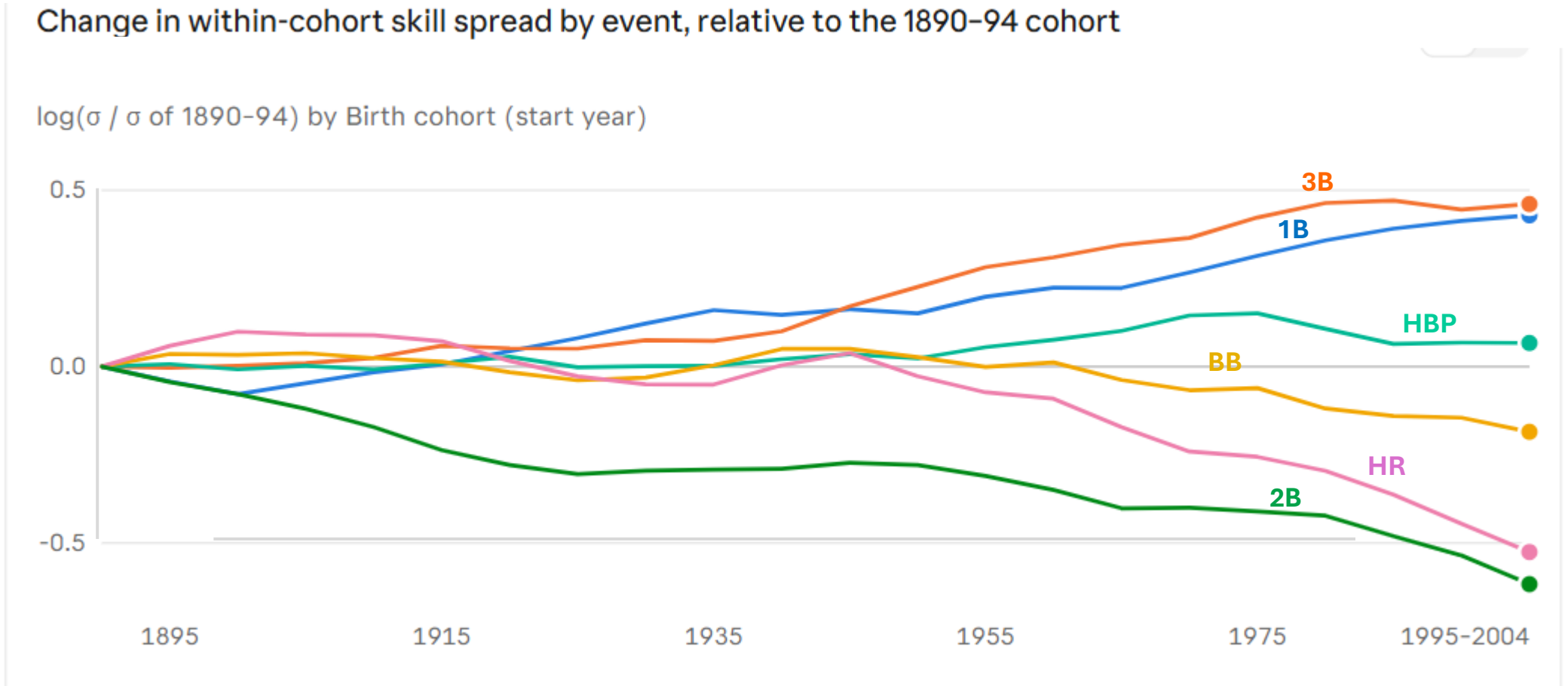
# Hierarchical Bayes model

Player metrics come from a hierarchical Bayesian model of event counts. Level 1 is the data: for each player-season (separate rows for home and road and, for switch-hitters, each side of the plate), the counts of singles, doubles, triples, home runs, walks, hit-by-pitch and outs, modelled as multinomial. Each event's log-odds against an out is the sum of the player's skill and the environment: a season effect, a shared aging curve, the park's effect for that batting side, and home advantage. Level 2 draws each player's skills from his birth cohort's distribution: an overall level, a linear birth-year term, the cohort's mean curvature, and the cohort's spread (the model's principal estimand—as in Berry et al. [1999], the principal research focus for this project is the Gould Conjecture). Level 3 holds the cohort-level walks for mean curvature and log spread. The season effect is a mean-reverting AR(1) path with steps at six rule-change seasons; park effects are estimated per park, event and batting side. The top level holds weakly informative priors on event levels, the birth-year term, the level and trend of skill spread, step sizes, home advantage and park spread.

