Environmental structure and competitive scoring advantages in team competitions
Supporting Information

Sears Merritt[1,*] and Aaron Clauset[1,2,3,†]

[1] *Department of Computer Science, University of Colorado, Boulder, CO 80309*
[2] *BioFrontiers Institute, University of Colorado, Boulder, CO 80303*
[3] *Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501*

[*] sears.merritt@colorado.edu
[†] aaron.clauset@colorado.edu

I. DETAILED DESCRIPTION OF DATA

Halo: Reach is a popular online game played by nearly 20 million individuals, and was the 3rd most popular US video game of 2010 [1]. It was publicly released by Bungie Inc., a former subdivision of Microsoft Game Studios, on 14 September 2010, and since then, players have generated more than 1 billion competitions. *Reach* is an example of the kind of virtual combat simulation known as a "first-person shooter" or FPS. Within the *Reach* system, players choose from among roughly seven primary game types and numerous subtypes, which are played on more than 33 terrain maps with 74 weapons (the precise number of maps and weapons has varied over time, as the publisher has periodically revised the online content through downloadable updates).

Instances of the game can be played alone, with or against other players via the Xbox Live online system. Participation in this system requires an account, which is distinguished by unique and publicly known "gamertag" or online pseudonym, chosen by the player. In the *Reach* system, both individual game and player summaries were made publicly available through the Halo Reach Stats API. Through this digital interface, we collected detailed data on the first 53 million competition instances (roughly 1TB of data).

Within our sample, there are three basic game types: *campaign games*, a sequence of story-driven, player-versus-environment (PvE) maps that many players complete first; *firefight games* (also PvE), in which a team of human-controlled players battle successive waves of computer-controlled enemies; and *competitive games*, a player-versus-player (PvP) game type, in which teams of the equal size (2, 4, 6 or 8 players) compete to either be the first to reach some fixed number of points or have the largest score after a fixed length of time. (The precise number of players per team, number of points required to win and length of a game depends on the game subtype.) Here, we focus on the most common type of competitive game, with teams of 4 players, a time limit of 600 seconds and a score limit of 50 points.

Among other information, each competition instance game file includes the sequence of scoring events at the per-second resolution and a list of players by team. Scoring events are annotated with the gamertag of the player generating the event, the number of points scored and the player giving up the points (if applicable).

Unlike professional sports, team composition and player resources in *Reach* competitions are not persistent across instances. The only attribute that persists is individual player skill, and thus each new instance is a kind of a "blank slate." To join a new instance, individual players or small groups (often friends [2]) first enter a general pool of available competitors. A Bayesian "matchmaking" algorithm, which seeks to build teams of equal skill [3], then fills teams in the new instance by drawing from this pool. This process substantially randomizes the pairing of individuals within teams and the pairing of teams across instances. Because of the matchmaking algorithm and the large size of the pools, a pair of non-friend players are highly unlikely to be paired again in a new instance; friends may elect to be matched as a unit by forming a "party," a special grouping that the matchmaking algorithm recognizes.

The non-persistence and the randomization are features absent from most studies of team performance or competition [4–6], and serve to mitigate the confounding effects of persistent teams and resources present in most competitive systems, e.g., professional sports. For our purposes, these features make *Reach* competitions a unique source of data for studying behavioral dynamics within competitions and how structural factors shape this behavior.

In competitive games, players move their avatars through the game map simultaneously, in real-time, navigating complex terrain, acquiring avatar modifications and encountering opponents. Teammates may interact through a private voice channel, or through visual signals. Points are scored by dealing sufficient damage to eliminate an opposing avatar and for each such success, a team gains a single point. Eliminated players must then wait several seconds before their avatar is placed back into the game at one of several specified "spawn" locations, equipped with "default" avatar resources that depend on the competition type being played.

For our analysis, we exclude all PvE games and all PvP games containing corrupt scoring event data. (Our analysis suggests no specific pattern to the corruption.) In our primary analyses, we further restricted our sample to PvP competitions (i) between two teams of 4 players and (ii) where no player exited the game early. This latter criterion was relaxed to calculate the relationship between dropouts and β (see Section VIII).

II. GENERATIVE MODEL FOR SCORING EVENT TIMING AND BALANCE

The timing and balance (which team receives the point) of scoring events within a competition are modeled by a conditionally independent Markov process, where an incremental change to a team's score s_r is given by

$$\Pr(\Delta s_r(t) > 0) = \Pr(\Delta s_r > 0 \,|\, \theta, \text{event}) \Pr(\text{event at } t \,|\, \theta) \ ,$$

where θ parameterizes the impact of non-ideal competitive features. That is, the probability that team r's score increases at some time t is the probability that a scoring event occurred at time t and that the resulting point was awarded to r. Furthermore, team labels r and b are arbitrary, and we choose r as our reference team below.

The generation of scoring events is given by a non-stationary Poisson process, in which the probability that a scoring event occurs at time t varies linearly with time:

$$\Pr(\text{event at } t \,|\, \lambda_0, \alpha\,) = \lambda_0 + \alpha\, t \ , \tag{1}$$

where λ_0 is the event background rate and α is the acceleration. When $\alpha = 0$, we recover the stationary Poisson process expected for ideal competitions.

In a real competition, we observe $n \leq T$ scoring events, for a competition lasting T units of time. Let $\{t_i\}$ denote the observed times of these events, and $\{u_j\}$ the times at which no event was observed. The model parameters λ_0 and α are then jointly estimated by directly maximizing the generative model's log-likelihood function:

$$\ln \mathcal{L} = \sum_{i=1}^{n} \ln(\lambda_0 + \alpha\, t_i) + \sum_{j=1}^{T-n} \ln(1 - \lambda_0 - \alpha\, u_j) \ . \tag{2}$$

To limit the biasing effect of the highly non-stationary behavior found in the early- and end-phases of competitions (see main text), we restrict our estimation to events occurring in the middle phase, specifically $50 \leq t \leq 300$. This heuristic provides robust conclusions: the estimated timing parameters are very close to those found using smaller middle-phase windows, and the global average trend within this window is roughly linear (Fig. S1A).

