

Social and Emotional Artificial Intelligence

Jesse Hoey

David R. Cheriton School of Computer Science



December 13, 2018

Microsoft Research, December 14th, 2018



Artificial Intelligence



Artificial Intelligence

AI lacks:

- **empathy,**
- **altruism,**
- **culture, ... and**
- **emotion.**



Artificial Intelligence



AI lacks:

- **empathy,**
- **altruism,**
- **culture, ... and**
- **emotion.**



Computational Human Intelligence



theoatmeal.com/blog/google_self_driving_car

From A.I. To Affective Computing



von Neumann 1944



Simon 1967



Picard 1997

1997: Rosalind Picard in *Affective Computing*

This book proposes that we give computers the ability to recognize, express, and in some cases, "have" emotions. Is this not absurd?

Low Road and High Road

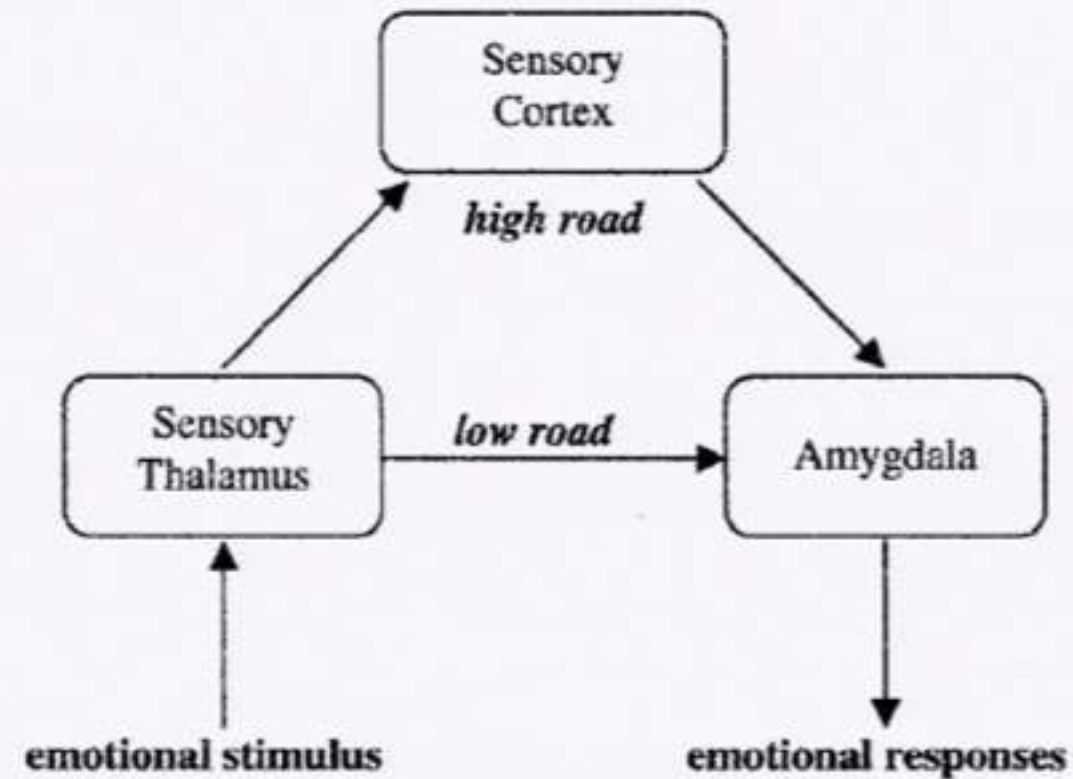
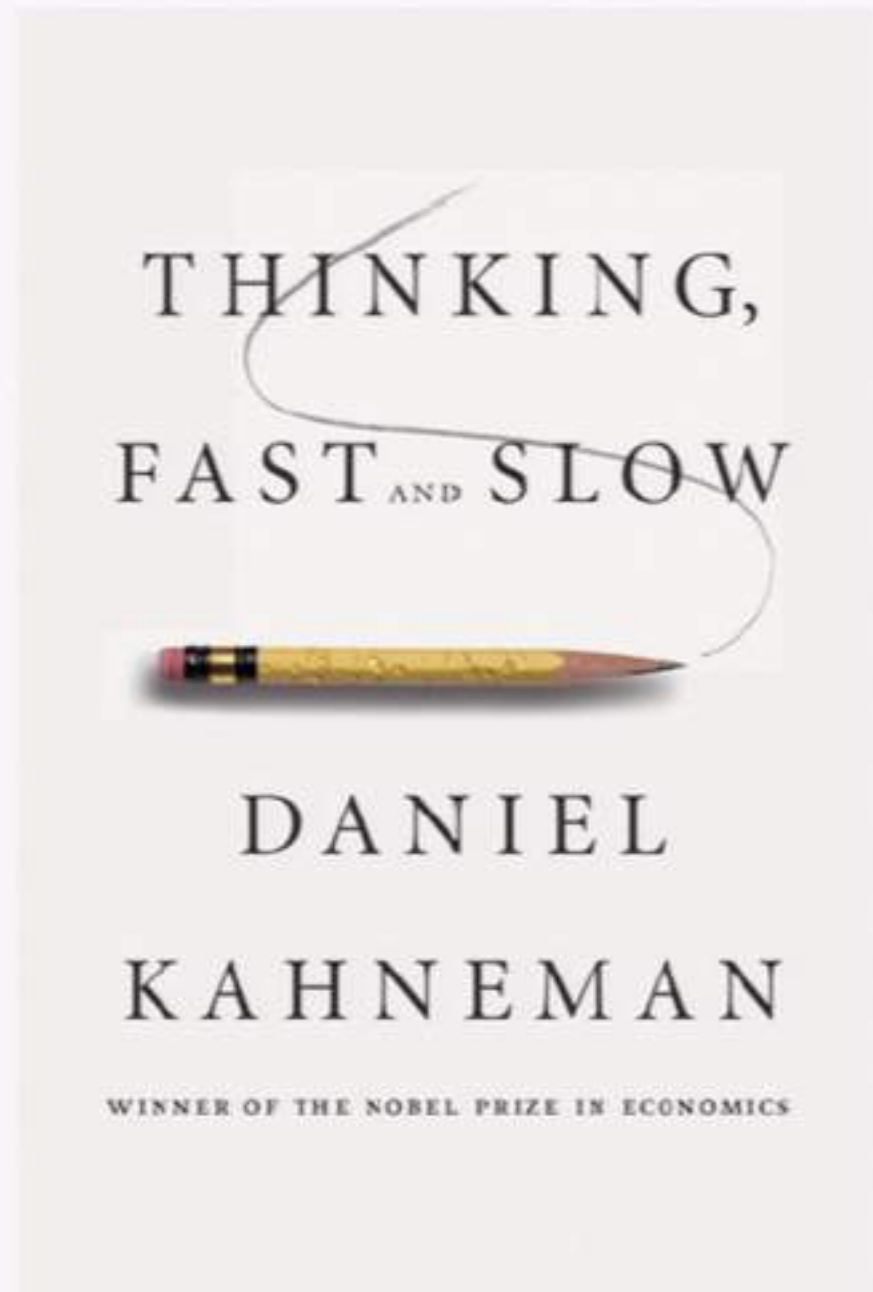