The model faces the age–period–cohort identification problem: any linear trend can be shared among season, aging and cohort with an identical fit. It is handled by convention: the season effect's linear trend is set to zero, so within-career change fixes the aging slope and between-cohort differences go *to the cohorts* as opposed to the period, which can be understood as a secular change in "difficulty" of game conditions bearing on the relevant offensive event. A persistent AR(1) prior on the season path closes the related curved ridge. The split is never interpreted; the within-cohort spread, the estimand critical to the Gould conjecture, does not depend on it. The pin, however, can be adjusted, to reflect any desired or posited allocation of the linear trend between the cohorts and "game difficulty."






**Decreasing trends signify compressing talent dispersion and hence increasing cohort skill. By the same token, increasing trend signifies growing dispersion and reduced cohort skill.**

$$p(\mathrm{y},\theta,u,\omega,x,A,\phi,\alpha,\beta,\tau,\delta,J,\lambda,\kappa) =$$
$$\underbrace{p(\mathrm{y}\mid\theta,\gamma,A,P,\lambda)}_{\text{level 1: data}}\ \underbrace{p(\theta\mid\alpha,\beta,\mu,\sigma)}_{\text{level 2: players}}\ \underbrace{p(u)p(\omega)}_{\text{level 3: cohorts}}\ \underbrace{p(x)p(A)p(\phi\mid\kappa)}_{\text{environment}}\ \underbrace{p(\alpha)p(\beta)p(\tau)p(\delta)p(J)p(\lambda)p(\kappa)}_{\text{top level}}$$

$$\mu = Qu, \qquad \sigma = \exp\left(\tau + \delta\frac{m-\bar{m}}{100} + Q\omega\right), \qquad \gamma = \Pi\left[\sum_j J_j \mathrm{1}[s \geq s_j] + x\right], \qquad P = P(\phi; w, h)$$

## Notation and levels

Indices: i player; c birth cohort; s season; a age (20–40); k park; d batting side; n observation (player-season, home or road); e event (1B, 2B, 3B, HR, BB, HBP), with the out as the reference outcome. Rows follow the factorization: level 1, level 2, level 3, environment, top level, then quantities calculated from them and fixed inputs.

| Symbol | Meaning | Level / factor | Distribution or definition |
|---|---|---|---|
| $y$ | Observed counts of the seven outcomes for each player-season, home and road, side of plate batted | Level 1: data | Multinomial(PA, π): PA = plate appearances (fixed); π = event probabilities, the softmax of the log-odds |
| $\theta_{i,e}$ | Player i's skill at event e: log-odds of e against an out, net of the aging curve's PA-weighted average level, the PA-weighted mean season and the average (unweighted) park. To report at a specific age (e.g. 27), add A at that age | Level 2: players (enters level 1) | $N(\alpha + \beta(b_i - \bar{b}) + \mu_c, \sigma_c)$ |
| $u_{c,e}$ | Random walk across birth cohorts; μ is its curvature | Level 3: cohorts | Random walk, step 0.08 (fixed) |
| $\omega_{c,e}$ | Random walk across birth cohorts giving the curvature of the log-spread | Level 3: cohorts | Random walk, step 0.05 (fixed) |
| $x_{s,e}$ | Season path before the rule-change steps and the pin | Environment | AR(1), ρ = 0.9, long-run SD 0.15 (fixed) |
| $A_{a,e}$ | Aging effect at age a (20–40), shared by all players | Environment (enters level 1) | Random walk over ages, step 0.03 (fixed); PA-weighted level removed |
| $\varphi_{k,d,e}$ | Park k's effect on event e for batting side d (left or right), constant over the park's life; none for HBP. Batters with unknown side get half of each side's effect | Environment | Normal with SD $\kappa_e$ (same κ for both sides), drawn separately for each side; sums to zero across parks for each event and side |
| $\alpha_e$ | Overall level of event e (the "grand" level) | Top level | N(−3, 1.5) |
| $\beta_e$ | Linear trend in skill by birth year; absorbs the unidentified trend, never interpreted | Top level | N(0, 0.05) per year |
| $\tau_e$ | Average log-spread of skill across cohorts | Top level | N(−1, 1) |
| $\delta_e$ | Linear trend in log-spread per century (negative = compression) | Top level | N(0, 1) per century |
| $J_{j,e}$ | Size of the season-path step at rule change j | Top level | N(0, 0.5) |
| $\lambda_e$ | Home-field advantage for event e | Top level (enters level 1) | N(0, 0.3) |
| $\kappa_e$ | Spread of park effects for event e | Top level | Half-normal(1) |
| $\mu_{c,e}$ | Cohort c's curvature away from the birth-year line | Calculated (used at level 2) | $\mu = Q\,u$ |
| $\sigma_{c,e}$ | Within-cohort SD of skill e: **the Gould estimand** | Calculated (used at level 2) | $\sigma = \exp(\tau + \delta(m_c - \bar{m})/100 + Q\,\omega)$ |
| $\gamma_{s,e}$ | Season effect: league environment for event e in season s (ball, rules, opposing pitching and fielding) | Calculated (used at level 1) | $\gamma = \Pi[\Sigma_j J_j \mathrm{1}[s \geq s_j] + x]$ |
| $P_{n,e}$ | Park effect the player actually faced in observation n | Calculated (used at level 1) | Home: home park's φ. Road: road parks' φ weighted by $w_{n,k}$ |
| $Q$ | Removes a walk's intercept and linear trend in cohort mid-year, every cohort counted equally (unweighted) | Operator | — |
| $\Pi$ | Removes the season path's PA-weighted mean and linear trend (the APC pin) | Operator | — |
| $b_i$, $\bar{b}$ | Player i's birth year, and the mean birth year | Data | — |
| $m_c$, $\bar{m}$ | Cohort c's mid birth year, and the unweighted mean of the cohort mid-years | Data | — |
| $s_j$ | Rule-change years: 1911, 1920, 1963, 1969, 2016, 2023 | Fixed | — |
| $\mathrm{1}[s \geq s_j]$ | 1 from year $s_j$ onward, 0 before (makes each step permanent) | Indicator | — |
| $w_{n,k}$ | Share of a player's road plate appearances in park k that season | Data | — |
| $h$ | 1 for home observations, 0 for road | Data | — |

