The development of talent in sports

den Hartigh, Ruud J. R.; Hill, Yannick; van Geert, Paul L. C.

Published in:
Complexity

DOI:
10.1155/2018/9280154

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
The Development of Talent in Sports: A Dynamic Network Approach

Ruud J. R. Den Hartigh, Yannick Hill, and Paul L. C. Van Geert

Department of Psychology, University of Groningen, 9712 TS Groningen, Netherlands

Correspondence should be addressed to Ruud J. R. Den Hartigh; j.r.den.hartigh@rug.nl

Received 1 March 2018; Accepted 8 July 2018; Published 29 August 2018

Academic Editor: Jordi Duch

Copyright © 2018 Ruud J. R. Den Hartigh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Understanding the development of talent has been a major challenge across the arts, education, and particularly sports. Here, we show that a dynamic network model predicts typical individual developmental patterns, which for a few athletes result in exceptional achievements. We first validated the model on individual trajectories of famous athletes (Roger Federer, Serena Williams, Sidney Crosby, and Lionel Messi). Second, we fitted the model on athletic achievements across sports, geographical scale, and gender. We show that the model provides good predictions for the distributions of grand slam victories in tennis (male players, \( n = 1528 \); female players, \( n = 1274 \)), major wins in golf (male players, \( n = 1011 \); female players, \( n = 1183 \)), and goals scored in the NHL (ice hockey, \( n = 6677 \)) and in FC Barcelona (soccer, \( n = 585 \)). The dynamic network model offers a new avenue toward understanding talent development in sports and other achievement domains.

1. Introduction

In 1869, Francis Galton published his work on the genetics of genius, in which he claimed that eminent individuals are born with the potential to excel in the future. He based this conclusion on his observation that elite performance tends to run in families at much higher rates than could be expected based on chance [1]. A few years later, De Candolle wrote a book in which he stated that environmental resources (e.g., family, education, and facilities) are the major factors explaining the emergence of excellence [2]. Galton later contrasted their viewpoints in the terminology of nature and nurture [3], which formed the starting point of the famous nature-nurture debate in psychology [4]. Now, over a century later, it remains a major challenge to understand how individual trajectories of talent development are shaped by the complex interplay between nature and nurture factors. A related and important question is why only very few individuals are ultimately able to demonstrate exceptional performance. Here, we briefly discuss different past and current perspectives, after which we explain that a novel dynamic network approach provides the theoretical principles and analytical tools to understand how talent develops. In doing so, we primarily focus on the domain of sports, in which talent development has received much attention from researchers, and in which rich sets of empirical data are available.

In order to define the kind of model that captures the process of talent development, the first important step is the conceptualization of talent and related concepts. Talent can be defined as an individual’s potential or capacity to excel in a particular domain that requires special skills and training [4]. The individual’s potential is a condition specified by all available factors contributing to the actual growth of a particular ability [4, 5]. An individual’s ability can then be defined as the manifest or actualized potential. This ability can vary from very low to exceptionally high, but actualized talent typically corresponds to the high ability range [4, 6, 7]. Although ability is not directly observable, an individual’s performance accomplishments, such as winning sports tournaments, is a stochastic function of ability that can be measured. Proceeding from this conceptualization, one can derive that talent development is the process through which potential turns into manifest abilities, which may result into measurable demonstrations of elite performance [4, 8].
So far, research on talent development has primarily centred around the question: How much do particular genetic and nurturing factors contribute to the development of elite performance [9–12]? Although some researchers have emphasized the importance of one particular factor, such as genetic endowment [13] or deliberate practice [14, 15], researchers have now reached consensus that various nature and nurture factors contribute to the development of talent [10–12, 16, 17]. The current challenge is to be able to answer the question: What kind of model or mechanism can account for the way in which combinations of nature and nurture variables shape the process of talent development, which for some athletes result in elite performance achievements? In the behavioural sciences, the standard model describes momentary associations between variables across samples that are large enough to represent the population of interest (e.g., elite athletes in a particular sport). The most obvious of such models is a regression model, which explains the interindividual variability of abilities, skills, or performances on the basis of the sum of factors that are associated to the athletic ability at hand. For instance, in a linear regression model, a level of ability, $A$, is the sum of levels of constituent components:

$$A = \alpha + \beta x + \gamma y + \cdots,$$

where the variables $x$, $y$, and so forth are the predictors, such as genetic endowment, physical factors, psychological factors such as commitment, and environmental factors such as family support, with $\alpha$, $\beta$, and $\gamma$ moderating the effects of the variables.