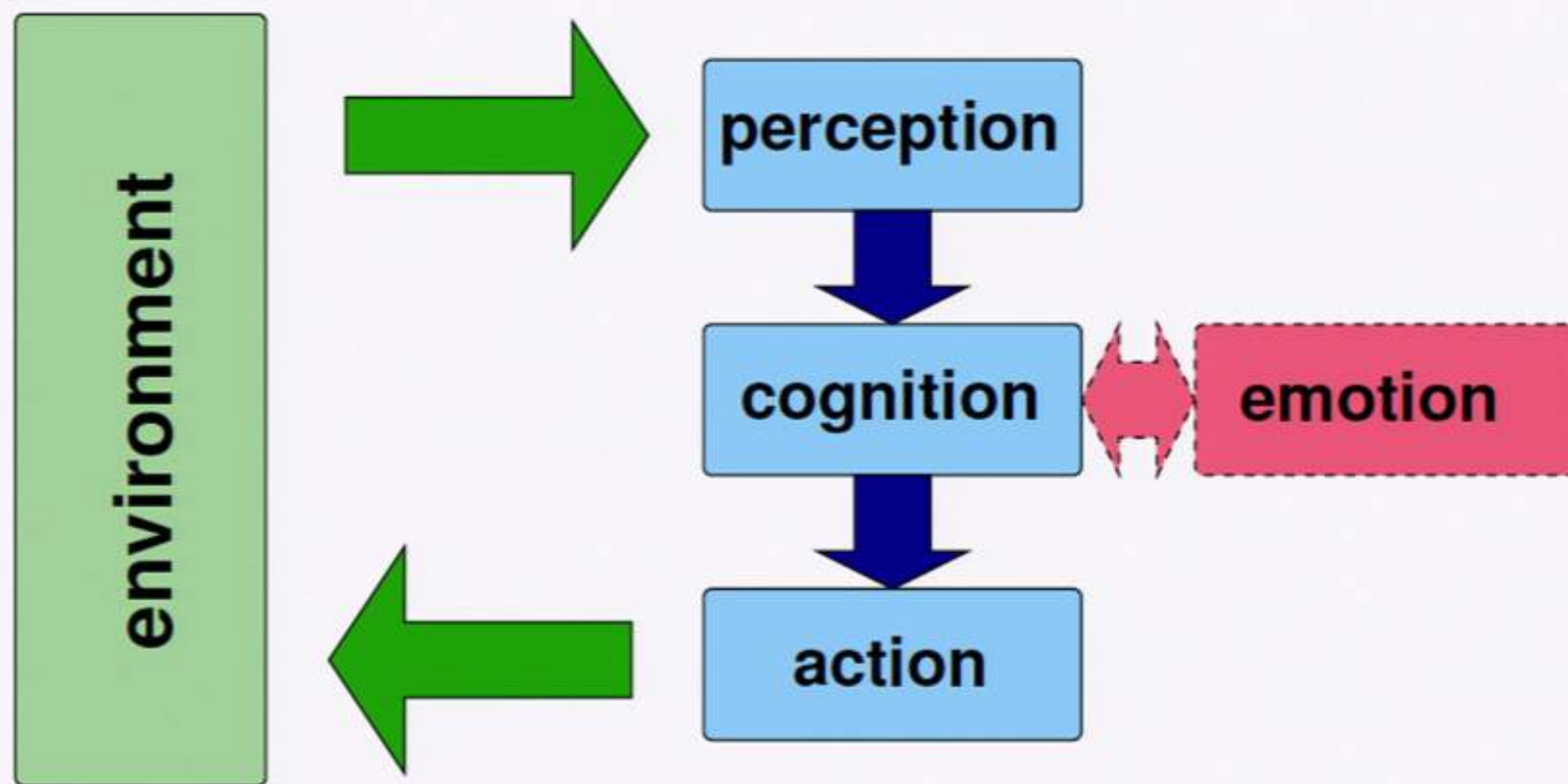
FIG. 3. Two separate pathways from sensory stimulus to emotional responses (adapted from LeDoux 1996, p. 164).

- From Zhu & Thagard "Emotion and Action".
Philosophical Psychology Vol 15 No 1, 2002.

