Psychometric Modeling of Decision Making Via Game Play

CIG 2013, Niagara Falls, Canada

Kenneth W. Regan[1] Tamal Biswas
University at Buffalo (SUNY)

12 Aug., 2013

[1] Sites: http://www.cse.buffalo.edu/~regan/chess/fidelity/ (my homepage links), http://www.cse.buffalo.edu/~regan/chess/ratings/ (not yet linked).

A Predictive Analytic Model

1. Domain: A set of decision-making situations t.
   Chess game turns

A Predictive Analytic Model

1. Domain: A set of decision-making situations t.
   Chess game turns
2. Inputs: Values v_i for every option at turn t.
   Computer values of moves m_i

A Predictive Analytic Model

1. Domain: A set of decision-making situations t.
   Chess game turns
2. Inputs: Values v_i for every option at turn t.
   Computer values of moves m_i
3. Parameters: $s, c, \ldots$ denoting skills and levels.
   Trained correspondence to chess Elo rating E

A Predictive Analytic Model

1. Domain: A set of decision-making situations t.
   Chess game turns
2. Inputs: Values v_i for every option at turn t.
   Computer values of moves m_i
3. Parameters: $s, c, \ldots$ denoting skills and levels.
   Trained correspondence to chess Elo rating E
4. Defines *fallible agent* $P(s, c, \ldots)$. **A Player.**

A Predictive Analytic Model

1. Domain: A set of decision-making situations t.
   Chess game turns
2. Inputs: Values v_i for every option at turn t.
   Computer values of moves m_i
3. Parameters: $s, c, \dots$ denoting skills and levels.
   Trained correspondence to chess Elo rating E
4. Defines *fallible agent* $P(s, c, \dots)$. **A Player.**
5. Main Output: Probabilities $p_{t,i}$ for $P(s, c, \dots)$ to select option i at turn t.

A Predictive Analytic Model

1. Domain: A set of decision-making situations t.
   Chess game turns
2. Inputs: Values v_i for every option at turn t.
   Computer values of moves m_i
3. Parameters: $s, c, \dots$ denoting skills and levels.
   Trained correspondence to chess Elo rating E
4. Defines *fallible agent* $P(s, c, \dots)$. **A Player.**
5. Main Output: Probabilities $p_{t,i}$ for $P(s, c, \dots)$ to select option i at turn t.
6. Derived Outputs:
   - Aggregate statistics: *move-match* MM, *average error* AE, ...
   - Projected confidence intervals for those statistics.
   - "Intrinsic Performance Ratings" (IPR's).

Data Sample

Houdini 3, 32-pv mode, basic search depth 17 ply = 8-1/2 moves.

```
FEN: 2r3k1/1p1r3p/p5pR/P3pp2/3Pq3/2P1P3/1P1Q1RPP/6K1 b - - 0 32
 dp/ex  value  diff  move and PV
 ...
 17/53  +0.18  0.37  32...exd4 33.exd4 Re7...
 17/53  +0.11  0.30  32...Rc4 33.g3 Ra4...
 17/53  +0.08  0.27  32...Qb1+ 33.Rf1 Qa2...
 17/53  +0.04  0.23  32...Qd5 33.Rh3 Re7...
 17/53  +0.04  0.23  32...Re7 33.Rh3 Qd5...
 17/53   0.00  0.19  32...Kg7 33.Rh3 Rc5...
 17/53  -0.19  0.00  32...Rc5 33.b4 Rc4...
```

Best move at bottom, 19 centipawn advantage to Black, to move.
These numbers and the move actually played (which was 32...Rc5) are the only chess-dependent inputs to the model.
Hence adaptable to any decision game with fungible values.

Two Skill Parameters, Universal?

- *Sensitivity* s divides eval-units to yield dimensionless quantities:

$$x_i = \frac{\Delta(v_1, v_i)}{s}.$$

- *Consistency* c magnifies high and low values of x_i.

Current model:

$$\frac{\log(1/p_1)}{\log(1/p_i)} = \exp(-x_i^c).$$

- Higher c makes the right-hand tinier, so p_i tinier, thus reducing the frequency of blunders. "Tactical"
- Lower s has a stronger effect on x_i when x_i is small, picking out slight differences. "Positional"
- Depth parameters are under development.

Isomorphism With a Rasch Application

Decision Making in Game Play

1. Values for move choices
2. Move-match (MM) score
3. Avg.-Error (AE) score
4. P-parameters
5. Model projections
6. Game criticality of position
7. "Intrinsic Perf. Rating" (IPR)
8. Moment statistics, confidence.

Multiple-Choice Tests

1. Point credits for (all) answers
2. Best-answer score
3. Partial-credit score
4. Aptitude parameters ("position")
5. Difficulty of question
6. Weight of question
7. Grade assessment
8. Grade distribution analysis.

Goal: Cross-fertilize the rich data and theory between psychometrics and games.

General Game Play Pertinence (mostly future)

1. How well does $P(s, c, \dots)$ simulate a "human" player of the given skill set?

General Game Play Pertinence (mostly future)

1. How well does $P(s, c, \ldots)$ simulate a "human" player of the given skill set?
2. Garry Kasparov's "Turing Test": he was able to distinguish games played by (older-and-weaker) computer-played games from human ones. *Would $P(s, c, \ldots)$ pass it?*

General Game Play Pertinence (mostly future)

1. How well does $P(s, c, \ldots)$ simulate a "human" player of the given skill set?
2. Garry Kasparov's "Turing Test": he was able to distinguish games played by (older-and-weaker) computer-played games from human ones. *Would $P(s, c, \ldots)$ pass it?*
3. Intrinsic estimates of position difficulty?
4. Relate human performance to difficulty statistically.

General Game Play Pertinence (mostly future)

1. How well does $P(s, c, \ldots)$ simulate a "human" player of the given skill set?
2. Garry Kasparov's "Turing Test": he was able to distinguish games played by (older-and-weaker) computer-played games from human ones. *Would $P(s, c, \ldots)$ pass it?*
3. Intrinsic estimates of position difficulty?
4. Relate human performance to difficulty statistically.
5. Influence of thinking time on skill.

General Game Play Pertinence (mostly future)

1. How well does $P(s, c, \ldots)$ simulate a "human" player of the given skill set?
2. Garry Kasparov's "Turing Test": he was able to distinguish games played by (older-and-weaker) computer-played games from human ones. *Would $P(s, c, \ldots)$ pass it?*
3. Intrinsic estimates of position difficulty?
4. Relate human performance to difficulty statistically.
5. Influence of thinking time on skill.
6. Behavior as a function of being ahead/behind/equal: Cognitive Bias or Rational Risk-Taking?

General Game Play Pertinence (mostly future)

1. How well does $P(s, c, \dots)$ simulate a "human" player of the given skill set?
2. Garry Kasparov's "Turing Test": he was able to distinguish games played by (older-and-weaker) computer-played games from human ones. *Would $P(s, c, \dots)$ pass it?*
3. Intrinsic estimates of position difficulty?
4. Relate human performance to difficulty statistically.
5. Influence of thinking time on skill.
6. Behavior as a function of being ahead/behind/equal: Cognitive Bias or Rational Risk-Taking?
7. Game quality with unevenly-matched players.

Chess Elo Ratings

- Based on results of games (only): win, lose, draw.
- Numbers have only relative meaning.
- A 200-point difference $\sim$ 75% expectation for the winner (now closer to 76%): "Class Unit" (László Mérő).

Chess Elo Ratings

- Based on results of games (only): win, lose, draw.
- Numbers have only relative meaning.
- A 200-point difference $\sim$ 75% expectation for the winner (now closer to 76%): "Class Unit" (László Mérő).
- USCF: 2400 = Senior Master, 2200 = Master, 2000 = Expert, 1800 = Class A, ..., 1200 = Class D, 1000 = Class E.
- "Beginner" might be 600, but kids push below USCF's 100 floor.

Chess Elo Ratings

- Based on results of games (only): win, lose, draw.
- Numbers have only relative meaning.
- A 200-point difference $\sim$ 75% expectation for the winner (now closer to 76%): "Class Unit" (László Mérő).
- USCF: 2400 = Senior Master, 2200 = Master, 2000 = Expert, 1800 = Class A, ..., 1200 = Class D, 1000 = Class E.
- "Beginner" might be 600, but kids push below USCF's 100 floor.
- Highest human (FIDE) rating is 2870, about fifty have 2700+.
- Computer programs have 3200+ (CCRL), even on cheap hardware.

Chess Elo Ratings

- Based on results of games (only): win, lose, draw.
- Numbers have only relative meaning.
- A 200-point difference $\sim$ 75% expectation for the winner (now closer to 76%): "Class Unit" (László Mérő).
- USCF: 2400 = Senior Master, 2200 = Master, 2000 = Expert, 1800 = Class A, ..., 1200 = Class D, 1000 = Class E.
- "Beginner" might be 600, but kids push below USCF's 100 floor.
- Highest human (FIDE) rating is 2870, about fifty have 2700+.
- Computer programs have 3200+ (CCRL), even on cheap hardware.
- Advantages of IPR:
  - independent of opponent's play
  - 50-100 games per year yield 1,500–3,000 relevant moves.

IPR Psychometric Procedure

A performance by a human player Q at international level is typically 9–10 games, giving a set T of about 250–300 analyzed game turns.

IPR Psychometric Procedure

A performance by a human player Q at international level is typically 9–10 games, giving a set T of about 250–300 analyzed game turns.

1. Run regression over T to find the closest agent $P(s, c, \dots)$.

IPR Psychometric Procedure

A performance by a human player Q at international level is typically 9–10 games, giving a set T of about 250–300 analyzed game turns.

1. Run regression over T to find the closest agent $P(s, c, \dots)$.
2. Calculate $a_e =$ the *projected* AE of P on a *fixed reference set* S of positions.

IPR Psychometric Procedure

A performance by a human player Q at international level is typically 9–10 games, giving a set T of about 250–300 analyzed game turns.

1. Run regression over T to find the closest agent $P(s, c, \dots)$.
2. Calculate a_e = the *projected* AE of P on a *fixed reference set* S of positions.
3. Read $\mathrm{IPR}(a)$ from the model's *training fit* to human players.

IPR Psychometric Procedure

A performance by a human player Q at international level is typically 9–10 games, giving a set T of about 250–300 analyzed game turns.

1. Run regression over T to find the closest agent $P(s, c, \dots)$.
2. Calculate a_e = the *projected* AE of P on a *fixed reference set* S of positions.
3. Read $\mathrm{IPR}(a)$ from the model's *training fit* to human players.

With unit weighting of decisions ("all questions equal value, regardless of criticality or difficulty"), the current best-fit regression to Elo rating is almost exactly:

$$\mathrm{IPR} = 3475 - a * 14{,}000.$$

IPR Psychometric Procedure

A performance by a human player Q at international level is typically 9–10 games, giving a set T of about 250–300 analyzed game turns.

1. Run regression over T to find the closest agent $P(s, c, \dots)$.
2. Calculate a_e = the *projected* AE of P on a *fixed reference set* S of positions.
3. Read $\mathrm{IPR}(a)$ from the model's *training fit* to human players.

With unit weighting of decisions ("all questions equal value, regardless of criticality or difficulty"), the current best-fit regression to Elo rating is almost exactly:

$$\mathrm{IPR} = 3475 - a * 14{,}000.$$

Error Bars *of measurement* are based on the run over T.

Deviation Test Procedure (for cheating):

Given player Q of rating E performing on positions T:

Deviation Test Procedure (for cheating):

Given player Q of rating E performing on positions T:

1. Choose parameters s_Q, c_Q appropriate to Q (corresp. to E).

Deviation Test Procedure (for cheating):

Given player Q of rating E performing on positions T:

1. Choose parameters s_Q, c_Q appropriate to Q (corresp. to E).
2. Compute projections m_Q, a_Q on T.

Deviation Test Procedure (for cheating):

Given player Q of rating E performing on positions T:

1. Choose parameters s_Q, c_Q appropriate to Q (corresp. to E).
2. Compute projections m_Q, a_Q on T.
3. Also get confidence intervals $\pm e_m$, $\pm e_a$ (also depend on s_Q, c_Q).
4. Compare with actual $\hat{m}$, $\hat{a}$ on T to get *z-scores*.

Deviation Test Procedure (for cheating):

Given player Q of rating E performing on positions T:

1. Choose parameters s_Q, c_Q appropriate to Q (corresp. to E).
2. Compute projections m_Q, a_Q on T.
3. Also get confidence intervals $\pm e_m$, $\pm e_a$ (also depend on s_Q, c_Q).
4. Compare with actual $\hat{m}$, $\hat{a}$ on T to get *z-scores*.

- Deviation test error itself is minimal, since s_Q, c_Q are from large-scale training fits. Hence e_m, e_a can set z-scores.

Deviation Test Procedure (for cheating):

Given player Q of rating E performing on positions T:

1. Choose parameters s_Q, c_Q appropriate to Q (corresp. to E).
2. Compute projections m_Q, a_Q on T.
3. Also get confidence intervals $\pm e_m$, $\pm e_a$ (also depend on s_Q, c_Q).
4. Compare with actual $\hat{m}$, $\hat{a}$ on T to get *z-scores*.

- Deviation test error itself is minimal, since s_Q, c_Q are from large-scale training fits. Hence e_m, e_a can set z-scores.
- Empirical testing on 10,000s of random 9-game subsets of training data, and actual player-performances, suggests adjustment factors.

Deviation Test Procedure (for cheating):

Given player Q of rating E performing on positions T:

1. Choose parameters s_Q, c_Q appropriate to Q (corresp. to E).
2. Compute projections m_Q, a_Q on T.
3. Also get confidence intervals $\pm e_m$, $\pm e_a$ (also depend on s_Q, c_Q).
4. Compare with actual $\hat{m}$, $\hat{a}$ on T to get *z-scores*.

- Deviation test error itself is minimal, since s_Q, c_Q are from large-scale training fits. Hence e_m, e_a can set z-scores.
- Empirical testing on 10,000s of random 9-game subsets of training data, and actual player-performances, suggests adjustment factors.
- Game decisions modeled as independent, but really have "Sparse Dependence." Adjustment reflects lower effective sample size $|T|$.

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-championship matches, almost all above 2700 strength.

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-championship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars ± 200–300):
    - Deep Thought in 1991: 2150.

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-championship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars $\pm$200–300):
   - Deep Thought in 1991: 2150.
   - Deep Blue in 1996–97: 2850–2900.

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-championship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars $\pm$200–300):
    - Deep Thought in 1991: 2150.
    - Deep Blue in 1996–97: 2850–2900.
    - Hydra in 2005: 3150

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-champonship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars $\pm$200–300):
    - Deep Thought in 1991: 2150.
    - Deep Blue in 1996–97: 2850–2900.
    - Hydra in 2005: 3150
    - Deep Fritz 10 on 4-core PC in 2006: 2980.

Results and Interpretations

1. Training done on games with both players within ±10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-champonship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars ±200–300):
   - Deep Thought in 1991: 2150.
   - Deep Blue in 1996–97: 2850–2900.
   - Hydra in 2005: 3150
   - Deep Fritz 10 on 4-core PC in 2006: 2980.
4. *Tournaments*, however, regularly have IPRs 20-30 below their average ratings.

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-champonship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars $\pm$200–300):
   - Deep Thought in 1991: 2150.
   - Deep Blue in 1996–97: 2850–2900.
   - Hydra in 2005: 3150
   - Deep Fritz 10 on 4-core PC in 2006: 2980.
4. *Tournaments*, however, regularly have IPRs 20-30 below their average ratings.
5. Perhaps owes to higher average rating difference in games?

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-championship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars ±200–300):
   - Deep Thought in 1991: 2150.
   - Deep Blue in 1996–97: 2850–2900.
   - Hydra in 2005: 3150
   - Deep Fritz 10 on 4-core PC in 2006: 2980.
4. *Tournaments*, however, regularly have IPRs 20-30 below their average ratings.
5. Perhaps owes to higher average rating difference in games?
6. Human IPR's rarely above 3000...

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-champonship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars $\pm$200–300):
   - Deep Thought in 1991: 2150.
   - Deep Blue in 1996–97: 2850–2900.
   - Hydra in 2005: 3150
   - Deep Fritz 10 on 4-core PC in 2006: 2980.
4. *Tournaments*, however, regularly have IPRs 20-30 below their average ratings.
5. Perhaps owes to higher average rating difference in games?
6. Human IPR's rarely above 3000... except for some players named Feller, Ivanov, Kotainy...

Results and Interpretations

1. Training done on games with both players within ± 10 of an Elo century-point, 2200, 2300, 2400, 2500, 2600, 2700.
2. IPRs *spot-on* with average rating in world-champonship matches, almost all above 2700 strength.
3. Even IPRs of computers make sense (though error bars ±200–300):
   - Deep Thought in 1991: 2150.
   - Deep Blue in 1996–97: 2850–2900.
   - Hydra in 2005: 3150
   - Deep Fritz 10 on 4-core PC in 2006: 2980.
4. *Tournaments*, however, regularly have IPRs 20-30 below their average ratings.
5. Perhaps owes to higher average rating difference in games?
6. Human IPR's rarely above 3000... except for some players named Feller, Ivanov, Kotainy... or most of the 2010 Azov Don Cup.

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."
- Issue: *error by whom?*

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."
- Issue: *error by whom?*
- Can be by a reference player P_0.

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."
- Issue: *error by whom?*
- Can be by a reference player P_0. (Can alternately define IPR as performance relative to P_0.)

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."
- Issue: *error by whom?*
- Can be by a reference player P_0. (Can alternately define IPR as performance relative to P_0.)
- But what level to use for P_0?

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."
- Issue: *error by whom?*
- Can be by a reference player P_0. (Can alternately define IPR as performance relative to P_0.)
- But what level to use for P_0?
- Can integrate error over whole $P(s, c, \dots)$ parameter space,

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."
- Issue: *error by whom?*
- Can be by a reference player P_0. (Can alternately define IPR as performance relative to P_0.)
- But what level to use for P_0?
- Can integrate error over whole $P(s, c, \ldots)$ parameter space, but how weighted? Or not a simple scalar...

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."
- Issue: *error by whom?*
- Can be by a reference player P_0. (Can alternately define IPR as performance relative to P_0.)
- But what level to use for P_0?
- Can integrate error over whole $P(s, c, \dots)$ parameter space, but how weighted? Or not a simple scalar...
- Instead try to correlate *observed* difficulty with intrinsic features of the game position...

Lessons for Estimating Difficulty in Games

- First idea: Difficulty of a position t = *expected error* on t, perhaps weighted by "how critical."
- Issue: *error by whom?*
- Can be by a reference player P_0. (Can alternately define IPR as performance relative to P_0.)
- But what level to use for P_0?
- Can integrate error over whole $P(s, c, \dots)$ parameter space, but how weighted? Or not a simple scalar...
- Instead try to correlate *observed* difficulty with intrinsic features of the game position... such as how much values *"swing"* as analysis depth changes.

Some Further Ideas

1. Characterize “styles” of both human players and ’bots in the $P(s, c, \ldots)$ space.

Some Further Ideas

1. Characterize “styles” of both human players and ’bots in the $P(s, c, \dots)$ space.
2. Is there a “Fischer Fingerprint”?

Some Further Ideas

1. Characterize "styles" of both human players and 'bots in the $P(s, c, \dots)$ space.
2. Is there a "Fischer Fingerprint"? Suppose 9 new games turn up, and someone claims they were played by Fischer in a previously-unknown tournament before 1970.

Some Further Ideas

1. Characterize "styles" of both human players and 'bots in the $P(s, c, \ldots)$ space.
2. Is there a "Fischer Fingerprint"? Suppose 9 new games turn up, and someone claims they were played by Fischer in a previously-unknown tournament before 1970.
   - With few parameters—and many players—probably someone else's games would be a closer match even if they *were* played by Fischer.

Some Further Ideas

1. Characterize "styles" of both human players and 'bots in the $P(s, c, \ldots)$ space.
2. Is there a "Fischer Fingerprint"? Suppose 9 new games turn up, and someone claims they were played by Fischer in a previously-unknown tournament before 1970.
    - With few parameters—and many players—probably someone else's games would be a closer match even if they *were* played by Fischer.
    - Similar issue with authorship disputes: Peter Millican re: J.K. Rowling in 19 July, 2013 interview with Canada's "Day 6" programme.

Some Further Ideas

1. Characterize "styles" of both human players and 'bots in the $P(s, c, \ldots)$ space.
2. Is there a "Fischer Fingerprint"? Suppose 9 new games turn up, and someone claims they were played by Fischer in a previously-unknown tournament before 1970.
   - With few parameters—and many players—probably someone else's games would be a closer match even if they *were* played by Fischer.
   - Similar issue with authorship disputes: Peter Millican re: J.K. Rowling in 19 July, 2013 interview with Canada's "Day 6" programme.
3. Distinguishing two far-apart styles is easier (e.g. human $\longleftrightarrow$ computer).

Some Further Ideas

1. Characterize "styles" of both human players and 'bots in the $P(s, c, \ldots)$ space.
2. Is there a "Fischer Fingerprint"? Suppose 9 new games turn up, and someone claims they were played by Fischer in a previously-unknown tournament before 1970.
   - With few parameters—and many players—probably someone else's games would be a closer match even if they *were* played by Fischer.
   - Similar issue with authorship disputes: Peter Millican re: J.K. Rowling in 19 July, 2013 interview with Canada's "Day 6" programme.
3. Distinguishing two far-apart styles is easier (e.g. human $\longleftrightarrow$ computer).
4. How to make (fallible) 'bots more human-realistic?

Some Further Ideas

1. Characterize "styles" of both human players and 'bots in the $P(s, c, \dots)$ space.
2. Is there a "Fischer Fingerprint"? Suppose 9 new games turn up, and someone claims they were played by Fischer in a previously-unknown tournament before 1970.
   - With few parameters—and many players—probably someone else's games would be a closer match even if they *were* played by Fischer.
   - Similar issue with authorship disputes: Peter Millican re: J.K. Rowling in 19 July, 2013 interview with Canada's "Day 6" programme.
3. Distinguishing two far-apart styles is easier (e.g. human $\longleftrightarrow$ computer).
4. How to make (fallible) 'bots more human-realistic?
5. Tame the *curve of fallibility*. . .

In Conclusion

Main tenet of the model:

In Conclusion

Main tenet of the model:

> *Human decision making (and physiological reactivity) ought to be governed in the large by relatively simple mathematical laws—*

In Conclusion

Main tenet of the model:

Human decision making (and physiological reactivity) ought to be governed in the large by relatively simple mathematical laws—laws that are independent of details of any particular game,

In Conclusion

Main tenet of the model:

Human decision making (and physiological reactivity) ought to be governed in the large by relatively simple mathematical laws—laws that are independent of details of any particular game, and hence ought to be revealed as common properties between games.

In Conclusion

Main tenet of the model:

> *Human decision making (and physiological reactivity) ought to be governed in the large by relatively simple mathematical laws—laws that are independent of details of any particular game, and hence ought to be revealed as common properties between games. And many activities in life are games.*

In Conclusion

Main tenet of the model:

> *Human decision making (and physiological reactivity) ought to be governed in the large by relatively simple mathematical laws—laws that are independent of details of any particular game, and hence ought to be revealed as common properties between games. And many activities in life are games.*

The results so far show that this expectation is plausible.