Keeping it Cool:
Emotional Biases in the English Lexicon

Amy Beth Warriner & Victor Kuperman

The Pollyanna Hypothesis

There is a universal human tendency to use evaluatively positive words more frequently, diversely and facilely than evaluatively negative words. Put even more simply, humans tend to "look on (and talk about) the bright side of life."

Boucher & Osgood, 1969

Type Frequency

Of all the words we have in our lexicon, how many are positive and how many are negative?

The distribution of items in our lexicon reflects what we pay attention to and distinguish between in our world.

Token Frequency

How often do we use the positive and negative words available in the lexicon? Is valence related to word frequency of occurrence?

We may purposely use more positive words to reflect or create positive experiences OR our increased use of certain words may create positive feelings from mere exposure.

McMaster
University

Previous Studies

Study	Words	Types	Tokens
Johnson, Thomson & Frincke (1960)	150	☑	☑
Zajonc (1968)	154		☑
Boucher & Osgood (1969)	100	☑	☑
Suitner & Maas (2008)	130		☑
Unkelbach, et al. (2010)	90		☑
Rozin, Berman & Royzman (2010)	14		☑
Augustine, Mehl & Larsen (2011)	1034	☑	☑
Garcia, Garas, & Schweitzer (2012)	1034		☑
Kloumann et al. (2012)	10,222	☑	☒

Comparing Datasets

Warriner, Kuperman & Brysbaert, 2013	Kloumann et al., 2012
13,915 words	10,222 words
Drawn primarily from the highest frequency items in SUBTLEX-US	Combined from the most frequent 5,000 words in four collections (Twitter, Google Books, New York Times, and music lyrics)
Restricted to lemmas and content words	Unrestricted – includes multiple spelling variants (bday, b-day), special characters and alphanumeric strings (#tcot, a3) and foreign words (hij, ziin)
Valence ratings collected from Amazon Mechanical Turk – strict rejection criteria and high correlations with previous ratings	Valence ratings collected from Amazon Mechanical Turk – no rejection criteria or correlation with previous studies reported

*Correlation between ratings is .919, but only 4,504 words overlapped

McMaster University

Emotion is (at least) Two-Dimensional

- Valence
  - Measured on a scale of 1 (how sad) to 9 (how happy) a word makes a person feel

- Arousal
  - Measured on a scale of 1 (how calm) to 9 (how excited) a word makes a person feel

