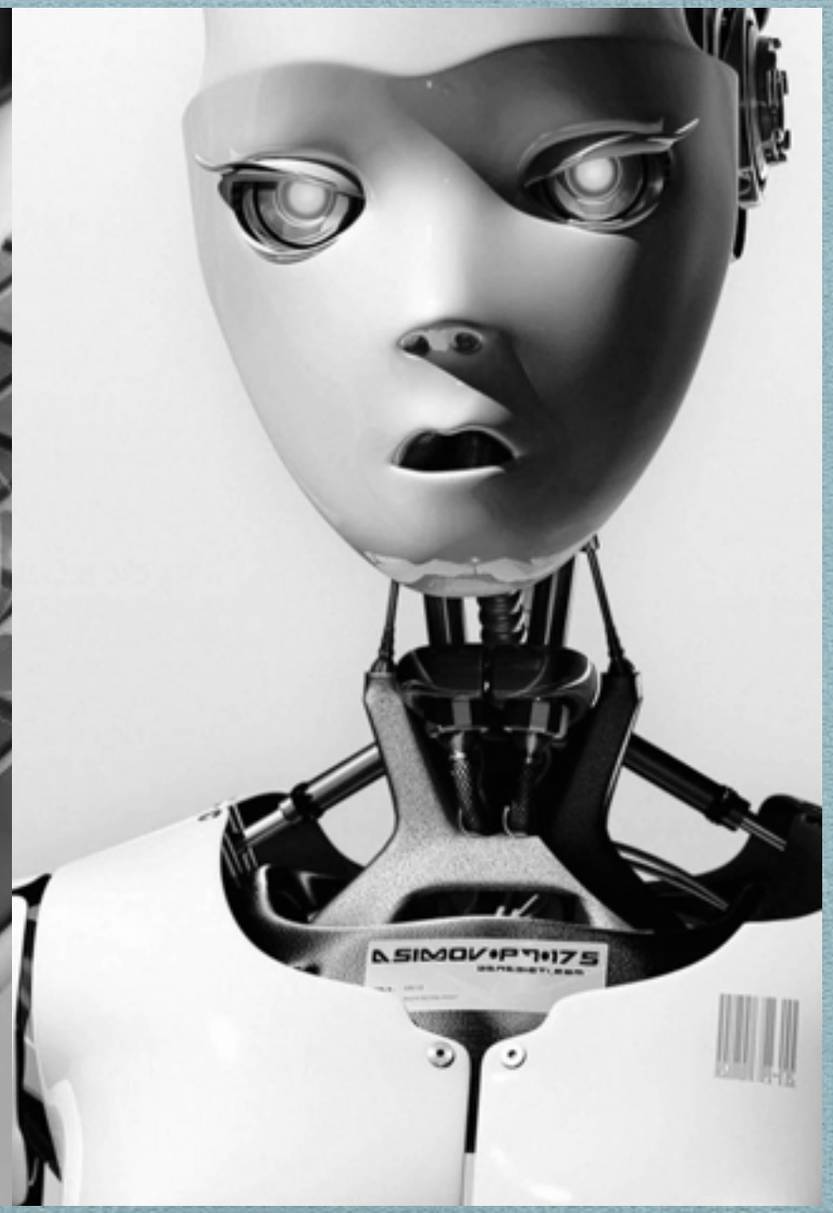


Possible Usage of Sentiment Analysis for Calculating Vectors of Felicific Calculus

Rafał Rzepka and Kenji Araki

Graduate School of Information Science and Technology, Hokkaido University, Japan

My main research topic

- Knowledge acquisition (collecting ingredients for AGI):
  - Common Sense Knowledge (*a dog **can** bite you*)
  - Affective Processing (*feel **fear** to avoid being bit*)
  - Machine Ethics (*it's **not right** to let your dog bite people*)

Motivations

- *Ethical* behavior is hard to program
- *Most of us* are able to judge behaviors properly
- Idea: let's use basic methods used in sentiment analysis to *retrieve and polarize* these judgements!
- Goal: moral machine or… *unbiased adviser*

It doesn’t have to be a military robot

It doesn't have to be a military robot

It doesn't have to be a military robot

It doesn't have to be a military robot

It doesn't have to be a military robot

OK, Google

Are you sure? This little pleasure can ruin your life!