Following the standard model, scientific projects across countries and types of sports have put a major focus on finding the physical, technical, tactical, psychological, practice, and environmental variables that distinguish groups of elite athletes from groups of sub- or nonelite athletes [18–24]. Outcomes of these projects increasingly suggest that the model underlying talent development is not a linear, uniform model that holds within samples of athletes. This suggestion is in accordance with the so-called ergodicity problem, according to which a model based on group data only generalizes to a model of individual processes if very specific conditions apply, which are hardly ever met in the behavioural and social sciences [25, 26]. For instance, a statistical model based on a typical sample of a great number of individuals may take the form of a linear regression model, describing the codistribution of the observations in the space of variables. Every individual model, on the other hand, is likely to take the form of idiosyncratically, dynamically coupled variables, which associate over time in ways that are fundamentally different from the statistical group model. When looking more closely at individual processes of talent development, research has increasingly shown that (i) an athlete’s ability level as well as possible determinants (e.g., physical qualities and commitment) change over time; (ii) genes, the environment, and other physical and psychological factors are intertwined in complex ways; and (iii) there is no average, linear developmental trajectory that holds across athletes. In addition, contrary to the assumption of standard models, the distribution of talent across the population is considered to be non-normal [10, 16, 27, 28].

To exemplify the four properties mentioned above, first, evidence for the dynamic development of talent can be derived from research tracking athletes’ performance histories [19], reports on athletes’ scores on correlates of sports performance (e.g., intermittent endurance capacity of soccer players [29]), and in-depth qualitative investigations [20]. Second, the property that genes, the environment, and other (physical and psychological) factors are intertwined is increasingly acknowledged in behavioural genetics and epigenetic models [28, 30]. Nature and nurture are thus inseparable in the development of certain traits or qualities, including sports talent. This is consistent with the idea that even environmental factors that are considered as signs of nurture, such as parental support, also carry a genetic component, given that parents’ genetic make-up is partly responsible for their creation of a stimulating home environment to develop talent [31–33]. Third, the complex interplay between nature and nurture factors may take different forms for different athletes, and researchers have shown that the road to the top is hardly ever a straight road [19, 34]. For instance, a study among elite Australian athletes showed that most athletes underwent different (nonlinear) trajectories from junior to senior, with less than 7% of all athletes demonstrating a pure linear trajectory [19]. A comparable conclusion could be drawn from longitudinal research projects in soccer, field hockey, basketball, artistic gymnastics, tennis, and speed skating, conducted in the Netherlands. In their studies, the researchers primarily searched for underlying predictors at the group level, but later concluded that athletes have their own unique developmental patterns that lead to excellent performance [35]. There are two kinds of explanations for these unique pathways, which may co-occur. The first is that the relationships between underlying variables are not static and linear but rather dynamic and complex [4, 7, 9, 16, 27, 36], and the second is that certain predictable or unpredictable events may occur that affect the further developmental trajectory of the individual athlete [20, 34, 37, 38]. One example of a predictable event is the transition from youth to professional, which can be a critical period in an athlete’s development [19, 38, 39]. Unpredictable events, such as trauma, may also occur and have a considerable impact on the athlete’s further trajectory [17, 20, 34, 38, 40, 41].

Finally, the fact that the distribution of talent across the population is not normal has been stressed repeatedly, mostly by Simonton [4, 42, 43]. Across the population, talent would be skewed with a heavy tail to the right. Although it is virtually impossible to directly measure talent (i.e., potential), it is possible to measure the expressions of an athlete’s ability (i.e., actualized talent) in terms of performance achievements. Assuming that the measurable achievements of athletes provide an indication of their actualized talent, research has indeed shown highly right-skewed distributions in different sports including American football, cricket, baseball, basketball, soccer, swimming, track and field, car racing, tennis, and skiing [16, 44–48]. These highly skewed distributions are often characterized by so-called power laws, in which the exceptional athletes can be found in the right
tal. This entails that, across the sample of athletes in any sports, there are very few who ultimately reach exceptional achievements at the professional level [20, 46, 49]. For instance, among the elite swimmers, Michael Phelps has won an incredible number of 28 Olympic medals whereas the great majority of professional swimmers never won an Olympic medal.

To advance the modelling of talent development, one should define the principles that can explain the properties above, driven by assumptions about the definition of talent and the nature of developmental processes. This means that talent should be modelled as a potential that develops through complex nature-nurture interactions [4, 7, 10, 32]. In addition, the model should account for the fact that (i) there is a potential that can grow and can be actualized, (ii) there are supporting and inhibiting factors that change across time, (iii) nature and nurture factors are intertwined and shape each other, and (iv) this developmental process is different for different athletes [4, 7, 10, 16, 27]. Based on these requisites, we propose that talent development can be understood from the perspective of dynamic networks (Figure 1).

In the 2000s, applications of different kinds of network models have become prominent across different scientific domains, including physics, economy, biology, and the social sciences [50]. The difference between dynamic network models and standard models is that the latter focus on associations between specific variables across a particular population (e.g., the association between commitment and performance in the population of soccer players [24]), whereas the dynamic network model focuses on the potentially explanatory properties of a dynamic network structure per se. Dynamic network models allow modelling of individual trajectories, and by modelling a representative sample of individual trajectories, the dynamic network model also offers a model of a population. Establishing a dynamic network model thereby lays the groundwork for future studies of person-specific network structures, in which the nature of the relevant network components can be specified. A key focus of the current article is to reveal what a basic dynamic network model of talent development may look like, validated by data from different sports.