For two teams r and b, the outcome of a scoring event (which team receives the point) is given by a biased Bernoulli process, in which the probability that an event increases the score of team i is

$$\Pr(s_i \text{ increases} \,|\, \theta\,) = \begin{cases} c & i = r \\ 1 - c & i = b \end{cases} ,$$

where $c \in [0, 1]$ represents the competitive advantage (outcome bias) of the r team. In our model system, 99.99% of scoring events yield a single point. Although we do not consider the possibility here, in general, the number of points produced by an event could be drawn from some distribution. Thus, the probability that the competition ends with final scores S_r and S_b is

$$\Pr(S_r, S_b \,|\, c) = c^{S_r}(1 - c)^{S_b} \ , \tag{3}$$

where c denotes the competitive advantage (scoring bias) of team r over team b.

Because team composition varies across competition instances, the competitive advantage of r is modeled as a random variable, drawn from some distribution $\Pr(c)$. The natural choice of the form of this distribution is a symmetric Beta distribution with parameter β, the conjugate prior for the Bernoulli scheme. (We note that the prior distribution must be symmetric about $c = 1/2$ because team labels are arbitrary.) This distributional assumption agrees well with the global empirical distribution of biases c (Fig. S1A inset).

The posterior probability of observing final scores $\{S_r, S_b\}_k$ in a competition instance k is given by their Bernoulli likelihood, weighted by the probability of c (Eq. (3)). Given N such instances, the total posterior probability of the observed final scores is

$$\begin{aligned} \Pr(\beta \,|\, \{S_r, S_b\}) &= \int_0^1 \left(\prod_{k=1}^{N} \Pr(\{S_r, S_b\}_k \,|\, c) \Pr(c \,|\, \beta) \right) \mathrm{d}c \\ &= \prod_{k=1}^{N} \left(\int_0^1 \frac{c^{S_{r_k}+\alpha-1}(1-c)^{S_{b_k}+\beta-1}}{\mathrm{B}(\beta,\beta)} \, \mathrm{d}c \right) \\ &= \prod_{k=1}^{N} \frac{\mathrm{B}(S_{r_k}+\beta, S_{b_k}+\beta)}{\mathrm{B}(\beta,\beta)} \ , \end{aligned} \tag{4}$$

where $\mathrm{B}(a, b)$ is the Beta function.

We estimate the competition balance parameter by numerically maximizing the logarithm of Eq. (4) with respect to β,

$$\ln \mathcal{L} = \sum_{k=1}^{N} \ln[\mathrm{B}(S_{r_k} + \beta, S_{b_k} + \beta)] - \ln[\mathrm{B}(\beta, \beta)] \ . \tag{5}$$

The resulting maximum likelihood estimate $\hat{\beta}$ provides a direct measurement of the overall balance within a set of competition instances: when $\beta \to \infty$, we recover the fair coin $c = 1/2$ expected for ideal competitions.

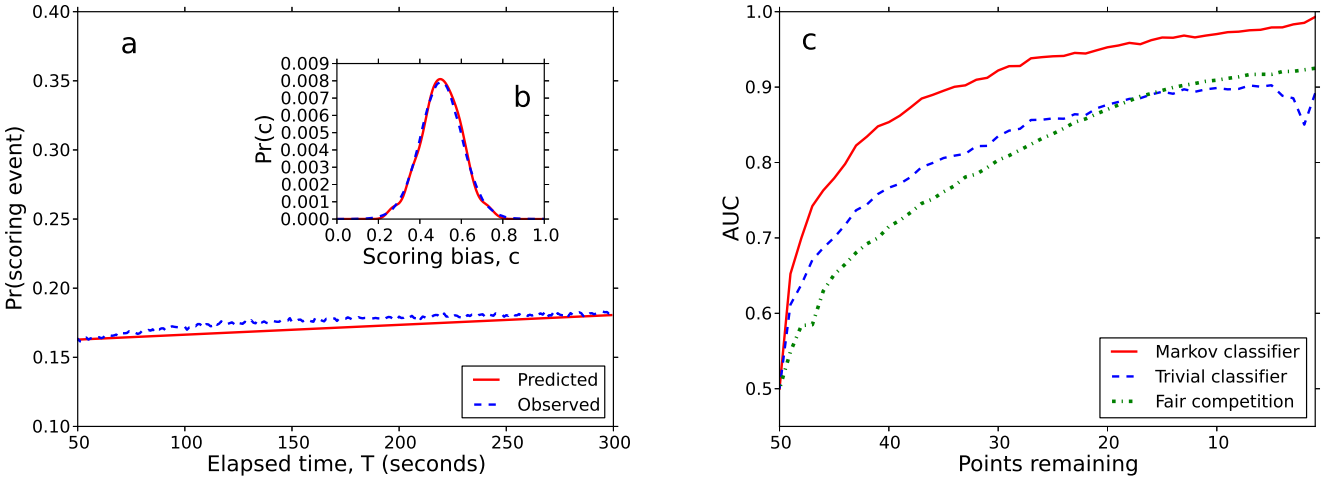

FIG. S1. (a) Global empirical and predicted scoring rates for competitions in *Halo: Reach*, over the window $[50, 300]$ seconds. (b) Global empirical and predicted distribution of competitive advantages (smoothed via a Gaussian kernel). (c) For all competitions, winner predictability (AUC) as a function of team r's points remaining, for three classifiers (see text).

parameter	estimate, global
β balance	29.50 ± 0.21
λ_0 base rate	0.1620 ± 0.0001
α acceleration	$7.00 \times 10^{-5} \pm 0.05 \times 10^{-5}$

TABLE S1. Estimated global scoring tempo and balance parameters, with bootstrap uncertainty estimate.

For a set of competition instances, numerically maximizing Eq. (2) with respect to λ_0 and α, and Eq. (5) with respect to β, produces maximum likelihood parameter estimates $\hat{\lambda}_0$, $\hat{\alpha}$, and $\hat{\beta}$. Uncertainty in these estimates is then calculated as the standard deviation of the bootstrap distribution [7], where we resample compelte competition instances with replacement. Table S1 gives the global parameters estimates and uncertainties, when applied to the full set of *Halo: Reach* competitions.