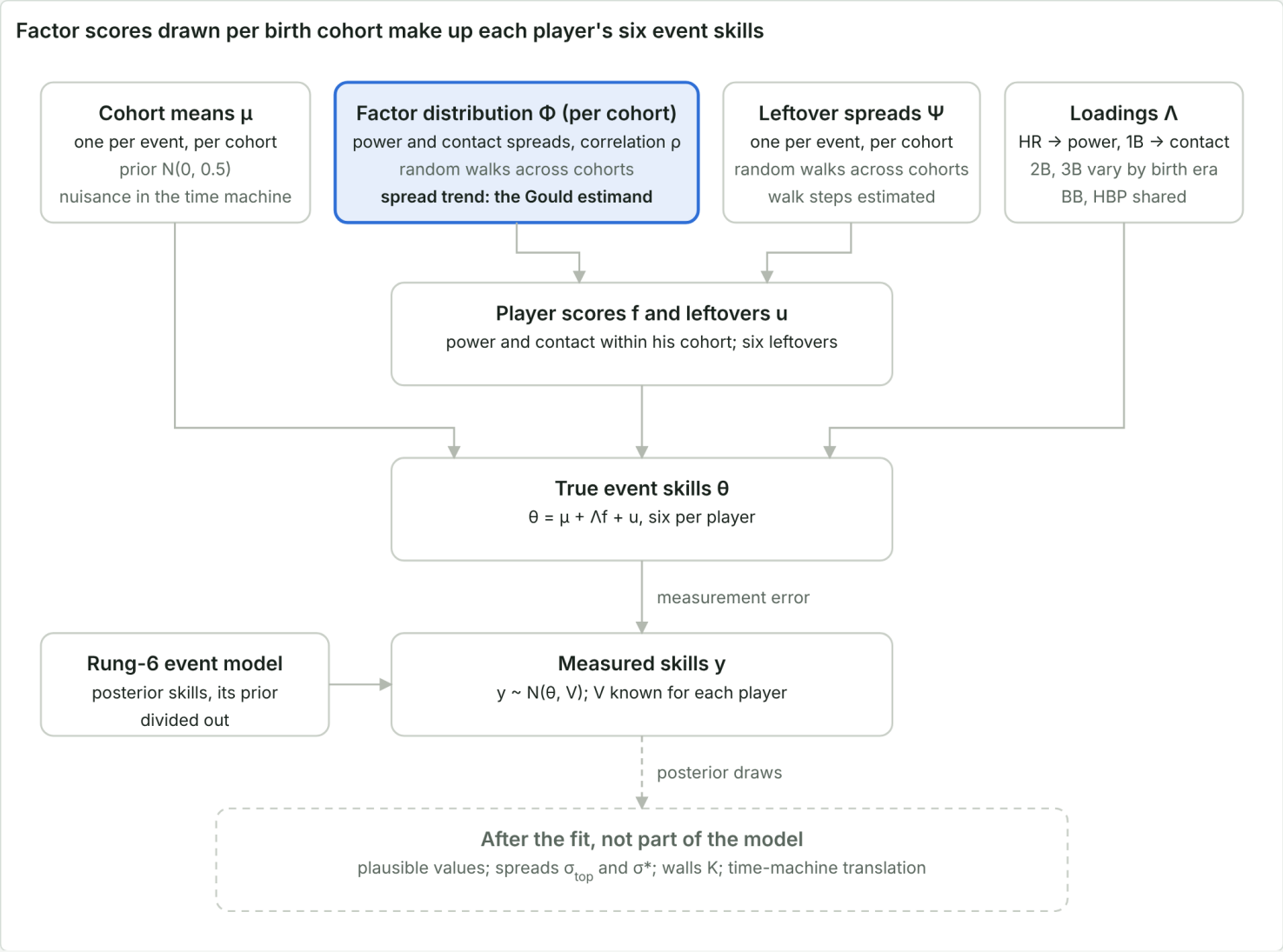
# Time machine transportation

## *part 1: offensive production factors*

The Time Machine performs **three essential operations** to "transport" players to different eras. The **first** is the translation of their performances into two factors, power and contact. This step is implemented to accommodate the contribution that evolving game conditions have made to the elements of offensive production over time. The weighting of the six offensive events (singles, doubles, triples, homeruns, walks, and HBP) are cohort specific and shift over time. In particular, triples shift from an element of "power" in the early part of the twentieth century to an element of "contact" (one that signifies the conjoining of this skill with speed) in the later part; likewise, doubles migrate from contact to power, leaving singles and triples as the principal "contact" elements by century's end.