The specific network model we present here is inspired by dynamic systems applications to human developmental processes [51–54]. Mathematically speaking, we proceeded from an extended logistic growth equation, according to which quantitative changes in developmental variables should be understood on the basis of dynamic relationships with other variables that are themselves subject to change [52, 53, 55, 56]. Here, talent is considered as a potential in terms of a mathematically defined growth, and ability is the actual level of a variable at a particular moment in time. The ability variable is embedded in a set of (changing) interconnected variables, defined as connected growers. The growers include stable resources, which correspond to the (epi)genetic contribution that may differ for different variables and different individuals [4, 7]. Furthermore, the network is a directed causal graph, which in most cases will be cyclic. The interactions among the variables in the network have a particular “weight” and can be direct but also indirect (e.g., if the athlete’s ability positively affects the support provided by the parents, which in turn positively affects the athlete’s coping skills, the athlete’s ability and coping skills are indirectly connected; see Figure 1).
These network properties can be mathematically defined as follows:

\[
\begin{align*}
\frac{\Delta L_A}{\Delta t} &= \left( r_A L_A \left( 1 - \frac{L_A}{K_{L_A}} \right) + \sum_{v=1}^{n} s_v L_A V_v \right) \left( 1 - \frac{L_A}{C_A} \right) \\
\frac{\Delta L_B}{\Delta t} &= \left( r_B L_B \left( 1 - \frac{L_B}{K_{L_B}} \right) + \sum_{v=1}^{n} s_v L_B V_v \right) \left( 1 - \frac{L_B}{C_B} \right)
\end{align*}
\]

(2)

where \( \Delta L_A/\Delta t \) corresponds to the change of the variable, \( K \) is the stable (genetic) factor, \( r \) is the growth rate associated with the stable factor, \( V \) corresponds to the other variable components in the network to which the component in question (e.g., \( L_A \)) is connected, and \( s \) represents the growth rate associated with the variable, supportive or inhibitory factors in the form of connection weights in the network. The \( C \) parameter corresponds to the limits of growth of a particular variable (i.e., the absolute carrying capacity), the specification of which is more important than its exact value [16]. This means that the function of the \( C \) parameter is to keep the variables within realistic (e.g., biophysical) limits, in the unlikely mathematical possibility that too many relationships are strongly positive and drive the system into an exponential explosion. The extended logistic growth equation gives rise to different, often nonlinear forms of development [16, 52, 53], which are typically observed in the domain of sports [9, 19, 27, 37].

In order to account for events, such as a transition from youth to professional, it should be possible to model a singular perturbation to an athlete’s ability level around the transition and expose him or her to new challenges and environments [36, 57, 58]. Following this transition, athletes may reach achievements or not (e.g., winning professional tournaments), which can be modelled by embedding a product model in the network model [16]. One such model is the ability-tenacity model, which is particularly relevant in domains where perseverance, commitment, and devotion are important [45], such as sports [9, 17, 24, 36]. This model also takes into account that the attainments of elite achievements are a function of a “chance” factor, which is typical for sports [35, 59, 60]. The specific formula to calculate an achievement of an athlete at each time point \( (P_i) \) therefore equals

\[ P_i = \varphi L_i T_i, \]

(3)

where \( \varphi \) is the likelihood parameter, \( L_i \) is the ability variable in the network, and \( T_i \) is the tenacity variable. Importantly, as the ability and tenacity components are directly and indirectly connected with the other network components, the resulting accomplishments are not just the result of these two variables, but are a stochastic function of the interaction-dominant network dynamics in which ability and tenacity are embedded.

In this study, we aimed to test whether a dynamic network model provides a valid theoretical foundation of talent development. Therefore, we simulated athlete-networks based on (2) and compared the outcomes of the simulations with current knowledge based on the extant literature and archival data that we collected. First, in its basic form, the model should generate the individual, nonlinear developmental trajectories for different athletes and include youth-to-professional transition events [9, 19, 27, 35]. Apart from the ability-development of the athletes, the model should be able to generate performance achievements that are a function of ability, tenacity, and a chance factor [49, 59, 60]. Ultimately, among the simulated athletes, only very few should demonstrate achievements that are disproportionately exceptional within the athletic population, as evidenced by a power law distribution [16, 46].

To empirically check the validity of the dynamic network model, we compared the model predictions based on computer simulations with data we collected from two major individual sports (i.e., tennis and golf) and two major team sports (i.e., (ice) hockey and soccer). More specifically, we compared the model predictions with cases of professional athletes (Federer, Williams, Crosby, and Messi) and with the distributions of performance attainments across sports, gender, and geographic scale (from worldwide to local).

2. Materials and Methods

2.1. Archival Data. For this study, we collected archival data from elite tennis players, golf players, (ice) hockey players, and soccer players. In tennis, the number of tournament victories is a direct indicator of a player’s achievements. In order to secure an even level of competition across the tournaments and to have comparable datasets for male and female players, we focused on the grand slam tournaments. Comparable to winning a grand slam in tennis is winning a major in golf. Major tournaments also host the highest-ranked players at the given point in time. Another parallel with tennis grand slams is that we can consider both male and female athletes for this sport.