III. PREDICTING COMPETITION OUTCOMES

For a set of competitions, the predictability of an instance's ultimate winner, after observing only part of the game, provides a second, non-parametric measure non-ideal dynamics. We model scoring as a Markov chain that terminates when a team reaches a score of 50. (In our data, 99% of competitive instances terminate according to this criteria; the remainder from the time limit.)

Suppose an instance has evolved so that teams r and b currently hold scores s_r and s_b. The probability that team r wins the competition is then

$$\begin{aligned}\Pr(r \text{ wins} \,|\, s_r, s_b) = &\Pr(r \text{ wins} \,|\, s_r + 1, s_b) \cdot \hat{c} + \\ &\Pr(r \text{ wins} \,|\, s_r, s_b + 1) \cdot (1 - \hat{c}) \ , \end{aligned} \tag{6}$$

where $\hat{c} = s_r/(s_r + s_b)$ is the current maximum likelihood estimate of r's scoring bias within this instance, and the two probability terms capture the probability that r wins if r (or b) wins the next point. (Because a team's score is cumulative, each state in the Markov chain has only two transitions.) Eq. (6) is then solved recursively by computing $\hat{c}$ for the current state and working backwards to the instances's current state from the winning states where $s_r = 50$ and $s_b < 50$.

We convert this Markov chain into a classifier by predicting that team r wins if $\Pr(r \text{ wins} \,|\, s_r, s_b) > 0.5$. The probability of correctly choosing the winning team in this case is equivalent to computing the AUC statistic over a set of instances. (AUC is defined as the area under the receiver-operating characteristic (ROC) curve [8], and is mathematically equivalent to the Mann-Whitney U test for distinguishing two classes of items.)

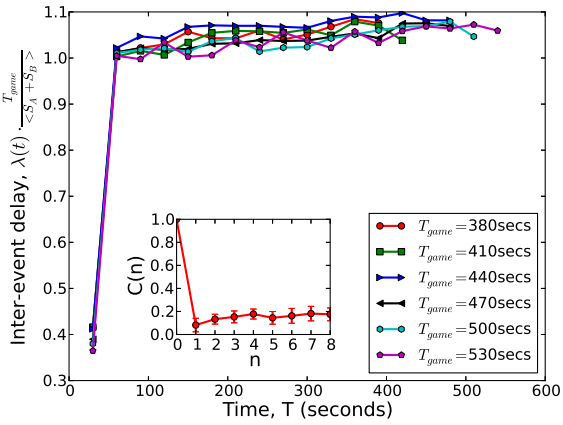

FIG. S2. Average normalized inter-arrival time between scoring events, computed in 30 second intervals, for cohorts of competitions lasting a specific amount of time. (inset) Auto-correlation function $C(n)$ for inter-event times.

Measuring the AUC as a function of the points remaining provides full information about the way the competition's predictability evolves over time. We convert this information into a point measure by computing, with 40 points remaining for r, the AUC for the Markov classifier, which we then divided by the corresponding AUC for an "ideal" classifier (with fixed $c = 1/2$). This provides a direct measure of how much more predictable a real competition's outcome is relative to the ideal model described in the main text.

Using the full data set, Figure S1b shows the full AUC-over-time curves, for the Markov classifier, the ideal classifier ($c = 1/2$), and for a trivial classifier in which at each moment we predict as the winner the team currently in the lead. Our Markov classifier outperforms the trivial classifier because it captures information about the size of the lead, i.e., it includes information about the bias c in the Bernoulli scoring process, and outperforms the ideal classifier because the competitions' dynamics are non-ideal.

IV. TEST OF THE MARKOV ASSUMPTION

We now test the accuracy of our Markov assumption in modeling the scoring dynamics of these competitions. If the arrival times of scoring events roughly follow a memoryless Poisson process, there will be little correlation between the sizes of subsequent delays. The correlation function $C(n)$ provides a direct measure of the accuracy of the Markov assumption, and is calculated as

$$C(n) = \frac{\langle T_i T_{i+n} \rangle - \langle T_i \rangle^2}{\langle T_i^2 \rangle - \langle T_i \rangle^2} \quad , \tag{7}$$

where T_i is the inter-event delay after event i, n is a shift size relative to i, and $\langle . \rangle$ indicates an average over i. A memoryless process matching the Markov assumption in our Bernoulli process will produce $C(n) \approx 0$ for $n > 0$; deviations indicate correlations (or anti-correlations) at the corresponding time scale.

First, a simple rescaling of the observed inter-event delays over the course of competitions of different lengths produces a data collapse (Fig. S2), illustrating relatively little memory in the system. Second, $C(n)$ for our entire sample of competitions (Fig. S2, inset) shows little correlation (memory) at any time scale. Thus, the Markov assumption seems largely justified.

V. MODEL GOODNESS-OF-FIT

We now test the plausibility of our generative model, i.e., how well it matches the underlying data, by comparing simulated competitions against the empirical data along specific statistical measures. This simulation is parametric and uses the estimated parameters from our generative model to define the corresponding probability distributions in the simulator. A close match between the synthetic scoring dynamics and the empirical data along multiple statistical measure is evidence that our generative model accurately captures the basic features of these competitions.

