

Crowdsourcing the Creation of an Emotion Lexicon



Saif Mohammad and Peter Turney
National Research Council Canada

Collaborator

Dr. Peter Turney
Principal Research Officer
Interactive Information Group
National Research Council Canada
Adjunct Professor at the University of Ottawa





Outline

- Emotions
- Crowdsourcing
- Annotation analysis



EMOTIONS

Examples



joy

sadness

surprise

hate

love

relief

shame

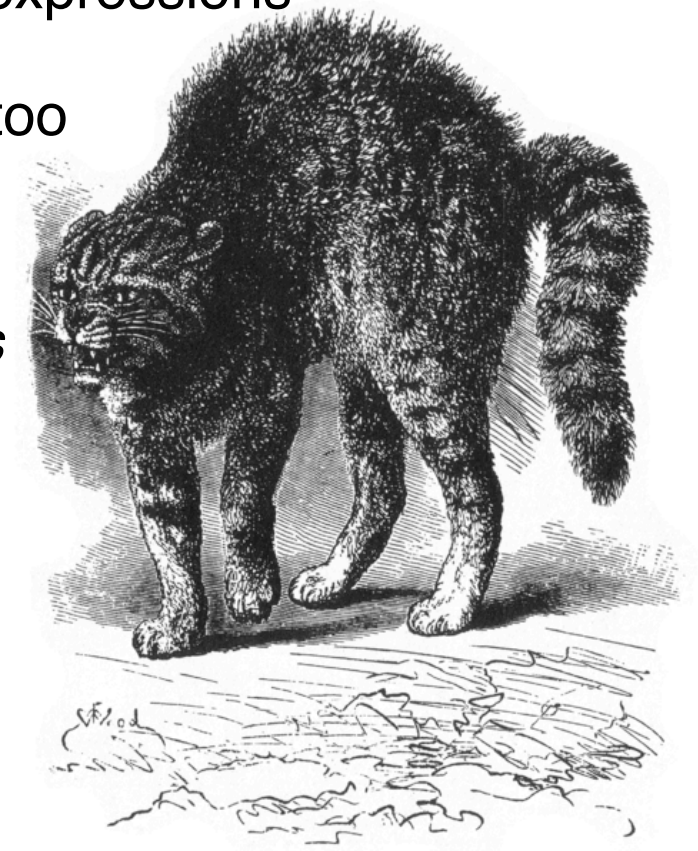
fear

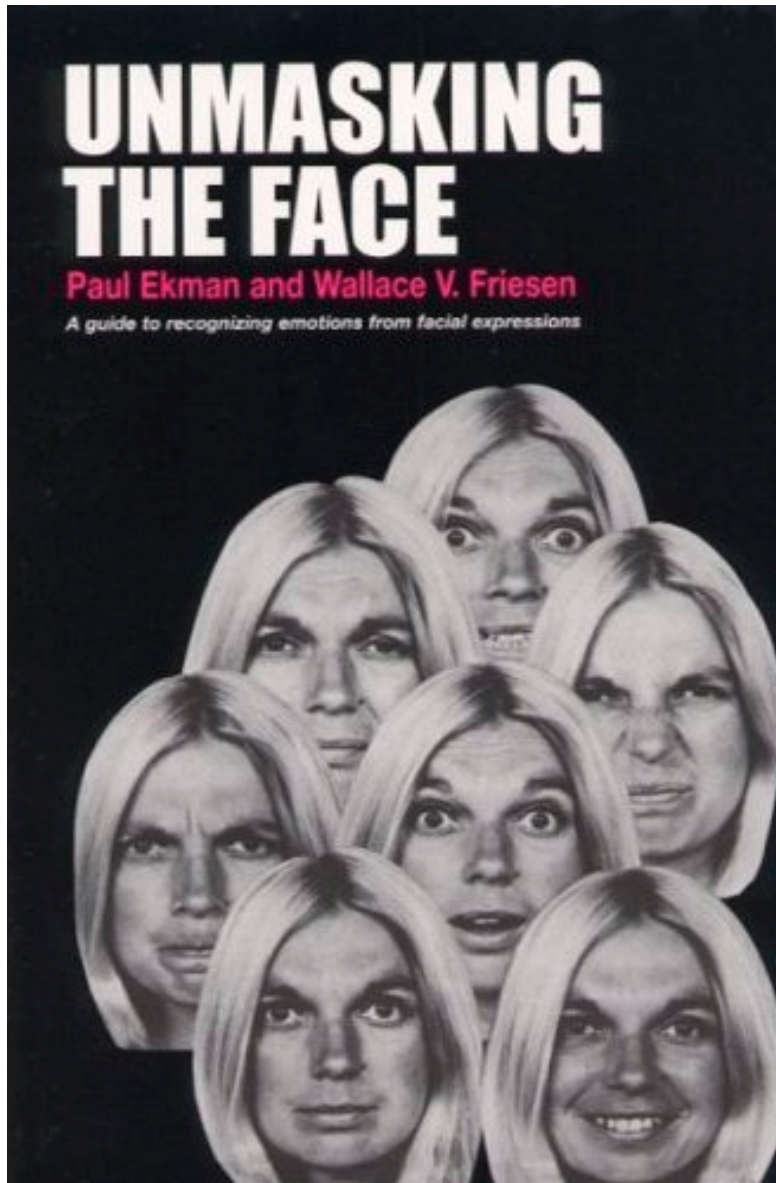
anxiety

And many many more.

Pervasiveness

- Cross-cultural
 - Paul Ekman's work on facial expressions
- Animals express/feel emotions too
 - Charles Darwin's *The Expressions of the Emotions in Man and Animals*





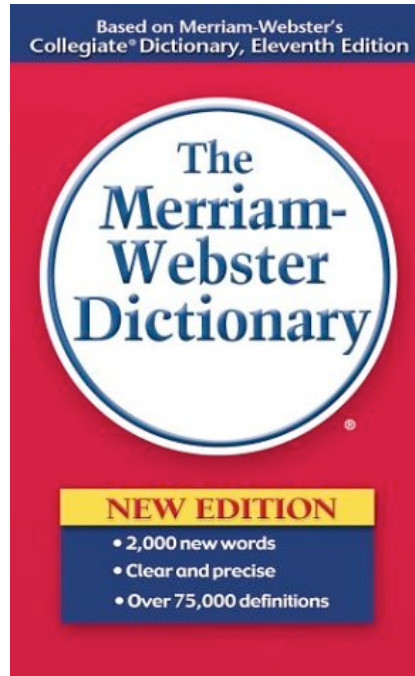
PENGUIN CLASSICS

CHARLES DARWIN

*The Expression of the Emotions
in Man and Animals*

Crowdsourcing the Creation of an
Emotion Lexicon.