Hockey is a team sport, in which six players are on the field for each team. Of these six players, one is the goaltender and the other five are so-called skaters. Due to the dynamic of the game and the relatively small rink size, each skater is involved in attacking as well as defending. This provides every skater with the opportunity to score goals. Since a team needs to score goals in order to win, scoring is a measurable expression of a player’s ability. We focus on the National Hockey League (NHL), USA, which is the highest level hockey competition worldwide. Similar to hockey, to determine performance achievements in soccer, we focus on the goals scored by field players.

To examine individual achievement trajectories, we zoomed into a few elite athletes with exceptional (measurable) achievements. These athletes were Roger Federer, who won an exceptional number of 18 grand slam titles in male tennis at the time of data collection, Serena Williams, who...
won an exceptional number of 23 grand slams in female tennis, Sidney Crosby, who scored an exceptional number of 338 goals in the NHL, and Lionel Messi, who scored an exceptional number of 312 league goals for FC Barcelona. In addition, we determined the population distributions of performance achievements in tennis, golf, hockey, and soccer. For tennis, we examined the distributions of grand slam titles for male (n = 1528) and female players (n = 1274). The samples included all players who played at least one single’s match in a grand slam tournament since the start of the open era of tennis tournaments (i.e., 1968) until present. Second, we focused on golf major titles for male (n = 1011) and female players (n = 1183). In order to have a homogeneous and comparable timeframe between men and women, the male count was restricted to the years 1968 (the year ladies major golf tournaments started) until present. The samples included all players who participated in at least one major. In the case of hockey, every skater (n = 6677, all male) who ever played in the National Hockey League (NHL), USA, until 2016, was taken into account. Finally, for soccer, we considered all field players (n = 585, all male), who played for FC Barcelona in the first Spanish Division since 1928.

The data for the different sports were retrieved from the sports’ official websites or the official website tracking the statistics of that sport. The data for tennis were collected through the Association of Professional Tennis’ website (http://www.atpworldtour.com, accessed at 16-02-2017) and the International Tennis Federation’s website (http://www.itftennis.com, accessed at 17-02-2017); for golf through the Professional Golfers’ Association of America’s website (http://www.pgatour.com, accessed at 17-04-2017); for hockey through official National Hockey League’s website (http://www.nhl.com, accessed at 21-02-2017); and for soccer through the La Liga website (http://www.laliga.es accessed at 22-02-2017).

2.2. Dynamic Network Model Settings. The dynamic network model was implemented in Visual Basic that runs under Microsoft Excel, which allowed us to simulate developmental trajectories of individual athletes. Table 1 shows the default settings of the parameters that we used in order to simulate athletes’ dynamic networks. These default settings correspond to the initial values of the parameters in (2), and the model further defines a probability of .25 that two components are directly connected, within a network consisting of 10 variables. This probability and the size of the network are defined a priori based on a previous theoretical paper on modelling excellent human performance [16]. The model corresponds to a neutral model, which means that the weights are on average zero, with a symmetrical distribution towards negative and positive values. In the network, we arbitrarily defined node 3 as the ability variable and node 4 as the tenacity variable.

In addition to the default settings that suffice to run simulations of the basic network, we inserted a transition from youth to professional. In order to model this, we applied a “perturbation” to the ability variable (i.e., node 3) at step 300, which in the simulation marks the transition point. More specifically, we modelled a drop around the transition point in node 3, defined as a decrease of 0.65 standard deviation (SD = 0.15), and we let three (out of the ten) variables enter the network around the transition period. The latter corresponds to the fact that athletes likely face new challenges and deal with new factors that might negatively or positively dynamically relate to their ability [20, 37].

Mathematically speaking, the parameters that we defined are dimensionless numbers. This means that they are numbers that do not directly correspond with the dimensionality of specific physical or psychological properties. The parameters are ratio numbers that specify a particular ratio or proportion of effect of one component on other components and on itself. The population described by the model is represented in the form of hypothetical distributions of these parameter values. An individual in this population is represented by any combination of parameter values randomly drawn from these distributions. The empirical verification of the dynamic network model is then based on the following predictions: (1) in any representative sample of parameter combinations, we will find resulting individual trajectories that correspond with observed individual trajectories of athletes, and (2) any representative sample of parameter combinations will generate a population of individual trajectories, the general properties of which correspond with the properties of an observed population of athletes.

In order to model the athletes’ achievements, we connected the dynamic network model with a product model. The likelihood that an achievement was generated for an athlete was based on the ability level, level of tenacity, and a likelihood parameter \( \varphi \) (see (3)) [45]. Because it is easier to score goals in hockey and soccer than it is to win grand slams or majors in tennis and golf, the \( \varphi \) parameter had the highest value in hockey (\( \varphi = 0.004 \)), followed by soccer (\( \varphi = 0.002 \)), and then by golf and tennis (\( \varphi = 0.0002 \)). Furthermore, in hockey and soccer it is possible to generate multiple achievements (i.e., goals) at a single time step, which is not possible for the achievements in terms of grand slam and major titles in golf and tennis. Therefore, the maximum number of products per time step was set to 3 in hockey and soccer and to 1 in golf and tennis.