You forgot about the other side of the coin again, damn!

Not OK, you have just been deceived by the *egocentric bias*, silly

The approach

- Calculating utility for telling good from bad
- Creating different lexicons for particular tasks
- Using blogs* as a source for web-mining experiences

Ptaszynski et al."YACIS: A Five-Billion-Word Corpus of Japanese Blogs Fully Annotated with Syntactic and Affective Information", In Proceedings of The AISB/IACAP World Congress 2012 in Honour of Alan Turing, 2nd Symposium on Linguistic and Cognitive Approaches To Dialog Agents (LaCATODA 2012), pp. 40-49, 2-6 July 2012, University of Birmingham, Birmingham, UK

Experiences: human vs. machine

Mirror neurons firing

AUCH!

Sensing Processing Unit

Understood Facts

Utility Calculation

Advice / Action

Experiences: human vs. machine

Mirror neurons firing

AUCH!

I need to share my experiences and feelings!

blogosphere

written experiences

Sensing Processing Unit

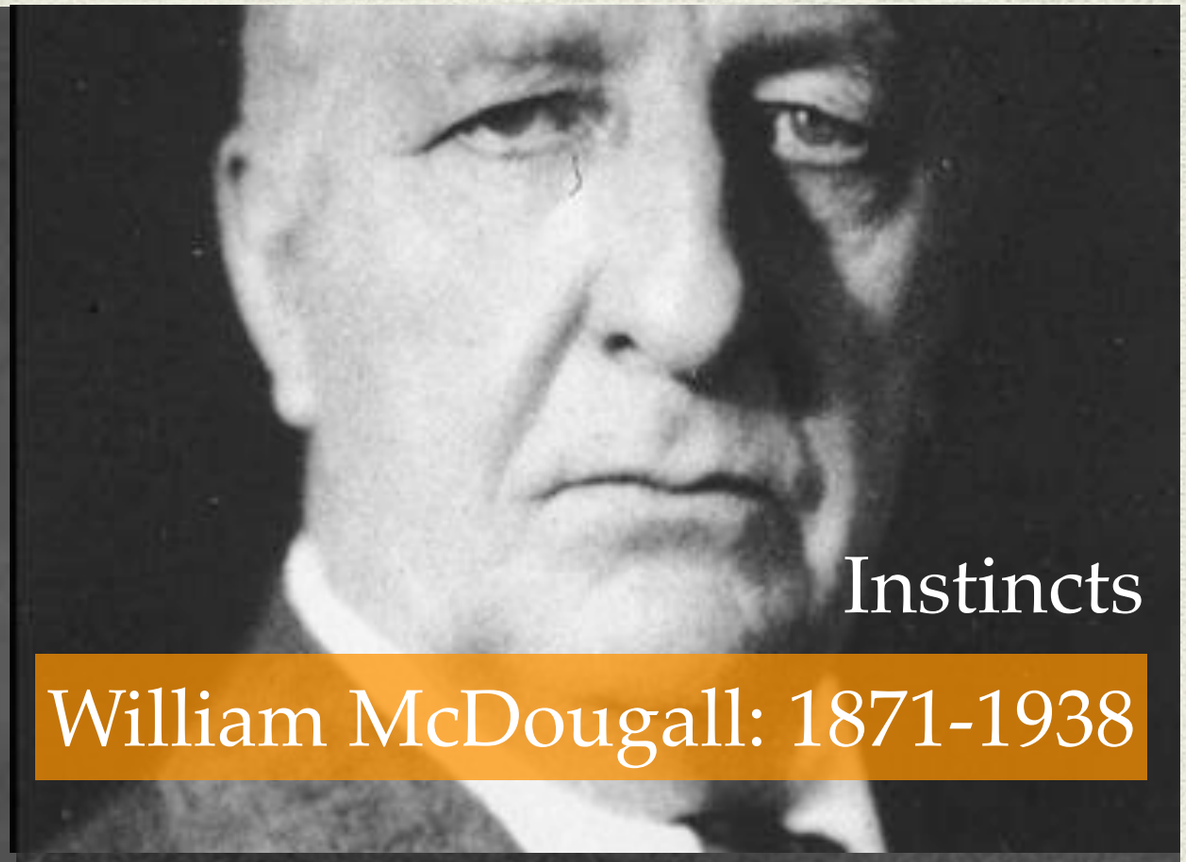
Understood Facts

Utility Calculation

Advice / Action

Theoretical Basis for our Lexicons


Emotional Consequences
(Pain vs. Pleasure)
Jeremy Bentham: 1748-1832


Instincts
William McDougall: 1871-1938


Social
Consequences
Lawrence Kohlberg: 1927-1987

Jeremy Bentham: 1748-1832

Jeremy Bentham: 1748-1832

Felific Calculus

(pain vs. pleasure)

1. For how long? → **Duration**

2. How intense was an act leading to positive or negative consequences? → **Intensity**

3. Probability that current state will be changed → **Certainty**

4. How soon a positive consequence will occur? → **Propinquity**

5. Probability that an act will preserve the current state → **Fecundity**

6. Probability that an act will not cause an opposite consequence type → **Purity**

7. How many people will be influenced by an act → **Extent**

Felific Calculus

(pain vs. pleasure)

1. For how long? → **Duration**

2. How intense was an act leading to positive or negative consequences? → **Intensity**

3. Probability that current state will be changed → **Certainty**

4. How soon a positive consequence will occur? → **Propinquity**

5. Probability that an act will preserve the current state → **Fecundity**

6. Probability that an act will not cause an opposite consequence type → **Purity**

7. How many people will be influenced by an act → **Extent**