NOT YET EXAMINED

High Valence, High Arousal
SEX

High Valence, Low Arousal
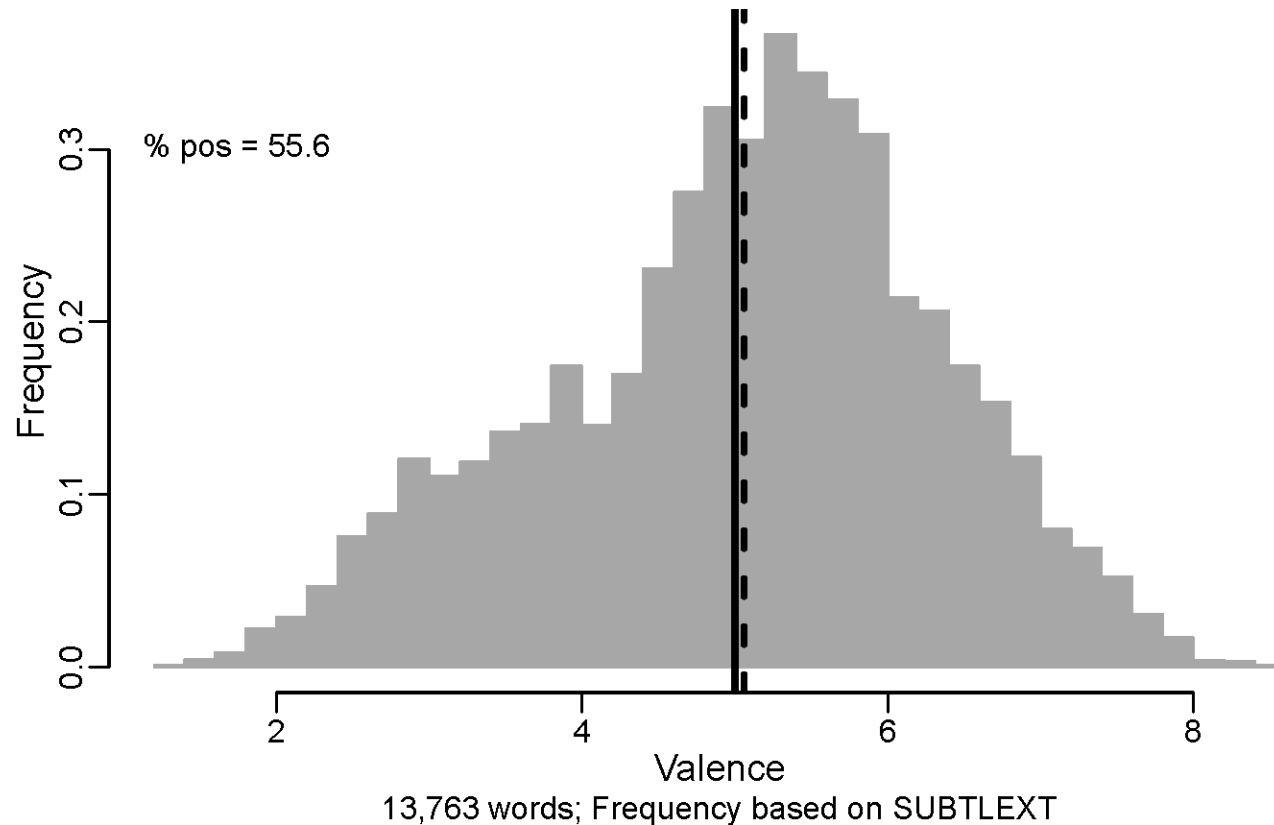**RAINBOW**

Low Valence, High Arousal
GUNPOINT

Low Valence, Low Arousal
DUST

McMaster University

OUR GOALS

1. To re-examine the positivity bias (valence) with respect to both type and token frequency

    - With a large set of carefully collected, restricted, and valid emotional ratings

    - Across a variety of corpora

2. To perform the exact same analyses for arousal and compare them to valence

Types and Tokens in Warriner et al, 2013

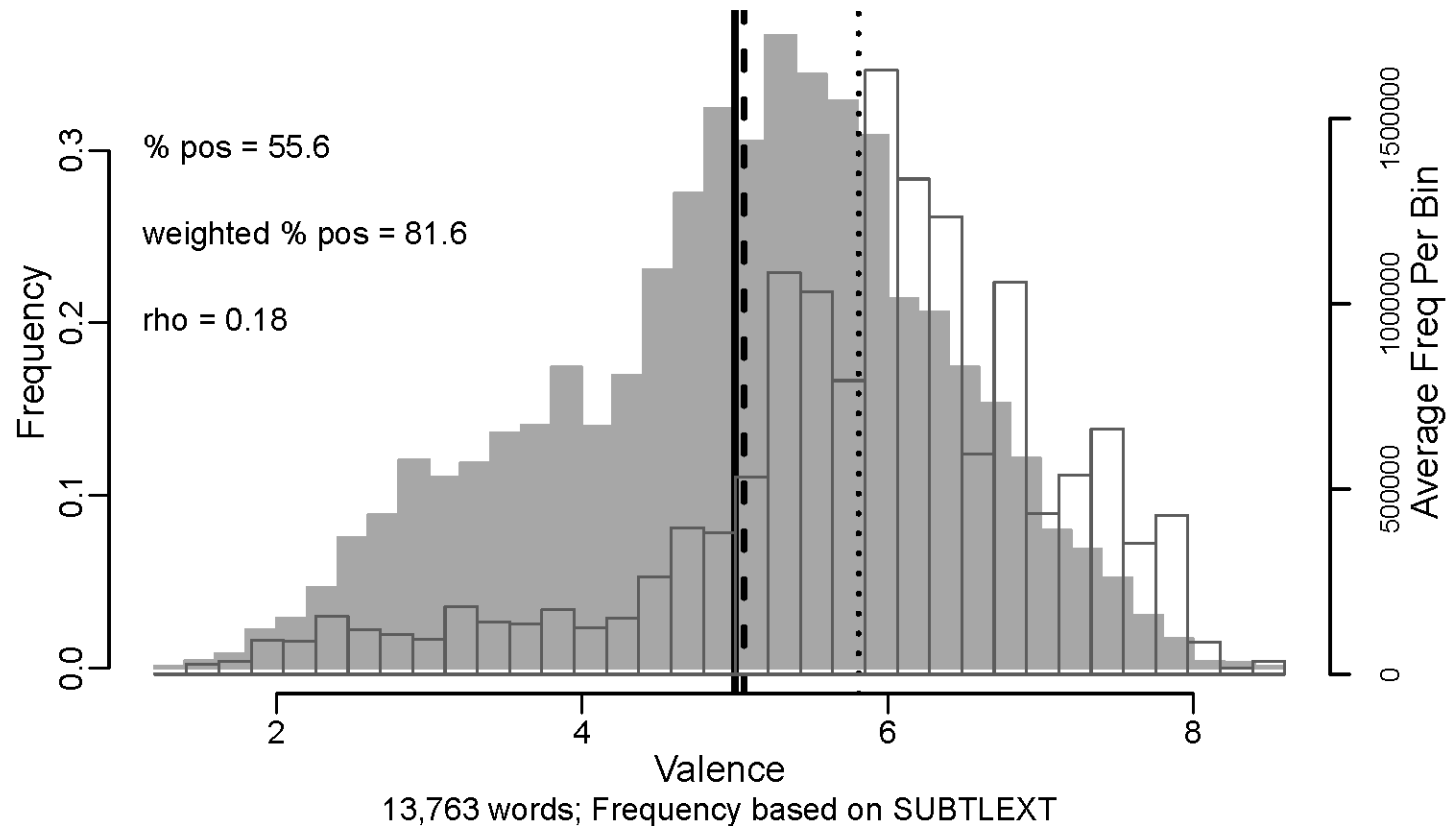# Types and Tokens in Warriner et al, 2013