Definition



A mental reaction subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body.

- subjectively experienced
- strong feeling
- directed toward (or evoked by) a specific object

Painting



The Destroyer
- *Frank Frazetta*

Sentence



(Phil, from the San Francisco Chronicle)
speaker/writer



When your cartoon can get you killed



listener/reader



Death threats over South Park episode
Event



Extremists
Participants



Trey Parker, Matt Stone
Participants

Our focus: words



evokes joy



When your **cartoon** can get you **killed**

killed



evokes sadness



Our goal

- Create and analyze an emotion lexicon through input from a large number of people.
- Examples:
 - **vampire** typically evokes fear
 - **vacation** typically evokes joy
 - **death** typically evokes sadness
 - **result** typically evokes anticipation



Motivation for emotion detection

- Devising automatic dialogue systems that respond appropriately to different emotional states of the user.
 - customer relation models
 - intelligent tutoring systems
 - emotion-aware games
- Tracking sentiment towards politicians, movies, products.
- Determining emotional intelligence.
- Assisting in writing e-mails, documents, and other text to convey desired emotion (and avoiding misinterpretation).

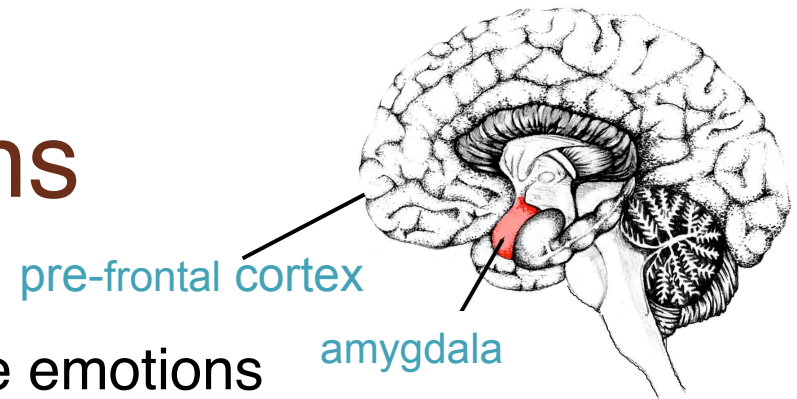


Motivation for emotion detection

(continued)

- Depicting flow of emotions in novels and other books
- Identifying what emotion a newspaper headline is trying to evoke
- Re-ranking and categorizing information/answers in a web 2.0 world
- Detecting how people use emotion-bearing-words and metaphors to persuade and coerce others
- Developing humanoid robots

Kinds of emotions



- Instinctual versus cognitive emotions
 - sensing and perceiving from the amygdala versus thinking and reasoning through the prefrontal cortex
- Basic versus complex
 - base emotions lead to more complex ones
- Categorization based on duration
 - some emotions occur over a period of seconds (e.g. surprise), whereas others can last years (e.g. love)



Base emotions

- Ekman: 6
 - joy, sadness, fear, anger, surprise, disgust
- Plutchik: 8
 - Ekman's 6 + anticipation + trust
 - 4 pairs of antonymous emotions
- More proposals by Parrot, Loyban, and others

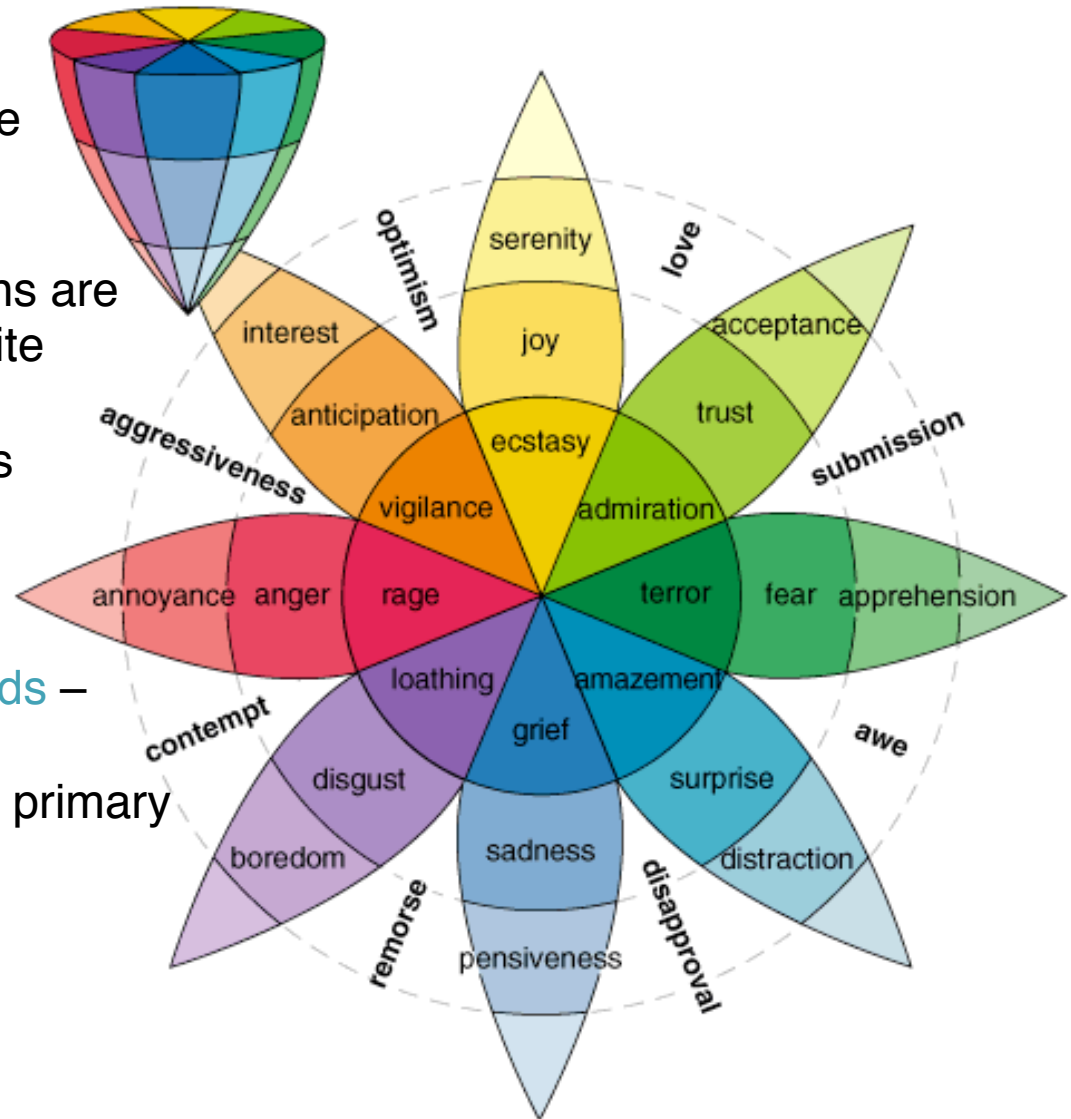
Robert Plutchik's psychoevolutionary theory

