Getting to Know Our Digital Assistants

Large data from chess hints what to expect and how to prosper

Kenneth W. Regan¹
University at Buffalo (SUNY)

TEDxBuffalo, 14 October, 2014

¹Includes joint work with Guy Haworth Giuseppe DiFatta, Bartlomiej Macieja, Tamal Biswas, and Jason Zhou. Websites: http://www.cse.buffalo.edu/~regan/
http://www.cse.buffalo.edu/~regan/chess/fidelity/
Personal Digital Assistants

- Software Agents, many with Personas...
Getting to Know Our Digital Assistants

Personal Digital Assistants

- Software Agents, many with Personas...
- GPS voices, Apple Siri, IBM Watson
Personal Digital Assistants

- Software Agents, many with Personas...
- GPS voices, Apple Siri, IBM Watson
- Google Now, SILVIA (platform), Microsoft Cortana

Cortana came from the game Halo.
Personal Digital Assistants

- Software Agents, many with Personas...
- GPS voices, Apple Siri, IBM Watson
- Google Now, SILVIA (platform), Microsoft Cortana
- Online Gaming Personas...
Getting to Know Our Digital Assistants

Personal Digital Assistants

- Software Agents, many with Personas...
- GPS voices, Apple Siri, IBM Watson
- Google Now, SILVIA (platform), Microsoft Cortana
- Online Gaming Personas...Cortana came from the game *Halo.*
Personal Digital Assistants

- Software Agents, many with Personas...
- GPS voices, Apple Siri, IBM Watson
- Google Now, SILVIA (platform), Microsoft Cortana
- Online Gaming Personas...Cortana came from the game *Halo*.
Cognitive Agents
Cognitive Agents

- **Read** lots of information
Cognitive Agents

- **Read** lots of information
- **Evaluate** options
Cognitive Agents

- **Read** lots of information
- **Evaluate** options
- **Write** recommendations
Cognitive Agents

- Read lots of information
- Evaluate options
- Write recommendations
- Guide our Decision Making
Getting to Know Our Digital Assistants

Cognitive Agents

- **Read** lots of information
- **Evaluate** options
- **Write** recommendations
- Guide our **Decision Making**
We may empower agents to act autonomously...
Getting to Know Our Digital Assistants

Autonomy, Folly, and Partnership?

- We may empower agents to act autonomously...drive a harder bargain...
Getting to Know Our Digital Assistants

Autonomy, Folly, and Partnership?

- We may empower agents to act autonomously...drive a harder bargain...
- ...even **walk away** when we might not.
Getting to Know Our Digital Assistants

Autonomy, Folly, and Partnership?

- We may empower agents to act autonomously...drive a harder bargain...
- ...even **walk away** when we might not.
- Automated Trading
Getting to Know Our Digital Assistants

**Autonomy, Folly, and Partnership?**

- We may empower agents to act autonomously...drive a harder bargain...
- ...even **walk away** when we might not.
- Automated Trading
- Folly of Over-Reliance:
We may empower agents to act autonomously... drive a harder bargain...

...even **walk away** when we might not.

- Automated Trading
- Folly of Over-Reliance: **Flash Crash**, 

**Autonomy, Folly, and Partnership?**
Getting to Know Our Digital Assistants

Autonomy, Folly, and Partnership?

- We may empower agents to act autonomously...drive a harder bargain...
- ...even **walk away** when we might not.
- Automated Trading
- Folly of Over-Reliance: **Flash Crash, Flash Floods**
Getting to Know Our Digital Assistants

**Autonomy, Folly, and Partnership?**

- We may empower agents to act autonomously...drive a harder bargain...
- ...even **walk away** when we might not.
- Automated Trading
- Folly of Over-Reliance: **Flash Crash, Flash Floods**
Getting to Know Our Digital Assistants

Autonomy, Folly, and Partnership?

- We may empower agents to act autonomously...drive a harder bargain...
- ...even walk away when we might not.
- Automated Trading
- Folly of Over-Reliance: Flash Crash, Flash Floods

(Photo from story on partnership using GPS to warn of disasters.)
Cheating With Computers at Chess

- Would have been laughed at in 1970s.
Cheating With Computers at Chess

- Would have been laughed at in 1970s.
- Not in 1980s: several programs achieved Master rating.
Cheating With Computers at Chess

- Would have been laughed at in 1970s.
- Not in 1980s: several programs achieved Master rating.
- 1997: World champ Garry Kasparov lost to IBM supercomputer.
Would have been laughed at in 1970s.

Not in 1980s: several programs achieved Master rating.

1997: World champ Garry Kasparov lost to IBM supercomputer.

2006: World champ Vladimir Kramnik lost to a home PC.
Cheating With Computers at Chess

- Would have been laughed at in 1970s.
- Not in 1980s: several programs achieved Master rating.
- 1997: World champ Garry Kasparov lost to IBM supercomputer.
- 2006: World champ Vladimir Kramnik lost to a home PC.
Cheating With Computers at Chess

- Would have been laughed at in 1970s.
- Not in 1980s: several programs achieved Master rating.
- 1997: World champ **Garry Kasparov** lost to IBM supercomputer.
- 2006: World champ **Vladimir Kramnik** lost to a **home PC**.
- 2010s: Your I-Pad, your smartphone, even Siri...
Cheating With Computers at Chess

- Would have been laughed at in 1970s.
- Not in 1980s: several programs achieved Master rating.
- 1997: World champ Garry Kasparov lost to IBM supercomputer.
- 2006: World champ Vladimir Kramnik lost to a home PC.
- 2010s: Your I-Pad, your smartphone, even Siri...
- A Big Problem
Testing Cheating With Style

- Not just “you played too well...”
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ...

Have analyzed almost entire history of top-level human chess...
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” … using chess programs...
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

- Fight Fire With
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

- Fight Fire With
Testing Cheating With **Style**

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

- **Fight Fire With** ![Fire](image1.png) ![Rybka](image2.png) (actually, **Rybka**
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

Fight Fire With (actually, Rybka)

And With Pretty Big Data:
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

Fight Fire With (actually,)

And With Pretty Big Data:
- $\frac{1}{2} \times 1M$ Games
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

- **Fight Fire With** (actually, **Rybka**)  

- **And With Pretty Big Data:**
  - $1/2 \times 1M$ Games
  - Over $3 \times 10M$ Moves
Getting to Know Our Digital Assistants

Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

- **Fight Fire With** ![Fire](image)
  (actually, ![Rybka](image))

- **And With Pretty Big Data:**
  - $\frac{1}{2} \times 1M$ Games
  - Over $3 \times 10M$ Moves
  - Over $100M$ Pages of Data
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

Fight Fire With (actually, Rybka)

And With Pretty Big Data:
- $\frac{1}{2} \times 1M$ Games
- Over $3 \times 10M$ Moves
- Over $100M$ Pages of Data

Have analyzed almost entire history of top-level human chess...
Testing Cheating With Style

- Not just “you played too well...” would be like, “You biked too fast.”
- Look for “Deep Patterns” ... using chess programs...

Fight Fire With  (actually, 

And With Pretty Big Data:

- \( \frac{1}{2} \times 1 \text{M Games} \)
- Over \( 3 \times 10 \text{M Moves} \)
- Over \( 100 \text{M Pages of Data} \)

Have analyzed almost entire history of top-level human chess... and computer chess.
The ___ of drug-resistant strains of bacteria and viruses has ___ researchers’ hopes that permanent victories against many diseases have been achieved.

(a) vigor .. corroborated
(b) feebleness .. dashed
(c) proliferation .. blighted
(d) destruction .. disputed
(e) disappearance .. frustrated

(source: itunes.apple.com)
2006 World Champ. “Toiletgate” Scandal

Veselin Topalov (right) accused Vladimir Kramnik of getting moves during the games via computer cable to his toilet.

(photo source: NY Times, 2006)
Getting to Know Our Digital Assistants

Statistical “Evidence” Cited

Coincidence Statistics of the moves of GM Kramnik with recommendations of the chess program Fritz 9
Elista, October 4, 2006

After very detailed analyze of all games of the match we would like to present to your attention coincidence statistics of the moves of GM Kramnik with recommendations of chess program Fritz 9.

First game:
From 75 moves: After the 12th move of Topalov Ba6, a novelty, from 65 remaining moves – 41 moves match with the first line of Fritz 9. (63% of matches)

Second game:
From 63 moves: After 17th move, where the theory ends, from remaining 46 moves – 40 moves match with the first line of Fritz 9. (87% of matches)

(…)

Coincidence Statistics of the moves of GM Kramnik with recommendations of the chess program Fritz 9

Elista, October 4, 2006

After very detailed analyze of all games of the match we would like to present to your attention coincidence statistics of the moves of GM Kramnik with recommendations of chess program Fritz 9.

**First game:**
From 75 moves: After the 12th move of Topalov Ba6, a novelty, from 65 remaining moves – 41 moves match with the first line of Fritz 9. (63% of matches)

**Second game:**
From 63 moves: After 17th move, where the theory ends, from remaining 46 moves – 40 moves match with the first line of Fritz 9. (87% of matches)

(...)

How to evaluate this kind of accusation?
Kramnik did match high—in Game 2—**BUT...**
Degrees of Freedom

- Kramnik did match high—in Game 2—**BUT**...
- Topalov had forced his hand....
Kramnik did match high—in Game 2—**BUT...**
Topalov had forced his hand....**only one real option** per move.
Kramnik did match high—in Game 2—**BUT...**

Topalov had forced his hand....**only one real option** per move.

**Main Qualitative Principle:**
Kramnik did match high—in Game 2—**BUT...**
Topalov had forced his hand....*only one real option* per move.

**Main Qualitative Principle:**

When there is only one way to stay afloat or stay ahead, chances are a good player and a computer will both find it.
Degrees of Freedom

- Kramnik did match high—in Game 2—**BUT...**
- Topalov had forced his hand....**only one real option** per move.
- **Main Qualitative Principle:**

  When there is only one way to stay afloat or stay ahead, chances are a good player and a computer will both find it.

- **Do computer players “force”?**
Kramnik did match high—in Game 2—**BUT...**
Topalov had forced his hand....**only one real option** per move.

**Main Qualitative Principle:**

> When there is only one way to stay afloat or stay ahead, chances are a good player and a computer will both find it.

**Do computer players “force”?**

**Oct. 12, 2006...**
Kramnik did match high—in Game 2—**BUT...**
Topalov had forced his hand....**only one real option** per move.

**Main Qualitative Principle:**

> When there is only one way to stay afloat or stay ahead, chances are a good player and a computer will both find it.

**Do computer players “force”?**

**Oct. 12, 2006...** last match day Fri. the 13th...
Kramnik did match high—in Game 2—**BUT**...
Topalov had forced his hand.... **only one real option** per move.

**Main Qualitative Principle:**

When there is only one way to stay afloat or stay ahead, chances are a good player and a computer will both find it.

**Do computer players “force”?**

**Oct. 12, 2006**... last match day Fri. the 13th... I was all set to
Kramnik did match high—in Game 2—**BUT**...
Topalov had forced his hand....**only one real option** per move.
**Main Qualitative Principle:**

> When there is only one way to stay afloat or stay ahead, chances are a good player and a computer will both find it.

**Do computer players “force”?**

**Oct. 12, 2006**... last match day Fri. the 13th... I was all set to announce my findings
**Degrees of Freedom**

- Kramnik did match high—in Game 2—**BUT**...
- Topalov had forced his hand....**only one real option** per move.
- **Main Qualitative Principle:**

  When there is only one way to stay afloat or stay ahead, chances are a good player and a computer will both find it.

- **Do computer players “force”?**
- **Oct. 12, 2006**...last match day Fri. the 13th...I was all set to **announce** my findings and **denounce** the accusation,
Kramnik did match high—in Game 2—**BUT**...
Topalov had forced his hand...**only one real option** per move.

**Main Qualitative Principle:**

When there is only one way to stay afloat or stay ahead, chances are a good player and a computer will both find it.

**Do computer players “force”?**

**Oct. 12, 2006**...last match day Fri. the 13th...I was all set to **announce** my findings and **denounce** the accusation, **BUT**...
The 2006 Buffalo October Storm
The 2006 Buffalo October Storm
Kramnik won anyway; by time power back, “that ship had sailed.”
Between Storms

- Kramnik won anyway; by time power back, “that ship had sailed.”
- Gave lots of time to make work *quantitative, deeper*...
Kramnik won anyway; by time power back, “that ship had sailed.”
Gave lots of time to make work quantitative, deeper...
And research Human Cognition, not just Cheating.
Kramnik won anyway; by time power back, “that ship had sailed.”

Gave lots of time to make work **quantitative, deeper**...

And research Human Cognition, not just Cheating.

Mid 2007 thru end 2010, no real cases...
Between Storms

- Kramnik won anyway; by time power back, “that ship had sailed.”
- Gave lots of time to make work quantitative, deeper...
- And research Human Cognition, not just Cheating.
- Mid 2007 thru end 2010, no real cases...

**2012**: some more cases...

**2013**: more and more cases...

Since June 2013 I am on a joint committee of FIDE and the Association of Chess Professionals on all aspects of cheating.

How is it possible to cheat at chess, anyway?
Kramnik won anyway; by time power back, “that ship had sailed.”

Gave lots of time to make work quantitative, deeper...

And research Human Cognition, not just Cheating.

Mid 2007 thru end 2010, no real cases...

**Jan. 2011**: case with top-100 player (games in 2010).

**2012**: some more cases...
Kramnik won anyway; by time power back, “that ship had sailed.”
Gave lots of time to make work **quantitative**, **deeper**...
And research Human Cognition, not just Cheating.
Mid 2007 thru end 2010, no real cases...
**Jan. 2011**: case with **top-100** player (games in 2010).
**2012**: some more cases...
**2013**: more and more cases...
Kramnik won anyway; by time power back, “that ship had sailed.”

Gave lots of time to make work *quantitative, deeper*...

And research Human Cognition, not just Cheating.

Mid 2007 thru end 2010, no real cases...

**Jan. 2011**: case with **top-100** player (games in 2010).

**2012**: some more cases...

**2013**: more and more cases...

Since June 2013 I am on a joint committee of **FIDE** and the **Association of Chess Professionals** on all aspects of cheating.
Kramnik won anyway; by time power back, “that ship had sailed.”

Gave lots of time to make work quantitative, deeper...

And research Human Cognition, not just Cheating.

Mid 2007 thru end 2010, no real cases...


2012: some more cases...

2013: more and more cases...

Since June 2013 I am on a joint committee of FIDE and the Association of Chess Professionals on all aspects of cheating.

How is it possible to cheat at chess, anyway?
Cheating Aside, What Can We Learn?

Skill Assessment

"Intrinsic Performance Rating" (IPR)
Not part of the cheating test.
Based just on quality of your moves, not results of games.
Opponent's performance not involved.

Prediction
Risk Assessment...

Fraud Detection
Predictive Analytics

Natural Human Tendencies

Natural Inhuman Tendencies...
Cheating Aside, What Can We Learn?

1. Skill Assessment

Intrinsic Performance Rating (IPR)

Not part of the cheating test.
Based just on quality of your moves, not results of games.
Opponent's performance not involved.

Prediction
Risk Assessment...
Fraud Detection
Predictive Analytics

Natural Human Tendencies

Natural Inhuman Tendencies...
Cheating Aside, What Can We Learn?

1. **Skill Assessment**
   - “**Intrinsic Performance Rating**” (IPR)
Cheating Aside, What Can We Learn?

1. **Skill Assessment**
   - “Intrinsic Performance Rating” (IPR)
   - Not part of the cheating test.
Cheating Aside, What Can We Learn?

1 Skill Assessment
   - “Intrinsic Performance Rating” (IPR)
   - Not part of the cheating test.
   - Based just on quality of your moves, not results of games.
Cheating Aside, What Can We Learn?

Skill Assessment

- “Intrinsic Performance Rating” (IPR)
- Not part of the cheating test.
- Based just on quality of your moves, not results of games.
- Opponent’s performance not involved.
Cheating Aside, What Can We Learn?

1. **Skill Assessment**
   - “Intrinsic Performance Rating” (IPR)
   - Not part of the cheating test.
   - Based *just* on quality of *your* moves, not results of games.
   - Opponent’s performance not involved.

2. **Prediction**
Cheating Aside, What Can We Learn?

1. **Skill Assessment**
   - “Intrinsic Performance Rating” (IPR)
   - Not part of the cheating test.
   - Based *just* on quality of your moves, not results of games.
   - Opponent’s performance not involved.

2. **Prediction**
   - Risk Assessment...
Cheating Aside, What Can We Learn?

1. **Skill Assessment**
   - “Intrinsic Performance Rating” (IPR)
   - Not part of the cheating test.
   - Based *just* on quality of your moves, not results of games.
   - Opponent’s performance not involved.

2. **Prediction**
   - Risk Assessment...Fraud Detection
Cheating Aside, What Can We Learn?

1. **Skill Assessment**
   - “Intrinsic Performance Rating” (IPR)
   - Not part of the cheating test.
   - Based *just* on quality of *your* moves, not results of games.
   - Opponent’s performance not involved.

2. **Prediction**
   - Risk Assessment...Fraud Detection
   - Predictive Analytics
Cheating Aside, What Can We Learn?

1. **Skill Assessment**
   - “Intrinsic Performance Rating” (IPR)
   - Not part of the cheating test.
   - Based *just* on quality of your moves, not results of games.
   - Opponent’s performance not involved.

2. **Prediction**
   - Risk Assessment...Fraud Detection
   - Predictive Analytics

3. **Natural Human Tendencies**
Getting to Know Our Digital Assistants

Cheating Aside, What Can We Learn?

1. **Skill Assessment**
   - “*Intrinsic Performance Rating*” (**IPR**)
   - Not part of the cheating test.
   - Based *just* on quality of *your* moves, not results of games.
   - Opponent’s performance not involved.

2. **Prediction**
   - Risk Assessment...Fraud Detection
   - Predictive Analytics

3. **Natural Human Tendencies**

4. **Natural Inhuman Tendencies...**
Win % Expectation Curve

Vs. all opponents
Fitted curve

Win Probability

Evaluation of Position
And When You’re Higher Rated
Would You Like it to be Your Move?

Position Evaluation vs. Win Expectation

- From Player's perspective
- From Opponent's perspective

Win Expectation

Position Evaluation (Stockfish DD)
Effect Absent in Computer Play

Position Evaluation vs. Win Expectation for CEGT

- CEGT-TTC: Player's perspective
- CEGT-TTC: Opponent's perspective

Win Expectation

Position Evaluation (Stockfish 4)
Managing a Time Budget

Scaled and Unscaled

Error Per Move

Time Control

Move Index

0 10 20 30 40 50 60 70 80

0.04 0.06 0.08 0.1 0.12 0.14 0.16 0.18 0.2
Minding Nickels and Dimes

Error Versus Advantage or Disadvantage

Humans, checked with four programs.

Computers
Are We Psychological or Rational?
Some Evidence for Psychological

Minima stay at 0.
Degrees of Forcing Play

Forcing Index (2500 perspective)

- Computer (avg.): 49
- Human: 53.3
Add Human-Computer Tandems

Forcing Index (2500 perspective)

- Computer (avg.): 49
- Computer+Human: 54.5
- Human: 53.3
Add Human-Computer Tandems

Forcing Index (2500 perspective)

- Computer (avg.): 49
- Computer + Human: 54.5
- Human: 53.3

Evidently the humans called the shots. How was the quality?
Adding 210 Elo was significant. Forcing but good teamwork.
2014 Freestyle Tournament Performance

Forcing Index (2500 perspective)

- Computer (avg.): 49
- Computer+Human: 54.5
- Human: 53.3

2014: tandems marginally better W-L, but quality not clear...
Add Topalov Forcing Kramnik

Forcing Index (2500 perspective)

- Computer (avg.): 49
- Computer+Human: 54.5
- Human: 53.3
- Kramnik (2006 g2): 74.5

Last bar goes way off the chart
Like “Spock” to our “Kirk”
Like “Spock” to our “Kirk”

”It is logical to cultivate multiple options.”

(photo sources: The Telegraph, it.wikipedia; lic. re-use/modify)
Summary For Us and PDAs

1. PDAs pick up every little difference: "Forest and Trees"
2. We should avoid overconfidence... and take counsel when "down."
3. Look before we Leap... Don't rush in... Measure risks.
4. Even at a purely calculational pursuit like chess, our brains still contribute.
5. Main takeaway: It should be natural to program PDAs so they enhance our freedom rather than constrain it. This could be the beginning of a beautiful relationship.
PDAs pick up every little difference: “Forest and Trees”
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence...
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence...and take counsel when “down.”
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence... and take counsel when “down.”
3. Look before we Leap...
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence... and take counsel when “down.”
3. Look before we Leap... Don’t rush in...
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence... and take counsel when “down.”
3. Look before we Leap... Don’t rush in... Measure risks.
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence... and take counsel when “down.”
3. Look before we Leap... Don’t rush in... Measure risks.
4. Even at a purely calculational pursuit like chess, our brains still contribute.
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence... and take counsel when “down.”
3. **Look before we Leap... Don’t rush in... Measure risks.**
Summary For Us and PDAs

1. PDAs pick up every little difference: "Forest and Trees"
2. We should avoid overconfidence... and take counsel when “down.”
3. Look before we Leap... Don’t rush in... Measure risks.
4. Even at a purely calculational pursuit like chess, our brains still contribute. (2014: maybe)
5. Main takeaway:
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence... and take counsel when “down.”
3. **Look before we Leap... Don’t rush in... Measure risks.**
5. **Main takeaway:**

It should be **natural** to program PDAs so they enhance our freedom rather than constrain it.
Summary For Us and PDAs

1. PDAs pick up every little difference: “Forest and Trees”
2. We should avoid overconfidence...and take counsel when “down.”
3. Look before we Leap...Don’t rush in...Measure risks.
4. Even at a purely calculational pursuit like chess, our brains still contribute. (2014: maybe)
5. Main takeaway:

   It should be natural to program PDAs so they enhance our freedom rather than constrain it.

This could be the beginning of a beautiful relationship...