The default parameter settings that we used for the simulations of populations of tennis players, hockey players, and golfers entered the model: (1) we made the statistical assumptions of our model against the observed data of the tour and (2) we calculated the proportion of coefficients and variable selection results, and the model was validated with the observed data of the three sports but not the tour. The model was run in Microsoft Excel, which allowed us to simulate developmental trajectories of individual athletes. Table 1 shows the default parameter values used for the dynamic model simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r ) (growth rate)</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>( s ) (connection weight with other variables)</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>( K ) (stable resources—genetics)</td>
<td>1.00</td>
<td>0.15</td>
</tr>
<tr>
<td>Connection probability with other variables</td>
<td>0.25</td>
<td>—</td>
</tr>
</tbody>
</table>

In order to model the athletes’ achievements, we connected the dynamic network model with a product model. The likelihood that an achievement was generated for an athlete was based on the ability level, level of tenacity, and a likelihood parameter \( \varphi \) (see (3)) [45]. Because it is easier to score goals in hockey and soccer than it is to win grand slams or majors in tennis and golf, the \( \varphi \) parameter had the highest value in hockey (\( \varphi = 0.004 \)), followed by soccer (\( \varphi = 0.002 \)), and then by golf and tennis (\( \varphi = 0.0002 \)). Furthermore, in hockey and soccer it is possible to generate multiple achievements (i.e., goals) at a single time step, which is not possible for the achievements in terms of grand slam and major titles in golf and tennis. Therefore, the maximum number of products per time step was set to 3 in hockey and soccer and to 1 in golf and tennis.

The default parameter settings that we used for the simulations of populations of tennis players, hockey players, and golfers entered the model: (1) we made the statistical assumptions of our model against the observed data of the tour and (2) we calculated the proportion of coefficients and variable selection results, and the model was validated with the observed data of the three sports but not the tour. The model was run in Microsoft Excel, which allowed us to simulate developmental trajectories of individual athletes. Table 1 shows the default parameter values used for the dynamic model simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r ) (growth rate)</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>( s ) (connection weight with other variables)</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>( K ) (stable resources—genetics)</td>
<td>1.00</td>
<td>0.15</td>
</tr>
<tr>
<td>Connection probability with other variables</td>
<td>0.25</td>
<td>—</td>
</tr>
</tbody>
</table>

In order to model the athletes’ achievements, we connected the dynamic network model with a product model. The likelihood that an achievement was generated for an athlete was based on the ability level, level of tenacity, and a likelihood parameter \( \varphi \) (see (3)) [45]. Because it is easier to score goals in hockey and soccer than it is to win grand slams or majors in tennis and golf, the \( \varphi \) parameter had the highest value in hockey (\( \varphi = 0.004 \)), followed by soccer (\( \varphi = 0.002 \)), and then by golf and tennis (\( \varphi = 0.0002 \)). Furthermore, in hockey and soccer it is possible to generate multiple achievements (i.e., goals) at a single time step, which is not possible for the achievements in terms of grand slam and major titles in golf and tennis. Therefore, the maximum number of products per time step was set to 3 in hockey and soccer and to 1 in golf and tennis.

The default parameter settings that we used for the simulations of populations of tennis players, hockey players,
3. Results

3.1. Developmental Trajectories of Athletes. In line with the literature on talent development, and with the fact that the extended logistic growth equation typically generates nonlinear developmental patterns, simulations of the dynamic network model revealed different trajectories of talent development for different athletes. Figure 2 displays the simulations of two athletes’ networks (graphs a and b) and shows that they reach comparable ability levels in different ways. Note also how the simulated athletes respond differently (yet ultimately adaptively) to the imposed perturbation when transitioning from youth to professional (i.e., step 300), whereas another simulation generated the realistic scenario of an athlete that could not adapt after the transition (graph c).

In order to check whether the model provides predictions that fit with the archival data we collected, we first determined whether the performance accomplishments generated by the model are in agreement with the data of specific athletes. To model these accomplishments, we assumed that athletes may accomplish an achievement (e.g., winning a tournament or scoring a goal) from the moment they transition from youth to professional. The probability that at a particular moment in time an achievement is accomplished is a function of the ability-tenacity model (3) [45].

Our first simulation corresponds to an athlete who reaches an ability level of 20.00, which is 17.74 standard deviations above the mean ability level ($M_{ability} = 13.6, SD = 1.27$). We connected the ability development of this athlete to a low $\varphi$ parameter (0.0002) to simulate grand slam victories in tennis, yielding 13 achievements ($M = 14.20, 95\% CI = 6.95 - 21.45$ at 1000 simulations with the same ability and tenacity levels). We compared the model prediction with the data of Roger Federer at the time of data collection. We found a good qualitative resemblance in terms of the simulated...
trajectory (see Figures 3(a) and 3(c)) and the total number of grand slams won (18), which falls within the simulated 95% confidence interval (CI). The second set of graphs corresponds to the Grand Slam titles of Serena Williams according to the actual and simulated data (see Figures 3(b) and 3(d)). The simulated athlete reaches a maximum ability level of 20.00 and is again connected to a low $\phi$ parameter (0.0002). The simulation resulted in a total of 20 achievements ($M = 15.59$, 95% CI = 8.16 – 23.02 at 1000 simulations with the same ability and tenacity levels). In reality, Williams had won 23 Grand Slams, which is included in the simulated 95% CI.