- Biologically primitive
- Evolved to increase “reproductive fitness”
- Triggers for behavior with high survival value
 - e.g., fear inspires fight-or-flight response



Plutchik's wheel of emotions

- Similar emotions are adjacent
- Contrasting emotions are diametrically opposite
- The radius indicates intensity
- In the white spaces are the **primary dyads** – emotions that are combinations of the primary emotions





Using Mechanical Turk for



LARGE-SCALE EMOTION ANNOTATION

Amazon's Mechanical Turk

- Requester
 - breaks task into small independent units – HITs
 - specifies:
 - compensation for solving each HIT
 - # of independent annotations required for each HIT
a.k.a. # of assignments/HIT
 - uploads HITs
- Turkers
 - attempt as many HITs as they wish
- Requester
 - inspects each assignment: approves or rejects



Amazon's Mechanical Turk: Features

- Inexpensive
 - \$1/hour is not uncommon
- Convenient
 - Web-based
 - Scripts to upload HITs and review assignments
- Takes care of certain ethics issues
 - Anonymity
 - No pressure on workers to solve HITs



Amazon's Mechanical Turk: Challenges

- Annotator time is precious
 - Minimum reading, minimum writing
 - Maximum information throughput
- Requestor time is precious
 - Automatic review and assimilation of annotations

One solution: Multiple choice questions, with examples instead of explanations.



Example question

How much does vampire evoke/produce the emotion fear?
(For example, horror and scary may strongly evoke fear.)

- vampire does not evoke fear
- vampire weakly evokes fear
- vampire moderately evokes fear
- vampire strongly evokes fear



Amazon's Mechanical Turk: Challenges

- Malicious annotations
 - Random selection or garbage data entry
 - Deliberate incorrect annotation
- Inadvertent and infrequent errors
 - Turker attempts HITs for unfamiliar words too

Emotion annotation: Challenges

- Words used in different senses and in different contexts can evoke different emotions.

High aspect ratio wings allow low speed flight.

The fight or flight response is crucial for survival.

- How to convey the target sense to the annotator?
 - definitions are long
 - need to discourage annotation for unfamiliar words





Our solution

Directions: Attempt HIT only if you are familiar with the word. Words in different senses may have different emotion associations. Question 1 will guide you to the intended sense.

Q1. Which word is closest in meaning (most related) to flight?

- buying
 - avoidance
 - doubt
 - boredom
-
- Near-synonym is taken from a thesaurus.
 - Categories in a thesaurus act as coarse senses
 - Three distracters are chosen at random

Emotion annotation: Challenges

- ✓ • Words used in different senses and in different contexts can evoke different emotions.

High aspect ratio wings allow low speed flight.

The fight or flight response is crucial for survival.



- ✓ • How to convey the target sense to the annotator?
 - definitions are long
 - need to discourage annotation for unfamiliar words



Identifying bad assignments

- If the word choice question is answered wrongly, then the whole assignment is discarded (answers to all questions in the HIT by the Turker are discarded)
- If an annotator gets more than 1 in 3 questions wrong, then we assume they are not following instructions.
 - We reject all their assignments.

Amazon's Mechanical Turk: Challenges

- ✓ • Malicious annotations
 - Random selection or garbage data entry
 - Deliberate incorrect annotation
- ✓ • Inadvertent and infrequent errors
 - Turker attempts HITs for unfamiliar words too

Will detect it 75% of the time.



Target n-grams

- Conditions:
 - Most frequent terms in the Google n-gram corpus
 - Must be in the thesaurus in just one or two categories
- Most frequent monosemous n-grams in each of the following categories:
 - noun unigrams (200)
 - noun bigrams (200)
 - verb unigrams (200)
 - verb bigrams (200)
 - adjective unigrams (200)
 - adjective bigrams (200)
 - adverb unigrams (200)
 - adverb bigrams (200)

Target n-grams (continued)


- Most frequent monosemous terms in the General Inquirer (GI) that are:
 - marked as positive (200)
 - marked as negative (200)
- Terms in WordNet Affect Lexicon (WAL) that have one or two senses and are:
 - marked as anger terms (107)
 - marked as disgust terms (25)
 - marked as fear terms (58)
 - marked as joy terms (109)
 - marked as sadness terms (86)
 - marked as surprise terms (39)

2176 terms in all.