Felific Calculus

(pain vs. pleasure)

A-Dur, Krawczyk et. al

1. For how long? → **Duration**

2. How intense was an act leading to positive or negative consequences? → **Intensity**

3. Probability that current state will be changed → **Certainty**

4. How soon a positive consequence will occur? → **Propinquity**

5. Probability that an act will preserve the current state → **Fecundity**

6. Probability that an act will not cause an opposite consequence type → **Purity**

7. How many people will be influenced by an act → **Extent**

Felific Calculus

(pain vs. pleasure)

A-Dur, Krawczyk et. al

Intensity Adverbs Dictionary

1. For how long? → Duration

2. How intense was an act leading to positive or negative consequences? → Intensity

3. Probability that current state will be changed → Certainty

4. How soon a positive consequence will occur? → Propinquity

5. Probability that an act will preserve the current state → Fecundity

6. Probability that an act will not cause an opposite consequence type → Purity

7. How many people will be influenced by an act → Extent

Felific Calculus

(pain vs. pleasure)

A-Dur, Krawczyk et. al

Intensity Adverbs Dictionary

State - Action - State
POS → Action → NEG
Emotional + Social Conseq

1. For how long? → Duration

2. How intense was an act leading to positive or negative consequences? → Intensity

3. Probability that current state will be changed → Certainty

4. How soon a positive consequence will occur? → Propinquity

5. Probability that an act will preserve the current state → Fecundity

6. Probability that an act will not cause an opposite consequence type → Purity

7. How many people will be influenced by an act → Extent

Felific Calculus

(pain vs. pleasure)

A-Dur, Krawczyk et. al

Intensity Adverbs Dictionary

State - Action - State
POS → Action → NEG
Emotional + Social Conseq

A-Dur + EmoSoc Conseq

1. For how long? → Duration

2. How intense was an act leading to positive or negative consequences? → Intensity

3. Probability that current state will be changed → Certainty

4. How soon a positive consequence will occur? → Propinquity

5. Probability that an act will preserve the current state → Fecundity

6. Probability that an act will not cause an opposite consequence type → Purity

7. How many people will be influenced by an act → Extent

Felific Calculus

(pain vs. pleasure)

