The emotional content of large-scale texts: The happiness of bloggers, song lyrics, and presidents.

Workshop on Challenges and Visions in the Social Sciences
ETH Zurich, August 18–23, 2008

Peter Dodds
Chris Danforth

Department of Mathematics & Statistics
Center for Complex Systems
Vermont Advanced Computing Center

University of Vermont
Outline

Measuring emotional content

Data sets

Analysis
  Songs
  Blogs
  SOTU

Winning: it’s not for everyone

Prediction

References
Happiness:

Greek philosophers held **Eudaimonia** as highest good. [9]

≃ flourishing, well-being, pleasure, ...

Socrates, Plato, Aristotle, Epicurus, ...

http://wikipedia.org
Happiness:

Bentham’s hedonistic calculus:

“[t]he greatest happiness of the greatest number is the foundation of morals and legislation”[17]

Priestly, John Stuart Mill, ...
United States’ Declaration of Independence:

“We hold these truths to be sacred & undeniable; that all men are created equal & independent, that from that equal creation they derive rights inherent & inalienable, among which are the preservation of life, & liberty, & the pursuit of happiness;”

http://wikipedia.org
Happiness:

Even the odd modern economist likes happiness:

What makes us happy?

Layard’s summary:

**Dominant factors:**

- Family relationships
- Financial situation
- Work
- Community and Friends
- Health
- Personal Values
- Personal Freedom

**Unimportant factors:**

- Age
- Gender
- Education
- Inherent intelligence
- Looks
What makes us happy?

Layard’s summary:

**Dominant factors:**
- Family relationships
- Financial situation
- Work
- Community and Friends
- Health
- Personal Values
- Personal Freedom

**Unimportant factors:**
- Age
- Gender
- Education
- Inherent intelligence
- Looks
Desiring happiness—not just for boffins:

- Average people routinely report being happy is what they want most in life [11, 12]
Desiring happiness—not just for boffins:

- Average people routinely report being happy is what they want most in life\textsuperscript{[11, 12]}

National indices of well-being:

- Bhutan
- France
- Australia
Emotional content

So how does one measure

1. happiness?
2. levels of other emotions?
Emotional content

So how does one measure

1. happiness?
2. levels of other emotions?

Just ask people how happy they are.
Emotional content

So how does one measure

1. happiness?
2. levels of other emotions?

Just ask people how happy they are.

► Experience sampling[^4, 6, 5] (Csikszentmihalyi et al.)
► Day reconstruction[^10] (Kahneman et al.)
Emotional content

So how does one measure

1. happiness?
2. levels of other emotions?

Just ask people how happy they are.

- Experience sampling\[^{4, 6, 5}\] (Csikszentmihalyi et al.)
- Day reconstruction\[^{10}\] (Kahneman et al.)

But self-reporting has drawbacks...

- relies on memory and self-perception
- induces misreporting\[^{13}\]
- costly
Measuring Emotional Content

We’d like to build an hedonometer:

▶ An instrument to ‘remotely-sense’ emotional states and levels, in real time or post hoc.
Measuring Emotional Content

We’d like to build an hedonometer:

- An instrument to ‘remotely-sense’ emotional states and levels, in real time or post hoc.

Ideally:

- Transparent
- Fast
- Based on written expression
- Uses human evaluation
- Non-reactive
- Complementary to self-reported measures
- Improvable
Measuring Emotional Content

We’d like to build an hedonometer:

▶ An instrument to ‘remotely-sense’ emotional states and levels, in real time or post hoc.

Ideally:

▶ Transparent
▶ Fast
▶ Based on written expression
▶ Uses human evaluation
▶ Non-reactive
▶ Complementary to self-reported measures
▶ Improvable

Some possibilities:

▶ Natural language processing (e.g., OpinionFinder)
▶ Declared mood levels in blogs (e.g., Livejournal) [14]
Measuring Emotional Content

- **Idea:** Gauge emotional content of an entity through human assessment via semantic differentials.
Measuring Emotional Content

- **Idea:** Gauge emotional content of an entity through human assessment via semantic differentials.