Questions:

1. Which word is closest in meaning (most related) to flight?
 - buying
 - avoidance
 - doubt
 - boredom
2. How positive (good, praising) is flight (for example, nice and **excellent** are strongly positive):
 - flight is not positive
 - flight is weakly positive
 - flight is moderately positive
 - flight is strongly positive



Questions (continued):

3. How negative (bad, criticizing) is flight
(for example, poor and pathetic are strongly negative):
 - flight is not negative
 - flight is weakly negative
 - flight is moderately negative
 - flight is strongly negative

4. How much does flight evoke/produce the emotion joy
(for example, happy and fun may strongly evoke joy):
 - flight does not evoke joy
 - flight weakly evokes joy
 - flight moderately evokes joy
 - flight strongly evokes joy



ANNOTATION ANALYSIS

Numbers

- 2176 (HITs) x 5 (assignments per HIT) = 10,880 assignments
- Annotators: 1012
- Turkers spent on average about 1 minute per HIT
- Hourly wage was about \$2.40 (about 4 cents per HIT)
- Total cost: US \$470 (cost per term: about 22 cents)
- More than 95% of the assignments had the correct answer for the word choice question.
 - The rest were discarded.
- 2081 terms had 3 or more valid assignments
 - on average 4.75 assignments per HIT

Emotive and non-emotive

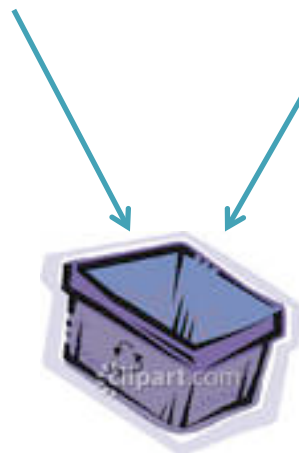
- Practical NLP applications may care for only two levels of intensity
- Example: vampire-fear