The factor analysis comprises its own hierarchical Bayesian model. *Level 1* is the data: the players' observed six offensive event performances. *Level 2* consists in the players' latent-skill estimates, calculated consistently with the HB latent-skill model ("rung 6"). *Level 3* are the players cohort-specific factor scores, fed by the top level, which stipulates the two factor solution and implements the random-walk determination of the spreads/distribuitons for each cohort factor.

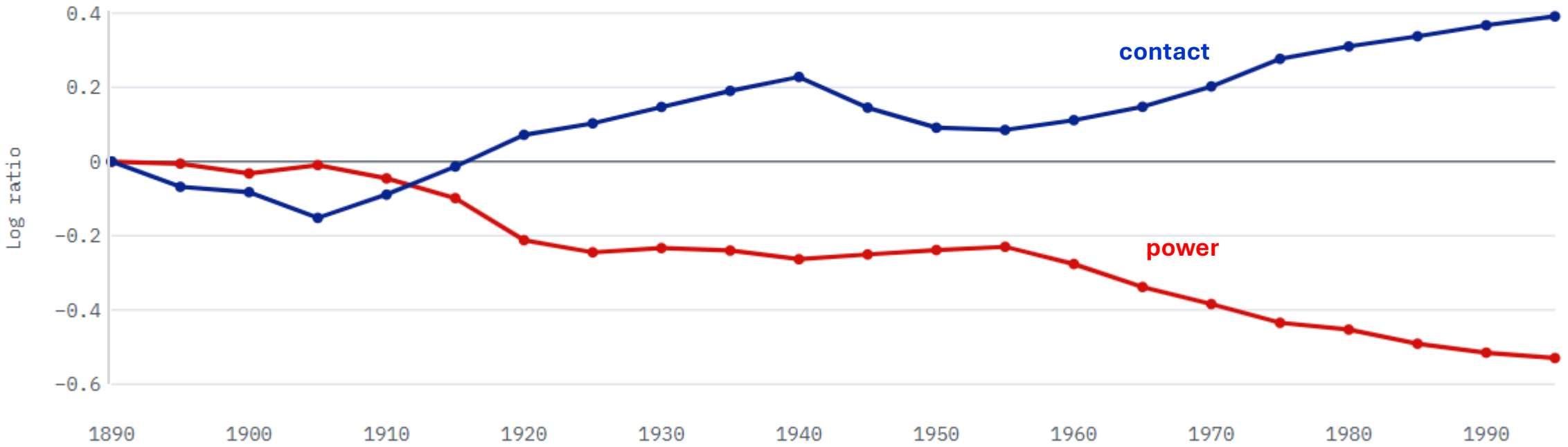
Consistent with the Gould Conjecture, the relative dispersion of cohorts along the factors is understood by the Time Machine to signfity how skillful cohorts are in relation to one another. Because the top performers converge on the "human performance wall" as a sport matures, cohorts characterized by tigher performance spreads are understood to be more talent rich than ones characterized by more diffuse spreads.