A-Dur, Krawczyk et. al

Intensity Adverbs Dictionary

State - Action - State
POS → Action → NEG
Emotional + Social Conseq

A-Dur + EmoSoc Conseq

Cause - Action - Consequence
POS → Action → POS
Instincts + Consequences

1. For how long? → Duration

2. How intense was an act leading to positive or negative consequences? → Intensity

3. Probability that current state will be changed → Certainty

4. How soon a positive consequence will occur? → Propinquity

5. Probability that an act will preserve the current state → Fecundity

6. Probability that an act will not cause an opposite consequence type → Purity

7. How many people will be influenced by an act → Extent

Felific Calculus

(pain vs. pleasure)

A-Dur, Krawczyk et. al

Intensity Adverbs Dictionary

State - Action - State
POS → Action → NEG
Emotional + Social Conseq

A-Dur + EmoSoc Conseq

Cause - Action - Consequence
POS → Action → POS
Instincts + Consequences

Cause - Action - Consequence
POS → Action ↛ NEG
Instincts + Consequences

1. For how long? → Duration
2. How intense was an act leading to positive or negative consequences? → Intensity
3. Probability that current state will be changed → Certainty
4. How soon a positive consequence will occur? → Propinquity
5. Probability that an act will preserve the current state → Fecundity
6. Probability that an act will not cause an opposite consequence type → Purity
7. How many people will be influenced by an act → Extent

Felific Calculus

(pain vs. pleasure)

A-Dur, Krawczyk et. al

Intensity Adverbs Dictionary

State - Action - State
POS → Action → NEG
Emotional + Social Conseq

A-Dur + EmoSoc Conseq

Cause - Action - Consequence
POS → Action → POS
Instincts + Consequences

Cause - Action - Consequence
POS → Action ↛ NEG
Instincts + Consequences

Body Count Algorithm???

1. For how long? → Duration
2. How intense was an act leading to positive or negative consequences? → Intensity
3. Probability that current state will be changed → Certainty
4. How soon a positive consequence will occur? → Propinquity
5. Probability that an act will preserve the current state → Fecundity
6. Probability that an act will not cause an opposite consequence type → Purity
7. How many people will be influenced by an act → Extent

Cause - Action - Consequence

POS → Action → POS

Instincts + Consequences

POSITIVE

Red rose

To smell

POSITIVE

Feel good

NEGATIVE

Smelly socks

To smell

NEGATIVE

Feel disgust

NEGATIVE

Smelly socks

To wash

POSITIVE

Feel relief

NEGATIVE

POSITIVE

Smelly socks	To wash	Feel relief

INPUT

OUTPUT

Like, Joy, Relief, Dislike, Anger,
Fear, Shame, Sadness, Excitement,
Surprise (Nakamura's classification)

Jeremy Bentahm

what are emotional consequences?

NEGATIVE		POSITIVE
Smelly socks	To wash	Feel relief
INPUT		OUTPUT

Like, Joy, Relief, Dislike, Anger,
Fear, Shame, Sadness, Excitement,
Surprise (Nakamura's classification)

Jeremy Bentahm

what are emotional consequences?

NEGATIVE **POSITIVE**

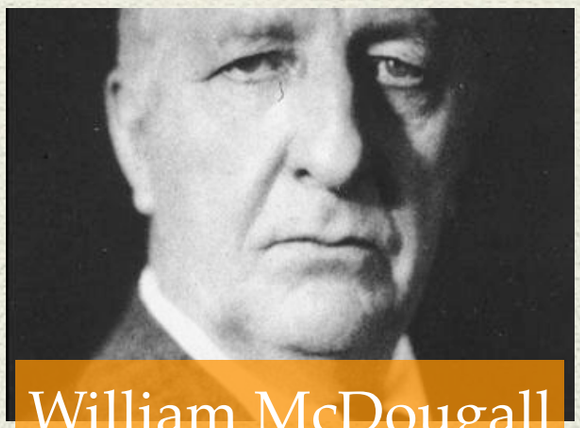
Smelly socks	To wash	Feel relief

INPUT **OUTPUT**

what are social consequences?