no fear
0 votes

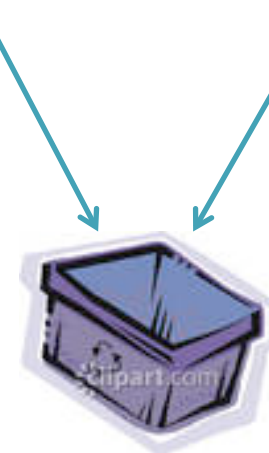
weak fear
1 vote

moderate fear
2 votes

strong fear
2 votes



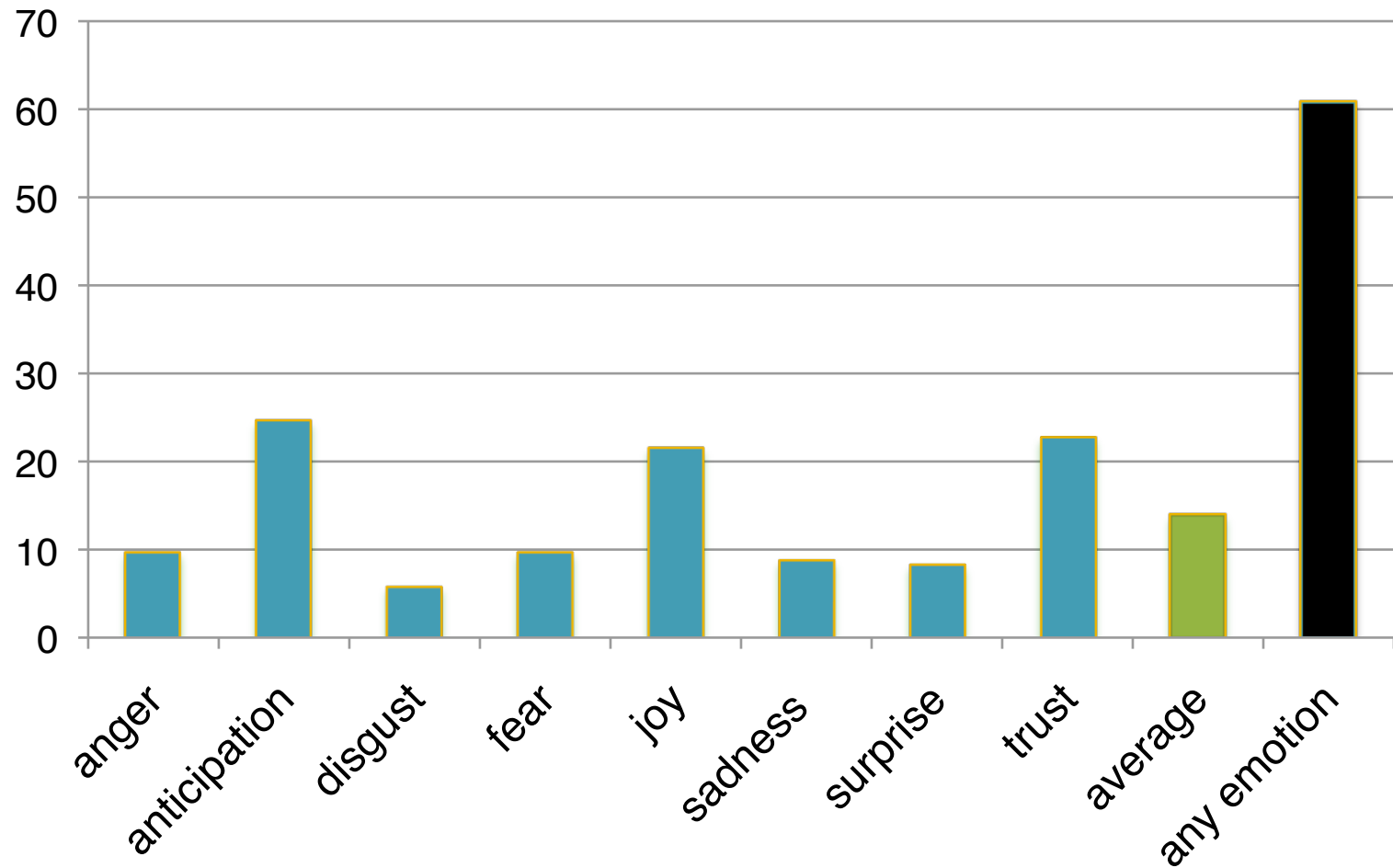
non-emotive
1 vote



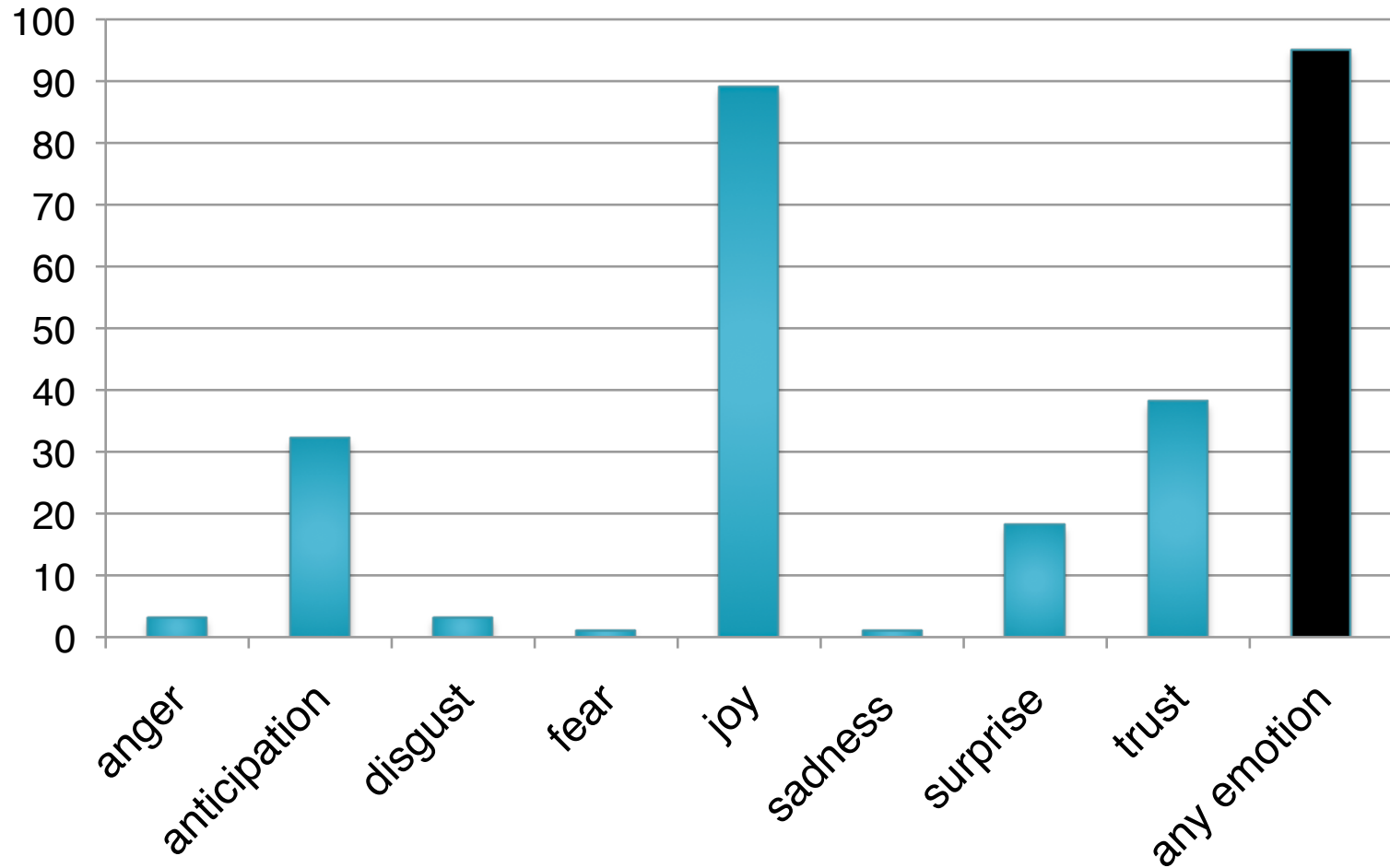
emotive
4 votes



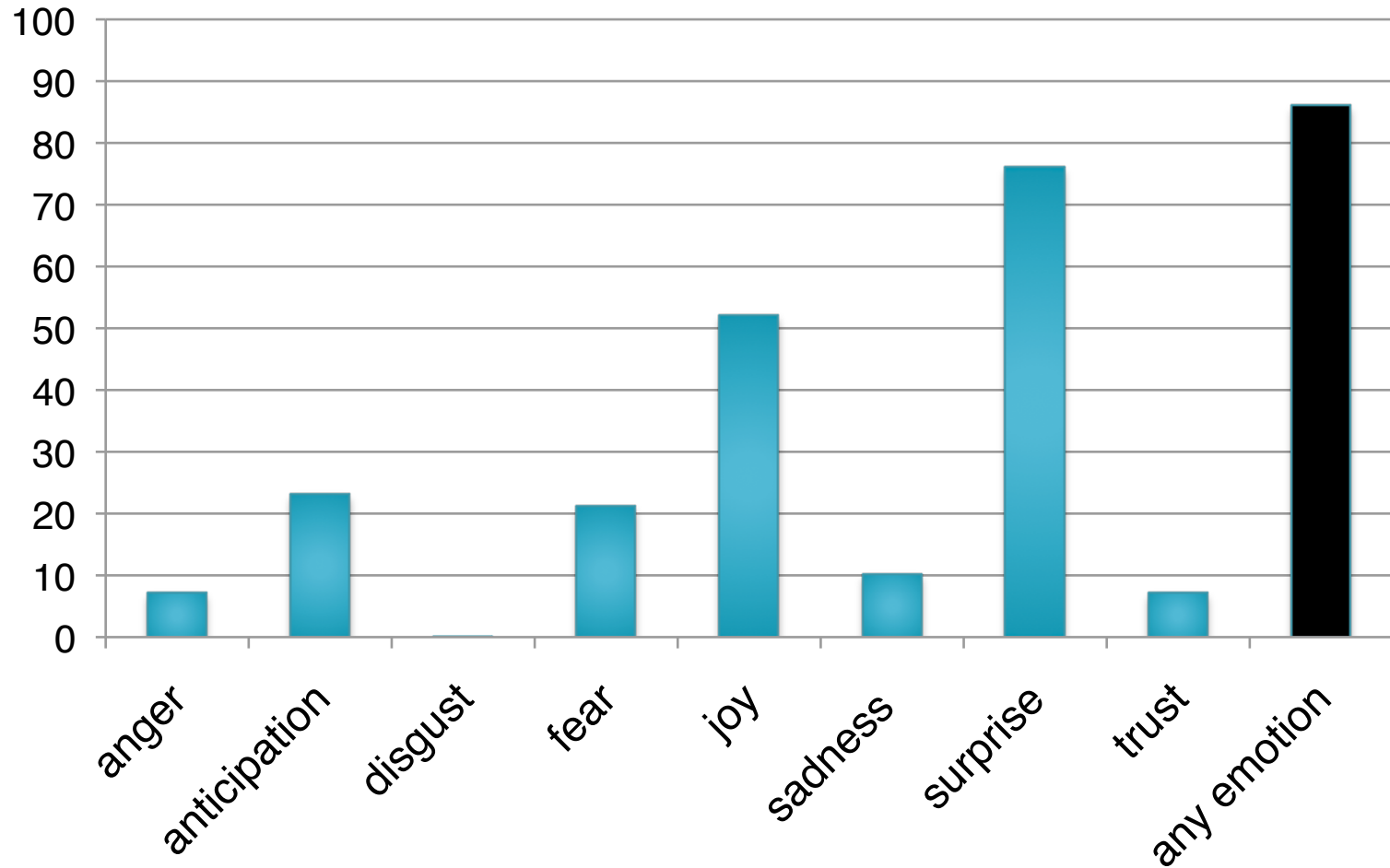
Percent of most frequent terms that are emotive