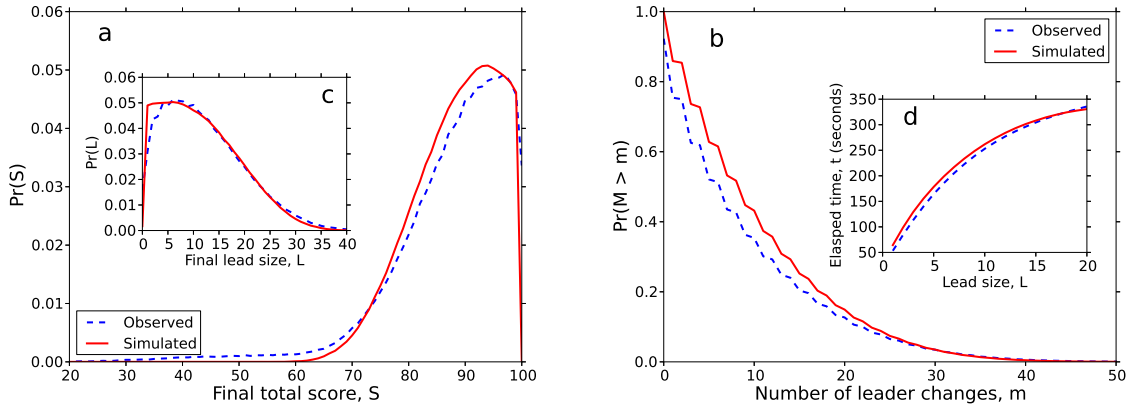

FIG. S3. Comparison of empirical (dashed blue) and simulated (parametric model, red) data for the (a) distribution of final total scores $S = S_r + S_b$, (b) distribution of the number of times the identity of the leading team changes m, (c) distribution of final lead sizes $L = |S_r - S_b|$, and (d) time t elapsed as leader given a lead size of L. The close agreement between data and simulation suggests that our generative model efficiently captures these competitions' dynamics.

The simulation framework is given in Algorithm S1. The competition clock is started at $t = 25$ seconds to account for the early-phase delay in the onset of scoring. The bias in the Bernoulli process is then chosen by drawing a value iid from the estimated Beta distribution with parameter $\hat{\beta}$. While neither of the termination criteria have been reached, delays between scoring events are drawn from the estimated linear non-stationary process with parameters $\hat{\lambda}_0$ and $\hat{\alpha}$. Finally, given that a scoring event occurs, with probability c, a single point is awarded to team r; otherwise, it is awarded to b.

Algorithm S1: COMPETITION SIMULATION()

$t \leftarrow 25$
$s_r \leftarrow s_b \leftarrow 0$
$c \leftarrow$ chooseScoringBias()
while $t < 600$ **and** $s_r < 50$ **and** $s_b < 50$
do
  $T \leftarrow$ interEventDelay()
  if $t + T < 600$
    then
      $\Delta s \leftarrow$ numPoints()
      updateScores(s_a,s_b, Δ s,c)
      $t \leftarrow t + T$
    else break

The goodness-of-fit of the model is measured by comparing the simulated and empirical distributions of (i) the final score S, (ii) the final lead size L (at termination), (iii) the number of leader changes m, and (iv) the amount of time t the leading team stays in the lead given a lead of size L. Notably, each of these four quantities is distinct (although related) to the aspects of the data used to estimate the parametric model's structure, and thus they make reasonable checks on the accuracy of the model. Figures S3a-d show the results of these tests, using 1 million simulated competitions, illustrating very good agreement on all dimensions between simulation and data. Thus, the basic structure of our generative model seems largely justified.

VI. ADDITIONAL RESULTS FOR HOW STRUCTURE SHAPES DYNAMICS

In the main text, we examined four pairs of competition types that each differed on one structural feature: team skill, environmental structure, policies, and resource quality. Figures S4a-d show the estimated distributions of $\Pr(c)$ (parameterized by $\hat{\beta}$) for these four pairs. For each group of instances, the model parameter β was estimated following Section II from the scoring events on the interval $t \in [30, 300]$ seconds of the competition. These times were chosen to exclude biases due to early- and end-phase boundary effects.

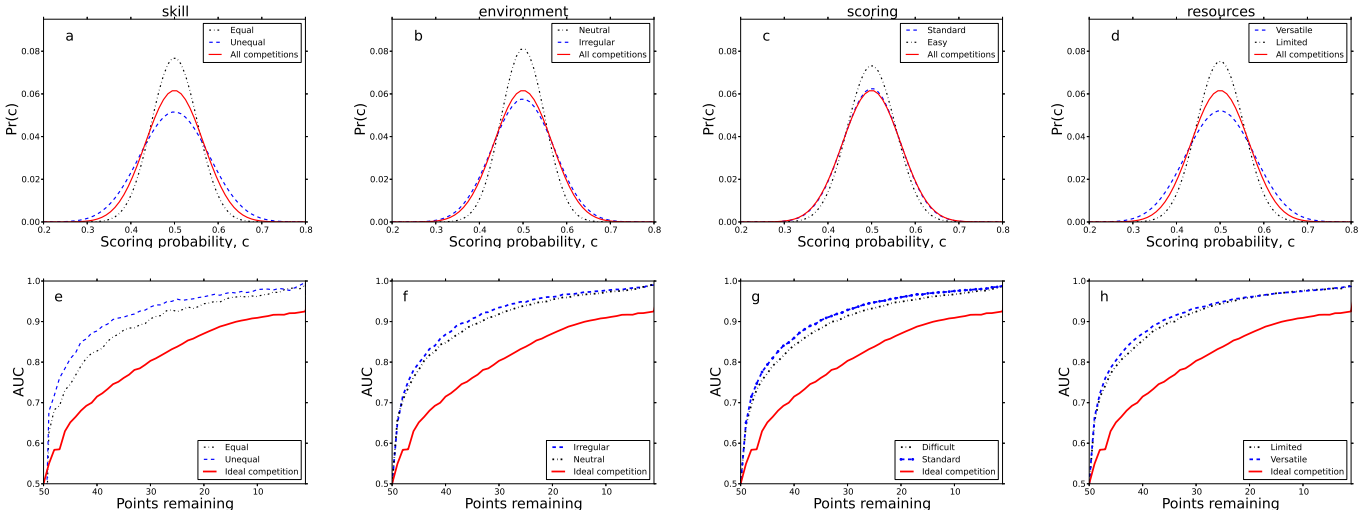

FIG. S4. For the four dimensions discussed in the main text, (a, b, c, d) estimated distribution of scoring biases $\Pr(c)$, and (e, f, g, h) the AUC as a function of points remaining in the competition.

Figures S4e-h show the AUC as a function of points remaining for same competitions, estimated following Section III. In each figure, we show for comparison the AUC curve for an ideal competition ($c = 1/2$). The large gap between the Markov classifier's AUC curve and the ideal curve demonstrates that these competitions are substantially more predictable than ideal competitions. This gap is largest early in the competition, where scores are still relatively far from the scoring limit. We also observe modest gaps between the AUC curves for members of each pair, illustrating that structural features do impact the predictability of competition outcomes.