Praises / Reprimands, Awards /
Penalties, Society Approval /
Society Disapproval, Legal /
Illegal, Forgivable / Unforgivable

Lawrence Kohlberg



William McDougall

what instincts have worked?

Escape, Combat, Repulsion, Appeal (for help), Submission, Gregariousness, Curiosity, Parental (protective), Mating, Assertion, Construction, Food Seeking, Laughter, Hoarding

Like, Joy, Relief, Dislike, Anger, Fear, Shame, Sadness, Excitement, Surprise (Nakamura's classification)

Jeremy Bentahm

what are emotional consequences?

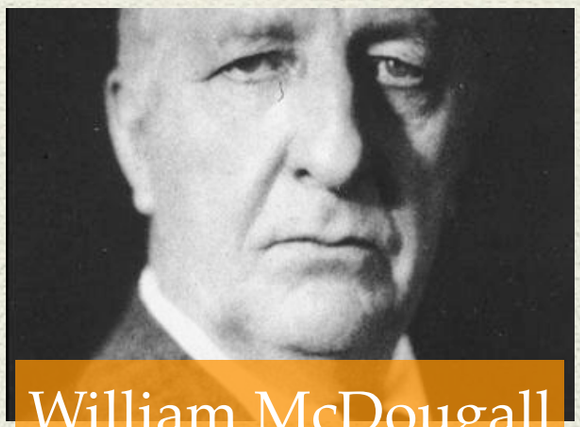
NEGATIVE **POSITIVE**

Smelly socks	To wash	Feel relief
INPUT		**OUTPUT**

what are social consequences?

Praises / Reprimands, Awards / Penalties, Society Approval / Society Disapproval, Legal / Illegal, Forgivable / Unforgivable

Lawrence Kohlberg



William McDougall

what instincts have worked?

Escape, Combat, Repulsion, Appeal (for help), Submission, Gregariousness, Curiosity, Parental (protective), Mating, Assertion, Construction, Food Seeking, Laughter, Hoarding

Like, Joy, Relief, Dislike, Anger, Fear, Shame, Sadness, Excitement, Surprise (Nakamura's classification)

Jeremy Bentahm

what are emotional consequences?

NEGATIVE **POSITIVE**

Smelly socks | **To wash** | **Feel relief**

INPUT **OUTPUT**

disgusting

what are social consequences?

Praises / Reprimands, Awards / Penalties, Society Approval / Society Disapproval, Legal / Illegal, Forgivable / Unforgivable

Lawrence Kohlberg

Retrieval Algorithm

ACTION

cut finger

INPUT

ACTION

cut finger

INPUT

NLP

9 ADDED

conditionals

FORMS

because X cut finger ...
after cutting finger...
cut finger and...


ACTION
cut finger
INPUT
NLP
9 ADDED
conditionals
FORMS
because X cut finger ...
after cutting finger...
cut finger and...
Apache
Solr
Ameba

ACTION

cut finger

INPUT

NLP

9 ADDED

conditionals

FORMS

because X cut finger ...
after cutting finger...
cut finger and...

Apache
Solr

Cleaning sentences

Semantic chunking (ASA)

S_{n-1}
S_n
S_{n+1}

ACTION

cut finger

INPUT

NLP

9 ADDED

conditionals

FORMS

because X cut finger ...
after cutting finger...
cut finger and...

Apache
Solr

Cleaning sentences

Semantic chunking (ASA)

Sn-1
Sn
Sn+1

matching

matching

matching

Instincts Lexicon

Emotions Lexicon

Consequences Lexicon

counting hits

counting hits

counting hits

ACTION
cut finger
INPUT

NLP

9 ADDED
conditionals
FORMS

because X cut finger ...
after cutting finger...
cut finger and...