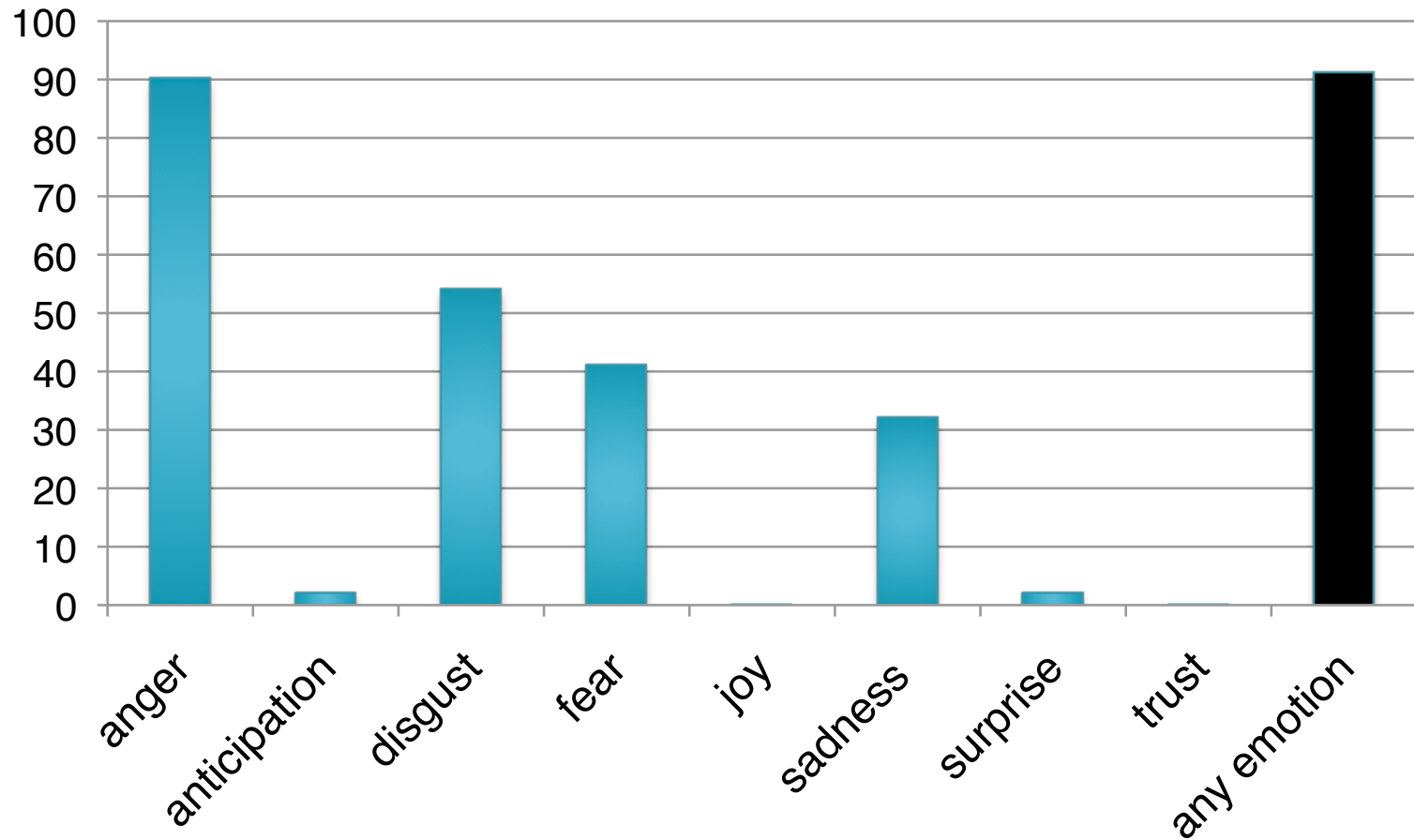
% of WAL joy terms evocative of different emotions as per the Turkers



% of WAL surprise terms evocative of different emotions as per the Turkers



% of WAL anger terms evocative of different emotions as per the Turkers





What was missed?

baffled

covetousness

exacerbate

gravel

pesky

pestering



Anger and Joy!

