}$ connections (synapses) between them. The connectivity between neurons and their targets is achieved during development by two fundamentally different programs: Molecular guidance cues and patterned neural activity. Molecular cues guide axons (the output element of a neuron) from specific regions to broadly defined target regions, and initiate the formation of synaptic connections

[1]Interdisciplinary Center for Neural Computation

[2]Center for the Study of Rationality,

The Hebrew University, Jerusalem 91904 Israel. Correspondence: iritn@math.huji.ac.il

(Mann, Holt *et al.* [1] 2002). However, such molecular cues are not always sufficient to establish the final pattern of synaptic connections, which depends at lease in part on patterned neural activity evoked by sensory input (Kandel, Scwartz *et al.* [2] 2000). Activity-dependent fine-tuning of neural circuitry is not limited to early development, but rather neural circuits are adaptable even in the mature individual and is thought to be the physiological basis of learning and memory (Martin *et al.* [3] 2000, Kandel *et al.* [4][5][6], 1978, 1999, 2000, Tiesen *et al.* [7], 1996).

1.1. **Connectivity changes during development.** One form of refinement of connectivity is the elimination of connections, which occurs in many parts of the developing nervous system (Crepel *et al.* [11] 1976, Mariani *et al.* [12] 1996, Hubel & Wiesel [13] [9] 1977, Purves & Lichtman [10] 1980, Sanes & Lichtman [8] 1999). In each of these areas, elaboration of synapses by the remaining neurons also occurs. Thus, while some inputs are being eliminated, others are becoming stronger, giving the impression that the elimination process may be interactive and competitive.

In general, the rules that govern these modifications are poorly understood, partly because of the difficulty of monitoring synaptic connections in the central nervous system over long periods of time. This is possible however, at the neuromuscular junction, a simple and accessible synapse between nerve and muscle, where elimination and formation of connections have been directly observed.

1.2. **The Neuromuscular system.** Synaptic communication in the brain relies mainly on chemical mechanisms. The neuromuscular junction is an ideal site for studying chemical signaling because it is relatively simple and also very accessible to experimentation. The muscle cell (called muscle-fiber) is large enough (typically with diameter of 50-100 $\mu m$, and length of $2-6$ cm) to accommodate the two or more microelectrods needed to make electrical measurements. Also, in the adult system, the muscle-fiber is normally innervated by just one axon, in contrast to the convergent connections on central nerve cells.

We believe the relative simplicity of the neuromuscular system, and the abundant experimental data accumulated on this system is an advantage, when applying a new theoretical approach. Thus we deal here, with the competition that arises between motoneurons (MNs) that innervate the same muscle (for review, see Lichtman & Colman [14] 2000, Walsh & Lichtman [15] 2003).

1.3. **The game MNs play.** A typical skeletal muscle consists of many thousands of fibers. The MNs innervating one muscle are usually clustered together into a nucleus within the ventral spinal cord (Figure 1). At birth, each muscle-fiber is innervated by several MNs, which innervate it in one specialized location called "the end-plate".

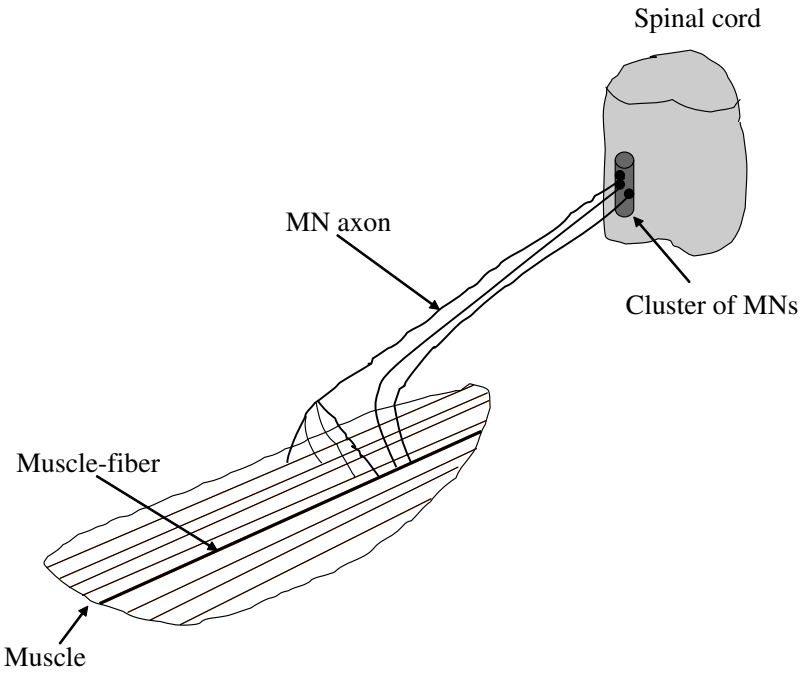


FIGURE 1. *A typical skeletal muscle consists of many thousands of fibers. The MNs innervating the muscle are clustered into a nucleus within the ventral spinal cord. At birth, each muscle-fiber is innervated by several MNs.*

During the first couple of weeks after birth, a competitive mechanism - called "synapse elimination" - abolishes all inputs but one, which we term "the winner at the muscle-fiber".

Synapse elimination proceeds as follows: One input gradually withdraws from its postsynaptic sites and another input elaborates new branches that take over these postsynaptic sites. Once an axon loses all its postsynaptic sites, it withdraws altogether from the muscle-fiber.

At birth, each MN innervates many muscle-fibers and therefore it engages in many competitions simultaneously, winning at (i.e., singly innervating) some muscle-fibers and losing at others. When the process ends, each muscle-fiber is innervated by a single MN but each MN innervates a group of muscle-fibers called a "muscle unit".

Synapse elimination is competitive in the sense that the fate of a connection by one MN depends on the presence or absence of connections by other MNs. The main two experimental results that support this interaction between MNs, are that when synapse elimination ends, each muscle-fiber is innervated by a single MN, where as if the elimination process of different MNs was not interconnected, there would be multiply innervated muscle-fibers or muscle-fibers with no connections at all. Another evidence for the competitive nature of synapse elimination is that if some of the axons are cut soon after birth, then at the end of synapse elimination, the remaining intact MNs innervate more muscle-fibers than they normally do (Thompson & Jansen [34] 1977, Fladby & Jansen[35] 1987).

The function of MNs is as follows: A MN receives electrical inputs from other neurons. If the aggregated input is higher than the MN's "activation threshold", then the MN produces a train of 5-10 electrical impulses called action potentials (APs) per second. The firing rate increases as the load on the muscle increases. The APs travel down the MN's axon and activate all the muscle-fibers in its muscle unit. MNs differ in their activation threshold. MNs having low thresholds are more active (i.e., fire more frequently) than MNs having higher activation thresholds (Henneman [17] 1957).

1.3.1. ***The Size Principle***. In the adult system (i.e., at the end of the competition period), MNs with successively higher activation thresholds have successively larger muscle units. This is called "*the size principle*" (Henneman [16][17] 1985, 1957, Cope & Pinter [36] 1995). As a result of the size principle, muscle units are being recruited in a fixed order, from smallest to largest. This is essential to optimal control of the muscle (Henneman & Olson [18] 1965, Solomonow *et al.* [19] 1990).

The size principle is well established and is considered as one of the most fundamental principles in the organization of motor-unit behavior. It is therefore important to understand how it evolves. In viewing the elimination period as a game in which MNs are "competing" to innervate a maximal number of muscle-fibers, the translation of the size principle is that less-active MNs (i.e., MNs with higher activation thresholds) win in more competitions than more-active MNs. But surprisingly, the majority of the experiments that have selectively manipulated the activity of MNs during the competition period - seem to point to the opposite conclusion - that it is the *more*-active MNs that are advantageous in this process (Ribchester & Taxt [22] 1983, Ridge & Betz [23] 1984, Balice-Gordon & Lichtman [24] 1994,

Lo & Poo [25] 1991). Only two experiments (Callaway *et al.* [20] [21] 1987, 1989) seem consistent with the size principle, (but inconsistent with the majority of the experimental data), by pointing to a competitive advantage of the less-active MNs. Due to this seemingly contradictory experimental data, the effect of activity on synapse elimination is considered "paradoxical" (Barber & Lichtman [44] 1999).

In addition, although there is no doubt that electrical activity plays a role in synapse elimination, researches disagree on its importance or centrality in determining the outcome of the competitions: Some give a decisive role to the different activity-levels of the competing MNs, whereas others believe that activity is just one of many influences in this competition, while the actual competition is governed by other factors, e.g., neurotrophic factors and their receptors (van Ooyen [26] 2001, van Ooyen & Willshaw [27] 1999).

In the game we define here, the MNs are players competing to innervate a maximal number of muscle-fibers. The (pre-programmed) strategies of the players are their activity-levels, and the payoffs are the sizes of their muscle units.

The goals of this work are to offer a theoretical explanation to how the size principle emerges from the game MNs play, to reconcile the seemingly contradictory experimental data, and to offer new experimentally-*testable* predictions regarding the magnitude of the size principle in different environments (i.e., different muscles).

## 2. ASSUMPTIONS OF THE MODEL

The assumptions of the model consist of the random initial connectivity between MNs and muscle-fibers, and the three rules of the game.

2.1. **Random Initial connectivity.** We denote the number of fibers in the muscle as $N$. Without loss of generality, we assume that the number of MNs, innervating the muscle, is even[3] and denote it by $2n$. We define Indicator random variables $I^N_{i,k}$ ( $i = 1, \ldots, N$, $k = 1, \ldots, 2n$, ) as follows:

$$I^N_{i,k} = \begin{cases} 1, & \text{if MN } k \text{ is connected to muscle-fiber } i; \\ 0, & \text{otherwise.} \end{cases}$$

[3]If it is odd, we ignore a MN with the median activity-level.

We assume that initially, MNs connect to muscle-fibers in a completely random fashion, i.e., each MN innervates each muscle-fiber, with probability $\gamma > 0$, independently of other connections (Willshaw [37] 1981).

Thus, $I_{i,k}^N$ are *i.i.d.r.v* (i.e., independent identically distributed random variables) satisfying that:

$$I_{i,k}^N \sim \mathbf{B}(1,\gamma), \quad \text{(a binomial distribution)}.$$

We define the activity-level of MN $k$, as $Y_k$, $(k = 1,\ldots,2n)$, and assume that the $Y_k$s are *i.i.d.r.v*, each having a continuous positive density function $f_Y$ over some given interval $(a,b)$. In order not to introduce irrelevant bias into the model, we assume that $f_Y$ is symmetrical on $(a,b)$, (for example, a normal or uniform distribution).

We divide the population of MNs that connect to the muscle, into two equal-sized teams: M-team (the More-active MNs) containing the MNs with activity-levels that are higher than the median activity-level, and L-team (the Less-active MNs), with activities that are lower than the median activity-level. Note that this division into teams does not imply that MNs cooperate in some way (for example, share resources), but rather MNs play individually, and the division into teams serves only for the analysis of the game (see a more detailed account of that in Section 2.5 ahead).

Given a muscle with $N$ fibers, we define the activity-level $X_i^N$ of muscle-fiber $i$ $(i = 1,\ldots,N)$, as the sum of activity-levels of the MNs connecting to it, namely:

$$X_i^N = \sum_{k=1}^{2n} Y_k I_{i,k}^N. \tag{1}$$

Thus $X_1^N, X_2^N, \ldots, X_N^N$ are identically distributed (but dependent, as they are all defined by $\{Y_1, Y_2, \ldots, Y_{2n}\}$), with a positive and continuous density function $f_X$ over the interval $(2na, 2nb)$.

We now present the three rules of the game. The first two rules define the initial conditions and the third rule defines the dynamics of the game.

2.2. **Rule 1: Temporal order of competitions.** Roughly, all competitions start at the same time (around birth), but end at different times. At about two weeks after birth, all competitions are over, namely all muscle-fibers are singly innervated. Enhancing activity to

a muscle has been shown in many experiments to accelerate synapse elimination [28]-[29], whereas reducing activity delays or prevents it [30]-[31]. We call this "the effect of activity on the *rate* of synapse elimination".

Applying this principle to the level of muscle-fibers implies that competitions at muscles-fibers with higher initial activity-levels end earlier than competitions at muscle-fibers with lower initial activity-levels. This confers an *order* upon the resolvent times of the $N$ competitions (at the $N$ muscle-fibers) according to their initial level of activity; from highest to lowest.

Hence, our game is composed of $N$ *successive* stages (corresponding to $N$ successive ending-times). At each stage exactly one competition is resolved. This is the competition at the muscle-fiber with the highest activity-level, which was not yet resolved. Thus, if we re-index the muscle-fibers, according to the time in which their competition ends, then the first rule of the game is: $X_1^N \geq X_2^N \geq \ldots \geq X_N^N$.

2.3. **Rule 2: Democratic prior winning probabilities.** Earlier, we divided the MNs into two equal-sized teams: M-team (the More-active MNs) and L-team (the Less-active MNs). Whenever a MN wins (singly innervates) a muscle-fiber, we say that its team won there.

Given the ordered sequence of $N$ muscle-fibers (from highest activity-level to lowest activity-level), we define "the prior winning probability of M-team" at each muscle-fiber $i$ as: $P_i^N$. The meaning of $P_i^N$ is that this is the winning probability of M-team at muscle-fiber $i$, has there been no dependence between the competitions at different muscle-fibers.

There are only two requirements the prior winning probabilities must obey. First, since a team wins whenever one of its members win, then the more connections a team has at a muscle-fiber, (relative to its competitor team), the more chances it has of winning there. We call this requirement "the monotonic requirement". We denote as $q_i^N$ the proportion of M-team connections at muscle-fiber $i$. Concentrating on the dependence of $P_i^N$ on $q_i^N$, we write:

$$P_i^N = \rho(q_i^N),$$

where $\rho : [0,1] \to [0,1]$ is called the "prior winning function". The mathematical formulation of the monotonic requirement is:

*The function $\rho$ is monotonically increasing in $q$.*

For simplicity, we assume that $\rho$ is *strictly* increasing.

The second requirement is called the "neutral" or "symmetric" requirement. It states that a priori, a more-active and a less-active MN, competing at a muscle-fiber, have the same chances of defeating each other. This is actually only a default assumption. Later on we relax this assumption, and show that even if a more-active MN has higher initial probability of defeating a less-active MN, then still, if this bias is not overwhelming, less-active MNs win in more competitions, (and a fortiori, if less-active MNs have better chances of eliminating more-active MNs, then they win the game). The mathematical formulation of the neutrality requirement is:

$$\rho(1-q) = 1 - \rho(q). \quad \forall q.$$

To see that this is indeed a formulation of the neutrality requirement, note that $\rho(1-q)$ is the winning probability of M-team at a muscle-fiber with a fraction of $1-q$ M-team connections, and $1-\rho(q)$ is the winning probability of L-team in the "mirror-case", namely at a muscle-fiber with a fraction of $1-q$ L-team connections.

Taken together, the above two requirements (neutrality and monotonicity) imply that the team, which has the larger fraction of connections at a muscle-fiber, has higher prior winning probability there, since by the neutrality requirement: $\rho\left(\frac{1}{2}\right) = \frac{1}{2}$, and so if $q > \frac{1}{2}$, then by the monotonic requirement: $\rho(q) > \rho(\frac{1}{2}) = \frac{1}{2}$.

Note that the requirements of rule 2, pose a very weak restriction, allowing a broad range of possible biological behaviors. Any sequence of probabilities that satisfies the minimalist requirements of rule 2 may serve as the sequence of prior winning probabilities. This generality endows our model with robustness, which is especially important in light of the uncertainties regarding the competitive mechanism of synapse elimination.

As mentioned above, the meaning of $P_i^N$ is that this is the winning probability of M-team at muscle-fiber $i$, has there been no dependence between the competitions at different muscle-fibers. Since this model assumes that the outcome of the competitions *do* depend on one another, the prior winning probabilities define the initial conditions of the game, and they are adjusted during the game, according to rule 3 ahead.

2.4. **Rule 3: Limitation of resources and posterior winning probabilities.** If a MN wins at a muscle-fiber, then from that time on, it must devote resources for maintaining *all* the synapses there, thus its effectiveness in competing at future muscle-fibers is reduced. This idea is based on experimental results (Kasthuri & Lichtman [38] 2003) , showing that at late stages of synapse elimination, a MN has a clear advantage over a competitor with a larger muscle unit. Thus the competitive vigor of local competitions depends on a globally distributed resource. An accompanying paper by Lichtman and Sanes tests the idea that the amount of neurotransmitter released is this global resource ( Lichtman, Sanes *et al.* [39] 2003). Even though the players of the game are the individual MNs, the mathematical formulation of the resource limitation is applied to the *teams*. This matter is discussed in the following section.

For each stage $1 \leq i \leq N$, we define a random variable $W_i^N$ as the difference between the number of winnings of M-team and the number of winnings of L-team, until (including) stage $i$. Note that $W_0^N \equiv 0$, and $W_i^N \in \{-i, \ldots, 0, \ldots i\}$.

We define $P(i|W_{i-1}^N)$ to be the actual (posterior) winning probability of M-team at the current $i$th muscle-fiber, given that so far it won in $W_{i-1}^N$ more competitions than L-team.

The formulation of rule 3 is:

$$P(i|W_{i-1}^N) = P_i^N - \mu(N, P_i^N)W_{i-1}^N, \tag{2}$$

where $P_i^N$ was defined as the prior winning probability of M-team at muscle-fiber $i$, and $\mu = \mu(N, P_i^N)$, which we call "the adjustment function", is a positive function, satisfying that:

$$0 \leq P_i^N - \mu(N, P_i^N)W_{i-1}^N \leq 1, \quad \forall N, i, W_{i-1}^N. \tag{3}$$

Note that in particular, $\mu$ must satisfy that:

$$\mu\left(N, P_i^N\right) \leq \frac{1}{2}, \quad \forall N, i, W_{i-1}^N,$$

since in order to satisfy (3) for $W_{i-1}^N = \pm 1$, , $\mu$ must satisfy that:

$$\mu\left(N, P_i^N\right) \leq \min\{P_i^N, 1 - P_i^N\} \leq \frac{1}{2}.$$

In addition, in order to satisfy Equation (3), $\mu$ must also satisfy that:

$$\mu\left(N, 0\right) = \mu\left(N, 1\right) = 0, \ \forall N,$$

namely, that when $P_i^N = 0$ or $P_i^N = 1$, then there is no adjustment. This is logical since for example when $P_i^N = 1$, this means that muscle-fiber $i$ is innervated exclusively by M-team members, and so no matter which MN wins, M-team wins there, thus $P(i|W_{i-1}^N)$ must also equal 1. For convenience, we require that, $\mu$ is *strictly* positive in $(0, 1)$. Namely, that for all $0 < P_i^N < 1 :\ \mu(\cdot, P_i^N) > 0$.

By Equation (2), if up to stage $i$, M-team won more than it lost (i.e., $W_{i-1}^N > 0$), then its posterior winning probability $P(i|W_{i-1}^N)$ at muscle-fiber $i$, is smaller than its prior winning probability $P_i^N$, expressing the limitation of resources. A simple example for $\mu$ is:

$$\mu\left(N, P_i^N\right) = \frac{P_i^N\left(1 - P_i^N\right)}{N}.$$

2.5. **The division into teams.** We divided the MNs into two equal-sized teams, M-team and L-team, according to their activity-levels (see above). Indeed, when a MN wins, this influences only its own resources, and thus it *directly* influences only its own future winning probabilities. However, the future winning probabilities of its team are also reduced, as now explained: The winning probability of a team, at any muscle-fiber equals the sum of winning probabilities of its members there. And since each MN has positive probability of innervating each muscle-fiber, then the team's winning probability at any muscle-fiber, is reduced as a result of previous winnings of any of its members. Hence, it is equally true to speak in terms of *teams* rather than individual MNs, and say that when a *team* wins, its future winning probabilities are reduced. In other words, the dynamics of adjusting the winning probabilities of individual MNs, confers a dynamics of adjusting the winning probabilities of the teams.

Deciding which of the two optional dynamics is more appropriate to describe "the game MNs play", reflects a classical issue in modeling. There is usually a tradeoff between the level of accuracy in description and the predictive (analytical) power of the model. Since in this work, we are not interested , in the detailed sizes of muscle units of each MN, we choose to define the dynamics at the level of the teams. This is indeed proved to be fruitful as it provides (as will be shown later) a simple mathematical formula for predicting the magnitude of the size principle under different initial conditions, as well as explanations for a variety of phenomena. Nevertheless, to ensure that our qualitative results are not dependent on the level of analysis, we also constructed an alternative dynamics defined at

the level of the individual MNs. We tested this dynamics by simulations (using *Matlab*), and indeed, also in this case, we confirmed the conclusions of our analysis, namely that less-active MNs strictly win the game. In summary, it is important to understand that the division into teams, does not define a game, in which members in the same team cooperate in any way (e.g., share common resource, or do not compete with each other), but rather MNs play individually and the division into teams merely reflects the level, at which the game is analyzed mathematically.