# COHORT SKILL COMPARISON

Again, cohorts are assessed in relation to one another based on estimated true dispersion or spread of player skills within each: a diffuse distribution signifies a less talent-rich pool than a more compact one. Cohort spreads are measured on the upper tail (75th–95th percentile) of both of its skill dimensions, contact and power.

From the 1890–94 to the 1995–2004 cohort, the power spread contracted 41% (log ratio –0.53), meaning power hitters became progressively more skillful in each cohort. For contact, the spread expanded 48% (log ratio 0.39)—i.e., contact hitters became less talent rich as a class.



# Time machine transportation

## *part 2: the walls…*

The **second** essential operation for Time Machine transportation comprises the performance "walls." The Gould Conjecture asserts that as an athletic endeavor matures, skill heterogeneity compresses as the top performers approach the "wall" of human performance. The Time Machine ranks cohorts based on estimates of the sill dispersion of their members; it transforms dispersion into quantitative measures of cohort skill disparities by determine  proximat its mean is to the "wall."

Cohort skill dispersions are measured in $\sigma^*$, a standard-deviation approximation extracted from gap between the cohort's $75^{th}$ and $95^{th}$ percentile. This measure of dispersion is used in lieu of conventional SD to avoid the distorting impact of long low-skill tails.

Every cohort's mean skill level (along each dimension, "power" and "contact") is assumed to be the same number of  $\sigma^*$s—*K*—from the "wall." Accordingly, for each cohort, $K \times \sigma^*$ supplies a "wall proximity" score.

*But where is the wall located?*

The Time Machine identifies two: the *close wall* and the *far wall*.

The close wall effectively assumes that the wall has in fact already been reached by someone in baseball history.  That someone is the player who is the greatest number of $\sigma^*$s (determined by draws from his factor-score posterior) above the mean in his cohort along one or the other dimension. They are Hank Greenberg (3.67) for power and Ty Cobb (3.70) for contact.

The far wall assumes that the highest level achievement is yet to be reached. In sum, it reflects a number of $\sigma^*$s that were it realized would mean that some player's latent-skill estimated performance at age 27 would break a single-season performance record in any of the six offensive events. Players only set major league records in extremely productive single seasons; a player projected to be *that good* for an average season at age 27 would be a giant among men.

It is impossible to say which of these "walls"—or which of any number of other plausible ones—is "correct."  The choice of which to use reflects the prior—the philosophical inclinantions or simply the imagantive preferences—of the Time Machine user.  In general, if one tends to think that the greatest ever players are better than today's greats, then the "near wall" will likely seem more plausible. If one thinks players have become more proficient at "what counts" over time, than the "far wall" will likely seem more sensible—although because today's players are now being selected more more for power profieciency than contact proficiency, their edge will show up principally along the former dimension.

# Time machine transportation

## *part 3: translation...*

The **third** and final element of BGW Time Machine travel involves simple mathetmical translation. It consists of operations that turn factor scores, cohort wall proximity scores, and HB Model cohort offensive-event dispersion levels into baseball performance metrics.

The user can decide what season to transport players to. Each season maps onto a *host cohort*—the one who contain players who would be 27 years old.

Each player's factor score on each dimension is then moved to that season. Effectively, each player's factor score is adjusted by a quantity equal to the difference between the host cohort's wall proximity score and the player's own cohort's wall proximity score. Thus, players who are transported to a cohort less skillful than their own—ones with higher (more remote) wall proximity scores—will get a score boost; by the same token, those transported to a more skillful cohort, will suffer a score degradation—one equal to how much further their own cohort is from the wall than is the host.

The resulting season-specific factor scores are then applied to each player's own estimated latent-skill for each offensive event (including an appropriate "leftover" or residual Bayesian adjstument to capture player deviations from the factor-predicted values) to generate baseball-scale outcomes.

***Run-production*** is itself a derivative metric based on estimated "runs above average" associated with weighted on-base average (wOBA), using weights for the season in question.

## *part 4: fantasy destinations*

The Time Machine is nonjudgmental. If you'd like to examine how players perform in acontexual environments, you are free to do so!

Select ***uber season*** and every player is evaluated in a timeless season. Based on 1900-2025 mean weights, the Uber season translates every player's age-27 peak skills into performance outcomes that privilege the game conditions of no single era.

Select **own era** and every player's performance is presented in numbers that reflect the HB-model latent-skill estimate of his age 27 performance in the season in which he was 27.