adjourn

credit card

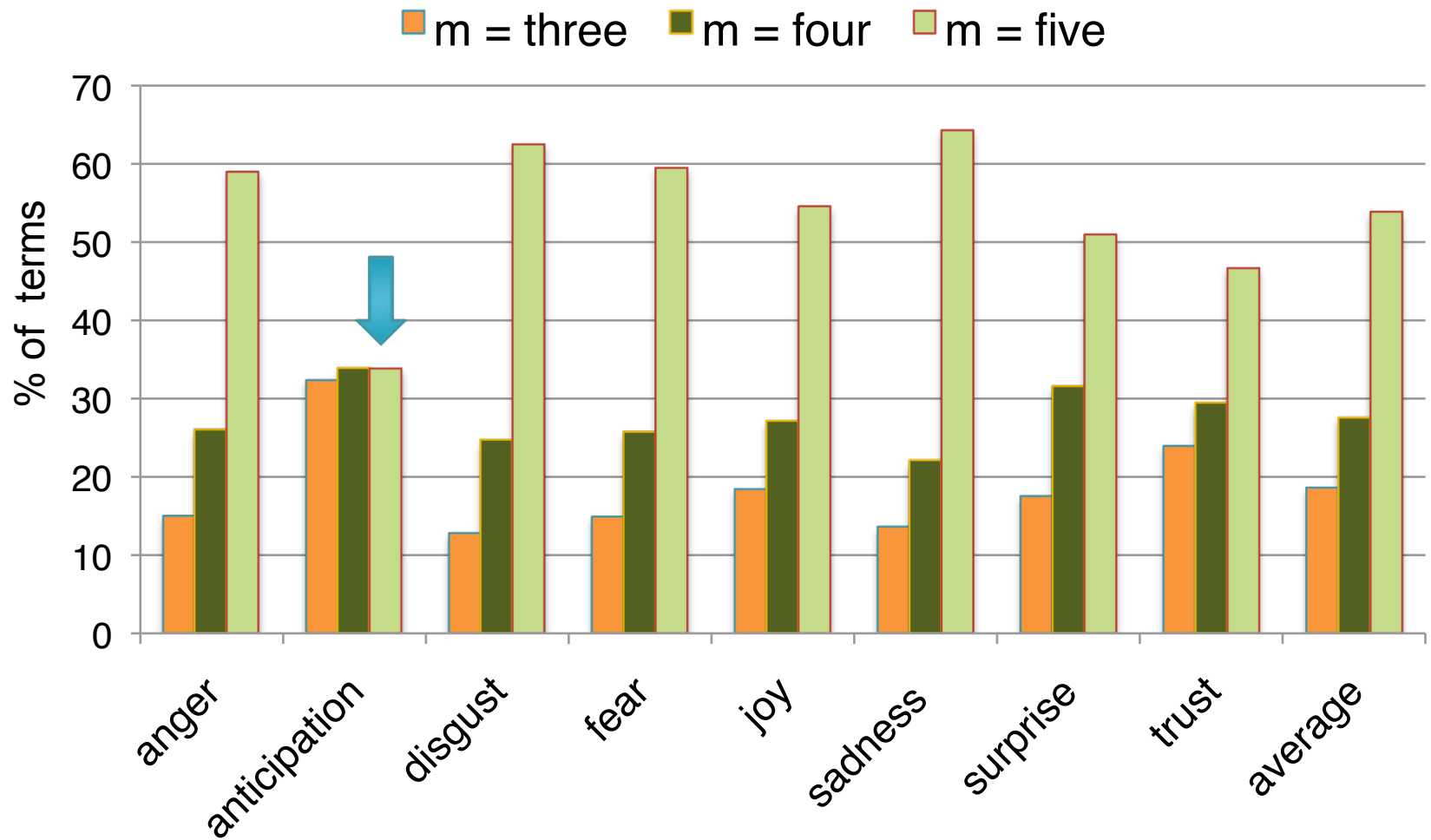
find out

gloat

spontaneously

surprised

Agreement at two intensity levels: Majority class (m) = 3, 4, 5



Conclusions

- Regular folks can produce high quality emotion annotations with proper guidelines and checks:
 - Annotations match those in GI and WAL
 - Reasonable degree of agreement
 - Anticipation and trust are sources of more disagreement
- A large number of commonly used terms are emotive:
 - About 61%
(evoke one or the other base emotion)

