Emotional Artificial Intelligence in Socio-Technical Systems

Jesse Hoey
David R. Cheriton School of Computer Science

October 10, 2017
National Academies, Washington, DC
• 6 million active users
• 20 million active repositories
• 10 million active issues
• 331 thousand active organizations
Fig2. - Negative behavior in open source
Source: opensourcesurvey.org

- **Rudeness**
  - Witnessed: 45%
  - Experienced: 25%
  - Error: 5%

- **Name calling**
  - Witnessed: 30%
  - Experienced: 15%
  - Error: 5%

- **Serious incidents**
  - Witnessed: 15%
  - Experienced: 10%
  - Error: 5%

- **Stereotyping**
  - Witnessed: 10%
  - Experienced: 5%
  - Error: 5%
Fig1. - Problems encountered in open source

Source: opensourcesurvey.org

- Incomplete or confusing documentation
- Unresponsiveness
- Dismissive responses
- Conflict
- Unexplained rejection
- Unwelcoming language or content
Overarching Goals:

- **Understand** group behaviour with models that are:
  - computational
  - social
  - cultural
  - emotional

- **Build** artificial agents to help teams be:
  - engaging
  - inclusive
  - effective
Osgood’s Semantic Differential

Fundamental Sentiments

Evaluation

Power

good

Evaluation

strong

weak

bad

hyper

Activity

asleep
Affect Control Theory

fundamental sentiments

social action

transient impressions

- Shared sentiments
- Shared *emotional dynamics*
- Shared *consistency* → Cooperation
THEMIS.COG: Social Programming Networks

Tobias Schröder
Potsdam

Kimberly B. Rogers
Dartmouth

Mei Nagappan
Waterloo
Support

Funding:
- American Alzheimer’s Assoc.
- Research Institute for Aging
- AGEWELL Canadian NCE
- Trans-Atlantic Partnership (TAP)
- Cheriton Faculty Fellowship
- NSERC
- SSHRC
- CIHR
- MITACS
- CCNA

More Information:
- Bayesian Affect Control Theory: bayesact.ca
- Jesse Hoey: jhoey@cs.uwaterloo.ca
Affect Control Theory

- Actor-Behaviour-Object
- fundamental sentiments: \( F \in [-4.3, 4.3]^9 \)
- transient impressions: \( T \in [-4.3, 4.3]^9 \)
- prediction \( T_{t+1} = MG(F_t, T_t) \) measured empirically
- deflection \( D = \sum_i w_i (f_i - \tau_i)^2 \)
- **Affect Control Principle**: actors work to experience transient impressions that are consistent with their fundamental sentiments
- Emotion \( \epsilon \propto f - \tau \)

Bayesian Affect Control Theory

- identities and behaviours as probability distributions
- external goals and planning
Affect Control Theory

Deflection:

nurse

\[ F: [2.9, 1.5, 0.2] \]

patient

\[ [1.1, -0.8, -0.9] \]
Affect Control Theory

Deflection: 1.3

**nurse**
- F: [2.9, 1.5, 0.2]
- T: [3.3, 1.2, 0.3]

**patient**
- [0.9, −0.7, −1.1]
- [1.4, −0.9, −0.7]

**comforts**
- [2.8, 2.1, 0.1]
- [2.3, 1.4, 0.4]
Affect Control Theory

Deflection: 13.4

\[
\text{nurse} \\
F: [2.9, 1.5, 0.2] \\
T: [-0.5, 0.9, 0.3]
\]

\[
\text{ignores} \\
[-1.9, -0.3, -0.9] \\
[-1.2, 0.4, -0.4]
\]

\[
\text{patient} \\
[0.9, -0.7, -1.1] \\
[0.4, -1.4, -0.8]
\]
GitHub annotations

Github pull request comments\(^1\) annotated for

- type of behaviour portrayed in the sentence: 12 IPA categories\(^2\)
  - Shows Solidarity
  - Shows tension release
  - Agrees
  - Gives Suggestion
  - Gives opinion
  - Gives orientation
  - Asks for orientation
  - Asks for opinion
  - Asks for suggestion
  - Disagrees
  - Shows Tension
  - Shows Antagonism

- Emotion displayed by the person writing the comment
  - Thanks
  - Sorry
  - Calm
  - Nervous
  - Careless
  - Cautious
  - Aggressive
  - Defensive
  - Happy
  - Angry

---

\(^1\) Georgios Gousios. The ghtorrent dataset and tool suite. MSR13
\(^2\) Bales 1950
### Examples of GitHub data

<table>
<thead>
<tr>
<th>Comment</th>
<th>IPA</th>
<th>Emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>*I’m responsible for this.,*Sorry.</td>
<td>shows tension</td>
<td>sorry, nervous</td>
</tr>
<tr>
<td>Limitation?, I would call that a show-stopper.</td>
<td>gives opinion, disagrees, shows antagism</td>
<td>aggressive, defensive</td>
</tr>
<tr>
<td>Great that you harmonized these, params to the python-layer style.</td>
<td>shows solidarity, agrees</td>
<td>thanks happy</td>
</tr>
<tr>
<td>Sorry., I started this file version by copying the Airy and forgot to edit this description.</td>
<td>gives orientation</td>
<td>sorry, careless</td>
</tr>
<tr>
<td>Wow, what a shame. Then I suggest we push this feature in to 2.1 instead of 2.0.2 since we can’t make it non-kludgy while retaining binary compat.</td>
<td>gives opinion, gives orientation</td>
<td>cautious, angry</td>
</tr>
</tbody>
</table>
Cheriton’s Three Laws of Automation

1. Everything that can be automated will be automated
2. Everything can be automated
3. Humans become (even) less competent at a task when it can be automated, making automation urgent.
Emotions: the new AI

- Artificial Intelligence: intelligence = rationality
- We now know that emotions are necessary for intelligence
- A low road gives “heuristic” social intelligence
- Encode a social order that allows us to work in a society

With infinite resources, are emotions necessary?