Select **neutral** and all cohort rankings disappear. Players' age 27 performances are presented in terms that reflect only their latent skill estimates relative to those of others in their own cohort.

# Factor model

$$p(y,\theta,f,u,\mu,\Lambda,s,r,h,\tau) = \underbrace{p(y \mid \theta)}_{\text{level 1: data}}\ \underbrace{p(\theta \mid \mu,\Lambda,f,u)\,p(f \mid s,r)\,p(u \mid h)}_{\text{level 2: players}}\ \underbrace{p(s \mid \tau)\,p(r \mid \tau)\,p(h \mid \tau)\,p(\mu)}_{\text{level 3: cohorts}}\ \underbrace{p(\Lambda)\,p(\tau)}_{\text{top level}}$$

$$\theta_i = \mu_c + \Lambda_{g(c)} f_i + u_i, \qquad \Phi_c = D_c\,R(\rho_c)\,D_c, \quad D_c = \operatorname{diag}\!\left(e^{s_{c,1}}, e^{s_{c,2}}\right), \quad \rho_c = \tanh r_c, \qquad \Psi_c = \operatorname{diag}\!\left(e^{2h_{c,e}}\right)$$

$\theta_i$ is a deterministic function of its parents, so level 2 amounts to $f_i \sim \mathcal{N}(0, \Phi_c)$ and $u_i \sim \mathcal{N}(0, \Psi_c)$. For each player $i$ in cohort $c$, the fit evaluates the marginal likelihood with $f_i$ and $u_i$ integrated out:

$$y_i \sim \mathcal{N}\!\left(\mu_c,\ \Lambda_{g(c)}\,\Phi_c\,\Lambda_{g(c)}^{\top} + \Psi_c + V_i\right)$$

# Inputs from HB latent-skill model (“rung 6”)

$$\hat{z}_i = L_i\,S_i^{-1} a_i, \qquad L_i = \left(S_i^{-1} - I\right)^{-1}, \qquad y_i = \tilde{D}_c\,\hat{z}_i, \qquad V_i = \tilde{D}_c\,L_i\,\tilde{D}_c, \qquad \tilde{D}_c = \operatorname{diag}(\tilde{\sigma}_{c,e})$$

Rung 6's skill prior is divided out so the factor model's own prior is not stacked on it (no double shrinkage). This makes the setup two-stage: rung 6's posterior enters as a normal summary, and nothing flows back.

# Slope of cohort dispersion trend (from posterior draw)

$$T_k = \hat{b}_k\,(m_C - m_1), \qquad \hat{b}_k = \frac{\sum_c w_c\,(m_c - \bar{m}_w)\,(s_{c,k} - \bar{s}_{k,w})}{\sum_c w_c\,(m_c - \bar{m}_w)^2}$$

# Time machine transportation

$$\sigma_{c,k}^{\text{top}} = \frac{q_{.95}\left(f_{\cdot,k}^{(d)}\right) - q_{.75}\left(f_{\cdot,k}^{(d)}\right)}{0.9703}, \qquad \log \sigma_{c,k}^{*} = \bar{\ell}_k + \frac{T_k}{S_k}\left(\log \sigma_{c,k}^{\text{top}} - \bar{\ell}_k\right)$$

A player in cohort $c$ is translated into a **target game** $G$, which supplies target spreads $\sigma_G^*$ and $\psi_G$, loadings $\Lambda_G$, baselines $c_G$, run values $v_G$ and league rates $\bar{\pi}_G$:

$$\text{Neutral: } f_k' = f_k \frac{\sigma_{G,k}^*}{\sigma_{c,k}^*} \qquad \text{Walls: } f_k' = f_k + K_k\left(\sigma_{G,k}^* - \sigma_{c,k}^*\right), \quad K_k = \max_{i:\,\mathrm{PA}_i \ge 3000} \operatorname{median}_d \frac{f_{i,k}^{(d)}}{\sigma_{c(i),k}^*}$$

$$u_e' = u_e \frac{\psi_{G,e}}{\psi_{c,e}}, \qquad \eta = c_G + \Lambda_G f' + u', \qquad \pi_e = \frac{e^{\eta_e}}{1 + \sum_{e'} e^{\eta_{e'}}}, \qquad \text{runs}_{600} = 600 \sum_e v_{G,e}\,(\pi_e - \bar{\pi}_{G,e})$$

**Season** $T$ (target cohort $t$, births in $T-27$): the game as it was that season.

$$\left(\sigma_G^*,\ \psi_G,\ \Lambda_G,\ c_G,\ v_G,\ \bar{\pi}_G\right) = \left(\sigma_t^*,\ \psi_t,\ \Lambda_{g(t)},\ c_T,\ v_T,\ \bar{\pi}_T\right)$$