- **Examples:**
  - hate ↔ love
  - rough ↔ smooth
  - up ↔ down

---

Osgood et al. (1957) identified a basis of 3 semantic differentials:

- **Valence:** bad ↔ good
- **Dominance:** weak ↔ strong
- **Arousal:** passive ↔ active

(also often: Evaluation, Potency, and Activity)
Measuring Emotional Content

- **Idea:** Gauge emotional content of an entity through human assessment via semantic differentials.

- **Examples:**
  - hate ↔ love
  - rough ↔ smooth
  - up ↔ down

- **Osgood et al. (1957)\textsuperscript{[15]}** identified a basis of 3 semantic differentials:
  - **Valence:** bad ↔ good
  - **Dominance:** weak ↔ strong
  - **Arousal:** passive ↔ active

(Also often: Evaluation, Potency, and Activity)
ANEW study

- ANEW = “Affective Norms for English Words”
ANEW study

- **ANEW** = “Affective Norms for English Words”

- Study: participants shown lists of isolated words
- Asked to grade each word’s valence, arousal, and dominance level
- Integer scale of 1–9
ANEW study

- **ANEW** = “Affective Norms for English Words”

- Study: participants shown lists of isolated words
- Asked to grade each word’s valence, arousal, and dominance level
- Integer scale of 1–9

- \( N = 1034 \) words—previously identified as bearing emotional weight
- Participants = College students (*cough*)
- Results published by Bradley and Lang (1999) \[^2\]
ANEW study—three 1–9 scales:

valence:
ANEW study—three 1–9 scales:

valence:

arousal:

dominance:
ANEW words—examples

<table>
<thead>
<tr>
<th>Valence v</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

- love/paradise/triumphant
- glory/luxury/trophy
- optimism/pancakes/church
- engine/paper/street
- derelict/neurotic/vanity
- fault/corrupt/lawsuit
- trauma/hostage/disgusted
- funeral/rape/suicide
Analysing text:

- Simplest measure for a text:

\[
\theta_{\text{avg}} = \sum_{i=1}^{N} p_i \theta_i
\]

where \( p_i \) is fractional abundance of word \( i \) and \( \theta \) is average valence, arousal, or dominance for word \( i \).

- Focus on valence, \( \theta = v \).

- Average valence typically falls between 5 and 7.
Unhappiness:

Some obvious problems/issues:

- Partial coverage of all words.
- Context is ignored.
Unhappiness:

Some obvious problems/issues:

- Partial coverage of all words.
- Context is ignored.
- You just don’t like it.
Unhappiness:

Some obvious problems/issues:

- Partial coverage of all words.
- Context is ignored.
- You just don’t like it. Really.
Unhappiness:

Some obvious problems/issues:

▶ Partial coverage of all words.
▶ Context is ignored.
▶ You just don’t like it. Really.

Clearly:

▶ Only suitable for large-scale texts.
Data sets:

Texts:
1. Song lyrics (1960–2007)
2. Song titles (1960–2008)

Sources:
- hotlyrics.com
- freedb.com
Happiness

Measuring emotional content

Data sets

Analysis
Songs
Blogs
SOTU

Winning: it’s not for everyone

Prediction

References

wefeelfine.org:
wefeelfine.org:

<table>
<thead>
<tr>
<th>Feeling</th>
<th>Gender</th>
<th>Age</th>
<th>Weather</th>
<th>Location</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>0s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>afghanistan</td>
<td>2006</td>
</tr>
<tr>
<td>argentina</td>
<td>2006</td>
</tr>
<tr>
<td>australia</td>
<td>2006</td>
</tr>
<tr>
<td>bahamas</td>
<td>2006</td>
</tr>
<tr>
<td>bangladesh</td>
<td>2006</td>
</tr>
<tr>
<td>belarus</td>
<td>2006</td>
</tr>
<tr>
<td>belgium</td>
<td>2006</td>
</tr>
<tr>
<td>brazil</td>
<td>2006</td>
</tr>
<tr>
<td>brunei</td>
<td>2006</td>
</tr>
<tr>
<td>darussalam</td>
<td>2006</td>
</tr>
<tr>
<td>bulgaria</td>
<td>2006</td>
</tr>
<tr>
<td>cambodia</td>
<td>2006</td>
</tr>
<tr>
<td>canada</td>
<td>2006</td>
</tr>
<tr>
<td>chile</td>
<td>2006</td>
</tr>
<tr>
<td>china</td>
<td>2006</td>
</tr>
<tr>
<td>colombia</td>
<td>2006</td>
</tr>
<tr>
<td>croatia</td>
<td>2006</td>
</tr>
<tr>
<td>czech</td>
<td>2006</td>
</tr>
<tr>
<td>republic</td>
<td>2006</td>
</tr>
<tr>
<td>denmark</td>
<td>2006</td>
</tr>
<tr>
<td>dominican</td>
<td>2006</td>
</tr>
<tr>
<td>republic</td>
<td>2006</td>
</tr>
<tr>
<td>estonia</td>
<td>2006</td>
</tr>
<tr>
<td>finland</td>
<td>2006</td>
</tr>
<tr>
<td>france</td>
<td>2006</td>
</tr>
<tr>
<td>gambia</td>
<td>2006</td>
</tr>
<tr>
<td>germany</td>
<td>2006</td>
</tr>
<tr>
<td>greece</td>
<td>2006</td>
</tr>
<tr>
<td>hungary</td>
<td>2006</td>
</tr>
<tr>
<td>iceland</td>
<td>2006</td>
</tr>
<tr>
<td>india</td>
<td>2006</td>
</tr>
<tr>
<td>indonesia</td>
<td>2006</td>
</tr>
<tr>
<td>iraq</td>
<td>2006</td>
</tr>
<tr>
<td>ireland</td>
<td>2006</td>
</tr>
<tr>
<td>israel</td>
<td>2006</td>
</tr>
<tr>
<td>italy</td>
<td>2006</td>
</tr>
</tbody>
</table>

Find Feelings
Some demographics for blog sentences:

Breakdown by # of sentences:

<table>
<thead>
<tr>
<th>Country</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>82.3</td>
</tr>
<tr>
<td>Canada</td>
<td>6.1</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>4.8</td>
</tr>
<tr>
<td>Australia</td>
<td>3.7</td>
</tr>
<tr>
<td>Philippines</td>
<td>0.4</td>
</tr>
<tr>
<td>Germany</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Some numbers:

<table>
<thead>
<tr>
<th>Counts</th>
<th>Song lyrics</th>
<th>Song titles</th>
</tr>
</thead>
<tbody>
<tr>
<td>All words</td>
<td>58,610,849</td>
<td>60,867,223</td>
</tr>
<tr>
<td>ANEW words</td>
<td>3,477,575 (5.9%)</td>
<td>5,612,708 (9.2%)</td>
</tr>
<tr>
<td>Individuals</td>
<td>~ 20,000</td>
<td>~ 632,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Counts</th>
<th>Weblogs</th>
<th>SOTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>All words</td>
<td>148,231,294</td>
<td>1,796,763</td>
</tr>
<tr>
<td>ANEW words</td>
<td>8,176,669 (5.5%)</td>
<td>61,926 (3.5%)</td>
</tr>
<tr>
<td>Individuals</td>
<td>~ 2,148,000</td>
<td>43</td>
</tr>
</tbody>
</table>
Most frequent ANEW words:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Song lyrics</th>
<th>Song titles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>love (7.37%)</td>
<td>love (7.39%)</td>
</tr>
<tr>
<td>2</td>
<td>time (4.18%)</td>
<td>time (4.19%)</td>
</tr>
<tr>
<td>3</td>
<td>baby (2.75%)</td>
<td>baby (2.75%)</td>
</tr>
<tr>
<td>4</td>
<td>life (2.59%)</td>
<td>life (2.60%)</td>
</tr>
<tr>
<td>5</td>
<td>heart (2.14%)</td>
<td>heart (2.15%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Weblogs</th>
<th>SOTU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>good (4.89%)</td>
<td>people (5.49%)</td>
</tr>
<tr>
<td>2</td>
<td>time (4.72%)</td>
<td>time (4.09%)</td>
</tr>
<tr>
<td>3</td>
<td>people (3.94%)</td>
<td>present (3.45%)</td>
</tr>
<tr>
<td>4</td>
<td>love (3.31%)</td>
<td>world (3.10%)</td>
</tr>
<tr>
<td>5</td>
<td>life (3.13%)</td>
<td>war (2.98%)</td>
</tr>
</tbody>
</table>
Outline