To compare the model predictions with hockey, in which athletes’ performances could be measured based on the number of goals they scored, we increased the value of the $\phi$ parameter to 0.004, and we set the number of achievements that can be produced at each time step equal to 3. Again, we took the archival data of an exceptional player, in this case Sidney Crosby. The simulation led to a maximal ability level of 11.97 (8.39 standard deviations above the mean) and a total of 337 achievements ($M = 352.02$, 95% CI = 303.71 – 388.87 for 1000 simulations with the same ability and tenacity levels). In reality, Crosby had scored 338 goals in the NHL, which falls within the simulated 95% CI.

3.2. Distributions of Performance Accomplishments. To test whether the distribution of athletes’ achievements follows a power law, in which very few athletes accomplish exceptional achievements across sports, gender, and geographical scale, we conducted our analyses on: grand slam titles in tennis for male and female players, major wins in golf for male and female players, goals scored in the National Hockey League (NHL) competition, and goals scored by FC Barcelona players. Then, we simulated these achievements for populations of tennis, golf, hockey, and soccer players.

For all analyses on the archival data, we found patterns close to a power law in the log-log plots for tennis, golf, hockey, and soccer (see Figures 5 and 6). The power laws are evidenced by the linear regression slopes in the log-log plots (see Tables 2 and 3), which provide a strong fit with the collected data ($R^2 = 0.94$ for male tennis, $R^2 = 0.89$ for

![Figure 3: Trajectories of performance accomplishments for Roger Federer and Serena Williams. Graph (a) corresponds with Federer’s actual trajectory of grand slam titles per year, and graph (c) with a simulated trajectory; graph (b) corresponds with Williams’ actual trajectory of grand slam titles per year, graph (d) with a simulated trajectory. In graphs (c) and (d), one year corresponds to 20 simulation steps.](image)
female tennis, $R^2 = 0.99$ for male golf, $R^2 = 0.97$ for female golf, $R^2 = 0.98$ for hockey, and $R^2 = 0.96$ for soccer). The results imply that the extremely skewed distributions hold across sports, gender, and geographical scale.

Simulating the performance accomplishments based on the dynamic network model, we find the same kinds of distributions as in the archival data. This is implied by the results that (i) the simulated number of players with zero accomplishments is close to the actual number of players with zero accomplishments, (ii) the simulated maximum number of accomplishments for an athlete within a given athletic population is close to the actual maximum number of accomplishments by an individual athlete, and (iii) the regression slopes (beta coefficients) of the log-log plots, which provide an estimate of the power parameter, show close resemblances between the simulated and archival data. Table 2 provides an overview of the results for the individual sports, and Figure 5 shows the log-log plots of the athletes’ achievements in the individual sports according to the archival and simulated data.

The results for hockey and soccer are shown in Table 3, and Figure 6 displays the log-log plots of the players’ achievements (i.e., goals scored) according to the archival and simulated data.

4. Discussion

Here, we proposed a dynamic network model of talent development and tested whether it explains the individual developmental patterns and achievements of elite athletes, as well as the distributions of achievements across populations of athletes in different sports. We therefore (i) defined the model principles based on the definition of talent and the literature on human developmental processes; (ii) ran simulations of the defined dynamic network model; (iii) collected performance attainments of specific cases in tennis (Federer and Williams), hockey (Crosby), and soccer (Messi) and compared their data with the patterns generated by simulations of our dynamic network model; and (iv) collected performance attainments across the population of elite athletes in tennis, golf, hockey, and soccer and compared the population distributions with those generated by the dynamic network model.

Regarding the ability-level trajectories, the dynamic network model generates nonlinear patterns that differ per individual athlete. This is in accordance with previous studies on talent development in sports [19, 35] and the nonergodicity of developmental processes [25, 26]. In order to model the process of talent development, we used a model of change in individuals. In such a model, the associations between the variables over the course of time differ quite fundamentally from statistical associations in a sample of individual cases and cannot be interpreted as random fluctuations around a common pattern present in all individual cases of a particular group (e.g., athletes in a particular sports). In addition, changes in an individual athlete’s trajectory are not driven by the nature of some specified variable. For
Figure 5: Log-log plots of the number of victories +1 against the number of athletes in the individual sports. The graphs correspond to actual and simulated grand slam titles by male players (a); actual and simulated grand slam titles by female tennis players (b); actual and simulated major titles by male golf players (c); and actual and simulated major titles by female golf players (d). Displayed simulated results are based on one simulation round of the population. For plots showing the raw actual and simulated data, see our research materials at https://hdl.handle.net/10411/ZTS6LQ.

Figure 6: Log-log plots of the number of goals +1 scored against the number of athletes in the team sports. The graphs correspond to actual and simulated goals scored by National Hockey League (NHL) players (a), and actual and simulated goals scored by soccer players from FC Barcelona (b). For plots showing the raw actual and simulated data, see our research materials at https://hdl.handle.net/10411/ZTS6LQ.
instance, the reason that an athlete may not adapt to a transition from youth to professional (Figure 2(c)) is not "located" in an underdeveloped variable specifying "ability to adapt," but rather lies in the structure of the connections between the variables in the athlete’s idiosyncratic network. These findings support the general observation that talent development in sports is a nonlinear process in which nature and nurture are intertwined [9, 10, 27, 36].