**Uber season** $U$: a game no one played, the average of the $N = 126$ seasons $T = 1900, \ldots, 2025$, one weight per season, each season $T$ with its target cohort $t(T)$:

$$c_U = \frac{1}{N}\sum_T c_T, \qquad v_U = \frac{1}{N}\sum_T v_T, \qquad \log \sigma_{U,k}^* = \frac{1}{N}\sum_T \log \sigma_{t(T),k}^*, \qquad \log \psi_{U,e} = \frac{1}{N}\sum_T \log \psi_{t(T),e}$$

$$\Lambda_U = \frac{1}{N}\sum_T \Lambda_{g(t(T))}, \qquad \bar{\pi}_{U,e} = \frac{\sum_i \mathrm{PA}_i\,\pi_{i,e}}{\sum_i \mathrm{PA}_i} \quad \text{(all 5,579 players, translated into } U\text{, per setting)}$$

| Symbol | Meaning | Level / factor | Distribution or definition |
|---|---|---|---|
| **MODEL** | | | |
| $y_i$ | Player $i$'s six event skills as rung 6 measured them, with rung 6's prior divided out: log-odds deviations from rung 6's cohort prediction | Level 1: data | $\mathcal{N}(\theta_i, V_i)$ |
| $V_i$ | Measurement covariance of $y_i$ (6×6), from the rung-6 posterior; differs by player | Level 1: data | Fixed input |
| $\theta_i$ | Player $i$'s true six event skills | Level 2: players (enters level 1) | $\mu_c + \Lambda_{g(c)} f_i + u_i$ |
| $f_i$ | Factor scores: power and contact, the player's position within his cohort | Level 2: players | $\mathcal{N}(0, \Phi_c)$ |
| $u_{i,e}$ | Event-specific leftovers: skill on event $e$ not explained by the factors | Level 2: players | $\mathcal{N}(0, \psi^2_{c,e})$, independent |
| $\mu_{c,e}$ | Cohort $c$'s mean on event $e$ (the cohort level; a nuisance in the time machine) | Level 3: cohorts | $\mathcal{N}(0, 0.5^2)$, independent across cohorts |
| $s_{c,k}$ | Log SD of factor $k$ in cohort $c$: **the Gould estimand at factor level** | Level 3: cohorts | Random walk: $s_{1,k} \sim \mathcal{N}(-1, 1)$; $s_{c,k} = s_{c-1,k} + \tau^s_k \varepsilon$ |
| $r_c$ | Power–contact correlation on the unconstrained scale; $\rho_c = \tanh r_c$ | Level 3: cohorts | Random walk: $r_1 \sim \mathcal{N}(0, 1)$; $r_c = r_{c-1} + \tau^r \varepsilon$ |
| $h_{c,e}$ | Log leftover SD for event $e$: $\psi_{c,e} = e^{h_{c,e}}$ | Level 3: cohorts | Random walk: $h_{1,e} \sim \mathcal{N}(-2, 0.75)$; $h_{c,e} = h_{c-1,e} + \tau^h_e \varepsilon$ |
| $\Lambda_g$ | Loadings (6×2) for birth era $g$: how much each event reflects power and contact | Top level | Markers fixed: HR $(1, 0)$, 1B $(0, 1)$. 2B and 3B rows free in each era (partial invariance, MP2); BB and HBP rows shared across eras. Free entries $\mathcal{N}(0, 1)$ |
| $\tau^s_k, \tau^r, \tau^h_e$ | Step sizes of the random walks (2 + 1 + 6): how much a cohort may differ from the one before | Top level | $\text{Half-normal}(0.1)$ each |
| $\varepsilon$ | Standardized steps of the walks (non-centred) | — | $\mathcal{N}(0, 1)$ |
| $\Phi_c$ | Factor covariance of cohort $c$ (2×2) | Calculated (used at level 2) | $D_c R(\rho_c) D_c$ |
| $\rho_c$ | Power–contact correlation of cohort $c$ | Calculated | $\tanh r_c$ |
| $\Psi_c$ | Leftover covariance of cohort $c$ (6×6 diagonal) | Calculated (used at level 2) | $\text{diag}(\psi^2_{c,e})$ |
| $T_k$ | Verdict trend of factor $k$: fitted change in log SD, first to last cohort | Calculated per draw | Weighted slope × span (above); median and 89% interval |
| **INPUTS AND DATA** | | | |
| $a_i, S_i$ | Rung-6 posterior mean and covariance of $z_i$, player $i$'s skills standardized by his cohort's spread (prior $\mathcal{N}(0, I)$) | Input (rung 6, run x1) | — |
| $\hat{z}_i, L_i$ | Data-only mean and covariance of $z_i$, rung 6's prior divided out | Input | $L_i = (S_i^{-1} - I)^{-1}$ (eigenvalue floor $10^{-3}$); $\hat{z}_i = L_i S_i^{-1} a_i$ |
| $\tilde{\sigma}_{c,e}$ | Rung-6 posterior median of cohort $c$'s skill spread on event $e$ | Input | — |
| $m_c$, $w_c$ | Cohort $c$'s mid birth year; its number of players (trend weights) | Data | — |
| $\text{PA}_i$ | Career plate appearances; players enter at 300, 1,000 or 2,000+ (settings) | Data | — |
| **TIME MACHINE** | | | |
| $f_i^{(d)}$ | Plausible values: a draw of $f_i$ from its posterior given draw $d$ of the parameters (unshrunk spread) | Calculated | 200 draws |
| $\sigma^{\text{top}}_{c,k}$ | Upper-tail spread: 75th–95th percentile gap of the cohort's plausible values, in SD-equivalent units | Calculated | Smoothed over three neighbouring cohorts |
| $\sigma^{*}_{c,k}$ | Verdict-anchored spread: $\sigma^{\text{top}}$'s shape, with its trend rescaled to $T_k$; used by all three settings | Calculated | $\bar{\ell}_k$ = player-weighted mean of $\log \sigma^{\text{top}}$; $S_k$ its trend |
| $K_k$ | Wall height in spreads above each cohort's mean | Calculated / stipulated | Near wall: highest standing reached; far wall: highest wall with no translated record breach |
| $f'_i$, $u'_i$ | Translated factor scores and leftovers | Calculated | Above |
| $c_{T,e}$ | Season $T$'s baseline log-odds for event $e$, calibrated so the season's actual players reproduce its league rates | Calculated | — |
| $v_{T,e}$, $\bar{\pi}_{T,e}$ | Season $T$'s run value (own RE24) and league rate for event $e$ | Data | 1900–07 FanGraphs values calibrated to RE24 |