VII. ADDITIONAL DETAILS OF MULTIVARIATE REGRESSION ANALYSIS

Here we describe additional details of our investigation of how resources, policy, environment, and skill features explain the variance in the values β, λ_0, α, and ρ observed in our data. To quantify the structure of a competition type $\vec{\eta}$, we defined 35 structural features that characterize the different combinations of environment, resources, policies, and teams. Table S4 gives the full list of features, with descriptions, classified into four types: resources (R), environment (E), policies (P), and skill (S). Applied to our data yields 125 unique competition types.

For all competition instances with a particular set of features, we estimated the coordinates $(\beta, \lambda_0, \alpha, \rho)$ following Sections II and III. Regression models were built on each coordinate independently, and robustness checks were conducted to verify these results (see below). Table S5 lists the statistically significant ($p \leq 0.1$) features and corresponding coefficients for all four of our models.

For competition balance β, we first used a linear model $\beta = \theta^T \mathbf{x}$, with a design matrix $\mathbf{x}$ composed of the previously defined 125 observations containing 35 features. Fitting this model via least squares produced $r^2 = 0.716$ ($p \ll 0.001$, F-test), but with strongly skewed residuals. We then fitted the model $\log \beta = \theta^T \mathbf{x}$ to the data, which produced $r^2 = 0.933$ ($p \ll 0.001$, F-test), a marked improvement, and more symmetric residuals. Examining the coefficients, we find that evenly matched teams using medium-to-long-range weapons, competing on large environments without strategic or defensible positions produce more balanced scoring outcomes (larger β).

For the base scoring rate λ_0, a simple linear model yields $r^2 = 0.955$ ($p \ll 0.001$, F-test), indicating that structural features explain almost all the observed variance. The estimated coefficients show that environmental structure features play a dominant role in setting λ_0. In particular, environments that are small, open, and circular correlate best with base scoring rate. In addition to the environment's spatial organization, evenly matched teams also correlate with higher scoring rates. Teams with more experience are likely to be familiar with all terrain options and methods for its exploitation. Environments that are small do not require competitors to spend much time seeking out scoring opportunities (other avatars). Lastly, environments that are open do not provide places to avoid encounters, thus increasing the tempo of competition.

For the acceleration α in the competition tempo, a linear model produces an $r^2 = 0.652$ ($p \ll 0.001$, F-test). We find that few of our features correlate with α, with the exception of long-range weapons and equally-skilled teams, which correlate with smaller α (more ideal competitions). This suggests that in competitions where players are experienced,

	$\log\beta$	λ_0	α	ρ
$\log\beta$	–	0.356	0.053	0.776
λ_0	0.356	–	0.003	0.398
α	0.053	0.003	–	–
ρ	0.776	0.398	–	–

TABLE S2. Coefficients of variation r^2 for pairs of dependent variables. Cells containing no data are either irrelevant or statistically insignificant ($p > 0.1$).

parameter	r^2	p-value
$\log\beta$	0.08	0.98
λ_0	0.12	0.84
α	0.12	0.8
ρ	0.08	0.98

TABLE S3. Regression results after randomly permuting the vectors of 35 independent variables and tuple of 5 scoring dynamics parameters, $(\log\beta, \lambda_0, \alpha, \rho)$.

there is less to learn and thus α is low. This agrees well with the results from λ_0, where more experience leads to a higher base scoring rate.

For the winner predictability ρ, a linear model produces an $r^2 = 0.885$ ($p \ll 0.001$, F-test). Notably, features related to neutral environments and equally-skilled teams correlated with less predictable (more ideal) outcomes. As expected from the correlation between β and ρ (Table S2), features that correlated with greater β typically also correlate with lower ρ.

Finally, we expected changes in policy to have an impact on scoring balance and tempo of events. However, we find that policy type features do not by themselves play a role in controlling these dynamics, once we control for other variables like skill, environmental structure and resources. Specifically, we find that the policy feature coefficients are insignificant in all of our models ($p > 0.1$) and thus we excluded from the results of our best-subset selection.

Tests of model robustness

To test the robustness of our results against spurious correlation, due to the high-dimensionality of our data, we conducted three additional analyses.

First, we consider colinearity among the dependent variables. Table S2 lists the pairwise coefficients of variation r^2, showing a high degree of correlation between ρ and $\log\beta$, modest correlation between $\log\beta$ and λ_0, but little else. To test whether these correlations impact our results, we conducted a MANOVA on a multiple multivariate regression model (Table S6). The results show that the same set of features reported in Table S5 are significant, suggesting that our original results are robust.

Second, we perform a stepwise AIC feature selection procedure to choose the best subset of features under mild regularization. With the exception of α, the results shown in Tables S7, S8, and S9 indicate that the selected features and their weights presented in the original regression analysis are robust. The best-subset selection for α produces a larger list of significant features than in the original model, but a slightly lower r^2. The most significant negative feature, long range resources, is robust to this procedure while equally skilled teams and other resource features are not.

Finally, we perform a randomization test by randomly permuting the dependent variables across the associated features and repeating the original multivariate regression. This randomization destroys any natural correlation between the features and the dependent variable. Table S3 shows the resulting coefficients of variation, none of which are statistically significant. These results further support the robustness of our original results.