However, we also went beyond general description of the trajectories of ability development and connected the dynamic network model to an ability-tenacity product model to examine athletes’ simulated performance attainments. Doing this, we were able to replicate the qualitative pattern of achievements of some exceptional athletes in different sports (i.e., Federer, Williams, Crosby, and Messi). Together, these results indicate that the dynamic network model can explain the individual trajectories of talent development, which would not be possible using traditional linear models, such as regression models applied to samples of athletes [61–63]. Indeed, a recent study attempted to generate the typical properties of excellent performance across domains (e.g., sports, science, and music) by simulating a model based on the standard statistical assumption that abilities are normally distributed across the population and result from additive effects of various relevant performance-related variables. No matter how the parameter values were tweaked, predictions did not come near the patterns found in the observed data across excellent performers [16].

Furthermore, in line with previous research [16, 44, 46–48], we found that athletes’ performance attainments in tennis, golf, hockey, and soccer conform to extremely skewed distributions at population level. This means that the exceptional athletes are in the extreme right tail of the highly skewed distributions and that the great majority of athletes accomplished considerably less. As our results show, a power law holds across sports (tennis, golf, hockey, and soccer), gender (male, female), and geographical scale (worldwide competition in tennis and golf; national competition in hockey, and within one club in soccer). The dynamic network simulations revealed interesting resemblances with the actual data in terms of the overall (power law) shape of the distributions, as well as more specific measures such as the number of professional athletes with zero countable achievements and the maximum number of achievements by one particular athlete in a given sports.

The resemblances between the performance accomplishment distributions based on the archival data and the model predictions were more evident for the individual sports than for the team sports. In particular, the predictions in hockey provided a distribution that was more

### Table 2: Achievements in individual sports according to archival and simulated data.

<table>
<thead>
<tr>
<th>Sport</th>
<th>Measure</th>
<th>Actual titles</th>
<th>Simulated titles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tennis (m)</td>
<td>Athletes with 0 titles</td>
<td>1439</td>
<td>1417.60 ± 9.38</td>
</tr>
<tr>
<td></td>
<td>Maximum number of titles</td>
<td>18</td>
<td>21.96 ± 7.62</td>
</tr>
<tr>
<td></td>
<td>Beta coefficient ($\beta_i$)</td>
<td>−3.32</td>
<td>−3.59 ± 0.21</td>
</tr>
<tr>
<td>Tennis (f)</td>
<td>Athletes with 0 titles</td>
<td>1231</td>
<td>1183.10 ± 8.90</td>
</tr>
<tr>
<td></td>
<td>Maximum number of titles</td>
<td>22</td>
<td>20.46 ± 8.99</td>
</tr>
<tr>
<td></td>
<td>Beta coefficient ($\beta_i$)</td>
<td>−3.26</td>
<td>−3.58 ± 0.24</td>
</tr>
<tr>
<td>Golf (m)</td>
<td>Athletes with 0 titles</td>
<td>911</td>
<td>937.92 ± 8.54</td>
</tr>
<tr>
<td></td>
<td>Maximum number of titles</td>
<td>14</td>
<td>16.40 ± 7.73</td>
</tr>
<tr>
<td></td>
<td>Beta coefficient ($\beta_i$)</td>
<td>−3.40</td>
<td>−3.64 ± 0.26</td>
</tr>
<tr>
<td>Golf (f)</td>
<td>Athletes with 0 titles</td>
<td>1098</td>
<td>1099.74 ± 16.08</td>
</tr>
<tr>
<td></td>
<td>Maximum number of titles</td>
<td>10</td>
<td>20.62 ± 8.89</td>
</tr>
<tr>
<td></td>
<td>Beta coefficient ($\beta_i$)</td>
<td>−3.64</td>
<td>−3.59 ± 0.25</td>
</tr>
</tbody>
</table>

Note. The measures include distributional characteristics of achievements for male (m) and female (f) tennis (grand slam titles), and for male (m) and female (f) golf (major titles). The averages and SDs under the simulated titles are based on 50 simulations of the entire populations.