Measuring emotional content

Data sets

Analysis

Songs
Blogs
SOTU

Winning: it’s not for everyone

Prediction

References
Lyrics—average valence

![Graph showing the average valence of lyrics over decades from 1960 to 2010. The graph indicates a general decrease in valence over time.]
Lyrics—measurement robustness

100 random subsets of 750 ANEW words
Lyrics—average valence of genres:

- Gospel/Soul (6.91)
- Pop (6.69)
- Reggae (6.40)
- Rock (6.27)
- Rap/Hip-Hop (6.01)
- Punk (5.61)
- Metal/Industrial (5.10)
Valence shift details:

Given two texts $a$ and $b$:

- Measure difference in average valence: $v_{avg}^{(b)} - v_{avg}^{(a)}$
Valence shift details:

Given two texts \( a \) and \( b \):

- Measure difference in average valence: \( v_{avg}^{(b)} - v_{avg}^{(a)} \)
- Break difference down by contributions from individual words:
  \[
  \Delta_i = 100 \times \left[ p_{i,b} - p_{i,a} \right] \frac{v_i - v_{avg}^{(a)}}{v_{avg}^{(b)} - v_{avg}^{(a)}}
  \]
  \[
  \sum_i \Delta_i = v_{avg}^{(b)} - v_{avg}^{(a)}
  \]
- Rank words by \( |\Delta_i| \)
Valence Shift
Word Graph:

Post versus pre 1980 for song lyrics

Per word valence shift $\Delta$

Word number $i$

love ↓
lonely ↓
hate ↑
pain ↑
baby ↓
dead ↑
dead ↑
home ↓
sick ↑
fear ↑
hit ↑
hell ↑
fall ↑
sin ↑
lost ↑
sad ↓
burn ↑
lie ↑
scared ↑
afraid ↑
music ↓
life ↑
god ↑
trouble ↓
loneliness ↓

References
## Top 50 of \(~\sim\) 20,000 artists:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Artist</th>
<th>Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All-4-One</td>
<td>7.15</td>
</tr>
<tr>
<td>2</td>
<td>Luther Vandross</td>
<td>7.12</td>
</tr>
<tr>
<td>3</td>
<td>S Club 7</td>
<td>7.05</td>
</tr>
<tr>
<td>4</td>
<td>K Ci &amp; JoJo</td>
<td>7.04</td>
</tr>
<tr>
<td>5</td>
<td>Perry Como</td>
<td>7.04</td>
</tr>
<tr>
<td>6</td>
<td>Diana Ross &amp; The Supremes</td>
<td>7.03</td>
</tr>
<tr>
<td>7</td>
<td>Buddy Holly</td>
<td>7.02</td>
</tr>
<tr>
<td>8</td>
<td>Faith Evans</td>
<td>7.01</td>
</tr>
<tr>
<td>9</td>
<td>The Beach Boys</td>
<td>7.01</td>
</tr>
<tr>
<td>10</td>
<td>Jon B</td>
<td>6.98</td>
</tr>
<tr>
<td>11</td>
<td>Dru Hill</td>
<td>6.96</td>
</tr>
<tr>
<td>12</td>
<td>Earth Wind &amp; Fire</td>
<td>6.95</td>
</tr>
<tr>
<td>13</td>
<td>Ashanti</td>
<td>6.95</td>
</tr>
<tr>
<td>14</td>
<td>Otis Redding</td>
<td>6.93</td>
</tr>
<tr>
<td>15</td>
<td>Faith Hill</td>
<td>6.93</td>
</tr>
<tr>
<td>16</td>
<td>NSync</td>
<td>6.93</td>
</tr>
</tbody>
</table>