13,763 words; Frequency based on SUBTLEXT

Scale Midpoint ——— Unweighted Mean — — · Weighted Mean ···········

Types and Tokens in Warriner et al, 2013

% calm = 80.7

Probability

Arousal

13,763 words; Frequency based on SUBTLEXT

Scale Midpoint ——— **Unweighted Mean** — — · **Weighted Mean** ·········

Types and Tokens in Warriner et al, 2013

% calm = 80.7

weighted % calm = 83.4

rho = 0.039

13,763 words; Frequency based on SUBTLEXT
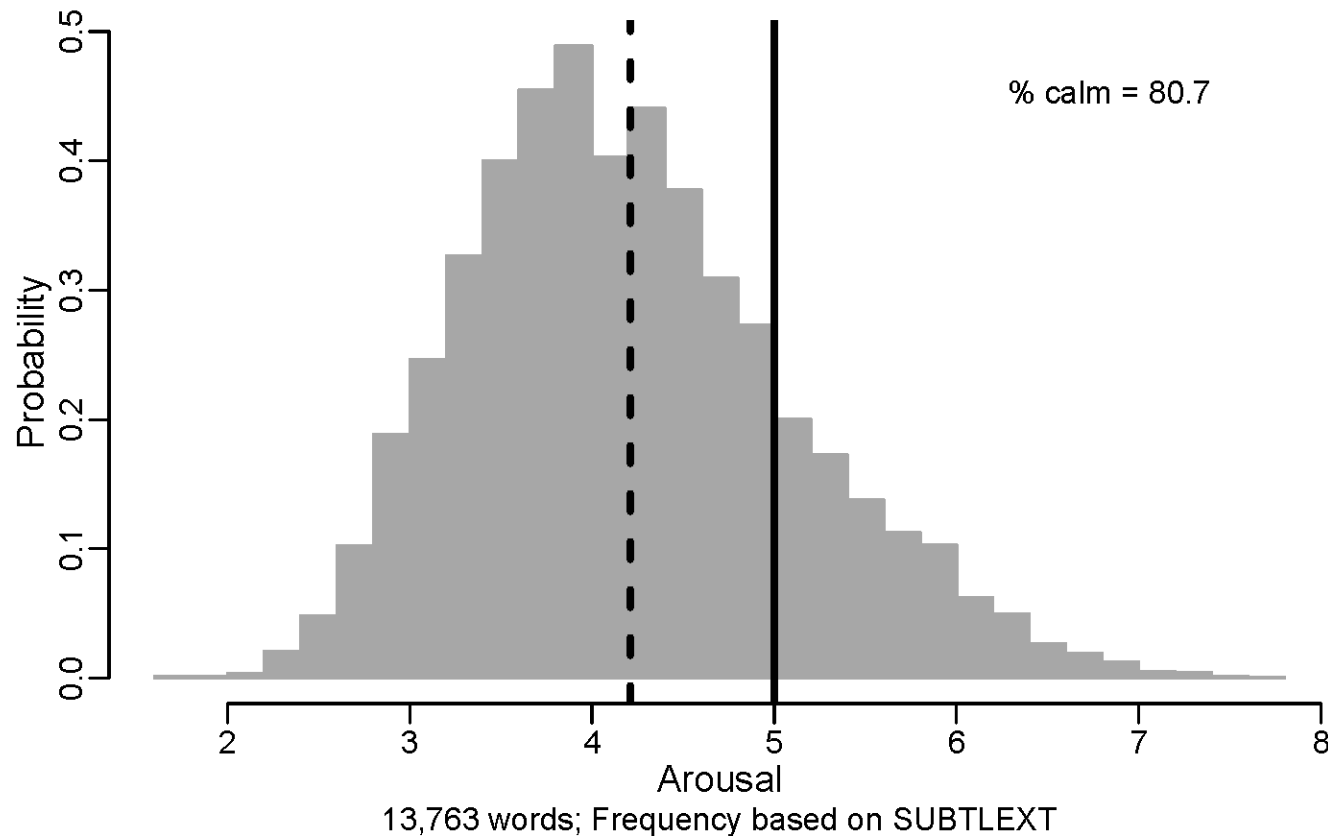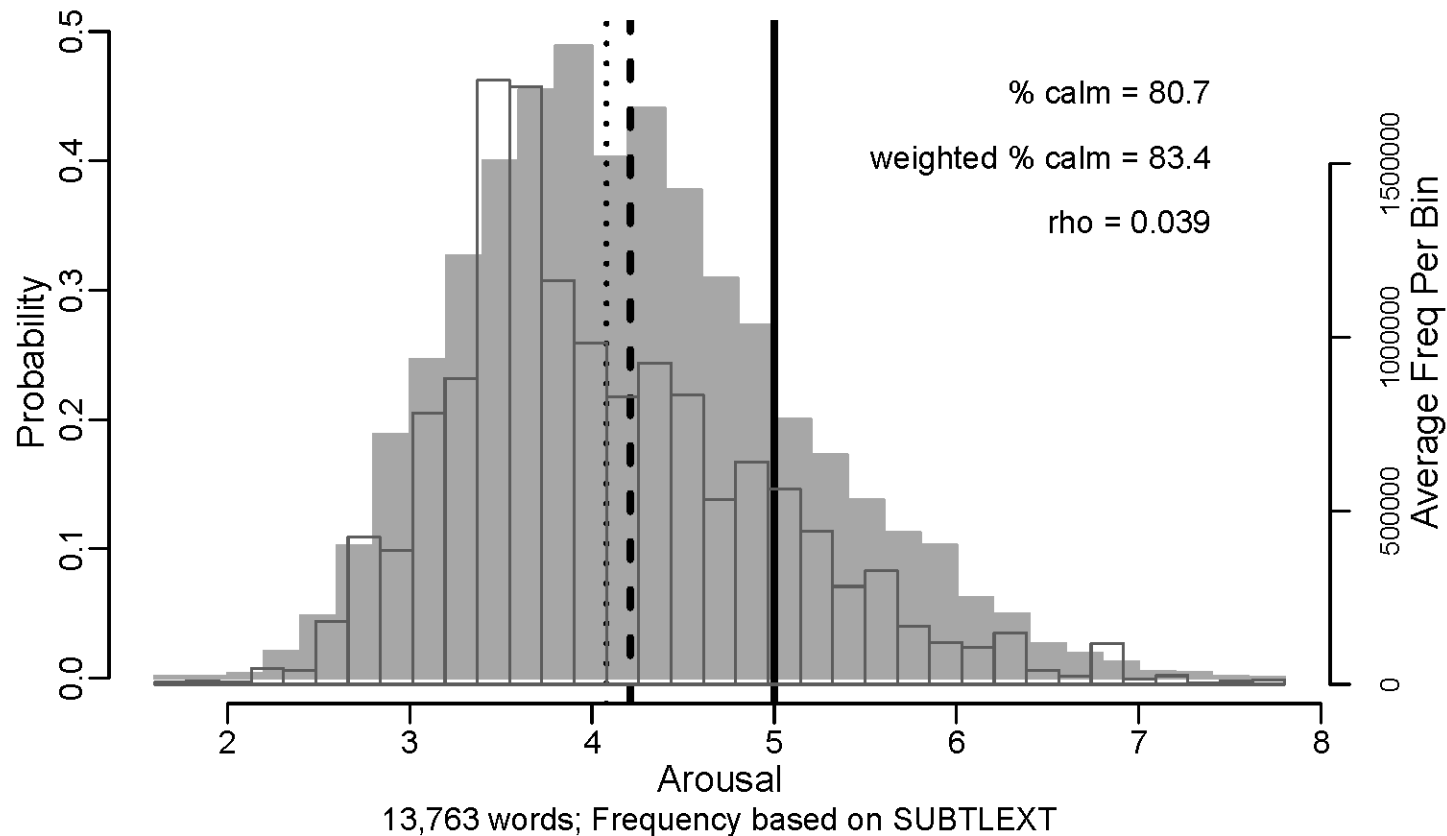
Scale Midpoint ——— Unweighted Mean — — · Weighted Mean ··········