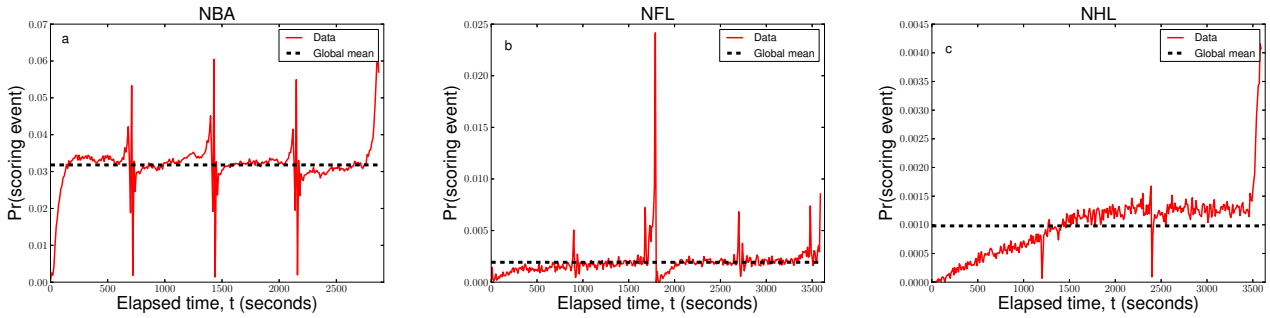

FIG. S5. Patterns in tempo dynamics for (a) professional basketball (NBA), (b) American football (NFL), and (c) hockey (NHL). For each sport, the probability of a scoring event at time t, in regular competition; The global average (dashed) patterns are also shown.

VIII. PLAYER PREFERENCE AND COMPETITION BALANCE

When competitions are predictable they become less enjoyable. In professional sports this manifests itself as fans leaving a stadium well before the end of a game when one team is winning by such a large amount that there is little chance that the trailing team will make a comeback.

In our model system, the same decision can occur for players themselves, who can effectively walk off the field by voluntarily exiting the competition early. For each of the competition types in our sample we calculated the competition dropout rate as

$$\omega = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{\text{at least 1 player quits early}\}, \tag{8}$$

where N is the number of instances of the given type.

From the first 25 million games, we extracted a total of 4.1 million competitive type games that did not contain corrupt data. From these 4.1 million games we selected only those where at least one player left the game early. Using the remaining 1.9 million games we then tested for a correlation between the dropout rate ω and the overall balance β. If players prefer more balanced competitions, as β increases (more ideal competitions), the dropout rate should decrease. A simple linear regression yields the equation $\ln\omega = 1.593 - 1.371\ln\beta$ ($r^2 = 0.43$, $p \ll 0.001$, t-test). These results corroborate our hypothesis, illustrating that the more predictable the scoring dynamics of a competition (small β), the more likely at least one player will exit early. Quantitatively, this relationship predicts that increasing competition balance β by a factor of 1.66 correlates with reducing the early exit probability ω by a factor of 2.

As a caveat, we note that there are several involuntary reasons a player may exit early, e.g., network issues, power loss, system error, being "booted" for excessive friendly fire, and several voluntary reasons unrelated to player engagement, e.g., to join friends in another game, to change competition types, etc. Most of these variables are inaccessible to us for analysis; however, we cannot conceive of a mechanistic relationship between most of these reasons and the scoring balance of a competition. Additional investigation may further illuminate the precise mechanism by which increase in β produce decreased exit rates.

IX. TEMPO PATTERNS IN PROFESSIONAL SPORTS

We study the timing of scoring events in professional basketball, hockey, and American football by analyzing data drawn from 10 consecutive seasons. For each sport, the data contain records of all scoring events that occur in each competition. Each record is annotated with the time at which the event occurred, to the nearest second, the player and corresponding team that won the event, and the event's point value [9].

The timing of scoring events in sports such as professional basketball, hockey, and American football, exhibit similar patterns to those observed in *Reach*. In particular, we observe three distinct, phases of play: an early phase, a middle or steady-state phase, and an end phase. In sports whose games are subdivided into distinct blocks of time (quarters in basketball and American football; thirds in hockey), these patterns are repeated within each block. At the beginning of a period, tempo grows towards a steady-state, a pattern that agrees with players moving from initial set locations on the court or playing field and "warming up" to a well-mixed state. Similar to *Reach*, during the middle, or steady-state phase, tempo remains roughly stationary and scoring events occur with nearly equal probability, illustrated in

Fig. S5 by the scoring rate holding at roughly the full-game mean (indicated by the dashed line). Finally, as each game approaches its final seconds of play, scoring rates increase dramatically, in agreement with teams engaging in more risky strategies (including aggressive clock management through timeouts) in order to score additional points.

While all sports exhibit the three phases observed in *Reach*, there are notable differences in tempo dynamics. Specifically, the end phase in professional sports never tapers, as observed in some competition types in *Reach*. Additionally, because sports have multiple periods of competition within a single game, we observe distinct end-like phases leading up to each. This pattern is most pronounced in football (particularly at the half-way point) and basketball (at the end of each quarter), but also appears to some degree in hockey. Additional features in these time series are attributable either to the particular rules of the game, e.g., in football, some quarters end by players resuming their set locations while other quarters do not, or to the particular dynamics of the game, e.g., in hockey, both players the puck can move very quickly across the rink, leading to very short delays between set positions and a well-mixed state. Finally, base rates λ_0 differ substantially across games, reflecting the fundamental difference in scoring rates in these different types of competitions.

[1] Entertainment Software Association, "Essential Facts about the Computer and Video Game Industry," (2011), http://bit.ly/kLHJ2Q, (access date February, 2012).
[2] W. Mason and A. Clauset (2013) 16th ACM Conference on Computer Supported Cooperative Work and Social Computing.
[3] R. Herbrich, T. Minka, and T. Graepel, Advances in Neural Information Processing Systems **20**, 569 (2007).
[4] M. Ruef, H. E. Aldrich, and N. M. Carter, American Sociological Review **68**, 195 (2003).
[5] T. T. Baldwin, M. D. Bedell, and J. L. Johnson, Acad. of Manag. Journal **40**, 1369 (1997).
[6] P. Balkundi and D. A. Harrison, Acad. of Manag. Journal **49**, 49 (2006).
[7] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, NY, 1993).
[8] J. A. Hanley and B. J. McNeil, Radiology **143**, 29 (1982).
[9] Data Source: STATS LLC, ©2013.