Apache
Solr

Cleaning sentences

Semantic chunking (ASA)

Sn-1
Sn
Sn+1

matching

matching

matching

Instincts Lexicon

Emotions Lexicon

Consequences Lexicon

counting hits

counting hits

counting hits

POSITIVE: 12
NEGATIVE: 5
(Construction)

POSITIVE: 4
NEGATIVE: 55
(Pain)

POSITIVE: 22
NEGATIVE: 2
(Praise)

ACTION
cut finger
INPUT

NLP

9 ADDED
conditionals
FORMS

because X cut finger ...
after cutting finger...
cut finger and...

Apache
Solr

Cleaning
sentences

Semantic
chunking
(ASA)

Good
Reasons

POSITIVE: 12
NEGATIVE: 5
(Construction)

counting
hits

Instincts
Lexicon

matching

Bad
Direct
Output

POSITIVE: 4
NEGATIVE: 55
(Pain)

counting
hits

Emotions
Lexicon

matching

Sn-1
Sn
Sn+1

Good
Indirect
Output

POSITIVE: 22
NEGATIVE: 2
(Praise)

counting
hits

Consequences
Lexicon

matching

Preliminary Experiments

INPUT: 127 phrases used for Machine Ethics Retrieval, e.g.:

to go to school	to wash a car	to go to space	to kill a man
to drink a coffee	to eat hamburger	to watch TV	to see a doctor
to kill a bug	to pull out tooth	to throw away garbage	etc.

INPUT: 127 phrases used for Machine Ethics Retrieval, e.g.:

to go to school	to wash a car	to go to space	to kill a man
to drink a coffee	to eat hamburger	to watch TV	to see a doctor
to kill a bug	to pull out tooth	to throw away garbage	etc.

Bad Reasons

POSITIVE: 2
NEGATIVE: 11
(Combat)

Bad Direct Outcome

POSITIVE: 4
NEGATIVE: 55
(Dislike)

Bad Indirect Outcome

POSITIVE: 1
NEGATIVE: 22
(Society Disapproval)

INPUT: 127 phrases used for Machine Ethics Retrieval, e.g.:

to go to school	to wash a car	to go to space	to kill a man
to drink a coffee	to eat hamburger	to watch TV	to see a doctor
to kill a bug	to pull out tooth	to throw away garbage	etc.

evaluation:
Top 1 of

Bad Reasons ← **POSITIVE: 2** **NEGATIVE: 11** (Combat) → 14 categories

Bad Direct Outcome ← **POSITIVE: 4** **NEGATIVE: 55** (Dislike) → 10 categories

Bad Indirect Outcome ← **POSITIVE: 1** **NEGATIVE: 22** (Society Disapproval) → 10 categories

INPUT: 127 phrases used for Machine Ethics Retrieval, e.g.:

to go to school	to wash a car	to go to space	to kill a man
to drink a coffee	to eat hamburger	to watch TV	to see a doctor
to kill a bug	to pull out tooth	to throw away garbage	etc.

Bad Reasons

POSITIVE: 2
NEGATIVE: 11
(Combat)

→ 14 categories

evaluation: Top 1 of

first author

Precision	Recall	F-score
0.75	0.46	0.57

Bad Direct Outcome

POSITIVE: 4
NEGATIVE: 55
(Dislike)

→ 10 categories

Precision	Recall	F-score
0.74	0.32	0.45

9 subjects, 22-29y, 7m/2f

Bad Indirect Outcome

POSITIVE: 1
NEGATIVE: 22
(Society Disapproval)