### Conclusion of the assumptions

- MNs connect to muscle-fibers at random: Each MN connects to each muscle-fiber with probability $\gamma$, independently of other connections.
- The game has $N$ successive stages. The competition that is resolved on stage $i$, has an activity-level of $X_i^N$, and **rule 1** states that: $X_1^N \geq X_2^N \geq \ldots \geq X_N^N$.
- The prior winning probability $P_i^N$ of M-team at stage $i$, satisfies that $P_i^N = \rho(q_i^N)$, $i = 1, 2, \ldots, N$, where $q_i^N$ is the proportion of M-team connections at muscle-fiber $i$, and the function $\rho$ is democratic, namely, it is a monotonically increasing function satisfying in addition, that: $\rho(1-q) = 1 - \rho(q),\ \forall q$ (**rule 2**).
- Resources are limited. Thus, the actual (posterior) winning probability $P(i|W_{i-1}^N)$ of M-team at stage $i$, depends on the difference $W_{i-1}^N$ in the number of winnings between the two teams, at stage $i$. Namely, **rule 3** states that:

$$P(i|W_{i-1}^N) = P_i^N - \mu(N, P_i^N)W_{i-1}^N,$$

where $P_i^N$ is the prior winning probability of M-team at muscle-fiber $i$, and $\mu = \mu\left(N, P_i^N\right)$ is an adjustment function.

## 3. RESULTS

### Biological consequences of the model

In this Discussion paper, we present only the main biological consequence of our model. The most important result is the emergence of the size principle from the game MNs play. Before providing the formal proof for the size principle (in Section 4.4 ahead), we present here the intuition underlying our proof.

3.1. **Why less-active MNs win more competitions?** Consider first the simpler case, in which all muscle-fibers have the same number of connections. In this case, competitions at muscle-fibers with more connections by M-team - which are therefore typically more active - tend to occur earlier than competitions at muscle-fibers with more connections by L-team (rule 1). Therefore by the democratic rule (rule 2), a priori, M-team is likely to win more than L-team, during earlier stages of the game, but just as well, L-team is likely to win more than M-team during later stages of the game. This difference in the expected times of winning (which comes from rule 1), is responsible for the symmetry breaking in favor of L-team, as we now explain: When a team wins at a muscle-fiber, this has a negative effect on its future winning probabilities, which will be reduced according to rule 3 of limited resources. Therefore, *it is better to win at later competitions* because this will reduce the winning probabilities only in the few competitions which are left until the end of the game. Since M-team is more likely to win in earlier competitions, this will affect many competitions, whereas later on, when L-team is more likely to win, this will effect fewer competitions and hence altogether L-team wins more competitions. In other words, since a winning means less available resources, it is better to win later and thus to have this resources longer, helping to win in more competitions. This is the basic intuition behind the game-theoretic result, which underlies our conclusion that less-active MNs win the game.

As explained above, the proof we give later, relies on our claim that M-team is more likely to win in earlier stages of the game and L-team is more likely to win in later stages of the game. The main difficulty in proving this comes from the fact that muscle-fibers *do not all* have the same total number of connections, thus a muscle-fibers innervated by *many* MNs all belonging to L-team may be more-active than a muscle-fiber innervated by, only few, MNs all from M-team. In this case, the competition at the former muscle-fiber is resolved earlier, even though the winning probability of M-team there is 0. (compared to 1 at the latter muscle-fiber). In other words, the *total* number of connections at a muscle-fiber is an important factor influencing the time the competition is resolved and enabling competitions with high prior winning probabilities (of M-team) to occur late, and competitions with low prior winning probabilities to occur earlier. We started with random connectivity (see Section **Assumptions of the Model** above). Namely, each MN connects to each muscle-fiber with probability $\gamma$, and with probability $1-\gamma$ does not connect to it. This yields a connectivity pattern illustrated in the simulation appearing in Figure 2. In this simulation (as in all

the simulations appearing in this work), the prior winning probability of M-team, at each muscle-fiber, is defined as the fraction of M-team connections there, thus the $y-$axis also denotes the *fraction* of M-team connections.

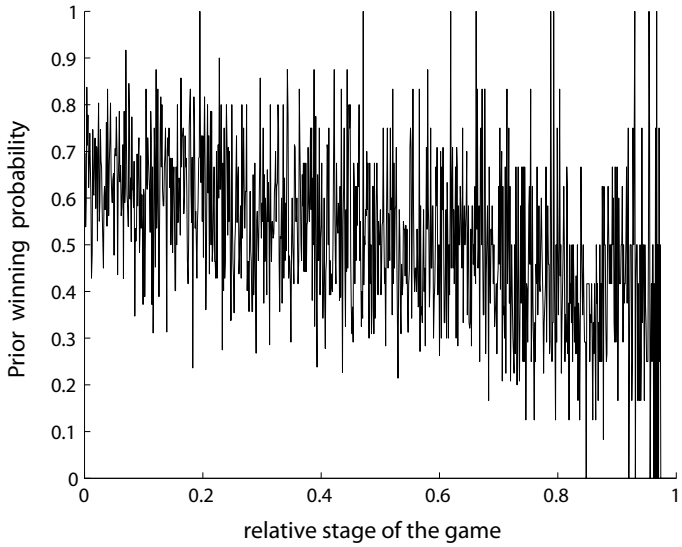


FIGURE 2. ***Initial conditions.*** *The $x$-axis represents the relative stage of the game, namely: $x = 1/N, 2/N, \ldots, 1$. The $y$-axis depicts the prior winning probability of M-team (which equals the fraction of M-team connections) at each stage $xN$, $N = 1000$.*

As seen, the sequence of prior winning probabilities $P^N_{xN}$ is noncontinuous and non monotonic. However, observe a decreasing tendency as $x$ moves from 0 to 1. This tendency expresses our claim that also under the more complicated case (when muscle-fibers do not necessarily have the same total number of connections), a priori, M-team is more likely to win in earlier stages of the game and L-team is more likely to win in later stages of the game. The intuitive argument for that is that our explanation for the simpler case, in which muscle-fibers have the same number of connections (see above), applies *separately* to each group of muscle-fibers having the same total number of connections.

3.2. **Economical implications.** As explained, more-active MNs are more involved in earlier competitions, and less-active MNs are more involved in later competitions. Thus a more-active MN can be viewed as "investing in early competitions", and similarly, a less-active MN can be viewed as "investing in late competitions". Thus, the general behavioral (or strategic) conclusion which follows from this game-theoretic result, and may apply also

to non-biological scenarios, e.g., economical settings, is that when resources are limited (the limitation being as in our setting), one should invest more in later competitions in order to win in more competitions. In contrast, one may erroneously conjecture that it is better to try winning in early competitions as well, guaranteeing winnings from the start, instead of taking a risk by waiting and letting the competitor collect these winnings. Which of the two intuitive arguments is the correct one, must thus be decided by mathematical proof, which is given in Section 4.4 ahead. In Figure 3a we show a simulation of the game (see **Methods**). As seen, at early stages M-team leads the game (as $W$ is positive) but eventually L-team wins in more competitions, expressing the size principle ($P < 10^{-107}$, one-tailed $t$-test).

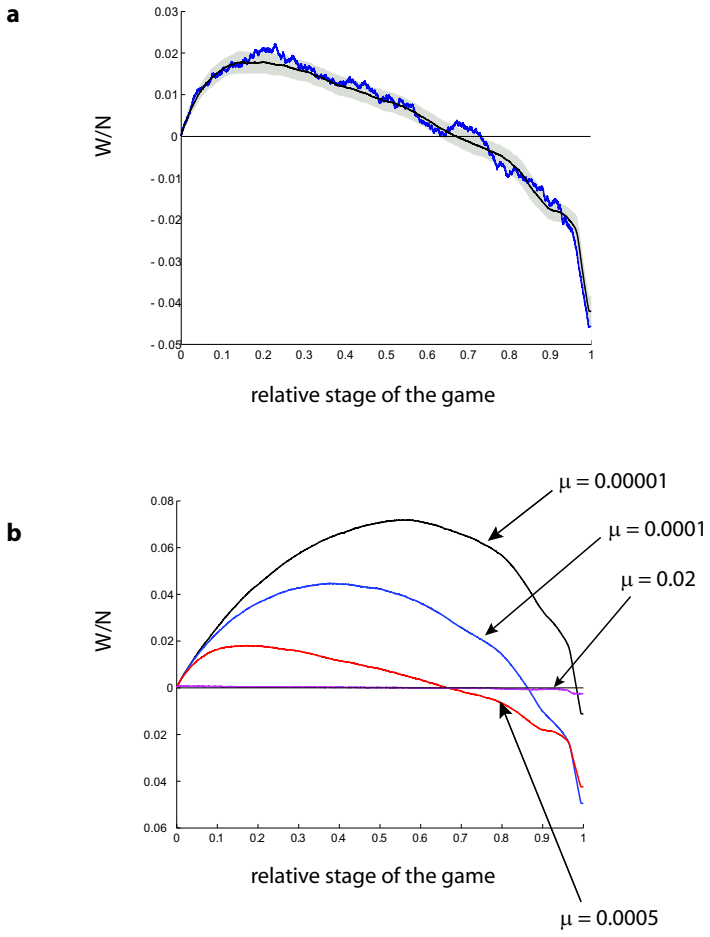


FIGURE 3. ***The emergence of the size principle as a result of the game MNs play.*** ***(a)*** *Simulation of a single game (thick line) and averaged over 100 games (thin smooth line). The $y$-axis shows the difference $W$ in the number of winnings, divided by the $N$ stages of the game (see* ***Methods****). Shaded area represents ± sd.* ***(b)*** *Averaged game for different values of adjustment size $\mu$.*

Figure 3b presents simulations for different values of $\mu$, where $\mu$ is constant along the game. Note that $\frac{1}{N}W_N^N < 0$ for each $\mu$. The strongest expression of the size principle (smallest $\frac{1}{N}W_N^N$) is obtained for $\mu = 0.0001$, which is not the largest or smallest $\mu$ plotted. To understand why large $\mu$ yield weak expression of the size principle, recall that the *early* winnings of M-team, are the cause for less *total* winnings of M-team. Thus a large $\mu$, which decreases the actual winning probabilities, prevents M-team from winning as much *already at early stages* of the game, and thus is less "punishing" later.

In addition, we show that even if the prior winning function is not neutral but rather it is strongly biased in favor of M-team (namely, $\rho$ satisfies: $\rho(1-q) \geq 1-\rho(q)\ \forall q$), then still L-team may win. In Figure 4, we present an example, in which a priori, M-team wins with probability of about 0.8 at muscle-fibers with an equal number of connections by both teams; yet, L-team still wins the game (This result holds for all $\mu > 0.0005$).

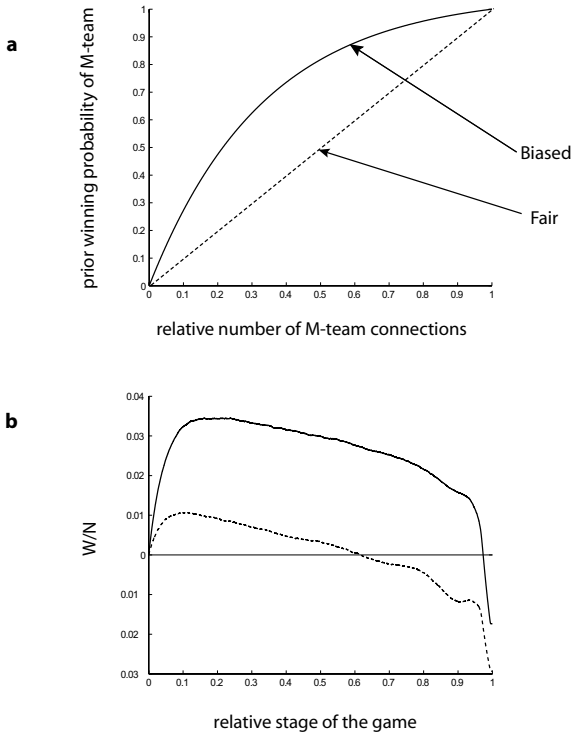


FIGURE 4. ***Even when the game is strongly biased in favor of more-active MNs, less-active MNs still win the game.*** ***(a)*** *Two prior winning functions. The solid line describes a function strongly biased in favor of M-team. The dashed line describes a neutral function.* ***(b)*** *Simulations of the game using the biased and neutral prior winning functions from a. Although M-team loses less when the prior winning function is biased in its favor, yet it still loses the game. ($\mu = 0.001$, $P < 10^{-9}$, one-tailed t-test).*

The mathematical treatment of biased games is presented in Appendix 7.2.

**The validity of our result.** The intuitive arguments presented above, are proved mathematically in Theorem 4.10 ahead. Specifically, we proved that for any adjustment function $\mu = \mu\left(N, P_i^N\right)$, and any democratic function $\rho$, the size principle emerges. This means that not knowing $\mu$, the specific rule by which the resource limitation is implemented, and not knowing the specific competitive rule $\rho$ that mediates the competition at the muscle-fiber, we were still able to prove that unless the competition is extremely biased in favor of more-active MNs, then the size principle emerges.

3.3. **Resolving the paradox of contradictory experimental data.** One body of research comes from experiments that were done at the *single* muscle-fiber (some *in vivo* and others *in vitro*) (Balice-Gordon & Lichtman [24] 1994, Lo & Poo [25] 1991, Vrbova *et al.* [28] 1978, Connold, Vrbova *et al.* [40] 1986, Dan & Poo [41] 1992, Liu *et al.* [43][42] 1994). This body of research results is consistent among itself, and suggests a competitive advantage for *more*-active MNs (namely, the stimulated MNs in the case of selective stimulation, and the unblocked MNs in the case of selective blocking). In order to relate to the results of these experiments, we divide the population of MNs into "manipulated" and "unmanipulated" teams, (e.g., "blocked" and "unblocked" teams), instead of the former "more-active" and "less-active" teams (see **Methods** for details).

3.3.1. *Selective blocking.* Selective blocking of a group of axons to a muscle has been shown to have an opposite effect in different experiments. In the experiments of Callaway [20][21] (1987, 1989), the muscle units of the blocked group were larger than in the control, whereas in the experiment of Ribchester and Taxt [22] (1983), the opposite was true. Indeed, according to our model, selective blocking is *expected* to have opposite effects. On the one hand, the blocked MNs are expected to lose in almost all the competitions that are resolved during the blocking period, as these competitions are strongly biased against them. This may explain the experimental results of Ribchester and Taxt, in which the blockade period was long (1-2 weeks) and probably included the resolution-time of most (if not all) competitions. On the other hand, selective blocking specifically delays the competitions at the muscle-fibers that are innervated by some blocked axons. This is not only predicted by rule 1 as the overall level of activity of these muscle-fibers is reduced, but has also been found empirically

by Callaway [21] (1989). This delay works in favor of the blocked group when activity is resumed, since, as explained earlier, *it is better to win at later competitions.* This may explain the experimental results of Callaway, in which the activity was blocked for a shorter time (4-5 days), and was then recovered. See Figure 5 for a simulation replicating both sets of results for $\mu = 0.0005$.

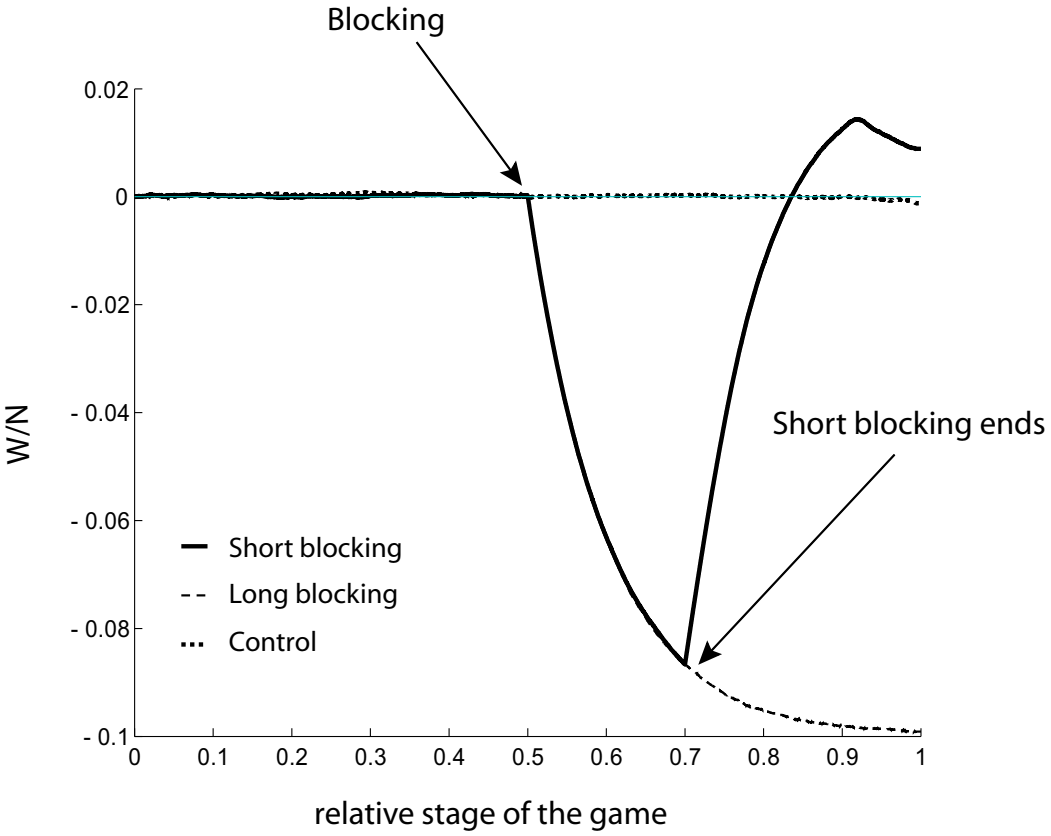


FIGURE 5. ***Resolving the paradox; reproducing the contradictory results from blocking experiments.*** *Simulations of the blocking procedure of Callaway[7][8] (solid line) and Ribchester & Taxt[9] (dashed line) against control (dotted line). Depicted is the averaged difference in the number of winnings between the blocked and un-blocked teams. Activity is blocked halfway through the game. In Callaway's experiments, activity resumes, and the blocked team wins significantly more than control as the solid line is higher than the dotted line at* $x = 1$, *(*$P < 10^{-27}$, *one-tailed t-test). In Ribchester & Taxt (dashed line), blocking continues until the end of the game and the blocked team loses. The same qualitative behavior is achieved for all* $0.00035 \leq \mu \leq 0.002$.

3.3.2. *Selective stimulation.* As a "mirror-case" of selective blocking, selective stimulation is also expected to have opposite effects; raising the winning probabilities of the stimulated MNs during the competition period, but also specifically bringing forward competitions at muscle-fibers that are innervated by stimulated axons (thus reducing their actual winning probabilities). Figure 6a shows a simulation that follows the stimulation procedure of Ridge and Betz [23] (1984) and produces the same qualitative result - that the stimulated MNs win more than the control. We predict that executing the same procedure *earlier* will prove to be less successful for the stimulated group. (Figure 6b). The reason for that is that this brings the competitions which stimulated MNs take part (and are most likely to win in), even *more* forward, reducing their winning probabilities in all the future competitions that are left until the end of the game.

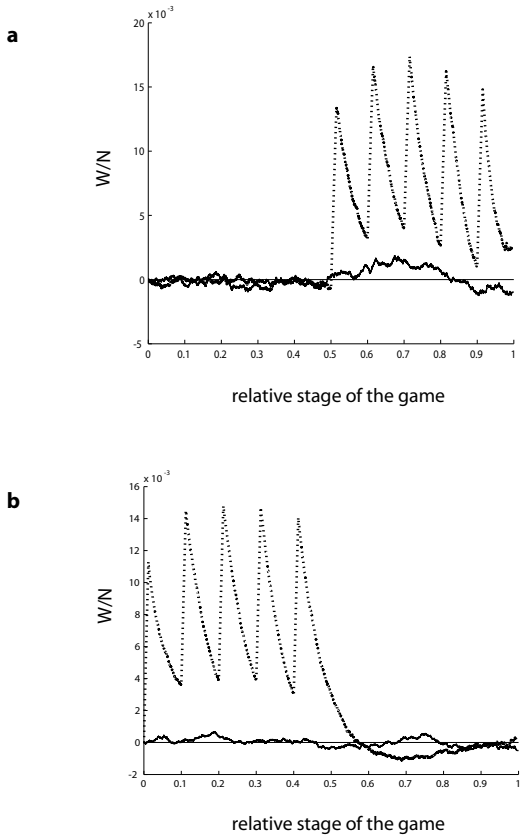


FIGURE 6. ***Further resolving the paradox; stimulated MNs win more than control.*** ***(a)*** *As in Ridge & Betz[10], 5 consecutive short stimulations were applied, each for a fraction of 0.015 stages (corresponding to 4 hrs of stimulation per day, for 5 days). Depicted, is the difference in the number of winnings between the stimulated and un-stimulated teams. The stimulated group (dotted line) won in significantly more competitions than control (solid line)* $P < 0.03$*, one-tailed t-test.* ***(b)*** *Executing the same procedure earlier is less successful for the stimulated group.*

3.4. **New Predictions.** The mathematical analysis of the model (presented in the subsequent sections), provides several new predictions that are *experimentally testable*. Here, we present the predictions, and in Appendix 7.3 we provide the mathematical proofs and explanations. We prove (in the following sections), that under normal conditions, less-active MNs indeed have larger muscle units than more-active MNs. But how much larger? Our work, as a theoretical work, cannot provide numerical assessments of that, however it can point to the factors that influence the magnitude of the size principle. Specifically, it compares different muscles, and predicts in which of them, the magnitude of the size principle is larger (namely, the less-active MNs have *much* larger muscle units than the more-active MNs). The first prediction though is different, as it relates to the identity of the winner (and not to the magnitude of the size principle).