	feature	code	domain	description
resources	loadout_1	R1	$\{0,1\}$	short range and medium range
	loadout_2	R2	$\{0,1\}$	low quality resources
	loadout_3	R3	$\{0,1\}$	long range and grenades
	loadout_4	R4	$\{0,1\}$	short and long range
	loadout_5	R5	$\{0,1\}$	medium range
	vehicles_revenant	R6	$\{0,1\}$	lightly armored vehicle
	vehicles_scorpion	R7	$\{0,1\}$	heavy tank vehicle
	vehicles_mongoose	R8	$\{0,1\}$	unarmored vehicle
	vehicles_ghost	R9	$\{0,1\}$	rapid attack vehicle
	weapons_short	R10	$\{0,1\}$	short range
	weapons_medium	R11	$\{0,1\}$	medium range
	weapons_long	R12	$\{0,1\}$	long range
	weapons_grenades	R13	$\{0,1\}$	grenade type
	weapons_rocket	R14	$\{0,1\}$	rocket launcher
	weapons_unsc	R15	$\{0,1\}$	high-quality only resources
	weapons_covenent	R16	$\{0,1\}$	low-quality only resources
	weapons_both	R17	$\{0,1\}$	high- and low-quality resources
skill	TrueSkill matchmaking	S1	$\{0,1\}$	equally skilled teams
	team size	S2	$\{0,1\}$	4- or 5-person teams
environmental structure	map_open	E1	$\{0,1\}$	open terrain
	map_vertical	E2	$\{0,1\}$	vertical environment
	map_circular	E3	$\{0,1\}$	circular terrain
	map_varied	E4	$\{0,1\}$	no clear organizing principle
	map_corridors	E5	$\{0,1\}$	indoor terrain
	map_bases	E6	$\{0,1\}$	defensible positions
	map_towers	E7	$\{0,1\}$	high ground
	map_transporters	E8	$\{0,1\}$	teleporters, jump pads and vents
	map_outdoor	E9	$\{0,1\}$	outdoor terrain
	map_size_small	E10	$\{0,1\}$	small or medium sized map
	map_size_large	E11	$\{0,1\}$	large arena
	map_size_perim	E12	$\mathbb{R}^+$	perimeter of map, seconds required to run in game
policies	rules_noradar	P1	$\{0,1\}$	HUD radar is off
	rules_noshields	P2	$\{0,1\}$	shield is off
	rules_headshot	P3	$\{0,1\}$	headshot required for kill (SWAT rules)
	rules_snipers	P4	$\{0,1\}$	sniper fight

TABLE S4. Competition features, abbreviations and verbal descriptions, grouped in four categories: resources (R), skill (S), environmental structure (E), and policy (P).

parameter	feature	θ	std. error	t value	$\Pr(>\|t\|)$	r^2
$\log\beta$	E5	1.849	0.320	5.764	$\ll 0.001$	0.933
	E1	1.391	0.371	3.745	$\ll 0.001$	
	E11	1.123	0.141	7.920	$\ll 0.001$	
	S1	0.822	0.034	23.828	$\ll 0.001$	
	E3	0.570	0.256	2.224	0.028	
	E9	0.481	0.076	6.265	$\ll 0.001$	
	R10	−0.354	0.134	−2.642	0.009	
	R8	−0.495	0.215	−2.303	0.023	
	R15	−0.580	0.233	−2.488	0.014	
	E6	−0.813	0.150	−5.414	$\ll 0.001$	
	E2	−1.861	0.252	−7.375	$\ll 0.001$	
	E7	−2.126	0.224	−9.467	$\ll 0.001$	
λ_0	E5	0.082	0.008	9.966	$\ll 0.001$	0.955
	E11	0.059	0.003	16.344	$\ll 0.001$	
	E1	0.045	0.009	4.774	$\ll 0.001$	
	E3	0.029	0.006	4.437	$\ll 0.001$	
	E9	0.023	0.001	12.028	$\ll 0.001$	
	R10	0.008	0.003	2.478	0.014	
	S1	0.005	0.001	6.010	$\ll 0.001$	
	E4	−0.009	0.004	−2.374	0.019	
	R8	−0.011	0.005	−1.995	0.048	
	R13	−0.011	0.004	−2.266	0.025	
	E6	−0.011	0.003	−2.845	0.005	
	R2	−0.015	0.008	−1.873	0.063	
	R1	−0.021	0.008	−2.680	0.008	
	R4	−0.030	0.008	−3.797	$\ll 0.001$	
	R15	−0.032	0.006	−5.444	$\ll 0.001$	
	E2	−0.081	0.006	−12.448	$\ll 0.001$	
	E7	−0.081	0.005	−13.991	$\ll 0.001$	
α	R12	-1.9×10^{-5}	8.1×10^{-6}	−2.449	0.016	0.652
	S1	-2.9×10^{-6}	1.7×10^{-6}	−1.692	0.093	
ρ	E7	0.138	0.022	6.295	$\ll 0.001$	0.885
	E2	0.123	0.024	4.989	$\ll 0.001$	
	R4	0.070	0.030	2.299	0.023	
	E6	0.061	0.014	4.175	$\ll 0.001$	
	R1	0.053	0.030	1.734	0.085	
	R15	0.046	0.022	2.030	0.044	
	R8	0.040	0.021	1.937	0.055	
	E4	0.031	0.015	2.018	0.046	
	R3	0.029	0.015	1.852	0.066	
	R14	−0.030	0.012	−2.366	0.019	
	E9	−0.036	0.007	−4.775	$\ll 0.001$	
	S1	−0.055	0.003	−16.413	$\ll 0.001$	
	E11	−0.089	0.013	−6.410	$\ll 0.001$	
	E5	−0.095	0.031	−3.020	0.003	

TABLE S5. Ordered multivariate regression model coefficients for all standard ("slayer") competitions regressed onto the estimated generative model parameters $\log\beta$, λ_0, α, and predictability measure ρ.