Testing Other Corpora

SOURCE	% pos	% calm	V ρ	A ρ	# words
TASA	56.4	81.0	0.233	-0.102	12,344
SUBTLEX	55.6	80.7	0.180	0.039	13,763
BNC	62.9	82.9	0.224	-0.035	4,812
COCA	64.0	80.6	0.185	-0.031	6,842
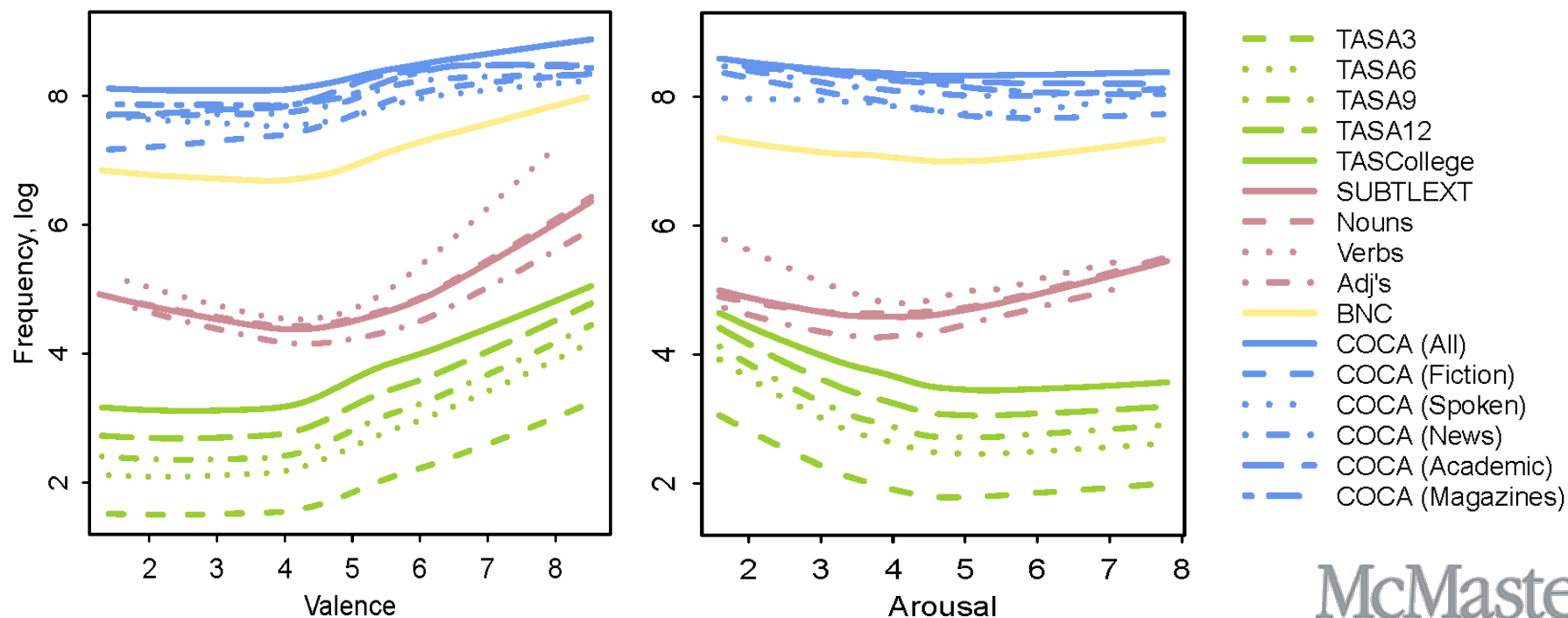
Re-analysis of Kloumann et al's Data
(our ratings; only overlapping words)

	Twitter	Google Books	New York Times	Music Lyrics
% pos	73.3	72.9	74.3	66.7
V ρ	0.176	0.128	0.066	0.149
% calm	77.0	84.4	82.6	77.4
A ρ	-0.019	-0.053	-0.054	-0.009
# words	2,443	2,704	2,354	2,458

Therefore our research confirms…

a POSITIVITY BIAS present in TOKENS (there is **nearly a balance** in the number of positive and negative types, but **we talk more** about positive ones)

VALENCE

a CALMNESS BIAS present in TYPES (there are **many more** calm than arousing types, and we **speak equally frequently** about both)

AROUSAL

McMaster University

Future Questions

- What is the direction of causation between real world phenomena and these biases? (i.e. social cohesion, risk aversion)

- Do these biases parallel semantic density in a way that explains behavioral measures?

- Are there gender, group, or individual differences in the presentation of these biases?

Thank you.
Any questions?