Recall that the random variable $W_N^N$ denotes the difference between the number of winnings of M-team and that of L-team, when the game ends. Note that $-1 \leq \frac{1}{N}W_N^N \leq 1$, hence, $\frac{1}{N}W_N^N$ could serve as a measure for the magnitude of the size principle in different muscles. For details see Appendix 7.4. The predictions are:

(1) ***An early winner, (namely a MN that singly innervates a muscle-fiber already during the first days of synapse elimination), has on average, a lower activation-threshold, than a MN winning later.*** See again Figure 1, which shows that at an early phase of synapse elimination, $\frac{1}{N}W_N^N$ is positive, reflecting more winnings by M-team than by L-team, but at later phases of the process $\frac{1}{N}W_N^N$ becomes negative, reflecting more winnings by the less-active L-team.

(2) ***Executing the same selective stimulation procedure as Ridge and Betz*** [23] ***(1984) but at earlier stage of synapse elimination weakens the advantage of the stimulated MNs seen in Ridge and Betz experiments (Figure 6b).***

(3) As mentioned earlier, enhancing activity of *all* the MNs, accelerates synapse elimination, whereas blocking the activity, delays it. But what is the effect of these manipulations on the size principe? Does enhancing the activity cause less-active MNs to win even more, or perhaps the other way around? This question was not answered previously. Our model predicts the following:

***Applying extremely strong stimuli - beyond the activation-thresholds of all the MNs, will abolish the size principle (namely, all muscle units are about the same size), whereas replacing the natural stimuli by weak stimuli - beneath***

***the median activation-threshold, would reverse the size principle (namely, the more-active MNs will have larger muscle units than the less-active MNs).***

(4) ***If the activation-thresholds of the MNs are very much alike, then the size principle is only weakly expressed.***

(5) ***If the initial innervation is manipulated into full innervation, namely each MN initially innervates each muscle-fiber, then the size principle vanishes.*** See Figure 7 at $\gamma = 1$.

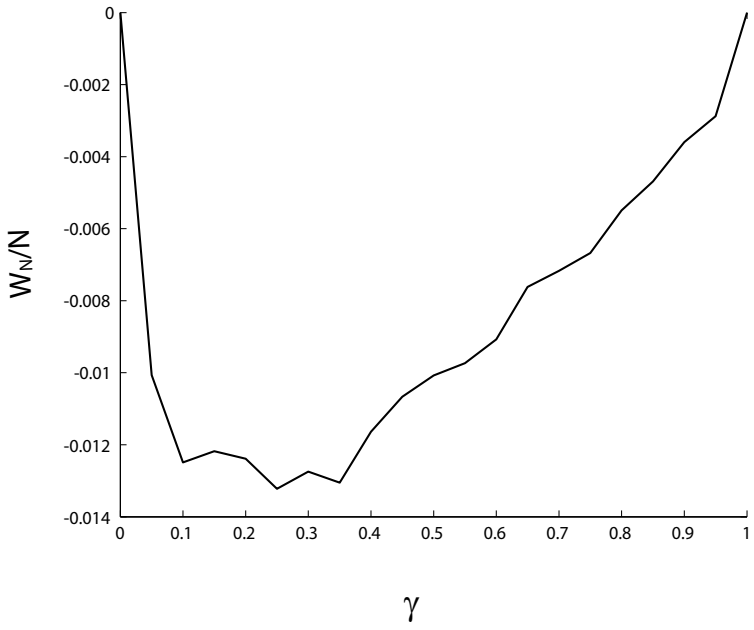


FIGURE 7. ***The expression of the size principle depends on the degree of innervation.*** *The value of $W_N/N$ (the difference in winnings at the end of the game, divided by the number of stages) is computed for different values of $\gamma$. $\mu\left(N, P_i^N\right) = \frac{100\gamma}{N}$.*

(6) ***The magnitude of the size principle is independent on the number of fibers in the muscle.*** This means that a large muscle (with many fibers) and a small muscle, will express the size principle to the same degree. Figure 8 shows simulations of the game for different number of muscle-fibers.

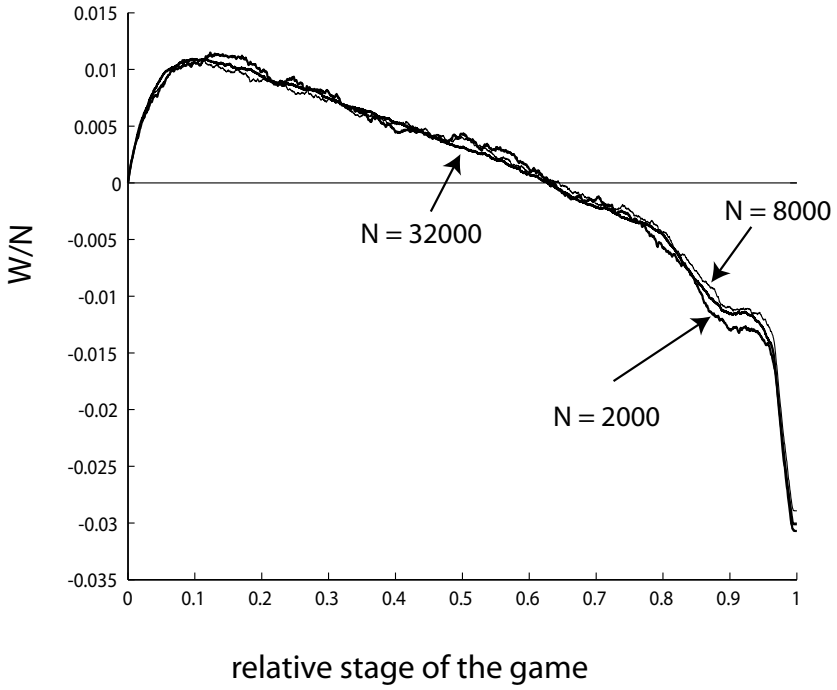


FIGURE 8. ***The magnitude of the size principle does not depend on the size of the muscle.*** *All along the game, there is no significant difference in W/N for different values of N, implying that the degree of expression of the size principle does not depend on the number of fibers in the muscle.* $\mu = \frac{10}{N}$.

## 4. MATHEMATICAL PROOFS

4.1. **Summary of main mathematical results.** The main goal is to prove the size principle (Theorem 4.10). We defined $W_N^N$ as the difference between M-team and L-team winnings, at the end of the game. Thus we wish to prove that $E(W_N^N) < 0$. We prove this for any number of innervating MNs, any averaged fraction $\gamma$ a MN innervates initially and any number of fibers $N$ in the muscle. Several other results are achieved for the limit: $N \to \infty$. These results apply to all muscles, since all muscles have many fibers ($N$ ranges from several thousands to millions). Thus $N \to \infty$ is a good enough approximation to reality.

Section 4.2 analyzes the initial conditions of the game, which are defined by the random initial connectivity, and rules 1 and 2. The proof of the size principle relies on our claim that M-team tends to win in early stages and L-team tends to win later on. This claim is proved in Lemma 4.3, and shown in Figure 9 by the decreasing tendency of the prior winning probabilities along the game.

In Proposition 4.5, we show that this result is robust in the sense that it does not vanish as $N \to \infty$.

In Section 4.3, we average over the prior winning probabilities. Namely, we analyze the *expectations* of the prior winning probabilities. We prove that unlike the sequence of prior winning probabilities, which is not continuous, the sequence of their expectations converges almost uniformly to a function that is continuous except at finitely many points (Proposition 4.6). This claim is illustrated in Figure 9.

We then turn to analyze the dynamics of the game, defined by the rule 3. Following two Lemmas providing an expression for the expectation $E(W_N^N)$, we prove the emergence of the size principle (Theorem 4.10).

Section 4.5 relates to nontrivial games. A nontrivial game is a game in which the strength of the dynamics, defined by the *total* size of adjustment is the same for any number of fibers $N$ in the muscle (i.e., any number of stages in the game). We show that this is equivalent to saying that the size of adjustment is of the form of $\mu\left(N, P_i^N\right) = \frac{f(P_i^N)}{N}$. As mentioned above, in Theorem 4.10 we proved the size principle by proving that for *any* adjustment function (trivial or nontrivial): $E(W_N^N) < 0$. In Theorem 4.12, we strengthen this result for nontrivial games and show that $\frac{1}{N}E(W_N^N)$ is bounded away from zero. Namely, that there exists $D > 0$, such that for all $N$ : $\frac{1}{N}E(W_N^N) < -D$.

In Section 4.6, we arrive at an equality enabling us to predict the magnitude of the size principle, which is measured by $\frac{1}{N}W_N^N$. For this we average on the initial conditions, namely we use the *expectations* of the prior winning probabilities.

We prove that:

$$\lim_{N\to\infty} \frac{1}{N} W_N^N = \int_0^1 (2p(s) - 1) e^{-2f(p(s))(1-s)}\, ds, \tag{4}$$

where $p(s)$ is the expectation of M-team's prior winning probability at muscle-fiber $s$, when $N \to \infty$, and the convergence of $\frac{1}{N}W_N^N$ is in probability (Theorem 4.13 and Corollary 4.14).

Equation (4), provides new predictions. These predictions were listed in Section 3.4 above, and are proved mathematically in Appendix 7.3.

The mathematical analysis that follows, relies on the assumptions of the model (see Section 2).

4.2. **Initial conditions - Rules 1 and 2 combined with random initial connectivity.** We partition the population of muscle-fibers into $2n+1$ sets $A_r$ according to the total number of MNs ($r = 0, 1, \ldots, 2n$) connecting to the muscle-fiber.

For each $r = 0, 1, \ldots, 2n$, we further divide the muscle-fibers in $A_r$ into subsets $A_{r,m}$, $m = 0, 1, \ldots, r$, according to the number of M-team connections at the muscle-fiber. Denote:

$$A_{r,m} = \left\{ i_\ell \ ; \ \sum_{k\in\text{M-team}} I^N_{i_\ell,k} = m \ , \ \sum_{k\in\text{L-team}} I^N_{i_\ell,k} = r-m \right\},$$

where $I^N_{i,k}$ was defined earlier, as the indicator variable receiving 1 *iff* MN $k$ connects to the random muscle-fiber $i$. Hence: $\sum_{m=0}^{r} A_{r,m} = A_r$.

Denote: $N_{r,m} = |A_{r,m}|$, namely, $N_{r,m}$ is the number of muscle-fibers having $m$ M-team connections, and $r-m$ L-team connections.

Denote the activity-levels of the muscle-fibers in $A_{r,m}$ as:

$$\left(X^N_{i_\ell}\,|r,m\right), \quad \ell = 1, \ldots, N_{r,m}$$

and the activity-levels of the muscle-fibers in $A_{r,r-m}$ as:

$$\left(X^N_{i'_\ell}\,|r,r-m\right), \quad \ell = 1, \ldots, N_{r,r-m}.$$

Note that $i_\ell$ and $i'_\ell$ depend on $m$ and $r$, but for the sake of convenience, we omit this dependence from the notation.

The muscle-fibers are ordered according to decreasing levels of activity and we did not disrupt this order, hence the above two sequences are decreasing. Namely:

$$\left(X^N_{i_1}\,|r,m\right) \geq \left(X^N_{i_2}\,|r,m\right) \geq \ldots \geq \left(X^N_{i_{N_{r,m}}}\,|r,m\right),$$

and,

$$\left(X^N_{i'_1}\,|r,r-m\right) \geq \left(X^N_{i'_2}\,|r,r-m\right) \geq \ldots \geq \left(X^N_{i'_{N_{r,m}}}\,|r,r-m\right).$$

The activity-level $(X|r,m)$ of a muscle-fiber with $m$ M-team and $r-m$ L-team connections, is determined by the MNs connecting to the muscle-fiber ($m$ from M-team and $(r-m)$ from L-team). Namely if $\{y_{k_m}\}$ are the activity-levels of the MNs innervating the muscle-fiber, then the activity-level of the muscle-fiber is:

$$x = \sum_{k_m} y_{k_m}.$$

Hence for any given vector of the MNs' activity-levels $Y = (Y_1, \ldots, Y_{2n}) = (y_1, \ldots, y_{2n})$, the number of possible values of $(X|r,m)$ is:

$$S(r,m) = \binom{n}{m}\binom{n}{r-m},$$

(as $m$ MNs are chosen from the $n$ members of M-team, and $r-m$ MNs are chosen from the $n$ members of L-team). Similarly, given $Y$, the number of possible values of $(X|r, r-m)$ is also $S(r,m)$.

*Remark* 4.1. Since $Y_k,\ k=1,\ldots,2n$ are continuous random variables, we assume that in any realization $y=(y_1,\ldots,y_{2n}):\ y_k \neq y_\ell,\ \forall k \neq \ell$. In addition, it also follows from the continuity of $Y_k,\ k=1,\ldots,2n$, that with probability 1, muscle-fibers that are innervated by different (non-identical) subsets of MNs, have different activity-levels.

Denote the possible values of $(X|r,m)$ and $(X|r,r-m)$ as

$$(x_1 \mid r,m) > \ldots > (x_{S(r,m)} \mid r,m), \tag{5}$$

and

$$(x_1 \mid r,r-m) > \ldots > (x_{S(r,m)} \mid r,r-m) \tag{6}$$

respectively.

Note that the above sequences of values, are actually random variables determined by: $Y=y$. Thus each pair: $m,r$, defines two sequences of random variables, appearing in Equations (5) and (6).

**Proposition 4.2.** *Given* $Y=y$, *for all* $m > \frac{r}{2}$ *and for each* $1 \leq s \leq S(r,m)$ :

$$(x_s \mid r,m) > (x_s \mid r,r-m).$$

The intuition behind this proposition is as follows: Although, the random variables

$$(x_1 \mid r,m),\ldots,(x_{S(r,m)} \mid r,m),$$

and

$$(x_1 \mid r,m),\ldots,(x_{S(r,r-m)} \mid r,r-m),$$

are all sums of activity-levels of $r$ MNs, the random variables in the first sequence are composed of $m$ MNs from the more-active MNs, whereas the random variables in the second sequence are composed from only $r-m<m$ MNs from M-team.

The proof of Proposition 4.2 appears in Appendix 7.5.

Recall that the random variable $i_\ell$, is the index of the $\ell$th muscle-fiber with $m$ M-team, and $r-m$ L-team connections, and similarly, the random variable $i'_\ell$, is the index of the $\ell$th muscle-fiber with $r-m$ M-team, and $m$ L-team connections.

Recall that the muscle-fibers are ordered according to decreasing level of activity. We wish to show that if $m > \frac{r}{2}$, then it is more probable that $i'_\ell > i_\ell$, than the other way around:

**Lemma 4.3.** *For all* $1 < i < j < N$, $\ell = 1, 2, \ldots,$ *and* $m > \frac{r}{2}$ :

$$P(\ i_\ell = i\ ,\ i'_\ell = j\ ) \geq P(\ i_\ell = j\ ,\ i'_\ell = i\ ), \tag{7}$$

*where an equality is achieved only when both sides of the inequality are zero.*

*Proof.* The proof follows from Proposition 4.2, the order rule (rule 1) and the random initial connectivity (i.e., each MN innervates each muscle-fiber with probability $\gamma$, independently on the other connections).

Because all MNs are equally likely to connect any given muscle-fiber, the random variable $(X \mid r, m)$ attains each of the values: $(x_1|r, m), \ldots, (x_{S(r,m)}|r, m)$, with the *same* probability of $\frac{1}{S(r,m)}$. Similarly, $(X \mid r, r-m)$ attains the values: $(x_1|r, r-m), \ldots, (x_{S(r,m)}|r, r-m)$, each with the same probability of $\frac{1}{S(r,m)}$. In other words, the random variable $(X \mid r, m)$ has a discrete uniform distribution over the set: $\Big\{(x_1|r, m), \ldots, (x_{S(r,m)}|r, m)\Big\}$, and similarly, the random variable $(X \mid r, r-m)$ has a discrete uniform distribution over a different set: $\Big\{(x_1|r, r-m), \ldots, (x_{S(r,m)}|r, r-m)\Big\}$.

Hence looking at the two corresponding $\ell$th order statistics $(X_{i_\ell} \mid r, m), (X_{i'_\ell} \mid r, r-m)$, $\ell = 1, 2, \ldots,$ we have that for each $1 \leq s \leq S(r, m)$, and for all $y$ :

$$p(s) := P\Big(\ (X_{i_\ell} \mid r, m) = (x_s \mid r, m) \ \Big|\ Y = y\ \Big) = P\Big(\ (X_{i'_\ell} \mid r, r-m) = (x_s \mid r, r-m) \ \Big|\ Y = y\ \Big).$$

It is important to realize that given $Y = y$, the activity-level of any random muscle-fiber, is independent on the activity-levels of other muscle-fibers. Thus, given $Y = y$, the two order statistics sequences:

$$\left(X^N_{i_1}|r, m\right) \geq \left(X^N_{i_2}|r, m\right) \geq \ldots \geq \left(X^N_{i_{N_{r,m}}}|r, m\right),$$

and,

$$\left(X^N_{i'_1}|r, r-m\right) \geq \left(X^N_{i'_2}|r, r-m\right) \geq \ldots \geq \left(X^N_{i'_{N_{r,r-m}}}|r, r-m\right),$$

are independent of one another[4].

In particular, given $Y = y$, $(X_{i_\ell} \mid r, m)$ is independent of $(X_{i'_\ell} \mid r, r-m)$, $\forall \ell, m, r$. Hence, by Proposition 4.2, for all $\ell = 1, 2, \ldots$ :

[4]The variables inside the *same* sequence, *do* depend on each other, as these sequences are ordered and the order creates a dependence between them.

$$P\Big(\ (X_{i_\ell} \mid r,m) > (X_{i'_\ell} \mid r, r-m) \ \Big|\ Y=y\Big) \geq$$

$$\sum_{s=1}^{S(r,m)} \sum_{t=s}^{S(r,m)} P\Big((X_{i_\ell} \mid r,m) = (x_s \mid r,m)\ ,\ (X_{i'_\ell} \mid r, r-m) = (x_t \mid r, r-m) \ \Big|\ Y=y\ \Big) =$$

$$\sum_{s=1}^{S(r,m)} P\Big((X_{i_\ell} \mid r,m) = (x_s \mid r,m) \Big|\ Y=y\Big) \sum_{t=s}^{S(r,m)} P\Big((X_{i'_\ell} \mid r, r-m) = (x_t \mid r, r-m) \Big|\ Y=y\Big) =$$

(8)

$$\sum_{s=1}^{S(r,m)} p(s) \sum_{t=s}^{S(r,m)} p(t) = \sum_{s=1}^{S(r,m)} p^2(s) + \sum_{s=1}^{S(r,m)} \sum_{t=s+1}^{S(r,m)} p(s)p(t).$$

Note that:

$$1 = \left(\sum_{s=1}^{S(r,m)} p(s)\right)^2 = \sum_{s=1}^{S(r,m)} p^2(s) + 2 \sum_{s=1}^{S(r,m)} \sum_{t=s+1}^{S(r,m)} p(s)p(t), \tag{9}$$

and so, dividing Equation (9) by 2, yields:

$$\frac{1}{2} = \frac{1}{2} \sum_{s=1}^{S(r,m)} p^2(s) + \sum_{s=1}^{S(r,m)} \sum_{t=s+1}^{S(r,m)} p(s)p(t),$$

Thus, going back to Equation (8), we get that:

$$P\Big((X_{i_\ell} \mid r,m) > (X_{i'_\ell} \mid r, r-m) \ \Big|\ Y=y\Big) \geq$$

$$\sum_{s=1}^{S(r,m)} p^2(s) + \sum_{s=1}^{S(r,m)} \sum_{t=s+1}^{S(r,m)} p(s)p(t) = \frac{1}{2} + \frac{1}{2} \sum_{s=1}^{S(r,m)} p^2(s) > \frac{1}{2}. \tag{10}$$

Thus for all $m > \frac{r}{2}$ and for all $y$ :

$$P\Big((X_{i_\ell} \mid r,m) > (X_{i'_\ell} \mid r, r-m) \ \Big|\ Y=y\Big) > \frac{1}{2}. \tag{11}$$

Consider the event $\{i_\ell, i'_\ell\} = \{i, j\}$. This event states that either $i_\ell$ or $i'_\ell$, is the $i$th activity-level, and the other, is the $j$th activity-level (in the sequence of all activity-levels: $X_1^N \geq X_2^N \geq \ldots \geq X_N^N$).

It follows from (11), that if $i < j$, then it is more probable that $(X_{i_\ell} \mid r,m)$ is the $i$th order statistics and $(X_{i'_\ell} \mid r, r-m)$ is the $j$th order statistics, than the other way around. Formally, for all $i < j$ :

$$P\Big(X_i^N = (X_{i_\ell} \mid r,m)\ ,\ X_j^N = (X_{i'_\ell} \mid r, r-m) \ \Big|\ \{i_\ell, i'_\ell\} = \{i,j\}\ ,\ Y=y\ \Big) >$$
$$P\Big(X_i^N = (X_{i'_\ell} \mid r, r-m)\ ,\ X_j^N = (X_{i_\ell} \mid r,m) \ \Big|\ \{i_\ell, i'_\ell\} = \{i,j\}\ ,\ Y=y\ \Big).$$

Since this is true for each $y$, then:

$$P\Big(X_i^N = (X_{i_\ell} \mid r, m)\ ,\ X_j^N = (X_{i'_\ell} \mid r, r-m) \ \Big|\ \{i_\ell, i'_\ell\} = \{i, j\}\ \Big) >$$
$$P\Big(X_i^N = (X_{i'_\ell} \mid r, r-m)\ ,\ X_j^N = (X_{i_\ell} \mid r, m) \ \Big|\ \{i_\ell, i'_\ell\} = \{i, j\}\ \Big).$$

This is equivalent to saying that for all $i < j$ :

$$P\Big(i_\ell = i\ ,\ i'_\ell = j \ \Big|\ \{i_\ell, i'_\ell\} = \{i, j\}\ \Big) > P\Big(i_\ell = j\ ,\ i'_\ell = i \ \Big|\ \{i_\ell, i'_\ell\} = \{i, j\}\ \Big). \tag{12}$$

Now, for all $i < j$ :

$$P\left(i_\ell = i\ ,\ i'_\ell = j\ \right) - P\left(i_\ell = j\ ,\ i'_\ell = i\ \right) =$$
$$P\Big(i_\ell = i\ ,\ i'_\ell = j \ \Big|\ \{i_\ell, i'_\ell\} = \{i, j\}\Big) P\Big(\{i_\ell, i'_\ell\} = \{i, j\}\ \Big)$$
$$-$$
$$P\Big(i_\ell = j\ ,\ i'_\ell = i \ \Big|\ \{i_\ell, i'_\ell\} = \{i, j\}\Big) P\Big(\{i_\ell, i'_\ell\} = \{i, j\}\ \Big) =$$

(13)

$$\Big\{P\Big(i_\ell = i\ ,\ i'_\ell = j \ \Big|\ \{i_\ell, i'_\ell\} = \{i, j\}\Big) - P\Big(i_\ell = j\ ,\ i'_\ell = i \ \Big|\ \{i_\ell, i'_\ell\} = \{i, j\}\Big)\ \Big\} P\Big(\{i_\ell, i'_\ell\} = \{i, j\}\Big).$$

By Equation (12), we have that the expression in the brackets in (13) is strictly positive. Hence for all $i < j$ satisfying $P\Big(\{i_\ell, i'_\ell\} = \{i, j\}\Big) > 0$, we get that:

$$P\left(i_\ell = i\ ,\ i'_\ell = j\ \right) - P\left(i_\ell = j\ ,\ i'_\ell = i\ \right) > 0,$$

and for all $\{i, j\}$, such that: $P\left(\ \{i_\ell, i'_\ell\} = \{i, j\}\ \right) = 0$ :

$$P\left(i_\ell = i\ ,\ i'_\ell = j\ \right) - P\left(i_\ell = j\ ,\ i'_\ell = i\ \right) = 0.$$

Note that:

$$P\Big(\{i_\ell, i'_\ell\} = \{i, j\}\Big) = P\left(i_\ell = i\ ,\ i'_\ell = j\ \right) + P\left(i_\ell = j\ ,\ i'_\ell = i\ \right),$$

and so: $P\Big(\{i_\ell, i'_\ell\} = \{i, j\}\Big) = 0$, *iff* $P\left(i_\ell = i\ ,\ i'_\ell = j\ \right) = 0$, and $P\left(i_\ell = j\ ,\ i'_\ell = i\ \right) = 0$, and with that, we complete the proof of the Lemma. □

Recall that $A_{r,m}$, was defined as the set of muscle-fibers with $m$ M-team, and $r - m$ L-team connections.

It follows from the Lemma above, that

$$P\left(\ i_\ell < i'_\ell\ \right) > \frac{1}{2}.$$

We now wish to strengthen this result by proving that for $N$ large enough, the proportion of stages separating between $i_\ell$ and $i'_\ell$ is at least $S$, $S > 0$. This is proved in Proposition 4.5, but first we prove that the proportion of muscle-fibers in $A_{r,m}$ is about the same as the proportion of muscle-fibers in

$A_{r,r-m}$ as $N \to \infty$. This follows from the assumption of random initial connectivity (namely, each MN connects to each muscle-fiber with probability $\gamma$, independently on other connections), and is proved in Lemma 4.4 ahead.

Denote:

$$a(r,m) = \frac{1}{2^r}\binom{r}{m}\binom{2n}{r}\gamma^r(1-\gamma)^{2n-r},$$

where $\gamma$ was defined earlier as the probability of a connection between any MN to any muscle-fiber.

**Lemma 4.4.**

$$P\left(\forall m \le r \le 2n \;;\; \lim_{N\to\infty}\frac{N_{r,m}}{N} = a(r,m)\right) = 1, \tag{14}$$

where $N_{r,m}$ was defined as the number of muscle-fibers with $m$ M-team and $r-m$ L-team connections.

*Proof.* Recall that $I^N_{i,k}$ are *i.i.d.r.v* s.t $I^N_{i,k} \sim \mathbf{B}(1,\gamma)$, (where $I^N_{i,k}$ was defined earlier as the indicator, having a value 1 *iff* MN $k$ connects to the random muscle-fiber $i$, and 0 otherwise). Since $I^N_{i,k}$ are independent, then:

$$\sum_{k=1}^{2n} I^N_{i,k} \sim \mathbf{B}\left(2n,\gamma\right).$$

Recall that $A_r$ is the set of muscle-fibers with a total of $r$ connections. Thus for any randomly chosen muscle-fiber $c$ :

$$P\left(c \in A_r\right) = P\left(\sum_{k=1}^{2n} I^N_{c,k} = r\right) = \binom{2n}{r}\gamma^r(1-\gamma)^{2n-r}, \tag{15}$$

and so:

$$EN_r = N\binom{2n}{r}\gamma^r(1-\gamma)^{2n-r}, \quad \text{where} N_r = |A_r|.$$

Now, $\left(N_{r,m} \mid N_r\right) \sim \mathbf{B}\left(N_r, \binom{r}{m}\frac{1}{2^r}\right)$, where $N_{r,m} = |A_{r,m}|$,

hence:

$$E\left(N_{r,m} \mid N_r\right) = \binom{r}{m}\frac{1}{2^r}N_r,$$

and so:

$$E\left(N_{r,m}\right) = EE\left(N_{r,m} \mid N_r\right) = \binom{r}{m}\frac{1}{2^r}EN_r =$$

$$N\frac{1}{2^r}\binom{r}{m}\binom{2n}{r}\gamma^r(1-\gamma)^{2n-r} = Na(r,m).$$

By the strong law of large numbers, applied to each average $\frac{N_{r,m}}{N}$, we get that:

$$P\left(\forall m \le r \le 2n \;;\; \lim_{N\to\infty}\frac{N_{r,m}}{N} = a(r,m)\right) = 1. \tag{16}$$

□

**Proposition 4.5.** *There exists $S > 0$, such that for all $\epsilon > 0$, there exists $N_\epsilon$, such that $\forall N > N_\epsilon$ :*

$$P\left(\forall \ell = 1, 2, \ldots, \quad \forall m \leq r \leq 2n \;\; ; \quad \frac{i'_\ell - i_\ell}{N} \geq S \right) \geq 1 - \epsilon, \tag{17}$$

where $i_\ell$ is the $\ell$th muscle-fiber with $m$ M-team and $r - m$ L-team connections, and similarly, $i'_\ell$ is the $\ell$th muscle-fiber with $r - m$ M-team and $m$ L-team connections. Recall that $i_\ell, i'_\ell$ depend on $m, r$, and this dependence was omitted from the notation for convenience only.

*Proof.* Recall that given any $Y = y$,

$$S(r, m) = \binom{n}{m}\binom{n}{r-m},$$

is the number of values attained by $(X \mid r, m)$, and is also the number of values attained by $(X \mid r, r - m)$. We denoted these values as

$$(x_1 \mid r, m) > \ldots > (x_{S(r,m)} \mid r, m),$$

and

$$(x_1 \mid r, r - m) > \ldots > (x_{S(r,m)} \mid r, r - m),$$

respectively, and noted that since each MN is equally likely to connect any given muscle-fiber, then $(X \mid r, m)$ attains each of the values $(x_1 \mid r, m), \ldots, (x_{S(r,m)} \mid r, m)$, with the *same* probability of $\frac{1}{S(r,m)}$, and similarly, the random variable $(X \mid r, r - m)$ attains the values: $(x_1 \mid r, r - m), \ldots, (x_{S(r,m)} \mid r, r - m)$, each with the same probability of $\frac{1}{S(r,m)}$.

Denote $N_x =$ the number of muscle-fibers with activity-level of $x$. Consider the random variable $\frac{N_{(x_s|r,m)}}{N_{r,m}}$. This variable is the proportion of muscle-fibers with $m$ and $r - m$, M-team and L-team connections, with activity-level of $(x_s \mid r, m)$.

Then by the strong law of large numbers:

$$P\Big(\forall m \leq r \leq 2n,\ 1 \leq s \leq S(r, m)\ ;\ \lim_{n_{r,m} \to \infty} \frac{N_{(x_s|r,m)}}{\Big(N_{r,m}\Big|N_{r,m} = n_{r,m}\Big)} = \frac{1}{S(r, m)} \ \Big) = 1. \tag{18}$$

Now,

$$\frac{N_{(x_s|r,m)}}{N} = \left(\frac{N_{(x_s|r,m)}}{N_{r,m}}\right)\left(\frac{N_{r,m}}{N}\right),$$

thus with probability 1 :

$$\lim_{N \to \infty} \frac{N_{(x_s|r,m)}}{N} = \lim_{N \to \infty} \frac{N_{(x_s|r,m)}}{\Big(N_{r,m}\Big|N_{r,m} = n_{r,m}\Big)} \lim_{N \to \infty} \frac{N_{r,m}}{N}.$$

By Lemma 4.4, when $N \to \infty$, then $N_{r,m} \to \infty$, with probability 1, thus, the above is equivalent to:

$$\lim_{N\to\infty} \frac{N_{(x_s|r,m)}}{N} = \lim_{\mathbf{n_{r,m}}\to\infty} \frac{N_{(x_s|r,m)}}{\left(N_{r,m}\middle| N_{r,m}=n_{r,m}\right)} \lim_{N\to\infty} \frac{N_{r,m}}{N}. \tag{19}$$

From Lemma 4.4, we have:

$$P\left(\forall m \le r \le 2n \ ; \ \lim_{N\to\infty} \frac{N_{r,m}}{N} = a(r,m) \ \right) = 1. \tag{20}$$

Substituting (18) and (20), in (19), we get that:

$$P\left(\forall m \le r \le 2n, \ 1 \le s \le S(r,m) \ ; \ \lim_{N\to\infty} \frac{N_{(x_s|r,m)}}{N} = \frac{a(r,m)}{S(r,m)} \ \right) = 1. \tag{21}$$

Denote:

$$2S = \min_{m,r} \left\{ \frac{a(r,m)}{S(r,m)} \right\} > 0.$$

Then there exists $N_\epsilon$, such that for all $N > N_\epsilon$ :

$$P\left(\forall m \le r \le 2n, \ 1 \le s \le S(r,m) \ ; \quad N_{(x_s|r,m)} > SN \ \right) \ge 1-\epsilon. \tag{22}$$

With probability 1, $(x_s \mid r,m) > (x_s \mid r, r-m)$ (see Proposition 4.2). Hence by the order rule (rule 1), the variables $(X|r,m)$ with value $(x_s \mid r,m)$, come before the variables $(X|r,r-m)$ with value $(x_s \mid r,r-m)$. According to Equation (22), with probability as close as we wish to 1, there are at least $SN$ variables with $(x_s \mid r,m)$, and thus there are at least $SN$ stages separating between $i_\ell$ and $i'_\ell$.

Thus, for all $\epsilon$, there exists $N_\epsilon$, such that for all $N > N_\epsilon$ :

$$P\left(\forall \ell = 1,2\ldots, \quad \forall m \le r \le 2n \ ; \quad i'_\ell - i_\ell \ \ge SN \ \right) \ge 1-\epsilon,$$

and so:

$$P\left(\forall \ell = 1,2\ldots, \quad \forall m \le r \le 2n \ ; \quad \frac{i'_\ell - i_\ell}{N} \ \ge S \ \right) \ge 1-\epsilon.$$

□

4.3. **The averaged initial conditions.** Denote: $P^N = (P_1^N, P_2^N, \ldots, P_N^N)$, where $P_i^N$ was defined earlier as the prior winning probability of M-team at muscle-fiber $i$.

Up to now, we related to the prior winning probabilities $P^N$. We now wish to "average" on the initial conditions, by relating instead, to $EP^N$ (the sequence of expectations). Consider $P^N$ and $EP^N$, as functions of the relative stages: $\frac{1}{N}, \frac{2}{N}, \ldots, 1$. Namely:

$$\left| P^N_{\left(\frac{i}{N}\right)N} - P^N_{\left(\frac{i-1}{N}\right)N} \right| = \left| P_i^N - P_{i-1}^N \right|.$$

The major difference between $P^N$ and $EP^N$, is that $P^N$ is noncontinuous, but rather, $\left|P_i^N - P_{i-1}^N\right|$ is very likely to be large. In contrast, the sequence $EP^N$, (as we will soon prove in Proposition 4.6), converges almost uniformly to a function $p$, that is continuous except at finitely many points. In Figure 9, we present a simulation showing the this difference between $P^N$ and $EP^N$.

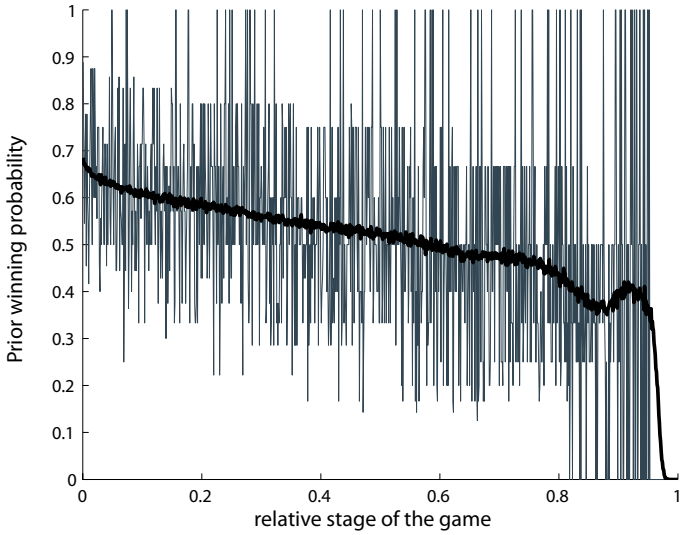


FIGURE 9. ***Initial conditions:*** $P^N$ ***compared with*** $EP^N$***.*** *Gray line presents one simulation of* $P^N$, $(N = 1000,\ \gamma = 0.05)$. *Black line presents the average over* 500 *simulations of* $P^N$, *thus is close to* $EP^N$. *Note the steep decrease to zero, near the end, which comes from muscle-fibers with no connections.*

**Proposition 4.6.** *There exists a function* $p : [0,1] \to [0,1]$, *that is continuous except at finitely many points, such that for each rational number* $s = \frac{i}{N} \in [0,1]$ :

$$\lim_{N\to\infty} EP_i^N = p(s),$$

*where the convergence is almost uniformly.*

*Proof.* Up to now we treated each pair $m, r$, separately, and denoted:

$$(x_1 \mid r, m), \ldots, (x_{S(r,m)} \mid r, m),$$

as the values attained by $(X \mid r, m)$.

We now wish to relate to the *full* sequence of muscle-fibers (i.e., all $m, r$, satisfying that: $0 \le m \le r \le 2n$).

Given the activity-levels $Y = y = (y_1, y_2, \dots, y_{2n})$ of all MNs, if $\{y_{k_1}, y_{k_2}, \dots\}$ are the activity-levels of the MNs innervating a muscle-fiber, then the activity-level of this muscle-fiber is:

$$x = \sum_l y_{k_l}.$$

Recall that $Y$ is a continuous random variable, hence, with probability 1, the activity-levels which correspond to different subsets of MNs, are different from each other (see Remark 4.1). Thus the number of possible values, an activity-level $X$ may attain, equals the number of subsets of $\{y_1, y_2, \dots, y_{2n}\}$, which is $2^{2n}$. We order these values such that:

$$x_{_1} > x_{_2} > \ldots > x_{_{2^{2n}}}.$$

Note that $x_{_1}, x_{_2}, \dots, x_{_{2^{2n}}}$, are random variables, since they are determined by $Y$, which is a random variable. Thus, given $Y = y$, we denote the values attained by $X$ as:

$$(x_1 \mid y) > (x_2 \mid y) > \ldots > (x_{_{2^{2n}}} \mid y).$$

We proved earlier, (see (21), in the proof of Proposition 4.5 above), that:

$$P\left(\forall m \le r \le 2n \ ; \ \lim_{N\to\infty} \frac{N_x}{N} = \frac{a(r,m)}{S(r,m)} \right) = 1, \tag{23}$$

where $N_x$ is the number of muscle-fibers with an activity-level of $x$, $x \in \{x_1, x_2, \dots, x_{2^{2n}}\}$. In particular, substituting $x = (x_t \mid y)$ in (23), we get that the average number $\frac{N_{(x_t|y)}}{N}$ of muscle-fibers, with activity-level of $(x_t \mid y)$, converges almost surely to:

$$p(x_t \mid y) := \frac{a\Big(\ r(x_t \mid y)\ ,\ m(x_t \mid y)\ \Big)}{S\Big(\ r(x_t \mid y)\ ,\ m(x_t \mid y)\ \Big)}, \tag{24}$$

where $r(x_t \mid y)$ denotes the total number of connections and $m(x_t \mid y)$ denotes the number of M-team connections, at a muscle-fiber with an activity-level of $(x_t \mid y)$.

Thus, given $Y = y$ :

$$P\left(\forall t; 1 \le t \le 2^{2n}; \quad \lim_{N\to\infty} \frac{N_{(x_t|y)}}{N} = p(x_t \mid y) \right) = 1, \tag{25}$$

where $N_{(x_t|y)}$ is the number of muscle-fibers with activity-level of $(x_t \mid y)$.

Stated in words: Given $Y = y$, and looking at the full sequence:

$$X_1^N \ge X_2^N \ge \ldots \ge X_N^N,$$

then roughly: The first (i.e., largest) $p(x_1 \mid y)N$ variables, equal $(x_1 \mid y)$, the following $p(x_2 \mid y)N$ variables, equal $(x_2 \mid y)$, and so on.

For all $t$, $(t = 1, 2, \ldots 2^{2n})$, denote:

$$u_y(t) = \sum_{l=1}^{t} p(x_l \mid y).$$

Note that:

$$0 = u_y(0) < u_y(1) < \ldots < u_y(2^{2n}) = 1.$$

For each $0 < s \leq 1$, denote as $t_s$, the value satisfying:

$$u_y(t_s - 1) < s \leq u_y(t_s).$$

Clearly, $t_s$ is uniquely determined by $s$.

Substituting $t = t_s$ in Equation (25), we get that:

$$P\left(\forall t_s; 1 \leq t_s \leq 2^{2n}; \quad \lim_{N\to\infty} \frac{N_{(x_{t_s}|y)}}{N} = p(x_{t_s} \mid y)\right) = 1,$$

Thus, for all $\delta > 0$, there exists $N_\delta$,
such that for all rational numbers $s = \frac{i}{N} \in \Big(u_y(t_s - 1) + \delta, u_y(t_s) - \delta\Big) \subset [0, 1]$ :

$$(X_i^N \mid Y = y) = (x_{t_s} \mid y), \tag{26}$$

for all $N > N_\delta$.

Note that $(X_i^N = x \mid Y = y)$ corresponds to a specific subset of MNs, satisfying:

$$x = \sum_v y_{k_v}.$$

Denote as $q(x \mid y)$, the fraction of M-team connections at a muscle-fiber with activity-level of $x$, given $Y = y$, and recall that $q_i^N$ is the fraction of M-team connections at muscle-fiber $i$.

Thus, if:

$$(X_i^N \mid Y = y) = (x_{t_s} \mid y),$$

then:

$$(q_i^N \mid Y = y) = q(x_{t_s} \mid y), \tag{27}$$

and so:

$$\rho(q_i^N \mid Y = y) = \rho(q(x_{t_s} \mid y)), \tag{28}$$

where $\rho$ is the prior winning function. Recall that by definition, $(P_i^N \mid Y = y) = \rho(q_i^N \mid Y = y)$, thus, by Equation (26), we get that for all $\delta > 0$, there exists $N_\delta$, such that for all rational numbers $s = \frac{i}{N} \in \Big(u_y(t_s - 1) + \delta, u_y(t_s) - \delta\Big) \subset [0, 1]$ :

$$(P_i^N \mid Y = y) = \rho(q(x_{t_s} \mid y)), \quad \text{for all } N > N_\delta. \tag{29}$$

Recall that:

$$u_y(t) = \sum_{l=1}^{t} p(x_l \mid y).$$

Now, by definition:

$$p(x_l \mid y) = \frac{a(r,m)}{S(r,m)},$$

where: $m = m(x_l \mid y)$, and $r = r(x_l \mid y)$. Thus $p(x_l \mid y)$ is determined by $m, r$, Because $0 \leq m \leq r \leq 2n$, then: $p(x_l \mid Y)$ attains a finite number of values.

Thus also $u_y(t)$ attains a finite number of values, and this finite set of values, is the *same* for all $y$, and all $t$. Denote these values as:

$$u_1, u_2, \ldots u_K.$$

Then by Equation (29), for all $s = \frac{i}{N} \in [0,1]$, satisfying that: $\min_{u_l} \left\{ \ |s - u_l| \ \right\} > \delta$ :

$$(P_i^N \mid Y = y) = \rho(q(x_{t_s} \mid y)), \tag{30}$$

for all $N > N_\delta$.

Thus:

$$\lim_{N\to\infty} E(P_i^N) = \int_y \rho(q(x_{t_s} \mid y)) f_Y(y)\, dy, \tag{31}$$

where $f_Y$ is the density function of $Y$, and the convergence is almost uniformly. Denote:

$$p(s) = \int_y \rho(q(x_{t_s} \mid y)) f_Y(y)\, dy, \tag{32}$$

thus we proved that $EP_i^N$ converges almost uniformly to $p(s)$.

The fact that the integral defining $p(s)$ is well defined, is established by the following proposition.

**Proposition 4.7.** *The function* $\rho(q(x_{t_s} \mid y)) f_Y(y)$ *is integrable.*

*Proof.* Consider $q(x_{t_s} \mid y)$ as a function of $y$. We claim that the set of discontinuity-points of $q(x_{t_s} \mid y)$, has a measure of 0, which implies the integrability of $q(x_{t_s} \mid y)$ :

Given $Y = y = (y_1, y_2, \ldots, y_{2n})$, with probability 1, the results:

$$(x_1 \mid y), (x_2 \mid y), \ldots, (x_{2^{2n}} \mid y),$$

are all different from each other, and we have:

$$(x_1 \mid y) > (x_2 \mid y) > \ldots > (x_{2^{2n}} \mid y). \tag{33}$$

Recall that each result $(x_\ell \mid y)$, corresponds to a specific subset of MNs. Hence the order of the different results, appearing in (33), is actually an order of the different subsets of MNs. For each $y$, satisfying (33), there is a neighborhood $B_y$, such that for all $y' \in B_y$, the change in the possible results $x$, is so small, so that the order of the different subsets is unchanged, and thus $(q \mid y)$ is

also unchanged. In other words, with probability 1, inside each $B_y$, $q$ is constant and therefore continuous. Thus, with probability 1, $q(x_{t_s}|y)$ is continuous in $y$. Thus, the set of discontinuity points of $q(x_{t_s}|y)$, has a measure of 0, and so $q(x_{t_s}|y)$ is integrable. Since $\rho$ and $f_Y$ are continuous, $\rho(q(x_{t_s}|y))f_Y(y)$ is also integrable. □

We now return to the proof of Proposition 4.6, and prove that the set of discontinuity-points of $p$ (as a function of $s$), has a measure of 0, where:

$$p(s) = \int_y \rho(q(x_{t_s} \mid y))f_Y(y)\,dy.$$

For all $s \notin \{u_1, u_2, \ldots, u_K\}$, denote:

$$c_s = \min_{u_\ell} \Big\{ |s - u_\ell| \Big\}.$$

Then for all $2\delta < c_s$, and for any rational number $v_1$, $v_1 \in (s-\delta, s+\delta)$, and for all $Y = y$ :

$$u_y(t_s - 1) + \delta < v_1 < u_y(t_s) - \delta.$$

Thus, by (26), for $\delta < c_s$, and for all $y$ :

$$(x_{t_{v_1}} \mid y) = (x_{t_s} \mid y).$$

Thus for any infinite sequence $v_m \to s$ :

$$\lim_{m\to\infty} p(v_m) = \lim_{m\to\infty} \int_y \rho(q(x_{t_{v_m}} \mid y))f_Y(y)\,dy = \int_y \rho(q(x_{t_s} \mid y))f_Y(y)\,dy = p(s).$$

Thus we proved that the $p$, is continuous except at $u_1, u_2, \ldots, u_K$. □

## 4.4. **DERIVATION OF THE SIZE PRINCIPLE.**

In the previous sections, we have defined the parameters (i.e., initial conditions) of the game: The activity levels $Y_1, \ldots, Y_{2n}$ of the MNs, the activity levels: $X_1^N \geq X_2^N \geq \ldots \geq X_N^N$ of the muscle-fibers, and the prior winning probability $P_i^N$ of M-team at muscle-fiber $i$, $1 \leq i \leq N$. We also introduced the first two rules of the game, which the initial conditions must obey.

The third rule of the game defines the dynamics of the game, which describes the change in the winning probabilities of the teams as a result of resource limitation:

$$P(i|W_{i-1}^N) = P_i^N - \mu(N, P_i^N)W_{i-1}^N, \tag{34}$$

where $P(i|W_{i-1}^N)$ is the (actual) winning probability of M-team at muscle-fiber $i$, given that so far it won $W_{i-1}^N$ more competitions than L-team. $P_i^N$ is the prior winning probability at muscle-fiber $i$ and $\mu$ is an adjustment function.

Before proving our main result - the emergence of the size principle - we first need to prove the following two Lemmas.

**Lemma 4.8.** *For all* $1 \leq i \leq N$ :

$$E\left(W_i^N \left| P^N\right.\right) = \left(1 - 2\mu\left(N, P_i^N\right)\right) E\left(W_{i-1}^N \left| P^N\right.\right) + 2P_i^N - 1,$$

where $P^N = \left(P_1^N, P_2^N, \ldots, P_N^N\right)$, is the vector of the prior winning probabilities of M-team.

*Proof.*

$$E\left(W_i^N \left| P^N, W_{i-1}^N\right.\right) = \left(W_{i-1}^N + 1\right)\left(P(i|W_{i-1}^N) \left| P^N\right.\right) + \left(W_{i-1}^N - 1\right)\Big(1 - \left(P(i|W_{i-1}^N) \left| P^N\right.\right)\Big),$$

hence:

$$E\left(W_i^N \mid P^N, W_{i-1}^N\right) = W_{i-1}^N + 2\left(P(i|W_{i-1}^N) \left| P^N\right.\right) - 1. \tag{35}$$

Substituting (34) in (35), we get that:

$$E\left(W_i^N \mid P^N, W_{i-1}^N\right) = \left(1 - 2\mu\left(N, P_i^N\right)\right)W_{i-1}^N + 2P_i^N - 1. \tag{36}$$

Hence:

$$E\left(W_i^N \mid P^N\right) = EE\left(W_i^N \mid P^N, W_{i-1}^N\right) = \left(1 - 2\mu\left(N, P_i^N\right)\right) E\left(W_{i-1}^N \left| P^N\right.\right) + 2P_i^N - 1,$$

where the outer expectation function (in the double expectation $EE$ above), is the expectation over $W_{i-1}^N$. □

**Lemma 4.9.** *For all* $1 \leq i \leq N$ :

$$E\left(W_i^N \left| P^N\right.\right) = \sum_{\ell=1}^{i}\left(2P_\ell^N - 1\right)\prod_{j=\ell+1}^{i}\left(1 - 2\mu\left(N, P_j^N\right)\right). \tag{37}$$

*Proof.* By induction on $i$. For $i = 1$ : Substituting $i = 1$ in Lemma 4.8, we get that since $W_0^N \equiv 0$ :

$$E\left(W_1^N \left| P^N\right.\right) = 2P_1^N - 1,$$

which is what we get by substituting $i = 1$ in (37). Assume (37) is true for $i - 1$, and we prove it for $i$ :

By Lemma 4.8:

$$E\left(W_i^N \left| P^N\right.\right) = \left(1 - 2\mu\left(N, P_i^N\right)\right) E(W_{i-1}^N \mid P^N) + 2P_i^N - 1.$$

Applying the induction hypothesis to $i-1$, we get:

$$E\left(W_i^N \left| P^N\right.\right) = \left(1-2\mu\left(N,P_i^N\right)\right)\sum_{\ell=1}^{i-1}\left(2P_\ell^N-1\right)\prod_{j=\ell+1}^{i-1}\left(1-2\mu\left(N,P_j^N\right)\right)+2P_i^N-1=$$

$$\sum_{\ell=1}^{i-1}\left(2P_\ell^N-1\right)\prod_{j=\ell+1}^{i}\left(1-2\mu\left(N,P_j^N\right)\right)+2P_i^N-1=$$

$$\sum_{\ell=1}^{i}\left(2P_\ell^N-1\right)\prod_{j=\ell+1}^{i}\left(1-2\mu\left(N,P_j^N\right)\right).$$

□

We are now ready to prove the emergence of the size principle from the game MNs play. The intuition behind this proof appears in **Biological consequences of the model**, above.

**Theorem 4.10** (**The size principle**). *Starting with random initial connectivity, for all $N$, and for any adjustment function $\mu=\mu\left(N,P_i^N\right)$ :*

$$E\left(W_N^N\right)<0.$$

*Proof.* By Lemma 4.9:

$$E\left(W_N^N \left| P^N\right.\right) = \sum_{i=1}^{N}\left(2P_i^N-1\right)\prod_{u=i+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right). \tag{38}$$

Recall that we have partitioned the population of muscle-fibers into $2n+1$ sets $A_r$ according to the total number of MNs ($r=0,1,\ldots,2n$) connecting to the muscle-fiber.

Thus:

$$E\left(W_N^N \left| P^N\right.\right) = \sum_{r=0}^{2n}\sum_{i\in A_r}\left(2P_i^N-1\right)\prod_{u=i+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right). \tag{39}$$

For each $r=0,1,\ldots,2n$, we further divided the muscle-fibers in $A_r$ into subsets $A_{r,m}$, $m=0,1,\ldots,r$, according to the number of M-team connections at the muscle-fiber. Namely:

$$A_{r,m}=\left\{i_\ell\ ;\ \sum_{k\in\text{M-team}} I_{i_\ell,k}^N=m\ ,\ \sum_{k\in\text{L-team}} I_{i_\ell,k}^N=r-m\right\},$$

where $I_{i,k}^N$ was defined earlier, as the indicator variable receiving 1 *iff* MN $k$ connects to muscle-fiber $i$. Note that:

$$i\in A_{r,m}\quad\Rightarrow\quad q_i^N=\frac{m}{r},$$

where $q_i^N$ was defined earlier as the fraction of M-team connections at muscle-fiber $i$. Recall that the prior winning probability $P_i^N$ of M-team at muscle-fiber $i$, was defined earlier as $P_i^N=\rho\left(q_i^N\right)$,

where $\rho$ is a democratic (i.e., monotonic and neutral) winning function. Hence:

$$i \in A_{r,m} \quad \Rightarrow \quad P_i^N = \rho\left(\frac{m}{r}\right). \tag{40}$$

Note that if $r$ is even then substituting $m = \frac{r}{2}$ in (40), we get that:

$$i \in A_{r,\frac{r}{2}} \quad \Rightarrow \quad P_i^N = \rho\left(\frac{1}{2}\right) = \frac{1}{2} \Rightarrow \quad 2P_i^N - 1 = 0.$$

Thus, by (39), the cases where $m = \frac{r}{2}$ add 0 to $E\left(W_N^N \left| P^N\right.\right)$, and will therefore be discarded from the sum in the following Equation.

Denote $\lceil x \rceil$ = the smallest integer that is larger or equals $x$. Then, continuing Equation (39):

$$E\left(W_N^N \left| P^N\right.\right) = \sum_{r=0}^{2n} \sum_{i \in A_r} \left(2P_i^N - 1\right) \prod_{u=i+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) =$$

$$\sum_{r=0}^{2n} \sum_{m=\lceil \frac{r+1}{2} \rceil}^{r} E\Big\{ \sum_{i_\ell \in A_{r,m}} \left(2P_{i_\ell}^N - 1\right) \prod_{u=i_\ell+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) +$$

$$\sum_{i'_\ell \in A_{r,r-m}} \left(2P_{i'_\ell}^N - 1\right) \prod_{u=i'_\ell+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) \Big\} =$$

$$\sum_{r=0}^{2n} \sum_{m=\lceil \frac{r+1}{2} \rceil}^{r} E\Big\{ \sum_{i_\ell \in A_{r,m}} \left(2\rho\left(\frac{m}{r}\right) - 1\right) \prod_{u=i_\ell+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) +$$

$$\sum_{i'_\ell \in A_{r,r-m}} \left(2\rho\left(1 - \frac{m}{r}\right) - 1\right) \prod_{u=i'_\ell+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) \Big\}.$$

Since $\rho\left(1 - \frac{m}{r}\right) = 1 - \rho\left(\frac{m}{r}\right)$ then:

$$E\left(W_N^N \left| P^N\right.\right) = \sum_{r=0}^{2n} \sum_{m=\lceil \frac{r+1}{2} \rceil}^{r} \left(2\rho\left(\frac{m}{r}\right) - 1\right)$$

$$E\Big\{ \sum_{i_\ell \in A_{r,m}} \prod_{u=i_\ell+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) - \sum_{i'_\ell \in A_{r,r-m}} \prod_{u=i'_\ell+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) \Big\}. \tag{41}$$

Recall that $N_{r,m} = |A_{r,m}|$, hence continuing Equation (41):

$$E\left(W_N^N \left| P^N\right.\right) = \sum_{r=0}^{2n} \sum_{m=\lceil \frac{r+1}{2} \rceil}^{r} \left(2\rho\left(\frac{m}{r}\right) - 1\right)$$

$$E\Big\{ \sum_{\ell=1}^{N_{r,m}} \prod_{u=i_\ell+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) - \sum_{\ell=1}^{N_{r,r-m}} \prod_{u=i'_\ell+1}^{N} \left(1 - 2\mu\left(N, P_u^N\right)\right) \Big\}. \tag{42}$$

Note that all $m$s appearing in the sum above, satisfy that: $\frac{m}{r} > \frac{1}{2}$, and so $\left(2\rho\left(\frac{m}{r}\right) - 1\right) > 0$. Hence, in order to prove that $E\left(W_N^N \left| P^N\right.\right)$ is negative, it is sufficient to prove that for each $m, r$ :

$$E\Big\{\sum_{\ell=1}^{N_{r,m}}\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\sum_{\ell=1}^{N_{r,r-m}}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}\leq 0,$$

and that there is at least one pair $m,r$ for which the above is strictly negative.

We compute the above sum, by considering all the events:

$$\left(N_{r,m}=s\ ,N_{r,r-m}=t\right),\quad 1\leq s,t\leq N.$$

$$E\Big\{\sum_{\ell=1}^{N_{r,m}}\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\sum_{\ell=1}^{N_{r,r-m}}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}=$$

$$\sum_{s,t}^{N}E\Big\{\sum_{\ell=1}^{s}\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\sum_{\ell=1}^{t}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}P\left(N_{r,m}=s\ ,N_{r,r-m}=t\right)=$$

$$\sum_{s<t}^{N}E\Big\{\sum_{\ell=1}^{s}\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\sum_{\ell=1}^{t}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}P\left(N_{r,m}=s\ ,N_{r,r-m}=t\right)$$

$$+$$

$$\sum_{s<t}^{N}E\Big\{\sum_{\ell=1}^{t}\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\sum_{\ell=1}^{s}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}P\left(N_{r,m}=t\ ,N_{r,r-m}=s\right)$$

$$+$$

(43)

$$\sum_{s=t}^{N}E\Big\{\sum_{\ell=1}^{s}\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\sum_{\ell=1}^{s}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}P\left(N_{r,m}=s\ ,N_{r,r-m}=s\right).$$

For all $0\leq m\leq r\leq 2n$, and for any muscle-fiber $a$ chosen at random ($a\in\{1,2,\ldots,N\}$) :

$$P(a\in A_{r,m})=P(a\in A_{r,m}\mid a\in A_r)P(a\in A_r)=$$

$$\frac{1}{2^r}\binom{r}{m}P(a\in A_r)=\frac{1}{2^r}\binom{r}{r-m}P(a\in A_r)=$$

$$P(a\in A_{r,r-m}\mid a\in A_r)P(a\in A_r)=P(a\in A_{r,r-m}).$$

Denote: $\mathsf{p}=P(a\in A_{r,m})=P(a\in A_{r,r-m})$,

and recall that the connections to the muscle-fibers are independent, thus: $N_{r,m}$ and $N_{r,r-m}$ are both distributed $\mathbf{B}(N,\mathsf{p})$, and so, for all $0\leq t\leq N$ :

$$P\left(N_{r,m}=t\right)=P\left(N_{r,r-m}=t\right).$$

In addition, for all $1\leq t\leq N$, the random variables:

$$\left(N_{r,m}\,|N_{r,r-m}=t\right),\ \text{and}\ \left(N_{r,r-m}\,|N_{r,m}=t\right),$$

are both distributed $\mathbf{B}(N-t,\mathsf{p})$. Hence:

$$P\left(N_{r,m}=s,N_{r,r-m}=t\right)=P\left(N_{r,m}=s\,|N_{r,r-m}=t\,\right)P\left(N_{r,r-m}=t\right)=$$

$$P\left(N_{r,r-m}=s\,|N_{r,m}=t\,\right)P\left(N_{r,m}=t\right)=$$

$$P\left(N_{r,m}=t,N_{r,r-m}=s\right). \tag{44}$$

Thus substituting this in Equation (43), and exchanging the order of the sums, we get that:

$$E\Big\{\sum_{\ell=1}^{N_{r,m}}\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\sum_{\ell=1}^{N_{r,r-m}}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}=$$

$$\sum_{s<t}^{N}\Big[\,E\sum_{\ell=1}^{s}\Big\{\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}+$$

$$E\sum_{\ell=1}^{t}\Big\{\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}\Big]P\left(N_{r,m}=s\,,N_{r,r-m}=t\right)$$

$$+$$

$$\sum_{s=1}^{N}E\sum_{\ell=1}^{s}\Big\{\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}\Big]P\left(N_{r,m}=s\,,N_{r,r-m}=s\right).$$

In order to prove that the above is strictly negative, it is sufficient to prove that for *each* $v=1,2,\ldots$ :

$$E\sum_{\ell=1}^{v}\Big\{\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}<0.$$

We consider all the possible events: $\left(i_\ell=i\ ,\ i'_\ell=j\right)$. Hence:

$$E\sum_{\ell=1}^{v}\Big\{\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}$$

$$=$$

$$\sum_{\ell=1}^{v}\sum_{i,j;i\neq j}^{N}\Big\{\prod_{u=i+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=j+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\ \Big|\ i_\ell=i,i'_\ell=j\Big\}P\left(i_\ell=i\ ,\ i'_\ell=j\right)$$

$$=$$

$$\sum_{\ell=1}^{v}\sum_{i<j}^{N}\Big\{\prod_{u=i+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=j+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\ \Big|\ i_\ell=i,i'_\ell=j\Big\}P\left(i_\ell=i\ ,\ i'_\ell=j\right)$$

$$+$$

$$\sum_{\ell=1}^{v}\sum_{i<j}^{N}\Big\{\prod_{u=j+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=i+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\ \Big|\ i_\ell=i,i'_\ell=j\Big\}P\left(\ i_\ell=j\ ,\ i'_\ell=i\ \right).$$

Looking at the products in the inner brackets of each of the two sums above, we can factor out a common divider, so that the above equals:

$$\sum_{\ell=1}^{v}\sum_{i<j}^{N}\Big\{\prod_{u=j+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big(\prod_{u=i+1}^{j}\left(1-2\mu\left(N,P_u^N\right)\right)-1\Big)\;\Big|\;i_\ell=i\;,\;i'_\ell=j\;\Big\}P\left(\;i_\ell=i\;,\;i'_\ell=j\;\right)$$

+

(45)

$$\sum_{\ell=1}^{v}\sum_{i<j}^{N}\Big\{\prod_{u=j+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big(1-\prod_{u=i+1}^{j}\left(1-2\mu\left(N,P_u^N\right)\right)\Big)\;\Big|\;i_\ell=j\;,\;i'_\ell=i\Big\}P\left(\;i_\ell=j\;,\;i'_\ell=i\;\right).$$

Denote:

$$b_{i,j}(r,m,\ell)=\Big\{\prod_{u=j+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big(\prod_{u=i+1}^{j}\left(1-2\mu\left(N,P_u^N\right)\right)-1\Big)\;\Big|\;i_\ell=i\;,\;i'_\ell=j\;\Big\}.$$

Recall that $i_\ell, i'_\ell$ depend on $m, r, \ell$.

Substituting $b_{i,j}(r,m,\ell)$ in each of the two sums in Equation (45), we get that the above equals:

$$\sum_{\ell=1}^{v}\sum_{i<j}^{N}b_{i,j}(r,m,\ell)\Big(\;P\left(\;i_\ell=i\;,\;i'_\ell=j\right)-P\left(\;i_\ell=j\;,\;i'_\ell=i\;\right)\;\Big). \tag{46}$$

We now show that the expression above, is negative when $m$ equals 0 or $r$, and is *strictly* negative otherwise (i.e., for all $0<m<r$).

For all $i,j$ satisfying $P\Big(\{i_\ell,i'_\ell\}=\{i,j\}\Big)>0$, we proved in Lemma 4.3, that:

$$\Big(\;P\left(\;i_\ell=i\;,\;i'_\ell=j\right)-P\left(\;i_\ell=j\;,\;i'_\ell=i\;\right)\;\Big)>0.$$

Note that for each $m,r$ there is at least one pair $i,j$ satisfying:

$$P\Big(\{i_\ell,i'_\ell\}=\{i,j\}\Big)>0.$$

In addition, for all $m,r$ :

$$b_{i,j}(r,m,\ell)=\prod_{u=j+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\{\prod_{u=i+1}^{j}\left(1-2\mu\left(N,P_u^N\right)\right)-1\;\Big\}\leq 0,$$

and for each $o<m<r$, we have: $0<\frac{m}{r}<1$, and so by the strict monotonicity of $\rho$, $(\;i_\ell=i\;,\;i'_\ell=j\;)$ implies that:

$$0<P_i^N=\rho\left(\frac{m}{r}\right)<1,$$

and also:

$$0<P_j^N=1-\rho\left(\frac{m}{r}\right)<1.$$

Since we required that $\mu(\cdot,p)$ is positive inside $(0,1)$, then: $\mu(\cdot,P_i^N),\mu(\cdot,P_j^N)>0$, and so:

$$b_{i,j}(r,m,\ell)<0.$$

Taken together, we proved that for all $m,r$, and for all $i,j$ :

$$b_{i,j}(r,m,\ell)\Big(\ P\left(\ i_\ell=i\ ,\ i'_\ell=j\right)-P\left(\ i_\ell=j\ ,\ i'_\ell=i\ \right)\ \Big)\leq 0, \tag{47}$$

and that for each $0<m<r$, there is at least one pair $i,j$ with a strict inequality in (47). □

We now finished proving that for any adjustment function $\mu=\mu(N,P_i^N)$, less-active MNs win in more competitions.

4.5. **Nontrivial games.** In the previous section, we proved the emergence of the size principle, for any adjustment function $\mu=\mu\left(N,P_i^N\right)$. Specifically, we proved that the expected difference $E(W_N^N)$ in the number of winnings, between the more-active and the less-active MNs, is negative.

Recall that the $N$ stages of the game reflect the $N$ successive times, in which the different competitions ended. Hence: $\frac{1}{N},\frac{2}{N},\ldots,\frac{N-1}{N},1$, denote the relative completion-times of the competitions. Thus doubling the number of stages $N$, is actually refining the time parameter, so that as $N\to\infty$ the time-parameter becomes continuous. Under this interpretation, it is natural to require that the "strength" of the dynamics, which is measured by the total size of adjustments, is *not* changed as a result of this refinement. This is the motivation for defining "nontrivial adjustment function" as adjustment functions that their strength do not depend on $N$.

Biologically, the meaning, of keeping the total size of adjustments unchanged as the number of muscle-fibers change, is that the *total influence* any one competition exerts over all the other competitions, is unchanged as the number of competitions grow. This is the natural assumption, since when a MN wins, this only influences the fraction $\gamma$ of competitions, in which this same MN takes part, and in this work $\gamma$ is fixed, thus in particular it does *not* depend on $N$. The reason for assuming that $\gamma$ is fixed follows from the fact that the number of MNs innervating a muscle, varies between 50 to 200, which could be considered fixed compared to the change in the number of fibers in the muscle, which ranges between few thousands to millions. Thus the *number* of muscle-fibers a MN innervates initially, is roughly proportional to the number of muscle-fibers, and so the *fraction* $\gamma$ of muscle-fibers a MN innervates initially, is fixed as the number of muscle-fibers $N$ change.

Mathematically, the requirement that the total size of adjustments $\mu\left(N,P_i^N\right)N$ will not change as a result of the growth in $N$, means that $\mu$ is of the order of magnitude of $\frac{1}{N}$. Thus, there exist:

$0 < \alpha < \beta$, such that for all $1 \leq i \leq N$ and $P_i^N$ :

$$\frac{\alpha}{N} \leq \mu\left(N, P_i^N\right) \leq \frac{\beta}{N}.$$

Moreover, note that a function $\mu$, which is smaller than the order of magnitude of $\frac{1}{N}$ defines a trivial game in the following sense: If $\mu$ is of smaller magnitude than $\frac{1}{N}$, then because:

$$P(i|W_{i-1}^N) = P_i^N - \mu\left(N, P_i^N\right) W_{i-1}^N,$$

then for all $\delta > 0$, there exists $N_\delta$ s.t. $\forall N > N_\delta$ : $\mu\left(N, P_i^N\right) < \frac{\delta}{N}$, and so:

$$\left|P(i|W_{i-1}^N) - P_i^N\right| = \left|\mu\left(N, P_i^N\right) W_{i-1}^N\right| < \frac{\delta}{N} \cdot N = \delta. \tag{48}$$

This is true for *any* small $\delta$ and it means, that the dynamics becomes weaker and negligible as the number of stages in the game, grow (i.e., $N \to \infty$). For these reasons, we refer to an adjustment function $\mu$ of smaller or larger magnitude than $\frac{1}{N}$, (and the game obtained from it), as *trivial*.

**Definition 4.11.** We say that an adjustment function $\mu = \mu\left(N, P_i^N\right)$, is *nontrivial*, if there is a function $f : [0,1] \to \mathbf{R}^+$, satisfying that for all $0 < p < 1$ : $f(p) > 0$, and: $f(0) = f(1) = 0$, such that:

$$\mu\left(N, P_i^N\right) = \frac{f\left(P_i^N\right)}{N}.$$

We call the game, defined by a nontrivial adjustment function, *a nontrivial game.*

For the sake of convenience, we assume that $f$ is continuous. Note that indeed, our definition of nontrivial adjustment function implies that the adjustment sizes are of the order of magnitude of $\frac{1}{N}$: If muscle-fiber $i$ has $m$ M-team connections and $r - m$ L-team connections, then:

$$P_i^N = \rho\left(\frac{m}{r}\right).$$

As $0 \leq m \leq r \leq 2n$, then $P_i^N$ attains a finite number of values, and so also $f\left(P_i^N\right)$ attains a finite number of values. For $0 < P_i^N < 1$, denote the minimal and maximal values that $f\left(P_i^N\right)$ attains, as $\alpha$ and $\beta$ respectively. Note that $\alpha, \beta > 0$, since $f$ is strictly positive in $(0,1)$.

Thus if $\mu$ is a nontrivial adjustment function, then for all $0 < P_i^N < 1$ :

$$\frac{\alpha}{N} \leq \mu\left(N, P_i^N\right) \leq \frac{\beta}{N}, \ \forall i, N, P_i^N.$$

In the previous section, we proved that, assuming random initial connectivity, any adjustment function, (trivial, or not) yields the size principle, namely: $E(W_N^N) < 0$ (see Theorem 4.10). We now wish to strengthen this result for nontrivial games.

**Theorem 4.12.** *Starting with random initial connectivity, For any nontrivial adjustment function* $\mu$, *there exists* $D > 0$, *such that for all* $N$ :

$$\frac{1}{N}E\left(W_N^N\right) \leq -D.$$

*Proof.* Using Equation (42), in the proof of Theorem 4.10, we have that:

$$\frac{1}{N}E\left(W_N^N\right) \leq$$

(49)

$$\frac{1}{N}\sum_{r=0}^{2n}\sum_{m=\lceil\frac{r+1}{2}\rceil}^{r}\left(2\rho\left(\frac{m}{r}\right)-1\right)E\Big\{\sum_{\ell=1}^{N_{r,m}}\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\sum_{\ell=1}^{N_{r,r-m}}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}.$$

By Lemma 4.4, for all $\delta > 0$, there exists $N_\delta$, such that for all $N > N_\delta$, with probability 1:

$$N_{r,m} \leq a(r,m)N + \delta N, \tag{50}$$

and also

$$N_{r,r-m} \leq a(r,m)N + \delta N,$$

where $a(r,m) = \frac{1}{2^r}\binom{r}{m}\binom{2n}{r}\gamma^r(1-\gamma)^{2n-r}$. Hence, since each item in the sum appearing in (49) above, is between $-1$ and 1, then the $2\delta N$ "excess" items, add at most $2\delta$ to the *averaged* sum $\frac{1}{N}E(W_N^N)$. Thus:

$$\frac{1}{N}E\left(W_N^N\right) =$$

$$\frac{1}{N}\sum_{r=0}^{2n}\sum_{m=\lceil\frac{r+1}{2}\rceil}^{r}\left(2\rho\left(\frac{m}{r}\right)-1\right)$$

$$\sum_{\ell=1}^{\lfloor a(r,m)N\rfloor}E\Big\{\prod_{u=i_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)-\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big\}+2\delta,$$

where $\lfloor a(r,m)N\rfloor$ is the largest natural number that is smaller or equals $a(r,m)N$.

Denote:

$$\mathbf{C} = \left\{\omega \text{ ; such that for all } m \leq r \leq 2n \text{ ; } \frac{i'_\ell - i_\ell}{N} \geq S\right\}.$$

According to Proposition 4.5, for all $\epsilon > 0$, there exists $N_\epsilon > 0$, such that $\forall N > N_\epsilon$ :

$$P(\mathbf{C}) \geq 1 - \epsilon.$$

Denote:

$$N_{\delta,\epsilon} = \max\{N_\epsilon, N_\delta\},$$

where $N_\delta$ appears in (50).

By the definition of $\mathbf{C}$, we get that if $\omega \in \mathbf{C}$ then $\forall N > N_{\delta,\epsilon}$ : $i'_\ell - i_\ell > 0$. We can therefore factor out a common divider, so that:

$$\frac{1}{N}E\left(W_N^N \,|\mathbf{C}\right) \leq$$

$$\frac{1}{N}\sum_{r=0}^{2n}\sum_{m=\lceil\frac{r+1}{2}\rceil}^{r}\left(2\rho\left(\frac{m}{r}\right)-1\right)$$

$$E\Big\{\sum_{\ell=1}^{\lfloor a(r,m)N\rfloor}\prod_{u=i'_\ell+1}^{N}\left(1-2\mu\left(N,P_u^N\right)\right)\Big[\prod_{u=i_\ell+1}^{i'_\ell}\left(1-2\mu\left(N,P_u^N\right)\right)-1\Big]\;\Big|\;\mathbf{C}\Big\}+2\delta. \tag{51}$$

Note that each item in the sum above is negative, since $\left(2\rho\left(\frac{m}{r}\right)-1\right) > 0$, for all $m > \frac{r}{2}$, $0 < \mu\left(N,P_i^N\right) < \frac{1}{2}$, and:

$$\left[\prod_{j=i_\ell}^{i'_\ell}\left(1-2\mu\left(N,P_j^N\right)\right)-1\right] < 0.$$

Hence, since $\mu$ is a nontrivial adjustment function, there exists $0 < \alpha, \beta$, such that for all $0 < P_j^N < 1$ :

$$\frac{\alpha}{N} \leq \mu\left(N,P_j^N\right) \leq \frac{\beta}{N}.$$

Thus:

$$\frac{1}{N}E\left(W_N^N \,|\mathbf{C}\right) \leq \frac{1}{N}\sum_{r=0}^{2n}\sum_{m=\lceil\frac{r+1}{2}\rceil}^{r}\left(2\rho\left(\frac{m}{r}\right)-1\right)$$

$$E\Big\{\sum_{\ell=1}^{\lfloor a(r,m)N\rfloor}\prod_{u=i'_\ell+1}^{N}\left(1-\frac{2\beta}{N}\right)\Big[\prod_{u=i_\ell+1}^{i'_\ell}\left(1-\frac{2\alpha}{N}\right)-1\Big]\;\Big|\;\mathbf{C}\;\Big\}+2\delta. \tag{52}$$

Thus:

$$\frac{1}{N}E(W_N^N \,|\mathbf{C}) \leq \frac{1}{N}\sum_{r=0}^{2n}\sum_{m=\lceil\frac{r+1}{2}\rceil}^{r}\left(2\rho\left(\frac{m}{r}\right)-1\right)$$

$$E\Big\{\sum_{\ell=1}^{\lfloor a(r,m)N\rfloor}\left(1-\frac{2\beta}{N}\right)^{N-i'_\ell}\left[\left(1-\frac{2\alpha}{N}\right)^{i'_\ell-i_\ell}-1\right]\;\Big|\;\mathbf{C}\;\Big\}+2\delta. \tag{53}$$

Thus using Proposition 4.5 again, we can exchange $i'_\ell - i_\ell$ appearing in (53), with $SN$ and so, for all $N > N_{\delta,\epsilon}$ :

$$\frac{1}{N}E\left(W_N^N \,|\mathbf{C}\right) \leq \frac{1}{N}\sum_{r=0}^{2n}\sum_{m=\lceil\frac{r+1}{2}\rceil}^{r}\left(2\rho\left(\frac{m}{r}\right)-1\right)\sum_{\ell=1}^{\lfloor a(r,m)N\rfloor}\left(1-\frac{2\beta}{N}\right)^{N}\Big[\left(1-\frac{2\alpha}{N}\right)^{SN}-1\Big]+2\delta \leq$$

$$\frac{1}{N}\sum_{r=0}^{2n}\sum_{m=\lceil\frac{r+1}{2}\rceil}^{r}\left(2\rho\left(\frac{m}{r}\right)-1\right)\sum_{\ell=1}^{\lfloor a(r,m)N\rfloor}e^{-2\beta}\left(e^{-2\alpha S}-1\right)+3\delta.$$

It is straightforward to show that for all $m > \frac{r}{2}$ :

$$\frac{m}{r} \geq \frac{n+1}{2n},$$

thus as $\rho$ is increasing, and $\left(e^{-2\alpha S}-1\right) < 0$, then for all $N > N_{\delta,\epsilon}$ :

$$\frac{1}{N} E\left(W_N^N \,|\mathbf{C}\right) \leq \left(2\rho\left(\frac{n+1}{2n}\right)-1\right) e^{-2\beta}\left(e^{-2\alpha S}-1\right) + 3\delta.$$

Denote:

$$-2D_1 = \left(2\rho\left(\frac{n+1}{2n}\right)-1\right) e^{-2\beta}\left(e^{-2\alpha S}-1\right).$$

Note that $-2D_1 < 0$ since $\rho\left(\frac{n+1}{2n}\right) > \rho\left(\frac{1}{2}\right) = \frac{1}{2}$, and $\left(e^{-2\alpha S}-1\right) < 0$.

Hence, for all $\delta < \frac{1}{3}D_1$ :

$$\frac{1}{N} E\left(W_N^N \,|\mathbf{C}\right) < -2D_1 + 3\delta < 0. \tag{54}$$

Now, denote as $\bar{\mathbf{C}}$, the complimentary set of $\mathbf{C}$. Then for $\delta < \frac{1}{3}D_1$ :

$$\begin{aligned}
&\frac{1}{N} E(W_N^N) = \frac{1}{N} E\left(W_N^N|\mathbf{C}\right) P(\mathbf{C}) + \frac{1}{N} E\left(W_N^N|\bar{\mathbf{C}}\right) P(\bar{\mathbf{C}}) \leq \\
&\frac{1}{N} E\left(W_N^N|\mathbf{C}\right)(1-\epsilon) + \epsilon \leq \\
&\frac{1}{N} E\left(W_N^N|\mathbf{C}\right) + \epsilon\Big(1 - \frac{1}{N} E\left(W_N^N|\mathbf{C}\right)\Big) \leq \frac{1}{N} E\left(W_N^N|\mathbf{C}\right) + 2\epsilon. \qquad (55)
\end{aligned}$$

Note that in the above, we used the fact that $\frac{1}{N} E\left(W_N^N|\mathbf{C}\right) < 0$, and $\frac{1}{N} E\left(W_N^N|\bar{\mathbf{C}}\right) \leq 1$.

Taken together, by (54) and (55), we proved that $\forall N > N_{\delta,\epsilon}$ :

$$\frac{1}{N} E(W_N^N) \leq -2D_1 + 3\delta + 2\epsilon,$$

Thus for all $\delta, \epsilon > 0$, such that: $3\delta + 2\epsilon < D_1$, we get that:

$$\frac{1}{N} E(W_N^N) \leq -D_1, \quad \forall N > N_{\delta,\epsilon}.$$

Now, by Theorem 4.10, we have that for all $N \leq N_{\delta,\epsilon}$ :

$$E\left(W_N^N\right) < 0.$$

Denote:

$$-D_2 = \max_{1 \leq N \leq N_{\delta,\epsilon}} \left\{\frac{1}{N} E W_N^N\right\},$$

and denote:

$$-D = \max\{-D_1, -D_2\},$$

then for *all* $N$ :

$$\frac{1}{N} E W_N^N \leq -D.$$

□

4.6. **Estimating the magnitude of the size principle.** Recall that the random variable $W_N^N$ denotes the difference between the number of winnings of M-team and that of L-team, when the game ends. Note that $-1 \leq \frac{1}{N} W_N^N \leq 1$, hence, we use $\frac{1}{N} W_N^N$ to measure the magnitude of the size principle in different muscles. The more negative it is, the more the size principle is expressed in the muscle.

In Theorem 4.10, we proved that under normal conditions $\frac{1}{N} E W_N^N$ is always negative, but the *value* of $\frac{1}{N} E W_N^N$, depends on the parameters of the game; $\rho$ and $\mu$.

We wish to arrive at an estimation for $\frac{1}{N} W_N^N$. For this, we simplify and average over the prior winning probabilities. Namely, instead of starting with the initial prior winning probabilities $P^N = (P_1^N, P_2^N, \dots, P_N^N)$, we start with their expectations: $EP^N = (EP_1^N, EP_2^N, \dots, EP_N^N)$.

We confine our interest to nontrivial games, and so $\mu$ is of the form:

$$\mu\left(N, EP_i^N\right) = \frac{f\left(EP_i^N\right)}{N},$$

where $f : [0,1] \to \mathbf{R}^+$, is a continuous function satisfying that for all $0 < p < 1$ : $f(p) > 0$, and: $f(0) = f(1) = 0$.

In Proposition 4.6, we proved that $EP^N$ converges almost uniformly to a function $p$, which is continuous except at finitely many points. Figure 9 shows the average of $P^N$ over 500 observations (black line). Thus this graph approximates the limit-function $p$.

**Theorem 4.13.** *If we start with initial conditions of $EP^N$, and with a nontrivial adjustment function, then:*

$$\lim_{N\to\infty} \frac{1}{N} E\left(W_N^N\right) = \int_0^1 (2p(s) - 1) e^{-2f(p(s))(1-s)}\, ds, \tag{56}$$

where $p(s) = \int_y \rho(q(x_{t_s} \mid y)) f_Y(y)\, dy$.

*Proof.* By Equation (38):

$$\frac{1}{N} E\left(W_N^N\right) = \frac{1}{N} E\left[\sum_{i=1}^{N} \left(2P_i^N - 1\right) \prod_{j=i+1}^{N} \left(1 - 2\mu\left(N, P_j^N\right)\right)\right].$$

Since we start with $EP^N$, and: $\mu\left(N, EP_i^N\right) = \frac{f(EP_i^N)}{N}$, then we get:

$$\frac{1}{N} E\left(W_N^N\right) = \frac{1}{N} \sum_{i=1}^{N} \left(2EP_i^N - 1\right) \left(1 - \frac{2f(EP_i^N)}{N}\right)^{N-i} = \frac{1}{N} \sum_{i=1}^{N} \left(2EP_i^N - 1\right) \left(1 - \frac{2f(EP_i^N)}{N}\right)^{N(1-\frac{i}{N})}.$$

By Proposition 4.6, for each rational number $s = \frac{i}{N} \in [0,1]$ :

$$\lim_{N\to\infty} EP_i^N = p(s),$$

where the convergence is almost uniformly. Thus for all $\nu$, there exists $N_\nu$, such that for all $N > N_\nu$ :

$$\left| \frac{1}{N} E\left(W_N^N\right) - \frac{1}{N} \sum_{i=1}^{N} \left(2p\left(\frac{i}{N}\right) - 1\right) \left(1 - \frac{2f(p(\frac{i}{N}))}{N}\right)^{N(1-\frac{i}{N})} \right| < \nu.$$

The function $p$ is continuous except at finitely many points, thus it is integrable, and since $\lim_{N\to\infty} \left(1 - \frac{2f(p(s))}{N}\right)^N = e^{-2f(p(s))}$, then for all $N > N_\nu$ :

$$\left| \frac{1}{N} E\left(W_N^N\right) - \int_0^1 (2p(s) - 1) e^{-2f(p(s))(1-s)}\, ds \right| < 2\nu. \tag{57}$$

This completes the proof of Theorem 4.13. □

We now wish to show that also $\frac{1}{N} W_N^N$ converges to the same limit (as its expectation).

**Corollary 4.14.** *For all* $\epsilon > 0$ :

$$\lim_{N\to\infty} P\left\{ \left| \frac{1}{N} W_N^N - \int_0^1 (2p(s) - 1) e^{-2f(p(s))(1-s)}\, ds \right| < \epsilon \right\} = 1.$$

*Proof.* For each $j = 1, 2, \ldots, N$, let $w_j$ be the random variable, having a value 1, *iff* M-team wins at stage $j$, and $-1$ otherwise. Hence:

$$W_i = \sum_{j=1}^{i} w_j.$$

Note that for each $i$ : $Var(w_i) \leq 2$. Because of the resource limitation (rule 3), each pair $w_i, w_j$ $(i \neq j)$, have a negative correlation. This is proved in Lemma 7.3 (see Appendix 7.6). Hence for all $i, j$, $(i \neq j)$ :

$$Var\left(\frac{1}{N} W_i^N\right) = Var\left(\frac{1}{N} \sum_{j=1}^{i} w_j\right) \leq$$

$$\frac{1}{N^2} \sum_{j=1}^{i} Var(w_j) \leq \frac{2i}{N^2} \leq \frac{2}{N}. \tag{58}$$

Thus, using Chebychev's inequality, we get that for all $\epsilon > 0$ :

$$P\left\{ \left| \frac{1}{N} W_N^N - \frac{1}{N} E W_N^N \right| < \frac{\epsilon}{2} \right\} \geq 1 - \frac{8}{N\epsilon^2}, \tag{59}$$

and so for all $\epsilon > 0$ :

$$\lim_{N\to\infty} P\left\{ \left| \frac{1}{N} W_N^N - \frac{1}{N} E W_N^N \right| < \frac{\epsilon}{2} \right\} = 1. \tag{60}$$

Thus by Equation (57), if we choose $2\nu < \frac{\epsilon}{2}$, then:

$$\lim_{N\to\infty} P\Big\{ \left| \frac{1}{N} W_N^N - \int_0^1 (2p(s)-1) e^{-2f(p(s))(1-s)}\, ds \right| < \epsilon \Big\} \geq$$
$$\lim_{N\to\infty} P\Big\{ \left| \frac{1}{N} W_N^N - \frac{1}{N} E W_N^N \right| < \frac{\epsilon}{2} \Big\} = 1.$$

□

## 5. DISCUSSION AND CONCLUSIONS

5.1. **Game theory and Biology.** In this work, we offer a new game theoretical approach to analyze biological competitive processes. Many competitive processes could be better understood by analyzing them on a shorter time-scale than the time-course considered in evolutionary dynamics. Instead of the change in the "fitness" of a player, which is the traditional payoff in evolutionary games, we define the payoff function, tailored to the specific questions addressed. Here, we address the question: How the size principle emerges from the game MNs play? Or in other words, why less-active MNs win in more competitions than more-active MNs? As the $N$ competitions end at different times, the game we define consists of $N$ successive stages. The payoff of a player, is the size of its muscle unit, and the strategy of a MN is its activation threshold (or its level of activity). In contrast to an evolutionary game, the strategies do not change during the game. The interaction between the players is expressed in the fact that the winning probability of a player at each stage (each competition), depends on the strategies (activation-thresholds) of the other players competing at that muscle-fiber, and also on the history, namely the number of previous winnings relative to its competitors.

We believe that this new approach provides a useful framework to analyze competitive processes in biological systems. In the first place, it provides a most appropriate framework for *thinking* about competition. The new game theoretical result of this work is that the *time* of winning is not (as may erroneously seem), a neutral factor, but rather it has a competitive value. We prove that in our setting, winning early is expensive, and thus, as resources are limited, one should invest more in later competitions in order to win in more competitions. This conclusion enabled us to explain the emergence of the size principle and to resolve the paradox of contradictory experimental data.

Another advantage of using this approach in biological competitions is that it may lead to conclusions even when the mechanisms underlying the competition is not fully understood or is under debate: Not knowing the adjustment function $\mu$ that nature uses, and not knowing the specific competitive rule $\rho$ that mediates the competition at the single muscle-fiber, we were still

able to prove that unless the competition is *extremely* biased in favor of more-active MNs, then the size principle emerges.

In recent years a new field named "Neuroeconomics" has evolved (Glimcher [32], Cassidi [33] 2006). Typically, neuroeconomics exploits experimental tools of neuroscience to answer questions arising in behavioral economics. This work is an attempt to do the opposite, namely to use a game theoretical approach in order to analyze competitive processes in the nervous system.

5.2. **The game MNs Play.** By using this new game theoretical approach, we succeeded to reconcile the paradoxical experimental data on this issue, and to explain the size principle, for any adjustment function $\mu$, and any prior winning function $\rho$ that is not extremely biased in favor of M-team.

Although the effect of activity on the rate of synapse elimination has been established more than 20 years ago, this effect was never attributed to be the source for the emergence of the size principle. Rather, the two models attempting to explain the emergence of the size principle assumed that the competitive mechanism, which generates such an activity-dependent phenomenon, must accordingly be governed by activity (Barber & Lichtman [44] 1999, Stollberg [45] 1995, see Appendix 7.7 for a detailed account of these models). In contrast, our model proves that the size principle can be explained merely as an *indirect* consequence of the effect of activity on the *rate* of synapse elimination, without further assuming a direct role of activity in the competitive mechanism itself.

The main idea is as follows: On the basis of the experimental data, we conclude that more-active MNs are more involved in early competitions and less-active MNs are more involved in later competitions. Thus the winner at early competitions is more likely to be more-active, and the winner at late competitions is more likely to be less-active. Once a MN wins at a muscle-fiber, it must devote some resources for maintaining it, thus it has less available resource for competing efficiently at other muscle-fibers. Therefore *it is better to win in later competitions* (as less-active MNs do) and lack this amount of resource in the *fewer* competitions that are left until the end of the game.

Using the same consideration, the model resolves the paradox of seemingly contradictory experimental data. For example, selective blocking of a group of axons to a muscle has been shown in one experiment, to result in smaller muscle units of the blocked group, whereas in another experiment, in which the blocking period was significantly shorter, it resulted in larger muscle units of the blocked group. Indeed, according to our model, selective blocking is *expected* to have such opposite effects. On the one hand, the blocked MNs are expected to lose in almost all the competitions that are resolved during the blocking period. This explains the experimental results in which the blocking period was long. On the other hand, selective blocking specifically delays the competitions at

the muscle-fibers that are innervated by some blocked axons (as the activity of these muscle-fibers is reduced by the blocking). This delay works in favor of the blocked group when activity is resumed, since, as explained, it is better to win at later competitions. This explains the experimental results in which the activity was blocked for a shorter time and was then recovered. We illustrate by a simulation, that indeed using the same adjustment function and prior winning function, but different blocking times, yielded these same opposite results.

5.3. **New testable predictions.** The evolutionary advantage of properties as the size principle, emerging as a consequence of competition, rather than being genetically hardwired, is that it endows the system with plasticity (or adaptation capabilities), such that the outcome may be fine-tuned to fit the environment. In accordance with this idea the present study provides six experimentally-testable predictions regarding the magnitude of the size principle in different environments (different muscles). We use the averaged difference in the number of winnings between the more-active and the less-active MNs, as a measure for the magnitude of the size principle. Briefly, our model predicts the following: 1) Enhancing the activity of all MNs, by applying stimuli that are beyond all activation-thresholds will abolish the size principle whereas applying a very weak stimuli will reverse the size principle. 2) Manipulating the initial innervation-ratio so that each MN innervates each muscle-fiber, also abolishes the size principle. 3) The size of the muscle (number of fibers in the muscle), does not influence the magnitude of the size principle. 4) Small variance of the activation-thresholds of the innervating MNs implies weak expression of the size principle. 5) An early winner is predicted to have lower activation-threshold than a later winner. 6) Executing the same selective stimulation procedure as Ridge and Betz [23] (1984) but at *earlier* stage of synapse elimination weakens the advantage of the stimulated MNs seen in Ridge and Betz experiments.

5.4. **Generalization of the size principle.** The rules of the game are based strictly on experimental results, thus providing testable new predictions listed above. At the same time, these rules are easily generalized so as to be applicable to a much wider scope of situations. As explained, more-active MNs are involved more in early competitions and less-active MNs are involved more in later competitions. Hence a more-active MN can be viewed as "investing more in early competitions", and a less-active MN can be viewed as "investing more in late competitions". Thus, the general strategic conclusion which follows from this game-theoretic result, and may apply also to non-biological scenarios, e.g., economical settings, is that when resources are limited, one should invest more in later competitions in order to win in more competitions. In contrast, one may erroneously conjecture that it is better to try winning in early competitions as well, guaranteeing winnings from the start, instead of taking a risk by waiting and letting the competitor collect these

winnings. Which of the two intuitive arguments is the correct one, is determined by our mathematical proof, which shows that in this setting, one should invest in later competitions in order to win in more competitions.

## 6. METHODS

### Simulations

Our claims are all proved mathematically, however, in order to illustrate these results, we use *Matlab* and simulate the game under different initial conditions.

6.1. **Initial conditions.** For consistency, in all simulations (unless stated differently), there are $N$=10,000 fibers in the muscle and 100 innervating MNs. Each MN innervates initially each muscle-fiber, with probability of $\gamma = 0.05$, and does not innervate it with probability of 0.95. Technically, this is done as follows: For each of the 100 MNs and each of the muscle-fibers, we randomly choose a number between 0 and 1. If the number is smaller than $\gamma$ =0.05, we say that the MN connects to the muscle-fiber, otherwise, we say it does not connect to it. The same is done for each of the MNs and each of the muscle-fibers, so finally we have a full description of all the initial connections between the MNs and muscle-fibers. Using $\gamma = 0.05$, yields an average of 5 connections per muscle-fiber.

The activity-levels of the 100 MNs are drawn from the uniform distribution over [0,1]. The 50 MNs with the higher activity-levels form M-team, and the reminder MNs form L-team.

The activity-levels of muscle-fibers are defined as the sum of activity-levels of their innervating MNs. The muscle-fibers are then ordered according to their activity-level, from highest to lowest.

The"prior winning probability" $P_i^N$ of M-team at each muscle-fiber $i$, is set as the initial proportion of M-team connections there. In other words $\rho(q) = q$. Thus we have the initial conditions of the game:

$$P_1^N, P_2^N, \ldots, P_N^N.$$

6.2. **The course of the game.** The game has 10,000 successive stages (corresponding to 10,000 muscle-fibers). At each stage, exactly one competition is being resolved. The adjustment size (unless stated differently), is $\mu = 0.0005$. At the first stage, the competition at the muscle-fiber with the highest activity-level is resolved, as follows:

**Stage 1** (muscle-fiber 1): We randomly choose a number between 0 and 1. If it is smaller than $P_1^N$, then we say that M-team won at the first muscle-fiber and set $W_1^N = 1$, else we say that L-team won at the first muscle-fiber and set $W_1^N = -1$ . **Stage 2:** The "actual winning probability" $P(2 \mid W_1^N)$ at muscle-fiber 2 is now adjusted according to $W_1^N$, as follows: $P(2 \mid W_1^N) = P_2^N - \mu W_1^N$

, where $P_2^N$ is the prior winning probability of M-team at muscle-fiber 2, and $\mu = 0.0005$. We randomly choose a number between 0 and 1. If it is smaller than the actual winning probability $P(2 \mid W_1^N)$ , we say that M-team won at the second muscle-fiber and compute $W_2^N$ according to the result. (Recall that $W_2^N$ is defined as the difference in the number of M-team winnings and L-team winnings up to stage 2. For example, if M-team won at the first stage and lost at the second one, then $W_1^N = 1$ and $W_2^N = 0$). Similarly, at each stage $i \geq 2$, we compute the actual winning probability, we randomly choose a number, compare it to the actual winning probability, and accordingly compute $W_i^N$ . Whenever the computed actual winning probability exceeds 1 (or becomes negative) we set it to be 1 (or 0).

6.3. **Figures.** In most figures, the points along the $x$-axis, denote the relative stage of the game, namely *1/N, 2/N, , 1*, where $N$ is the number of stages in the game. Unless stated differently, the $y$-axis describes *W/N*, which is the difference $W$ in the number of winnings along the game, divided by the number of stages *N*. In most simulations, the game is played a 100 times and the averaged process is shown.

6.4. **Statistical analysis.** For each game, we look at the random variable $\omega = \frac{1}{N} W_N^N$, (which is the difference in winnings at the end of the game divided by the number of stages $N$). All statistical tests are operated on the average of this random variable, over the $s = 100$ games played, namely: $\frac{1}{100} \sum_{i=1}^{100} \omega_i$. Since the games are independent identically distributed random variables, and we estimate the variance of $\omega$, from the 100 games played, then we use the $t$-tests for all our statistical testing. Significance-level is 0.05.

6.5. **Simulations of selective manipulation of activity.** In order to relate to the results of these experiments, we divide the population of MNs into "manipulated" and "unmanipulated" teams, (e.g., "blocked" and "unblocked" teams), instead of the former "more-active" and "less-active" teams. Like before, activity-levels of the MNs are distributed symmetrically over [0,1], but during a blocking period, the activity-levels of the MNs belonging to the "blocked" team, and the prior winning probabilities of the blocked team are set to zero (and similarly, during stimulation periods, the activity-levels of the stimulated MNs and the prior winning probabilities of the stimulated team are set to 1). In addition, at the start and end of each manipulation period, the temporal order of competitions is updated (as the manipulation changes the activity-levels of the muscle-fibers).

## References


[1] Mann, F., Ray, S., Harris, W.A. and Holt, C.E. Topographic Mapping in Dorsoventral Axis of the Xenopus Retinotectal System Depends on Signaling through Ephrin-B Ligands. *Neuron* **35**, 461-473, (2002).

[2] Kandel, E. R., Schwartz, J. H., and Jessel, T. M. Principles of neural science. *McGraw-Hill*, Fourth edition (2000).

[3] Martin, S. J., Grimwood, P. D., and Morris, R. G. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649-711 (2000).

[4] Castellucci V. F, Carrew T. J., Kandel E. R. Cellular analysis of long-term habituation of the gill-withdrawal reflex in Aplysia californica. *Science* **202**, 1306-1308 (1978).

[5] Squir LR, Kandel ER. Memory: mind to molecules. *New York: Sci. Am. Lib.* **1** (1999).

[6] Bailey, C.H, Giustetto, M., Huang, Y., Hawkins, R.D. and Kandel, E. 2000. Is Hetrosynaptic Modulation Essential for Stabilizing Hebbian Plasticity and Memory?. *Nature Reviews. Neurosci* **1**, 11-20 (2000).

[7] Tsien J. Z., Huerta P. T., and Tonegawa S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87**, 1327-1338 (1996).

[8] Sanes, J. R., and Lichtman, J. W. Development of the vertebrate neuromuscular junction. *Annu. Rev. Neurosci.* **22**, 389-442 (1999).

[9] Hubel, D. H., Wiesel, T. N., Stryker, M. P. Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature.* **269(5626)**, 328-30 (1977).

[10] Purves, D., and Lichtman, J.W. Elimination of synapses in the developing nervous system. *Science* **210**, 153-157 (1980).

[11] Crepel, F., Mariani, J., and Delhaye-Bouchaud, N. Evidence for a multiple innervation of purkinje cells by climbing fibers in the immature rat cerebellum. *J. Neurobiol.* **7**, 567-578 (1976).

[12] Lohof, A.M., Delhaye-Bouchaud, N., and Mariani, J. Synapse elimination in the central nervous system: functional significance and cellular mechanisms. *Rev. Neurosci.* **7**, 85-101 (1996).

[13] Lichtman, J.W. The reorganization of synaptic connexions in the rat submandibular ganglion during post-natal development. *J. Physiol. (Lond.)* **273**, 155-177 (1977).

[14] Lichtman, J.W. and Colman, H. Synapse elimination and indelible memory. *Neuron* **25**, 269-78 (2000).

[15] Walsh, M.K. and Lichtman, J.W. In vivo time-lapse imaging of synaptic takeover associated with naturally occurring synapse elimination. *Neuron* **37**, 67-73 (2003).

[16] Henneman, E. The size-principle: a deterministic output emerges from a set of probabilistic connections. *J Exp Biol* **115**, 105-12 (1985).

[17] Henneman, E. Relation between size of neurons and their susceptibility to discharge. *Science* **126**, 1345-7 (1957).

[18] Henneman, E. and Olson, C.B. Relations between Structure and Function in the Design of Skeletal Muscles. J Neurophysiol 28, 581-98 (1965).

[19] Solomonow, M., Baten, C., Smit, J., Baratta, R., Hermens, H., D'Ambrosia, R., Shoji, H. Electromyogram power spectra frequencies associated with motor unit recruitment strategies. *J Appl Physiol* **68**, 1177-85 (1990).

[20] Callaway, E.M., Soha, J.M. and Van Essen, D.C. Competition favouring inactive over active motor neurons during synapse elimination. *Nature* **328**, 422-6 (1987).

[21] Callaway, E.M., Soha, J.M. and Van Essen, D.C. Differential loss of neuromuscular connections according to activity level and spinal position of neonatal rabbit soleus motor neurons. *J Neurosci* **9**, 1806-24 (1989).

[22] Ribchester, R.R. and Taxt, T. Motor unit size and synaptic competition in rat lumbrical muscles reinnervated by active and inactive motor axons. *J Physiol* **344**, 89-111 (1983).

[23] Ridge, R.M. and Betz, W.J. The effect of selective, chronic stimulation on motor unit size in developing rat muscle. *J Neurosci* **4**, 2614-20 (1984).

[24] Balice-Gordon, R.J. and Lichtman, J.W. Long-term synapse loss induced by focal blockade of postsynaptic receptors. *Nature* **372**, 519-24 (1994).

[25] Lo, Y.J. and Poo, M.M. Activity-dependent synaptic competition in vitro: heterosynaptic suppression of developing synapses. Science 254, 1019-22 (1991).

[26] van Ooyen, A. Competition in the development of nerve connections: a review of models. *Network* **12**, R1-47 (2001).

[27] van Ooyen, A. and Willshaw, D.J. Competition for neurotrophic factor in the development of nerve connections. *Proc Biol Sci* **266**, 883-92 (1999).

[28] O'Brien, R.A., Ostberg, A.J. and Vrbova, G. Observations on the elimination of polyneuronal innervation in developing mammalian skeletal muscle. *J Physiol* **282**, 571-82 (1978).

[29] Vyskocil, F. and Vrbova, G. Non-quantal release of acetylcholine affects polyneuronal innervation on developing rat muscle fibres. *Eur J Neurosci* **5**, 1677-83 (1993).

[30] Thompson, W., Kuffler, D.P. and Jansen, J.K. The effect of prolonged, reversible block of nerve impulses on the elimination of polyneuronal innervation of new-born rat skeletal muscle fibers. *Neuroscience* **4**, 271-81 (1979).

[31] Ribchester, R.R. and Taxt, T. Repression of inactive motor nerve terminals in partially denervated rat muscle after regeneration of active motor axons. *J Physiol* **347**, 497-511 (1984).

[32] Glimcher, P. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics.* MIT Press. (2003).

[33] Cassidy, J. *Mind Games. What neuroeconomics tells us about money and the brain.* New Yorker. (2006).

[34] Thompson, W. and Jansen, J.K. The extent of sprouting of remaining motor units in partly denervated immature and adult rat soleus muscle. *Neuroscience* **2**, 523-35 (1977).

[35] Fladby, T. and Jansen, J.K. Postnatal loss of synaptic terminals in the partially denervated mouse soleus muscle. *Acta Physiol Scand* **129**, 239-46 (1987).

[36] Cope, T. C., Pinter, M. J. The size principle: Still working after all these years. *NIPS* **10**, 280-285 (1995).

[37] Willshaw, D.J. The establishment and the subsequent elimination of polyneuronal innervation of developing muscle: theoretical considerations. *Proc R Soc Lond B Biol Sci* **212**, 233-52 (1981).

[38] Kasthuri, N. and Lichtman, J.W. The role of neuronal identity in synaptic competition. *Nature* **424**, 426-30 (2003).

[39] Buffelli, M., Burgess, R. W., Feng, G., Lobe, C. G., Lichtman, J. W., Sanes, J. R. Genetic evidence that relative synaptic efficacy biases the outcome of synaptic competition. *Nature* **424**, 430-4 (2003).

[40] Connold, A.L., Evers, J.V. and Vrbova, G. Effect of low calcium and protease inhibitors on synapse elimination during postnatal development in the rat soleus muscle. *Brain Res* **393**, 99-107 (1986).

[41] Dan, Y. and Poo, M.M. Hebbian depression of isolated neuromuscular synapses in vitro. *Science* **256**, 1570-3 (1992).

[42] Liu, Y., Fields, R.D., Festoff, B.W. and Nelson, P.G. Proteolytic action of thrombin is required for electrical activity-dependent synapse reduction. *Proc Natl Acad Sci U S A* **91**, 10300-4 (1994).

[43] Liu, Y., Fields, R.D., Fitzgerald, S., Festoff, B.W. and Nelson, P.G. Proteolytic activity, synapse elimination, and the Hebb synapse. *J Neurobiol* **25**, 325-35 (1994).

[44] Barber, M.J. and Lichtman, J.W. Activity-driven synapse elimination leads paradoxically to domination by inactive neurons. *J Neurosci* **19**, 9975-85 (1999).

[45] Stollberg, J. Synapse elimination, the size principle, and Hebbian synapses. *J Neurobiol* **26**, 273-82 (1995).

[46] Brown, M.C., Jansen, J.K. and Van Essen, D. Polyneuronal innervation of skeletal muscle in new-born rats and its elimination during maturation. *J Physiol* **261**, 387-422 (1976).

[47] Gordon, T., Purves, R.D. and Vrbova, G. Differentiation of electrical and contractile properties of slow and fast muscle fibres. *J Physiol* **269**, 535-47 (1977).

[48] Vrbova, G., Gordon, T. and Jones, R. *Nerve Muscle Interaction. London: Chapman & Hall,* 119-145 (1978).

[49] Summerbell, D and Stirling, K. V. The innervation of dorsoventrally reversed chick wings: evid ence that motor axons do not actively seek out their appropriate targets. *J. Embryol. Exp. Morph.* **61**, 233-247 (1981).

[50] Gordon, T., Vrbova, G and Willcock, G. The influence of innervation on differentiating tonic and twitch muscle fibers of the chicken. *J. Physiol., Lond.* **319,** 261-269 (1981).

[51] Vrbova, G., Navarrete, R. and Lowrie, M. Matching of muscle properties and motoneurone firing patterns during early stages of development. *J Exp Biol* **115**, 113-23 (1985).

[52] Thompson, W.J., Sutton, L.A. and Riley, D.A. Fibre type composition of single motor units during synapse elimination in neonatal rat soleus muscle. *Nature* **309**, 709-11 (1984).

## 7. APPENDICES

### 7.1. More examples for $\rho$.

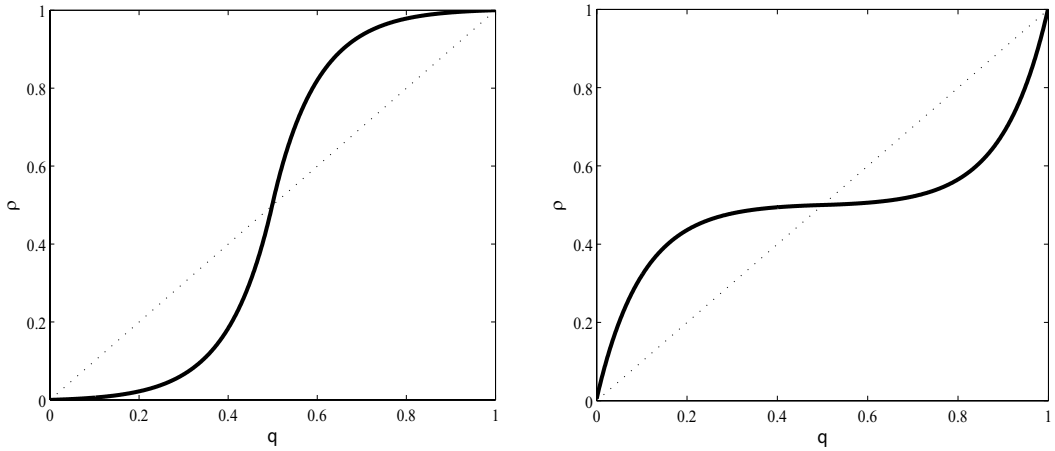


FIGURE 10. ***Two examples for a democratic prior winning function*** $\rho$. *The $x$-axis denotes the fraction of M-team connections at a muscle-fiber. The $y$-axis presents the prior winning probability of M-team at the muscle-fiber. The left hand graph represents a function which is more sensitive to small differences between the fractions of connections of the two teams.*

### 7.2. Games that are biased in favor of M-team.

Up to now, we showed that if a priori, each MN innervating a muscle-fiber has the same chances of winning it (defining a neutral game), then L-team win the game. Now, we prove that even if activity drives competition at the neuromuscular junction, giving *some* head-start to more-active MNs, then still L-team win. Mathematically this means that even if the prior winning function is not neutral, but rather it is biased in favor of M-team to some extent then still L-team win.