Current work



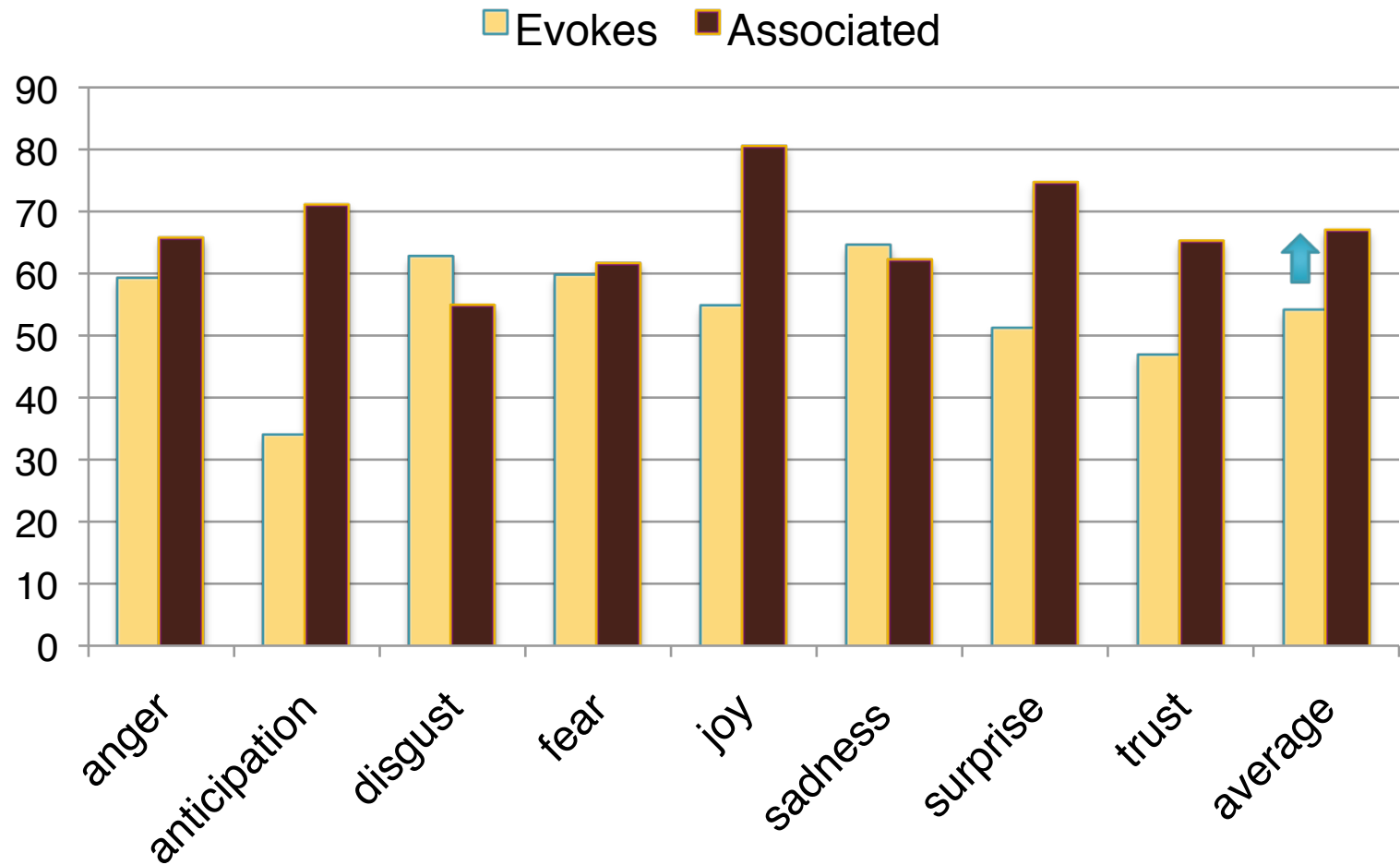
~~evoked~~ associated with joy

When your **cartoon** can get you **killed**



~~evoked~~ associated with sadness

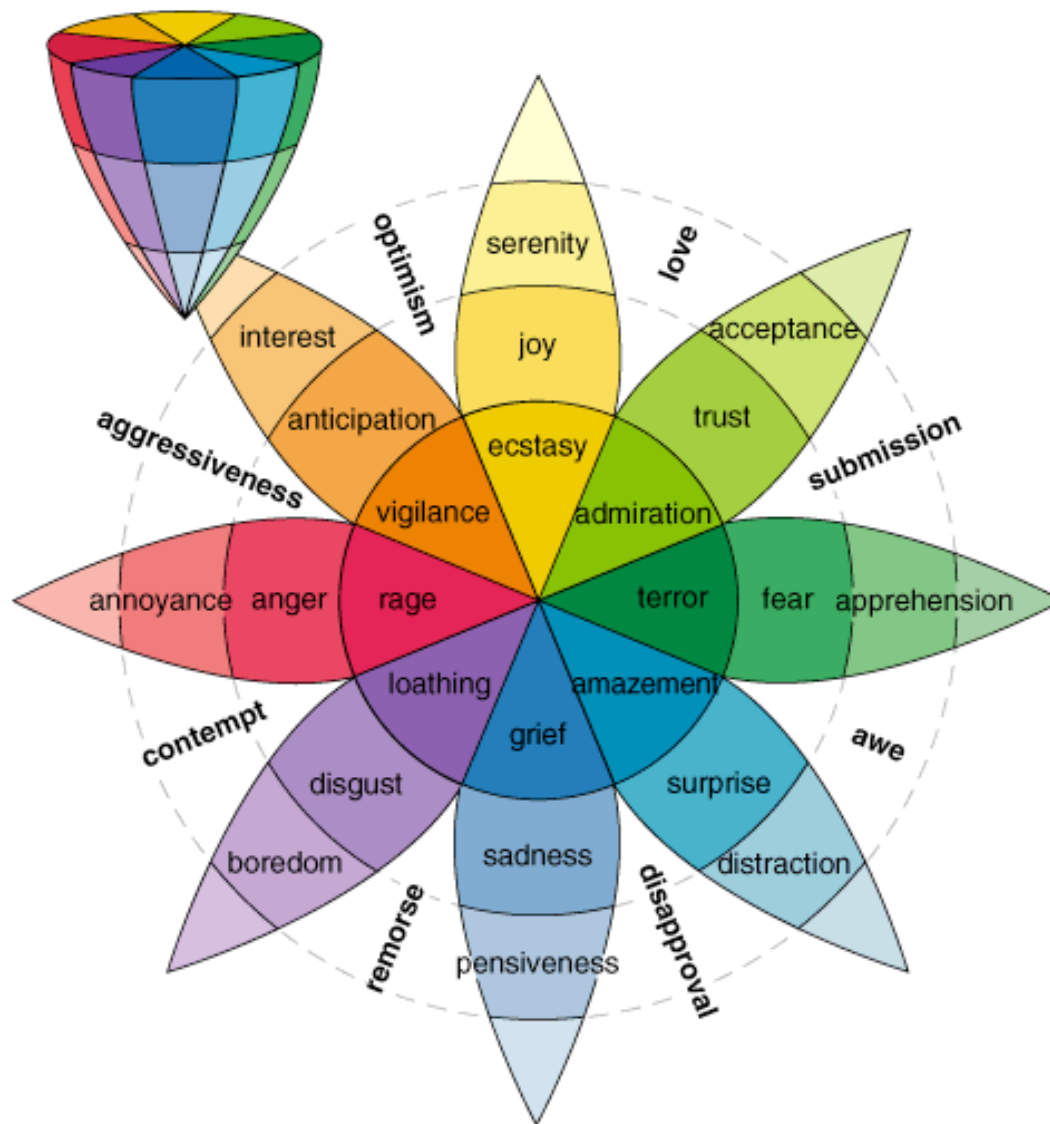
% of terms where all 5 agree





Current and future work

- Determining which terms have strong color associations and if there is a correlation with emotions.
- Determine how much near synonyms vary in emotional content.
- Empirically verify if complex emotions are indeed combinations of basic emotions.
- Create a much larger lexicon (40,000 terms, say).
 - Make lexicon publicly available.
- Incorporate a game and scoring in the annotation.
- Use lexicon in applications.



Questions.

Example HIT

Answer questions pertaining to “vampire”.

Directions:

- Return/skip HIT if you do not know the meaning of the word.
- Attempt HITS only if you are a native speaker of English, or fluent in English.
- Certain "check questions" will be used to make sure the annotation is responsible and reasonable. Assignments that fail these tests will be rejected. If you fail more than 33% of these check questions, then it will be assumed that you are not following instructions 1 and/or 2 above, and ALL your assignments will be rejected.
- Only those HITs with ALL questions answered reasonably and responsibly will be approved.
- Expected date the assignments will be approved: April 20, 2010.
- Confidentiality notice: Your responses are confidential...