| | | | |
|---|---|---|---|
| **TIME MACHINE** | | | |
| $f_i^{(d)}$ | Plausible values: a draw of $f_i$ from its posterior given draw $d$ of the parameters (unshrunk spread) | Calculated | 200 draws |
| $\sigma^{\text{top}}_{c,k}$ | Upper-tail spread: 75th–95th percentile gap of the cohort's plausible values, in SD-equivalent units | Calculated | Smoothed over three neighbouring cohorts |
| $\sigma^{*}_{c,k}$ | Verdict-anchored spread: $\sigma^{\text{top}}$'s shape, with its trend rescaled to $T_k$, the factor model's own trend in its estimated factor SDs (not rung 6's per-event $\sigma$); used by all three settings | Calculated | $\bar{\ell}_k$ = player-weighted mean of $\log \sigma^{\text{top}}$; $S_k$ its trend |
| $K_k$ | Wall height in spreads above each cohort's mean | Calculated / stipulated | Near wall: highest standing reached; far wall: highest wall with no translated record breach |
| $f'_i$, $u'_i$ | Translated factor scores and leftovers | Calculated | Above |
| $G$ | Target game: what a translation places a player in. Supplies $\sigma^*_G, \psi_G, \Lambda_G, c_G, v_G, \bar{\pi}_G$ | Representation choice | A season $T$, or the Uber season $U$ |
| $c_{T,e}$ | Season $T$'s baseline log-odds for event $e$, calibrated so the season's actual players reproduce its league rates | Calculated | — |
| $v_{T,e}$, $\bar{\pi}_{T,e}$ | Season $T$'s run value (own RE24) and league rate for event $e$ | Data | 1900–07 FanGraphs values calibrated to RE24 |
| $c_U$, $v_U$ | Uber season's baseline log-odds and run values | Calculated | Means over the 126 seasons, one weight per season ($c$ averaged in log-odds) |
| $\sigma^*_U$, $\psi_U$ | Uber season's target factor and leftover spreads | Calculated | Geometric means over the seasons' target cohorts $t(T)$ |
| $\Lambda_U$ | Uber season's loadings | Calculated | Mean over the seasons of $\Lambda_{g(t(T))}$ |
| $\bar{\pi}_U$ | Uber season's league rates, for centring runs | Calculated | Model population's PA-weighted mean of $\pi$ in $U$, per setting |