### Table 3: Achievements in team sports according to archival and simulated data.

<table>
<thead>
<tr>
<th>Sport</th>
<th>Measure</th>
<th>Actual goals</th>
<th>Simulated goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Hockey League</td>
<td>Athletes with 0 goals</td>
<td>1456</td>
<td>1327.40 ± 27.99</td>
</tr>
<tr>
<td></td>
<td>Maximum number of goals</td>
<td>894</td>
<td>899.42 ± 2.96</td>
</tr>
<tr>
<td></td>
<td>Beta coefficient ($\beta_i$)</td>
<td>−1.08</td>
<td>−1.16 ± 0.01</td>
</tr>
<tr>
<td>FC Barcelona</td>
<td>Athletes with 0 goals</td>
<td>244</td>
<td>193.02 ± 10.24</td>
</tr>
<tr>
<td></td>
<td>Maximum number of goals</td>
<td>312</td>
<td>463.68 ± 25.19</td>
</tr>
<tr>
<td></td>
<td>Beta coefficient ($\beta_i$)</td>
<td>−1.27</td>
<td>−1.25 ± 0.05</td>
</tr>
</tbody>
</table>

Note. Measures correspond to distributional characteristics of achievements (goals scored) for male athletes in hockey (NHL) and soccer (FC Barcelona). The averages and SDs under the simulated goals are based on 50 simulations of the entire populations.
complexes, the structure of the network characterizes particular
components. Furthermore, according to recent advances in network
dynamics, the effects of one component on another component
can be difficult to draw a line defining which players clearly have (no) attacking tasks on the field. Interestingly,
if one would only take only the attackers of FC Barcelona,
one would again find a strongly skewed distribution. This
supports the claim that distributions of the power-law kind
hold across all kinds of scales of analysis (see the research
materials at https://hdl.handle.net/10411/ZTS6LQ).

4.1. Theoretical and Applied Implications. A dynamic net-
work model seems to underlie the development of talent in
sports, which ultimately results in exceptional achievements
for very few athletes. This conclusion has important implica-
tions at both a theoretical and practical level. At a theoretical
level, an important step is to move away from a focus on
unravelling the underlying variables of talent development
and to embrace the complex interactions that exist across
performer, environment, practice, and training [17, 27, 37].
Exceptional growth of a particular ability in a specific person
is achieved by a wide variety of connection patterns,
which is in line with empirical findings showing that the
dynamics of talent development is highly idiosyncratic and
differs among individuals [7, 9, 27]. Novel challenges in the
direction of investigating dynamic talent networks are get-
ing a grip on the variables involved in individual networks,
as well as posing network-oriented research questions to be
further investigated. The first challenge can be addressed by
conducting longitudinal research in which individual pat-
terns of development are accounted for [37, 64]. Different
personal and environmental variables that are important to
an (youth) athlete’s development can be specified and
tracked over time. Importantly, a major focus should be on
how changes in the variables are embedded in the network
and spread their influence. Although such applications do
not exist yet in the domain of talent development, important
steps are currently made in the domain of clinical psychology
[65–67]. For instance, in a recent study on mental health
monitoring, researchers collected online diary data from the
Dutch population and used autoregressive modelling to
detect directed relationships as they exist between variables
in individual networks [67]. Although the statistical tech-
niques applied were still proceeding from a linear model, this
approach is an important first step to capturing individual
developmental patterns based on empirical data.

With respect to the point of posing network-oriented
research questions, the focus should be on the structure and
dynamics of the network. For instance, what would happen
when values of coupling parameters could change as a result
of long-term effects of one component on another compo-
nent? Furthermore, according to recent advances in network
sciences, the structure of the network characterizes particular
key features, such as resilience [68]. Recent research has
made interesting advances in defining a universal resilience
function that depends on the dynamics and topology of a
network [69]. This may open the door to future studies aimed
at examining whether particular talent networks are more or
less resilient to perturbations such as youth-to-senior transi-
tions or different setbacks during a career. Understanding
the link between network configurations, the development of
talent, and overcoming setbacks can be accomplished by com-
bining computer simulations with data from athletes’ diaries,
for example.

From an applied perspective, talent detection programs
in research and practice around the world are still largely
based on the assumption that talent can be detected in certain
variables “in the individual” and that it can be discovered at
an early age [9, 14, 27, 36, 70, 71]. Given the current knowl-
edge on talent development, the archival data we collected,
and the patterns simulated by the dynamic network model,
one may cast major doubt on this kind of practice. From the
dynamic network perspective, various kinds of direct
and indirect multiplicative relationships between dynamic
variables may exist and lead to different developmental
trajectories. Accordingly, a recent study based on com-
puter simulations already showed that, across achievement
domains, dynamic network predictions reveal that early
detection of later ability levels is virtually impossible [16].
Furthermore, a meta-analytic study recently stated that there
is no clear set of variables that can predict career success in
sports [72].

5. Conclusions

The dynamic network model provides a comprehensive
framework to understand the theoretical principles underly-
ing the development of talent. The model suggests that talent
emerges from intra- and interindividual variations in the
composition of individual dynamic networks. Having dem-
onstrated that the foundation of the dynamic network model
explains empirical observations across a variety of sports, it is
now time to explore and test the variety of practical applica-
tions of the dynamic network perspective.

Data Availability

The basic dynamic network model, the manual of the model,
the archival data, and the simulated data are available at
https://hdl.handle.net/10411/ZTS6LQ.

Conflicts of Interest

The authors declare that there is no conflict of interest
regarding the publication of this paper.

Acknowledgments

Publication of this article was funded by the Heymans Insti-
tute for Psychological Research, University of Groningen,
and a research grant awarded to Yannick Hill by the Spar-
kasse Bank.
References


[36] A. Abbott and D. Collins, "Eliminating the dichotomy between theory and practice in talent identification and development:


