Truth is a Lie
Rules & Semantics from Crowd Perspectives
“in embracing the diversity of human beings, we will find a surer way to true happiness”

Malcolm Gladwell, 2006

in his TED talk “Choice, happiness and spaghetti sauce”
“If we treat human brains as processors in a distributed system, each can perform a small part of a massive computation”

Luis von Ahn, 2006
“Games with a Purpose”, Computer Magazine
“The challenge for Watson was to answer questions posed in every nuance of natural language. The first step in accomplishing this was to have a lot of questions and their correct answers. Any cognitive system must take this first step.”

Chris Welty, 2014

Take Home Message

True and False is not enough,
There is diversity in human interpretation.

CrowdTruth introduces a spatial representation of meaning that through crowdsourcing harnesses disagreement.

Using CrowdTruth, untrained workers can be just as reliable as highly trained experts.
Is this song ....

Literate
Poignant
Wistful
Bittersweet
Autumnal
Brooding

Rollicking
Cheerful
Fun
Sweet
Amiable
Good-natured

Passionate
Rousing
Confident
Boisterous
Rowdy

Aggressive
Fiery
Tense
Anxious
Intense
Volatile

Humorous
Silly
Campy
Whimsical
Witty
Wry
Is this song ....

Literate
Poignant
Wistful
Bittersweet
Autumnal
Brooding

Rollicking
Cheerful
Fun
Sweet
Amiable
Good-natured

Passionate
Rousing
Confident
Boisterous
Rowdy

Aggressive
Fiery
Tense
Anxious
Intense
Volatile

Humorous
Silly
Campy
Whimsical
Witty
Wry
?- category("That’s not my name", x)

Literate
Poignant
Wistful
Bittersweet
Autumnal
Brooding

Rollicking
Cheerful
Fun
Sweet
Amiable
Good-natured

Passionate
Rousing
Confident
Boisterous
Rowdy

Aggressive
Fiery
Tense
Anxious
Intense
Volatile

Humorous
Silly
Campy
Whimsical
Witty
Wry

(Lee and Hu 2012)
ANTIBIOTICS are the first line treatment for indications of TYPHUS.

Patients with TYPHUS who were given ANTIBIOTICS exhibited several side-effects.

With ANTIBIOTICS in short supply, DDT was used during World War II to control the insect vectors of TYPHUS.
TREATS(Antibiotics, Typhus)

Is this *True*?
What is the treatment for typhus?

Antibiotic therapy is recommended for both endemic and epidemic typhus infections because early treatment with antibiotics (for example, azithromycin, doxycycline, tetracycline, or chloramphenicol) can cure most people infected with the bacteria. Consultation with an infectious-disease expert is advised especially if epidemic typhus or typhus in pregnant females is diagnosed. Delays in treatment may allow renal, lung, or nervous system problems to develop. Some patients, especially the elderly, may die.
What is the treatment for lupus?

In more severe cases of Lupus, medications that modulate the immune system (mainly corticosteroids, immunosuppressants) are used to control the disease and prevent recurrence of symptoms (known as flares). Depending on the dosage, **people who require steroids may develop Cushing's syndrome**. This may subside if and when the large initial dosage is reduced, but **long-term use of even low doses can cause elevated blood pressure and cataracts**. Numerous **new immunosuppressive drugs are being actively tested**. Rather than suppressing the immune system nonspecifically, as corticosteroids do, they target the responses of individual [types of] immune cells.

**TREATS(Corticosteroids, Lupus)**

Is this **True**?
Human Interpretation

“I will never understand people”

“They are the worst”
human subjectivity, ambiguity & uncertainty of expression are part of human semantics
current practices in computer science (AI) are based on fallacies
disagreement is bad
this all results in simplification of context
this means our systems understand only simplified black & white world
reality is complex diverse

but ...
current gold standard practices are not adequate any more
there is a gap we need to bridge
interestingly ...
in humanities ... diversity is well accepted
in sciences ... 
diversity is taboo
despite numerous examples ... string theory debate
and ... global warming debate
evolutionary bio. debate

and...

time

punctuated equilibrium

gradualism
What can we do?
“wisdom of crowds”

- collective decisions of large groups of people
- a group of error-prone decision-makers can be surprisingly good at picking the best choice
- when thumbs up or thumbs down - the chance of picking the right answer needs to be > 50%
- the odds that a most of them will pick the right answer is greater than any of them will pick it on their own
- performance gets better as size grows

1785
Marquis de Condorcet
“wisdom of crowds”

• asked 787 people to guess the weight of an ox
• none got the right answer
• their collective guess was almost perfect

1906
Sir Francis Galton
1942: Ballistics calculations and flight trajectories

WWII Math Rosies
transcribe raw flight data from celluloid film & oscillograph paper

NASA’s Computer Room
“Who Wants to Be a Millionaire?”

- help from **individual expert** or **audience poll**
- majority of the **audience** right 91% of the time
- **individuals** right only 65% of the time

2004
James Surowiecki

“wisdom of crowds”
Crowdsourcing
Human Computing
the wise crowd

diversity of opinion
independent perspectives
multitude of contexts
gives the big picture

James Surowiecki
current practices in computer science (AI) are based on fallacies
today’s real-world complex problem solving needs understanding of multiple perspectives

current practices in computer science (AI) are based on fallacies
single
truth
single truth

multiple truths
wisdom of crowds
disagreement is bad
disagreement is bad
“best collective decisions are result of disagreement, not consensus or compromise”

James Surowiecki
can we harness it?
Is this song ....

Literate  Poignant  Wistful  Bittersweet  Autumnal  Brooding

Passionate  Rousing  Confident  Boisterous  Rowdy

Aggressive  Fiery  Tense  Anxious  Intense  Volatile

Rollicking  Cheerful  Fun  Sweet  Amiable  Good-natured

Humorous  Silly  Campy  Whimsical  Witty  Wry
If "One Truth" & "No Disagreement"

<table>
<thead>
<tr>
<th>Worker</th>
<th>Mood-C1</th>
<th>Mood-C2</th>
<th>Mood-C3</th>
<th>Mood-C4</th>
<th>Mood-C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>W3</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>W5</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>W6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
If “Many Truths” & “Disagreement”

<table>
<thead>
<tr>
<th>Worker</th>
<th>Mood-C1</th>
<th>Mood-C2</th>
<th>Mood-C3</th>
<th>Mood-C4</th>
<th>Mood-C5</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W2</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W3</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>W4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>W6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>W7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W8</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>W10</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>
Disagreement as Signal

can indicate low quality workers

can indicate alternative interpretations

can indicate ambiguity in the categorisation

<table>
<thead>
<tr>
<th>Worker</th>
<th>Mood-C1</th>
<th>Mood-C2</th>
<th>Mood-C3</th>
<th>Mood-C4</th>
<th>Mood-C5</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>W10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>
Does this sentence express TREATS(Antibiotics, Typhus)?

ANTIBIOTICS are the first line treatment for indications of TYPHUS.

Patients with TYPHUS who were given ANTIBIOTICS exhibited several side-effects.

With ANTIBIOTICS in short supply, DDT was used during World War II to control the insect vectors of TYPHUS.

Disagreement is signal here too
If "One Truth", "No Disagreement" & "Experts"

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Treat</th>
<th>Prevent</th>
<th>Cause</th>
<th>Side Effect</th>
<th>Diagnose</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>S7</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>S8</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>S10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>
• we rejected 3 fallacies: one truth, experts rule and disagreement is bad
• by encouraging disagreement
• we exposed a richer set of possibilities
• that help in identifying, processing & understanding context
7 Myths

myths directly influence the practice of collecting human annotated data; Need to be revised with a new theory of truth (CrowdTruth)

- One truth: data collection efforts assume one correct interpretation for every example
- All examples are created equal: ground truth treats all examples the same – either match the correct result or not
- Detailed guidelines help: if examples cause disagreement - add instructions to limit interpretations
- Disagreement is bad: increase quality of annotation data by reducing disagreement among the annotators
- One is enough: most of the annotated examples are evaluated by one person
- Experts are better: annotators with domain knowledge provide better annotations
- Once done, forever valid: annotations are not updated; new data not aligned with old

Lora Aroyo, Chris Welty: Truth is a Lie: 7 Myths about Human Annotation, AI Magazine 2014.

http://lora-aroyo.org @laroyo
“It’s time to break free.”
a new approach to understanding semantics
harnessing the power & diversity of the crowd

crowdtruth.org
informs a vector space model of truth instead of a boolean, fuzzy or statistical model
triangle of reference

(Ogden & Richards, 1923)
example

annote

annotation choices

众包注释者

三角参考系

Category 1
Category 3
Category 5

passionate, rollicking, literate, humorous, aggressive, does not fit into any of the 5 clusters
rousing, cheerful, fun, poignant, wistful, campy, quirky, tense, anxious, clusters
confident, sweet, amiable, bittersweet, whimsical, witty, intense, volatile, clusters
boisterous, good-natured, autumnal, brooding, wry, visceral
rowdy


http://lora-aroyo.org

@laroyo
ANTIBIOTICS are the first line treatment for indications of TYPHUS.

Treat
Cause
Prevent

TREATMENTS expressed in references interpreted by crowd annotators.

Example: ANTIBIOTICS are used to treat TYPHUS.

<table>
<thead>
<tr>
<th>treat</th>
<th>prevent</th>
<th>side effect</th>
<th>diagnose</th>
<th>cause</th>
<th>other</th>
</tr>
</thead>
</table>

STEP 1: Select the valid RELATION(s) *(required)*

- [TREATS]
- [PREVENTS]
- [DIAGNOSED_BY_TEST_OR_DRUG]
- [CAUSES]
- [LOCATION]
- [SYMPTOM]
- [MANIFESTATION]
- [CONTRAINDICATES]
- [ASSOCIATED_WITH]
- [SIDE_EFFECT]
- [IS_A]
- [PART_OF]
- [OTHER]
- [NONE]
OUR NEW THEORY OF TRUTH
Feeling the way the **CHEST** expands (**PALPATION**), can identify areas of the lung that are full of fluid.

Unclear relationship between the two arguments reflected in the **disagreement**.
Redness (HYPERAEMIA), irritation (chemosis) and watering (epiphora) of the eyes are symptoms common to all forms of CONJUNCTIVITIS.

Is HYPERAEMIA — related to CONJUNCTIVITIS?

| 0 | 0 | 0 | 1 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

cause symptom

Clearly expressed relation between the two arguments reflected in the agreement
## Relation Similarity

|   | 1 | 3 | 4 | 7 | 6 |   |   |   |   |   | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |   |
| 1 |   |   | 13|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 4 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 1 |   |   |
| 7 |   |   |   |   |   |   |   |   |   |   |   | 1 |   |   |   |   |   |   |   |   |   |   |
| 6 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 1 |   |   |   |   |
| 3 |   | 3 |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 1 |   |   |   |   |   |
| 8 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 1 |   |
| 5 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 1 |
| 1 |   |   |   |   |   |   |   |   | 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 1 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | 1 |   |   |

http://lora-aroyo.org
@laroyo
Sentence-Relation Score

Measures how clearly a sentence expresses a relation

Unit vector for relation $R_6$

Cosine = .55  confidence

Sentence Vector

http://lora-aroyo.org  @laroyo
Worker Quality

Worker-Sentence Disagreement

Worker's sentence vector

Sentence Vector

AVG (Cosine)

Measured per worker averaged over all sentences (whether worker consistently disagrees with everyone else)
Worker Similarity

Worker-Worker Disagreement

Worker 1
sentence vector

Worker 2
Vector

Measured for pairs of workers / sentence averaged over all sentences
(idenitfy communities of though)
### Aggregated View of Annotations of 1 Selected Job(s)

#### High Quality Annotations.

#### Unit Clarity

After filtering low quality

#### Type of Annotations of the 1 selected job(s)

Click a category to see the distribution of annotations.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>1</th>
<th>0</th>
<th>3</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>4</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Unit Clarity Score

- unit 2179 in job 6: 0.92847669088526
- unit 2178 in job 6: 0.9760921603773
- unit 2174 in job 6: 0.9712858623726
- unit 2168 in job 6: 0.9735555485044
- unit 2128 in job 6: 0.57735026918963
- unit 1897 in job 6: 0.86824314212446
- unit 1892 in job 6: 0.70710678118655
- unit 1871 in job 6: 0.672867 unit 1656 in job 6: 0.816496 clarity: score: 0.9970544
- unit 1825 in job 6: 0.942806
- unit 1657 in job 6: 0.737264
- unit 1656 in job 6: 0.99705448550158
- unit 1659 in job 6: 0.56308682468615
- unit 1630 in job 6: 0.77459666924148
- unit 1617 in job 6: 0.94345635304973
- unit 1617 in job 6: 0.85280286542244
- unit 1617 in job 6: 0.738548945876
- unit 1578 in job 6: 0.62554324217122
- unit 1518 in job 6: 0.8574922571254
- unit 1488 in job 6: 0.97072534399435
- unit 1286 in job 6: 0.92884769088526
- unit 1265 in job 6: 0.9760921603773
- unit 1242 in job 6: 0.80178372573727
- unit 1218 in job 6: 0.57735026918963
- unit 1206 in job 6: 0.75047877438646
- unit 1200 in job 6: 0.78262379212493
- unit 1176 in job 6: 0.68824720161169
- unit 1155 in job 6: 0.63245553203368
- unit 1148 in job 6: 0.68599434057094
- unit 1119 in job 6: 0.69631062382279
- unit 1102 in job 6: 0.73720978077449
- unit 1099 in job 6: 0.894427190999992
- unit 1076 in job 6: 0.57735026918963

---

**Links:**
- [Web & Media Group](http://lora-aroyo.org)
- [Twitter](https://twitter.com/laroyo)
Example Results

Platforms distribution for 34 Unit(s) in the 1 selected job(s)
Click a category to see the distribution of judgements per unit

Judgements on 34 Unit(s) in 1 Selected job(s)
Select an area to zoom. To see detailed information select individual units. From legend select/deselect features.

CrowdTruth Average Unit Clarity: the value is defined as the maximum unit annotation score achieved on any annotation for that unit. High agreement over the annotations is represented by high cosine scores, indicating a clear unit. Click to select/deselect.

job 6 avg clarity after filter  # of high quality judgements of job 6

http://lora-aroyo.org  @laroyo
Example Results

Quality of 66 Worker(s) of the 1 selected job(s)
Click a category to see the distribution of judgements per worker.

Judgements of 66 Worker(s) in 1 Selected job(s)
Select an area to zoom. To see detailed information select individual units. From legend select/deselect features.

Worker 14584835
platform : cf
flagged : true
platform worker trust : 1
avg worker cosine across jobs : 0.49
avg worker agreement across jobs : 0.37

Job 6:
# of high quality judgements of this worker: 10
# of low quality judgements of this worker: 0

Worker ID (Red color: Workers identified at least ones on the platform as low quality)
- Job 6 avg agreement after filter - Job 6 avg cosine after filter - # of high quality judgements of job 6
- # of low quality judgements of job 6
Example Results

Overview of 6 Jobs
Overview of jobs having # of ReLex jobs like "relex"

Select an area to zoom. To see detailed information select individual jobs. Right click for table view. From legend select features. Adjust Y-Axis by dragging the labels (double click to return to default).

CrowdTruth Average Worker Agreement score. Higher scores indicate better quality workers. Click to select/deselect.

<table>
<thead>
<tr>
<th>metrics</th>
<th>payment</th>
<th>workers</th>
<th>judgements</th>
<th>units</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

http://lora-aroyo.org  @laroyo
experimental results show that CrowdTruth is a better way of capturing context, a more accurate way to predict & explain truth.
Training a Relation Extraction Classifier

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>Cost per sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CrowdTruth</td>
<td>0.642</td>
<td>$0.66</td>
</tr>
<tr>
<td>Expert Annotator</td>
<td>0.638</td>
<td>$2.00</td>
</tr>
<tr>
<td>Single Annotator</td>
<td>0.492</td>
<td>$0.08</td>
</tr>
</tbody>
</table>

“wisdom of the crowd” provides training data that is at least as good if not better than experts only with proper analytic framework for harnessing disagreement from the crowd.
Learning Curves

above 400 sent.: crowd consistently over baseline & single
above 600 sent.: crowd out-performs experts
Take Home Message

True and False is not enough,
There is diversity in human interpretation

CrowdTruth introduces a spatial representation of meaning that through crowdsourcing harnesses disagreement

Using CrowdTruth untrained workers can be just as reliable as highly trained experts
To be AND not to be: quantum intelligence?

Lora Aroyo & Chris Welty

https://www.youtube.com/watch?v=CyAl_IVUdzM