A function $\rho$, was said to be neutral, if it satisfies:

$$\rho(1-q) = 1 - \rho(q), \ \forall q. \tag{61}$$

Similarly, we now define a function $\rho'$ to be "biased in favor of M-team", if it satisfies:

$$\rho'(1-q) \geq 1 - \rho'(q), \ \forall q. \tag{62}$$

It follows from Theorem 4.12 and Theorem 4.13, that for each democratic function $\rho$ :

**Corollary 7.1.** *There exists $D > 0$, such that:*

$$\int_0^1 (2p(s) - 1)e^{-2f(p(s))(1-s)}\, ds \leq -D < 0, \tag{63}$$

where: $p(s) = \lim_{N\to\infty} E\rho(q_i^N)$.

*Proof.* By Theorem 4.12, there exists $D > 0$, such that for all $N$ :

$$\frac{1}{N} E\left(W_N^N\right) \leq -D.$$

Thus, by Theorem 4.13:

$$\int_0^1 (2p(s) - 1)e^{-2f(p(s))(1-s)}\, ds = \lim_{N\to\infty} \frac{1}{N} E\left(W_N^N\right) \leq -D < 0.$$

□

In fact, Equation (63) provides the mathematical condition which is necessary and sufficient for the emergence of the size principle. Thus, even if $\rho$ is biased in favor of M-team, as long as $\rho$ yields $p$ satisfying (63), then the size principle emerges. In Corollary 7.1, we proved that any democratic function $\rho$ (i.e., monotonic and neutral), satisfies the Equation.

Denote $\rho$ as a monotonic and neutral function, and $\rho'$ as a monotonic but *biased* function (i.e., $\rho$ satisfies (61), and $\rho'$ satisfies (62)).

Recall that:

$$p(s) = \lim_{N\to\infty} E\rho(q_i^N). \tag{64}$$

Hence, for all $\delta > 0$, there exists $\epsilon > 0$, such that if $\rho(s) \leq \rho'(s) \leq \rho(s) + \delta,\ \forall s$, then:

$$\left|p'(s) - p(s)\right| < \epsilon,\ \forall s, \tag{65}$$

where $p'$ satisfies Equation (64) for $\rho'$. Now, for all $\epsilon > 0$, satisfying (65), it is also true that:

$$\left| \int_0^1 (2p'(s) - 1)e^{-2f(p'(s))(1-s)}\, ds - \int_0^1 (2p(s) - 1)e^{-2f(p(s))(1-s)}\, ds \right| < \epsilon.$$

By Equation (63), there is $D > 0$, such that:

$$\int_0^1 (2p(s) - 1)e^{-2f(p(s))(1-s)}\, ds \leq -D < 0,$$

thus by choosing $\epsilon < \frac{D}{2}$, we get that also for the biased game defined by $\rho'$:

$$\int_0^1 (2p'(s) - 1)e^{-2f(p'(s))(1-s)}\, ds < 0.$$

In other words, each democratic function $\rho$ could be changed so that instead of being neutral, (i.e., $\rho(1 - q) = 1 - \rho(q),\ \forall q$), it is biased in favor of M-team (i.e., $\rho(1 - q) \geq 1 - \rho(q),\ \forall q$), and

*still* satisfies Equation (63), namely still yields the result that L-team win the game. In Figure 4, we show that the bias (in favor of M-team), for which L-team still wins, may be very large indeed.

7.3. **Mathematical explanations for the predictions.** We proved that:

$$\lim_{N\to\infty}\frac{1}{N}W_N^N=\int_0^1(2p(s)-1)e^{-2f(p(s))(1-s)}\,ds. \tag{66}$$

This equation provides the following five predictions. These predictions were presented earlier in Section 3.4. We now provide mathematical proofs and explanations.

(1) *An early winner, (namely a MN that already singly innervates a muscle-fiber during the first days of synapse elimination), has on average, a lower activation-threshold, than a MN winning later*: This results from rule 1 and 2: By rule 1, Competitions at muscle-fibers with more M-team connections, which are therefore more-active, tend to be resolved early and competitions at muscle-fibers with more connections by L-team tend to be resolved later on. Thus by the democratic rule (rule 2), the winner at early competitions is more likely to be a more-active MN, and the winner at late competitions is more likely to be a less-active MN. The dynamics of the game (defined by rule 3), further strengthen this effect.

(2) *Executing the same selective stimulation procedure as Ridge and Betz* [23] *(1984) but at earlier stage of synapse elimination weakens the advantage of the stimulated MNs seen in Ridge and Betz experiments.* The reason for that is that this brings the competitions which stimulated MNs take part (and are most likely to win in), even *more* forward, reducing their winning probabilities in all the future competitions that are left until the end of the game.

(3) *Applying extremely strong stimuli - beyond the activation-thresholds of all the MNs - will abolish the size principle (namely, all muscle units are about the same size), whereas replacing the natural stimuli by weak stimuli - beneath the median activation-threshold - would reverse the size principle (namely, the more-active MNs will have larger muscle units than the less-active MNs)*: Our model predicts that in most cases, (excluding only the extreme cases mentioned above), it is impossible to predict the effect of such manipulation on the *outcome* (i.e., the sizes of muscle units) of the game since it affects it in opposite ways: For example, enhancing activity causes L-team to fire more frequently, contributing a bias in favor of L-team thus resulting in a stronger expression of the size principle. On the other hand, it also reduces the variability in the behaviors of the two teams (making them more alike), which according to (3) ahead, results in a *weaker* expression of the size principle.

However, it is possible to predict the outcome in the two extreme cases: Applying extremely strong stimuli - beyond the activation-threshold of all players - will cause all MNs to fire each time. This is reflected in our model by identical activity-levels, resulting in $p \equiv \frac{1}{2}$, where $p$ is the expectation of the prior winning probability function when $N \to \infty$. Substituting $p \equiv \frac{1}{2}$, in Equation (66), results in a zero expression of the size principle. The other extreme case, is applying extremely weak stimuli - beneath the activation-threshold of L-team. This excludes L-team members from the game ($p \equiv 1$), thereby producing a positive value in Equation (66) (i.e., reversing the size principle).

(4) *If the activation-thresholds of the MNs are very much alike, then the size principle is only weakly expressed,* since in this case the activity-levels of muscle-fibers (and thus the order of the competitions) are determined more according to a decreasing *total* number of competitors at a muscle-fiber, and is less dependent on the number of M-team competitors, namely $p \approx \frac{1}{2}$.

(5) *If the initial innervation is manipulated into full innervation, namely that each MN initially innervates each muscle-fiber, then the size principle vanishes.* The proportion $\gamma$ of muscle-fibers each MN innervates initially, affects the process in several ways: First, When a MN wins, this only affects the other competitions, in which this same MN takes part. Therefore the more muscle-fibers a MN innervates initially (i.e., a large $\gamma$ ), the larger $\mu$ should be. Hence, $\mu$ is proportional to $\gamma$, (and thus $f$, appearing in (66), is proportional to $\gamma$). However, the dependence of the process on $\mu$ is rather complicated as shown in Figure 3. Secondly, the larger $\gamma$ is, the more similar will be the activity-levels of the muscle-fibers. Due to the multiple effects $\gamma$ has on the process, we cannot, in general, predict whether increasing $\gamma$, will increase or decrease the expression of the size principle. However, in the extreme case of full innervation ($\gamma = 1$), each muscle-fiber is innervated by each MN, thus $p \equiv 0.5$, and thus, by Equation (66), there is a zero expression of the size principle (see Figure 7).

(6) *The magnitude of the size principle is independent on the number of fibers in the muscle.* The number of fibers, in a typical skeletal muscle, is large (many thousands), hence a large muscle (large $N$), and a smaller muscle, have roughly the same $\frac{1}{N}E\left(W_N^N\right)$, which is close to: $\int_0^1 (2p(s)-1)e^{-2f(p(s))(1-s)}\,ds$. In other words, our model predicts that the degree of expression of the size principle *does not* depend on the number of fibers in the muscle (see Figure 8).

7.4. **Estimating the Magnitude of the Size Principle.** Given a muscle, a sample of a subgroup of the MNs innervating this muscle, is drawn at random. Denote the sample size as $2S$. The activation-thresholds of the $2S$ MNs are estimated, and accordingly, the MNs are divided into 2 equal sized teams; the $S$ MNs with the lowest activation-thresholds (i.e., the most active MNs) form the M-team, and the reminder $S$ MNs form L-team. Now, the size of the muscle unit (i.e., the number of muscle-fibers belonging to the muscle unit) of each MN in the sample, is assessed.

To compute $\frac{1}{N}W_N^N$, denote:

$$U_1^M, U_2^M, \ldots, U_S^M, \quad \text{and} \quad U_1^L, U_2^L, \ldots, U_S^L,$$

as the sizes of muscle units of the MNs in M-team and L-team respectively. Then:

$$\frac{1}{N}W_N^N = \frac{\sum_{i=1}^S U_i^M - \sum_{i=1}^S U_i^L}{\sum_{i=1}^S U_i^M + \sum_{i=1}^S U_i^L}.$$

The more negative $\frac{1}{N}W_N^N$, the more the size principle is expressed in the muscle, because it means that the muscle units of the less-active MNs are *much* larger than the muscle units of the more-active MNs. When $\frac{1}{N}W_N^N$ is close to zero, we say that the size principle vanishes or only weakly expressed. If $\frac{1}{N}W_N^N > 0$, we say that the size principle is reversed.

7.5. **Proof of Proposition 4.2.** Given $Y = y = (y_1, y_2, \ldots, y_{2n})$, order the MNs according to decreasing activity-levels. Thus: $y_1 > y_2 > \ldots > y_{2n}$, $k = 1, \ldots, n \in$M-team, $k = n+1, \ldots, 2n \in$L-team.

We prove Proposition 4.2, by proving the following more general proposition:

Let $a, b, n \geq 0$ be integers with $a + b \leq n$. Let $y_1 > y_2 > \cdots > y_{2n}$ be real numbers. Let $\Gamma_M$ be the set of all functions $f : \{1, \ldots, 2n\} \to \{0, 1\}$ such that:

$\sum_{i=1}^n f(i) = a + b$ and $\sum_{i=n+1}^{2n} f(i) = a$. Similarly let $\Gamma_L$ be the set of all functions $f : \{1, \ldots, 2n\} \to \{0, 1\}$ such that:

$\sum_{i=1}^n f(i) = a$ and $\sum_{i=n+1}^{2n} f(i) = a + b$.

For $f$ in $\Gamma_M$ or $\Gamma_L$, denote:

$$x_f = \sum_{i=1}^{2n} f(i) y_i.$$

Denote

$$S = |\Gamma_M| = |\Gamma_L| = \binom{n}{a+b}\binom{n}{a},$$

and denote $\ell = y_n - y_{n+1}$.

We name the elements of $\Gamma_M$ by $f_1, f_2, \ldots, f_S$ such that:

$$x_{f_1} \geq x_{f_2} \geq \ldots \geq x_{f_S},$$

and we name the elements of $\Gamma_L$ by $g_1, g_2, \ldots, g_S$ such that

$$x_{g_1} \geq x_{g_2} \geq \ldots \geq x_{g_S}.$$

**Proposition 7.2.** *For all* $1 \leq s \leq S$ :

$$x_{f_s} - x_{g_s} \geq b\ell.$$

Note that this proves Proposition 4.2, when taking: $a = r - m \quad and b = 2m - r > 0$.

*Proof.* The proof will be divided into two steps.

*Step 1:* We define a bijection $\Gamma_M \to \Gamma_L$ denoted by $f \mapsto f'$ and will prove that this bijection satisfies:

$$x_f - x_{f'} \geq b\ell,$$

for each $f \in \Gamma_M$. We define it as follows. Denote $h = n + \frac{1}{2}$, (it is the midpoint separating between the sets $\{1, \ldots, n\}$ and $\{n+1, \ldots, 2n\}$). Given $f \in \Gamma_M$, denote $k(f)$ as the minimal $0 \leq k \leq n$ such that $\sum_{i<h-k} f(i) = \sum_{i>h+k} f(i)$. Note that the minimum is taken on a non-empty set since $k = n$ satisfies the condition with the sum being empty and so 0 on both sides. Now we define

$$f'(i) = \begin{cases} f(2n+1-i), & \text{if } h - k(f) < i < h + k(f); \\ f(i), & \text{otherwise.} \end{cases}$$

The map $f \mapsto f'$ is indeed a bijection, its inverse $\Gamma_L \to \Gamma_M$ being defined by the same rule (note that $k(f') = k(f)$). We now prove by induction on $n$ that:

$$x_f - x_{f'} \geq b\ell.$$

We assume the truth of the claim for $n-1$ and any $a, b$ and proceed to establish it for $n$. We separate two cases.

- **If** $b = 0$ : Then for any $f \in \Gamma_M$, $k(f) = 0$ so $f' = f$ and so $x_f - x_{f'} = 0 = b\ell$.
- **If** $b > 0$ : Then $k(f) \geq 1$. We apply the induction hypothesis to the sequence

  $$y_1, \ldots, y_{n-1}, y_{n+2}, \ldots, y_{2n}$$

  (i.e. the elements $y_n, y_{n+1}$ are dropped). We denote the restriction of $f$ to this set of $2(n-1)$ indices by $f^-$. We denote $\sum_{i<h-1} f(i) - \sum_{i>h+1} f(i)$ by $b^-$, and $y_{n-1} - y_{n+2}$ by $\ell^-$, so $\ell^- \geq \ell$. It is clear that $(f')^- = (f^-)'$ since $k(f) \geq 1$. By the induction hypothesis

  $$x_{f^-} - x_{(f^-)'} \geq b^-\ell^-.$$

  Now if $f(n) = f(n+1)$ then also $f'(n) = f'(n+1) = f(n)$, and $b^- = b$ and so:

  $$x_f - x_{f'} = x_{f^-} - x_{(f^-)'} \geq b^-\ell^- \geq b\ell.$$

If $f(n) = 1$ and $f(n+1) = 0$, then $b^- = b - 1$ and so:

$$x_f - x_{f'} = x_{f^-} - x_{(f^-)'} + \ell \geq b^-\ell^- + \ell = (b-1)\ell^- + \ell \geq b\ell.$$

If $f(n) = 0$ and $f(n+1) = 1$ then $b^- = b + 1$ and so:

$$x_f - x_{f'} = x_{f^-} - x_{(f^-)'} - \ell \geq b^-\ell^- - \ell = (b+1)\ell^- - \ell \geq b\ell.$$

This completes the induction.

*Step 2:* To deduce the statement of the Proposition from step 1, we need to show the following. Let

$$x_1 \geq x_2 \geq \cdots \geq x_S$$

and

$$x'_1 \geq x'_2 \geq \cdots \geq x'_S$$

be two sequences and assume there is a bijection: $F : \{1, \ldots, S\} \to \{1, \ldots, S\}$, and a constant $c \geq 0$, such that:

$$x_s \geq x'_{F(s)} + c,$$

for all $1 \leq s \leq S$. Then: $x_s \geq x'_s + c$ for all $1 \leq s \leq S$.

We prove this by induction on $S$.

If $F(1) = 1$, then $x_1 \geq x'_1 + c$ and we can drop both, and complete the claim by the induction hypothesis, using the restriction of $F$ to $\{2, \ldots, S\}$. If $F(1) = k$ with $k > 1$, then let $t > 1$ be the index such that $F(t) = 1$. We have:

$$x_1 \geq x_t \geq x'_{F(t)} + c = x'_1 + c.$$

Again we may drop $x_1, x'_1$, and complete the claim by the induction hypothesis, this time using a new function $G$ on $\{2, \ldots, S\}$ defined as follows:

$$G(s) = \begin{cases} F(s) + 1, & \text{if } F(s) < k; \\ F(s), & \text{if } F(s) > k. \end{cases}$$

Indeed $G$ has the required property since for $s$ with $F(s) < k$, $x_s \geq x'_{F(s)} + c \geq x'_{F(s)+1} + c = x'_{G(s)} + c$, and for $s$ with $F(s) > k$, $x_s \geq x'_{F(s)} + c = x'_{G(s)} + c$. $\square$

7.6. **Proof of Lemma 7.3.**

**Lemma 7.3.** *For all* $i, j,\ i \neq j$ :

$$Cov(w_i, w_j) \leq 0.$$

*Proof.* For all $i < j$, denote:

$$W^N_{j-1;i} = W^N_{j-1} - w_i.$$

In this proof we abbreviate and use $\mu = \mu\left(N, P^N_i\right)$.

By definition:

$$P(w_j = 1 \mid w_i) = P^N_j - \mu W^N_{j-1;i} - \mu w_i.$$

Thus:

(1) $P(w_j = 1 \mid w_i = 1) = P^N_j - \mu W^N_{j-1;i} - \mu.$
(2) $P(w_j = 1 \mid w_i = -1) = P^N_j - \mu W^N_{j-1;i} + \mu.$
(3) $P(w_j = 1) = P^N_j - \mu W^N_{j-1;i} - \mu E(w_i).$

As: $-1 \leq E(w_i) \leq 1$, we get from (1),(2) and (3), that:

$$P(w_j = 1 \mid w_i = 1) \leq P(w_j = 1) \leq P(w_j = 1 \mid w_i = -1).$$

Thus also:

$$2P(w_j = 1 \mid w_i = 1) - 1 \ \leq \ 2P(w_j = 1) - 1 \ \leq \ 2P(w_j = 1 \mid w_i = -1) - 1,$$

and so:

$$E(w_j \mid w_i = 1) \ \leq \ E(w_j) \ \leq \ E(w_j \mid w_i = -1).$$

Hence:

$$E(w_i w_j) =$$
$$E(w_j \mid w_i = 1)P(w_i = 1) - E(w_j \mid w_i = -1)P(w_i = -1) \leq$$
$$E(w_j)P(w_i = 1) - E(w_j)P(w_i = -1) = E(w_j)\Big(P(w_i = 1) - P(w_i = -1)\Big) =$$
$$E(w_j)E(w_i).$$

Thus:

$$Cov(w_i, w_j) = E(w_i w_j) - E(w_j)E(w_i) \leq 0.$$

□

7.7. **Other models of competition in the neuromuscular system.** Several models have been proposed for the competition at the neuromuscular system (van Ooyen [26] 2001). However the main goal of most of these models was to explain the change from Polyneuronal to mononeural innervation during development. Namely, to explain why mononeural innervation is a stable state, which once reached, the elimination processes ends. Only two models (Stollberg [45] 1995, Barber and Lichtman [44] 1999) aimed to explain the emergence of the size principle. These models were also the only ones to try reconciling the paradoxical experimental data regarding the effects of activity on the sizes of muscle-units. In both of these models, the size principle emerges as a consequence of competition.

Stollberg [45] considered a "correlational learning rules", in which the relative synaptic strength increases when synapse and muscle-fiber are either both active or both inactive. A crucial assumption in Stollberg's Hebbian model is the existence of an early period, in which the collective strength of connections between a muscle-fiber and its innervating MNs is low. In contrast it has been experimentally shown (Brown, Jansen & van Essen [46] 1976) that from the start, as multiple innervation of muscle-fibers is established, each input is strong enough to cause *on its own* the activation of the muscle-fiber. Thus in particular, from the beginning of synapse elimination, the collective strengths of connections between any muscle-fiber and its innervating MNs is *high*. This contradicts the main assumption in Stollberg's model.

Barber and Lichtman [44] (1999) suggested that at the level of single muscle-fibers, more-active MNs have greater ability to eliminate less-active competitors, but at the same time more-active MNs restrict to a greater extent, their own competitive abilities at other muscle-fibers. Accordingly, they define two parameters; that "rewards" activity and that "punishes" it. Barber and Lichtman argue that a competitive advantage of more-active MNs early in the competition is overcome at later stages by greater synaptic efficacy of MNs firing at a lower rate. While it is clear why activity is advantageous earlier (when resources are more available) and disadvantageous later, it is not explained why the later disadvantage *overcomes* the earlier advantage. In other words, it is not explained why in total, less-active MNs win in more competitions. Barber and Lichtman achieve this by choosing the punishing signal to be 10 times larger than the rewarding signal.

Regarding the reconciliation of seemingly contradictory data, Barber and Lichtman qualitatively produce the results from the blocking experiments of Callaway [20][21] (1987, 1989), however they do not relate to the blocking experiments of Ribchester and Taxt [22] (1983) nor to the stimulation experiments of Ridge and Betz [23] (1984), which paradoxically yield opposite results to the results of Callaway.