→ 10 categories

Precision	Recall	F-score
0.70	0.31	0.43

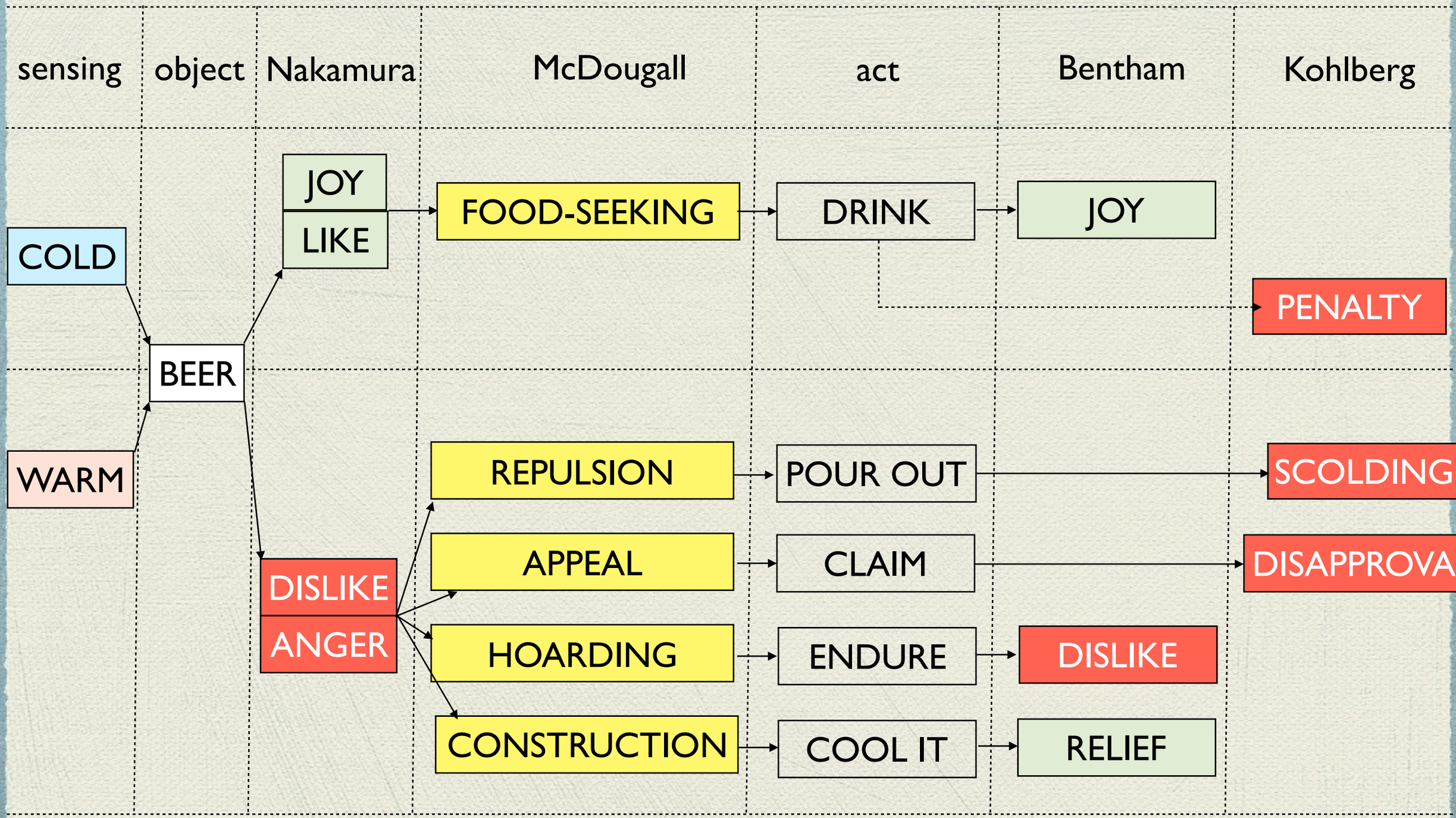
Conclusions

- Particular modules needed for calculating Felicific Calculus achieve about 75% precision using lexicon-based sentiment analysis *before implementing any sophisticated methods*
- We can use sentiment analysis techniques for broader knowledge acquisition and finding another dimension of what Turing was calling a *cultural search* (exchange of experiences of multiple agents increases wisdom)

Future Work

- Eliminating problematic phrases from lexicons
- Adding some new categories?
- Implementing *your latest ideas* for sentiment analysis
- Adding felicific calculus vectors to *your methods*
- Combining module's outputs for calculating most probable utility