feature	df	Wilks	approx. F	num. df	den. df	Pr($> F$)
R1	1	0.533	21.617	4	99	$\ll 0.001$
R2	1	0.339	48.147	4	99	$\ll 0.001$
R3	1	0.352	45.541	4	99	$\ll 0.001$
R4	1	0.716	9.802	4	99	$\ll 0.001$
R8	1	0.167	123.322	4	99	$\ll 0.001$
R10	1	0.302	57.109	4	99	$\ll 0.001$
R11	1	0.418	34.459	4	99	$\ll 0.001$
R12	1	0.383	39.799	4	99	$\ll 0.001$
R13	1	0.817	5.536	4	99	$\ll 0.001$
S1	1	0.112	194.402	4	99	$\ll 0.001$
R15	1	0.224	85.703	4	99	$\ll 0.001$
E1	1	0.455	29.610	4	99	$\ll 0.001$
E2	1	0.358	44.342	4	99	$\ll 0.001$
E3	1	0.606	16.076	4	99	$\ll 0.001$
E4	1	0.811	5.742	4	99	$\ll 0.001$
E5	1	0.246	75.711	4	99	$\ll 0.001$
E6	1	0.399	37.133	4	99	$\ll 0.001$
E7	1	0.842	4.623	4	99	0.001
E9	1	0.401	36.896	4	99	$\ll 0.001$
E11	1	0.239	78.378	4	99	$\ll 0.001$

TABLE S6. MANOVA results of multiple multivariate regression model, providing a robustness check on the results given in Table S5.

parameter	feature	θ	std. error	t value	$\Pr(>\|t\|)$	r^2
	E5	1.803	0.229	7.867	$\ll 0.001$	
	E1	1.320	0.228	5.779	$\ll 0.001$	
	E11	1.126	0.124	9.029	$\ll 0.001$	
	S1	0.822	0.034	24.153	$\ll 0.001$	
	E3	0.480	0.122	3.919	$\ll 0.001$	
	E9	0.479	0.069	6.888	$\ll 0.001$	
	R13	0.154	0.069	2.243	0.027	
$\log \beta$	R14	0.119	0.074	1.598	0.113	0.933
	R1	-0.322	0.054	-5.952	$\ll 0.001$	
	R3	-0.232	0.092	-2.505	0.013	
	R12	-0.310	0.110	-2.822	0.005	
	R10	-0.367	0.113	-3.232	0.001	
	R8	-0.472	0.181	-2.596	0.01	
	R4	-0.504	0.062	-8.081	$\ll 0.001$	
	R15	-0.644	0.092	-6.931	$\ll 0.001$	
	E6	-0.827	0.130	-6.353	$\ll 0.001$	
	E2	-1.860	0.207	-8.957	$\ll 0.001$	
	E7	-2.093	0.193	-10.840	$\ll 0.001$	
	E5	0.084	0.006	13.770	$\ll 0.001$	
	E11	0.061	0.002	20.759	$\ll 0.001$	
	E3	0.029	0.003	8.648	$\ll 0.001$	
	E9	0.024	0.001	12.383	$\ll 0.001$	
	R10	0.008	0.003	2.794	0.006	
	R3	0.005	0.002	2.080	0.039	
	S1	0.005	0.001	6.085	$\ll 0.001$	
λ_0	E1	0.048	0.005	8.880	$\ll 0.001$	0.954
	R13	-0.009	0.002	-3.979	$\ll 0.001$	
	E4	-0.008	0.002	-3.178	0.001	
	R8	-0.011	0.004	-2.467	0.015	
	E6	-0.012	0.003	-3.860	$\ll 0.001$	
	R2	-0.015	0.005	-2.939	0.004	
	R1	-0.022	0.005	-4.191	$\ll 0.001$	
	R4	-0.031	0.005	-5.852	$\ll 0.001$	
	R15	-0.034	0.004	-8.469	$\ll 0.001$	
	E7	-0.080	0.004	-16.695	$\ll 0.001$	
	E2	-0.081	0.005	-14.457	$\ll 0.001$	

TABLE S7. Ordered multivariate regression model coefficients for all standard ("slayer") competitions regressed onto $\log \beta$, λ_0, selected via stepwise AIC, providing a second check on the robustness of the results in Table S5.

parameter	feature	θ	std. error	t value	$\Pr(>\|t\|)$	r^2
ρ	E7	0.124	0.010	11.934	$\ll 0.001$	0.882
	E2	0.111	0.011	9.943	$\ll 0.001$	
	R4	0.067	0.010	6.444	$\ll 0.001$	
	E6	0.052	0.005	8.998	$\ll 0.001$	
	R1	0.049	0.010	4.958	$\ll 0.001$	
	R8	0.046	0.016	2.779	0.006	
	R15	0.045	0.006	7.335	$\ll 0.001$	
	E4	0.039	0.007	5.456	$\ll 0.001$	
	R2	0.037	0.010	3.533	$\ll 0.001$	
	R3	0.027	0.008	3.420	$\ll 0.001$	
	E9	−0.034	0.006	−4.912	$\ll 0.001$	
	R14	−0.036	0.006	−5.971	$\ll 0.001$	
	S1	−0.055	0.003	−16.763	$\ll 0.001$	
	E5	−0.076	0.010	−7.429	$\ll 0.001$	
	E11	−0.081	0.006	−12.389	$\ll 0.001$	

TABLE S8. Ordered multivariate regression model coefficients for all standard ("slayer") competitions regressed onto ρ selected via stepwise AIC, providing a second check on the robustness of the results in Table S5.

parameter	feature	θ ($\times 10^{-5}$)	std. error ($\times 10^{-6}$)	t value	$\Pr(>\|t\|)$	r^2
α	R3	1.570	2.583	6.077	$\ll 0.001$	0.637
	R11	1.446	3.328	4.345	$\ll 0.001$	
	R2	1.432	2.965	4.832	$\ll 0.001$	
	E5	1.105	2.114	5.226	$\ll 0.001$	
	E3	0.454	2.368	1.918	0.057	
	S1	−0.294	1.689	−1.746	0.083	
	R1	−0.470	2.529	−1.859	0.065	
	R15	−1.591	2.583	−6.157	$\ll 0.001$	
	R8	−1.868	7.159	−2.609	0.010	
	R12	−2.551	2.538	−10.053	$\ll 0.001$	

TABLE S9. Ordered multivariate regression model coefficients for all standard ("slayer") competitions regressed onto α selected via stepwise AIC, providing a second check on the robustness of the results in Table S5.