(criterion: \(\geq\) 50 songs and \(\geq\) 1000 ANEW words)
Bottom 50 of $\approx 20,000$ artists:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Artist</th>
<th>Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Slayer</td>
<td>4.80</td>
</tr>
<tr>
<td>2</td>
<td>Misfits</td>
<td>4.88</td>
</tr>
<tr>
<td>3</td>
<td>Staind</td>
<td>4.93</td>
</tr>
<tr>
<td>4</td>
<td>Slipknot</td>
<td>4.98</td>
</tr>
<tr>
<td>5</td>
<td>Darkthrone</td>
<td>4.98</td>
</tr>
<tr>
<td>6</td>
<td>Death</td>
<td>5.02</td>
</tr>
<tr>
<td>7</td>
<td>Black Label Society</td>
<td>5.05</td>
</tr>
<tr>
<td>8</td>
<td>Pig</td>
<td>5.08</td>
</tr>
<tr>
<td>9</td>
<td>Voivod</td>
<td>5.14</td>
</tr>
<tr>
<td>10</td>
<td>Fear Factory</td>
<td>5.15</td>
</tr>
<tr>
<td>11</td>
<td>Iced Earth</td>
<td>5.16</td>
</tr>
<tr>
<td>12</td>
<td>Simple Plan</td>
<td>5.16</td>
</tr>
<tr>
<td>13</td>
<td>Machine Head</td>
<td>5.17</td>
</tr>
<tr>
<td>14</td>
<td>Metallica</td>
<td>5.19</td>
</tr>
<tr>
<td>15</td>
<td>Dimmu Borgir</td>
<td>5.20</td>
</tr>
<tr>
<td>16</td>
<td>Mudvayne</td>
<td>5.21</td>
</tr>
</tbody>
</table>

(criterion: $\geq 50$ songs and $\geq 1000$ ANEW words)
Outline

Measuring emotional content

Data sets

Analysis
  Songs
  Blogs
  SOTU

Winning: it’s not for everyone

Prediction

References
Blogs—Overall trend

- Christmas Eve & Day, 2005
- Valentine’s Day, 2007
Per word valence shift $\Delta$

Mean valence: 6.06 vs. 5.86
2007/02/14 compared to 2007/02

Per word valence shift $\Delta$

Mean valence: 6.00 vs. 5.83

$\Sigma_{j=1}^{i} \Delta_{j}$
2006/09/11 compared to 2006/09

Mean valence: 5.72 vs. 5.81

Per word valence shift $\Delta$
Blogs—Age

- Self-report studies find little variation in happiness with age\textsuperscript{[7, 8]}
Blogs—Age

- Self-report studies find **little variation** in happiness with age \(^{[7, 8]}\)
- Surprising: Expect a rise and fall.
Blogs—Age

- Self-report studies find little variation in happiness with age [7, 8]
- Surprising: Expect a rise and fall.
- A ‘challenge’ for theory...
Blogs—Age

- Self-report studies find little variation in happiness with age\textsuperscript{[7, 8]}
- Surprising: Expect a rise and fall.
- A ‘challenge’ for theory...
- Related to the Easterlin Paradox: Money doesn’t buy happiness
Blogs—Age

- Self-report studies find little variation in happiness with age [7, 8]
- Surprising: Expect a rise and fall.
- A ‘challenge’ for theory...
- Related to the Easterlin Paradox: Money doesn’t buy happiness
Blogs

Average valence as a function of the age bloggers report they will turn in the year of their posting.
14 year olds compared to born 1960–1969

Mean valence: 5.55 vs. 5.96
Blogs—Latitude

Near equator—social factors

- Decrease in ‘good’ and ‘people’

Near poles—social/psychological/climate

- Offset by decrease in ‘hurt’ and ‘pain.’
- More ‘bed’ and ‘sleep.’
Blogs—day of the week

Very gentle weekly cycle:

Monday is not so bad for bloggers...
Female compared to Male

Per word valence shift $\Delta$

Word number $i$

Mean valence: 5.89 vs. 5.91
Outline

Measuring emotional content

Data sets

Analysis
  Songs
  Blogs
  SOTU

Winning: it’s not for everyone

Prediction

References
Presidential happiness:
Measuring Emotional Content

Goal: Improve on ANEW study

- Perform survey as an online game.
  - Local: university level
  - Intermediate: representative groups
  - Global: open on the Web
Measuring Emotional Content

Goal: Improve on ANEW study

- Perform survey as an online game.
  - Local: university level
  - Intermediate: representative groups
  - Global: open on the Web

Measure emotional content of

- Many more words
- Phonemes and letters
- Sentences
Where do superstars come from?

Rosen (1981): “The Economics of Superstars”

Examples:

- Full-time Comedians (≈ 200)
- Soloists in Classical Music
- Economic Textbooks (the usual myopic example)
- Highly skewed distributions (again)...

Rosen (1981): “The Economics of Superstars”
Superstars

Rosen’s theory:

- Individual quality $q$ maps to reward $R(q)$
- $R(q)$ is ‘convex’ ($d^2 R/dq^2 > 0$)
- Two reasons:
  1. Imperfect substitution:
     A very good surgeon is worth many mediocre ones
  2. Technology:
     Media spreads & technology reduces cost of reproduction of books, songs, etc.
- No social element—success follows ‘inherent quality’
Superstars

Adler (1985): “Stardom and Talent”

- Assumes extreme case of equal ‘inherent quality’
- Argues desire for coordination in knowledge and culture leads to differential success
- Success is then purely a social construction
Chase et al. (2002): “Individual differences versus social dynamics in the formation of animal dominance hierarchies” [3]

The aggressive female Metriaclima zebra (■):

Pecking orders for fish...
Dominance hierarchies

- Fish forget—changing of dominance hierarchies:

- 22 observations: about 3/4 of the time, hierarchy changed
Music Lab Experiment

Salganik, Dodds, and Watts (2006) 
“An experimental study of inequality and unpredictability in an artificial cultural market”[18]
Music Lab Experiment

### Happiness

#### Measuring Emotional Content

- **Data Sets**
- **Analysis**
  - Songs
  - Blogs
  - SOTU

#### Winning: It's Not for Everyone

- **Prediction**
- **References**

### Music Lab Experiment

**Frame 52/65**

<table>
<thead>
<tr>
<th>Title</th>
<th>Artist</th>
<th># of Downloads</th>
<th>[# of Downloads]</th>
<th>[# of Downloads]</th>
<th>[# of Downloads]</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;enough is enough&quot;</td>
<td>HARTSFIELD</td>
<td>20</td>
<td>12</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>&quot;for the sky&quot;</td>
<td>DEEP ENOUGH TO DIE</td>
<td>17</td>
<td>13</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>&quot;the end in end&quot;</td>
<td>THE BROKEN PROMISE</td>
<td>19</td>
<td>14</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>&quot;the belief above the answer&quot;</td>
<td>THIS NEW DAWN</td>
<td>12</td>
<td>24</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>&quot;walk away&quot;</td>
<td>MORAL HAZARD</td>
<td>6</td>
<td>20</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>&quot;steped in an orange peel&quot;</td>
<td>THE CALEFACIoN</td>
<td>8</td>
<td>17</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>&quot;breakdown&quot;</td>
<td>NOT FOR SCHOLARS</td>
<td>27</td>
<td>36</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>&quot;the time of the ball&quot;</td>
<td>SECRETARY</td>
<td>5</td>
<td>28</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>&quot;seductive tune, melodic breakdown&quot;</td>
<td>ART OF KANY</td>
<td>10</td>
<td>31</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>&quot;sensation anxiety&quot;</td>
<td>HYDRAULIC SANDWICH</td>
<td>20</td>
<td>17</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>&quot;the song of the sea&quot;</td>
<td>EMBER SKY</td>
<td>25</td>
<td>15</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>&quot;the song of the sea&quot;</td>
<td>SALUTE THE DAWN</td>
<td>13</td>
<td>12</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>&quot;eternal&quot;</td>
<td>RYAN ESMAKER</td>
<td>14</td>
<td>11</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>&quot;in the still&quot;</td>
<td>BEERBONG</td>
<td>12</td>
<td>17</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>&quot;best mistake&quot;</td>
<td>HALL OF FAME</td>
<td>19</td>
<td>51</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>
Music Lab Experiment