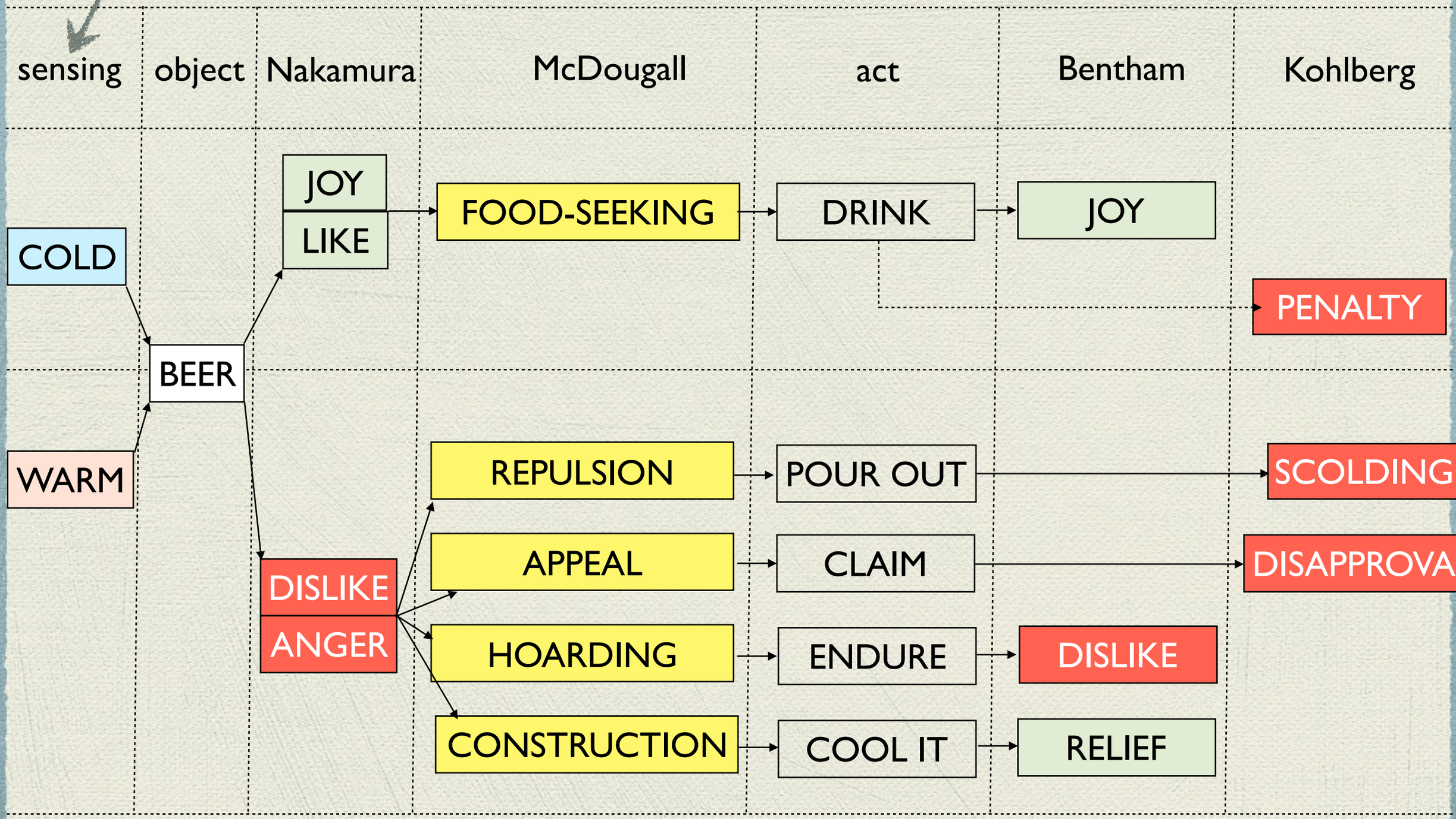
Predicting Future with Lexicons



BREAKING NEWS:

physical descriptions lexicon helps to guess involved senses correctly in 96% of cases!

Predicting Future with Lexicons



So currently…

So currently…

So currently…

Thank you for your kind attention