Experiment 1

<table>
<thead>
<tr>
<th>Title</th>
<th>Artist</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song 1</td>
<td>Artist 1</td>
<td>2010</td>
</tr>
<tr>
<td>Song 2</td>
<td>Artist 2</td>
<td>2015</td>
</tr>
<tr>
<td>Song 3</td>
<td>Artist 3</td>
<td>2020</td>
</tr>
</tbody>
</table>

Experiments 2–4

<table>
<thead>
<tr>
<th>Title</th>
<th>Artist</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song 1</td>
<td>Artist 1</td>
<td>2010</td>
</tr>
<tr>
<td>Song 2</td>
<td>Artist 2</td>
<td>2015</td>
</tr>
<tr>
<td>Song 3</td>
<td>Artist 3</td>
<td>2020</td>
</tr>
</tbody>
</table>

Frame 53/65
Music Lab Experiment

▶ Variability in final rank.
Inequality as measured by Gini coefficient:

\[ G = \frac{1}{(2N_s - 1)} \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} |m_i - m_j| \]
Music Lab Experiment

Unpredictability

\[ U = \frac{1}{N_s \binom{N_w}{2}} \sum_{i=1}^{N_s} \sum_{j=1}^{N_w} \sum_{k=j+1}^{N_w} |m_{i,j} - m_{i,k}| \]
Music Lab Experiment

Sensible result:

▶ Stronger social signal leads to greater following and greater inequality.
Music Lab Experiment

Sensible result:

- Stronger social signal leads to greater following and greater inequality.

Peculiar result:

- Stronger social signal leads to greater unpredictability.
Music Lab Experiment

Sensible result:

► Stronger social signal leads to greater following and greater inequality.

Peculiar result:

► Stronger social signal leads to greater unpredictability.

Very peculiar observation:

► The most unequal distributions would suggest the greatest variation in underlying ‘quality.’
Music Lab Experiment

Sensible result:

▶ Stronger social signal leads to *greater following and greater inequality*.

Peculiar result:

▶ Stronger social signal leads to greater *unpredictability*.

Very peculiar observation:

▶ The most unequal distributions would suggest the greatest variation in underlying ‘quality.’
▶ But success may be due to social construction through *following*.
Music Lab Experiment

Sensible result:

► Stronger social signal leads to greater following and greater inequality.

Peculiar result:

► Stronger social signal leads to greater unpredictability.

Very peculiar observation:

► The most unequal distributions would suggest the greatest variation in underlying ‘quality.’
► But success may be due to social construction through following.
► ‘Payola’ leads to poor system performance.
Music Lab Experiment—Sneakiness

Inversion of download count
Inversion of download count

The ‘pretend rich’ get richer ...
Inversion of download count
- The ‘pretend rich’ get richer ...
- ... but at a slower rate
Closing aside on Prediction:

Alan Greenspan (September 18, 2007):

http://wikipedia.org
Closing aside on Prediction:

Alan Greenspan (September 18, 2007):

“I’ve been dealing with these big mathematical models of forecasting the economy ...

http://wikipedia.org
Closing aside on Prediction:

Alan Greenspan (September 18, 2007):

“I’ve been dealing with these big mathematical models of forecasting the economy ...

If I could figure out a way to determine whether or not people are more fearful or changing to more euphoric,

http://wikipedia.org
Closing aside on Prediction:

Alan Greenspan (September 18, 2007):

“I’ve been dealing with these big mathematical models of forecasting the economy ...

If I could figure out a way to determine whether or not people are more fearful or changing to more euphoric,

I don’t need any of this other stuff.

http://wikipedia.org
Closing aside on Prediction:

Alan Greenspan (September 18, 2007):

“I’ve been dealing with these big mathematical models of forecasting the economy ...”

If I could figure out a way to determine whether or not people are more fearful or changing to more euphoric, I don’t need any of this other stuff.

I could forecast the economy better than any way I know.”
Economics, Schmeconomics

Greenspan continues:
Economics, Schmeconomics

Greenspan continues:
“The trouble is that we can’t figure that out. I’ve been in the forecasting business for 50 years.

Jon Stewart:
“You just bummed the @*!# out of me.”
Economics, Schmeconomics

Greenspan continues:
“The trouble is that we can’t figure that out. I’ve been in the forecasting business for 50 years. I’m no better than I ever was,
Economics, Schmeconomics

Greenspan continues:

“The trouble is that we can’t figure that out. I’ve been in the forecasting business for 50 years. I’m no better than I ever was, and nobody else is.
Economics, Schmeconomics

Greenspan continues:
“The trouble is that we can’t figure that out. I’ve been in the forecasting business for 50 years. **I’m no better than I ever was, and nobody else is.** Forecasting 50 years ago was as good or as bad as it is today.”
Economics, Schmeconomics

Greenspan continues:
“The trouble is that we can’t figure that out. I’ve been in the forecasting business for 50 years. I’m no better than I ever was, and nobody else is. Forecasting 50 years ago was as good or as bad as it is today. And the reason is that human nature hasn’t changed.”
Economics, Schmeconomics

Greenspan continues:
“The trouble is that we can’t figure that out. I’ve been in the forecasting business for 50 years. I’m no better than I ever was, and nobody else is. Forecasting 50 years ago was as good or as bad as it is today. And the reason is that human nature hasn’t changed. We can’t improve ourselves.”
Greenspan continues:
“The trouble is that we can’t figure that out. I’ve been in the forecasting business for 50 years. I’m no better than I ever was, and nobody else is. Forecasting 50 years ago was as good or as bad as it is today. And the reason is that human nature hasn’t changed. We can’t improve ourselves.”

Jon Stewart:
“You just bummed the @*!# out of me.”

▶ From the Daily Show (_Frame 60/65) (September 18, 2007)
References

M. Adler.
Stardom and talent.
pdf

M. Bradley and P. Lang.
Affective norms for english words (anew): Stimuli, instruction manual and affective ratings.
pdf

I. D. Chase, C. Tovey, D. Spangler-Martin, and M. Manfredonia.
Individual differences versus social dynamics in the formation of animal dominance hierarchies.
pdf
References II


R. A. Easterlin. 
Explaining happiness. 
[PDF](#)

W. T. Jones. 
The Classical Mind. 

D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone. 
A survey method for characterizing daily life experience: The day reconstruction method. 
[PDF](#)

R. Layard. 
Happiness. 
References IV

S. Lyubomirsky.  
*The How of Happiness.*  

C. Martinelli and S. W. Parker.  
Deception and misreporting in a social program.  

G. Mishne and M. de Rijke.  
Capturing global mood levels using blog posts.  
*AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs,* 2005.

C. Osgood, G. Suci, and P. Tannenbaum.  
*The Measurement of Meaning.*  
University of Illinois, Urbana, IL, 1957.
References V

S. Rosen.
The economics of superstars.

B. Russell.
*A History of Western Philosophy.*

M. J. Salganik, P. S. Dodds, and D. J. Watts.
An experimental study of inequality and unpredictability in an artificial cultural market.