Low Road and High Road

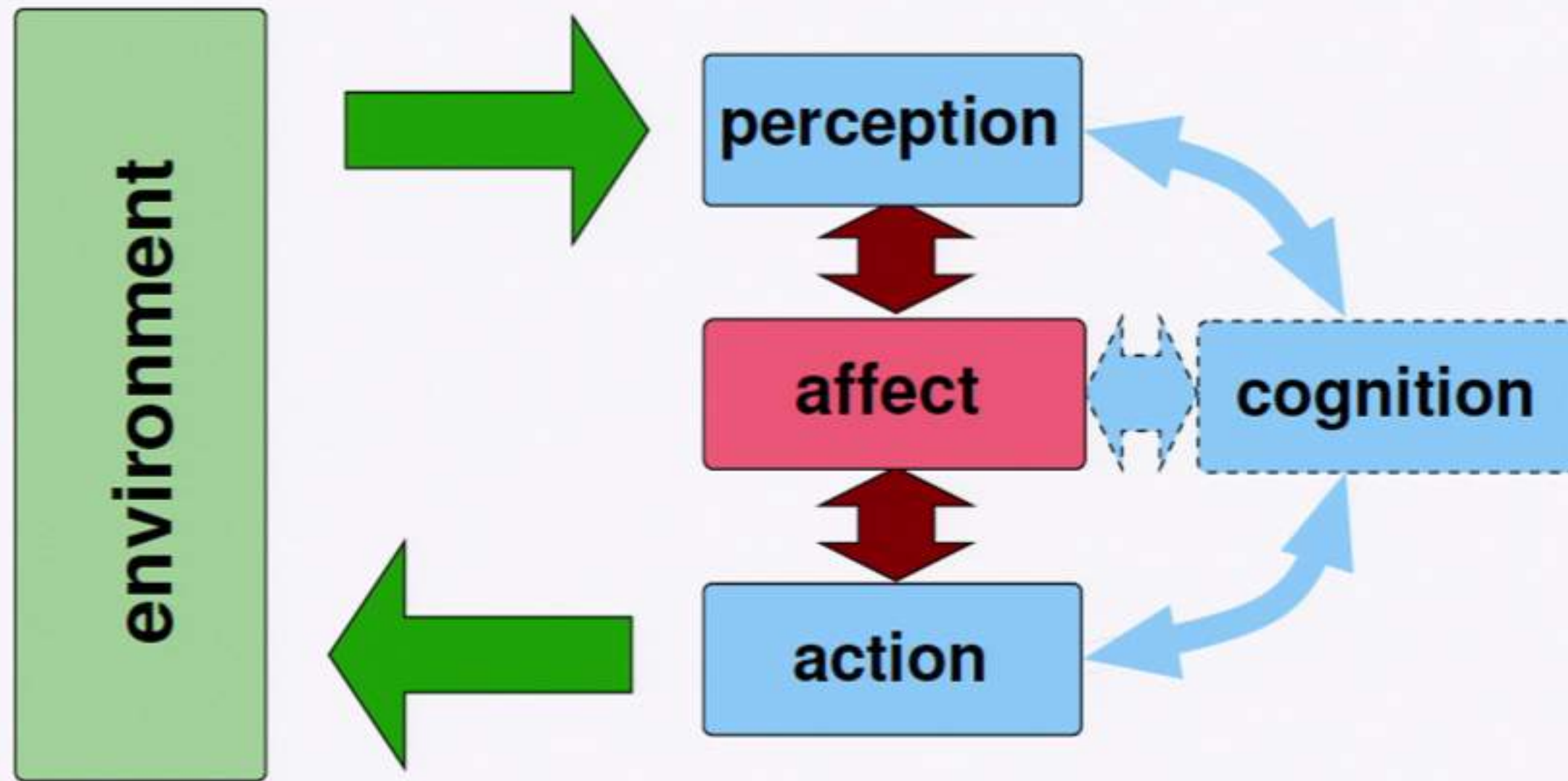


- **System I (Fast):** Operates automatically and quickly, without voluntary control
- **System II (Slow):** effortful mental activity, complex calculations
- When system I “runs into trouble”, it calls upon System II



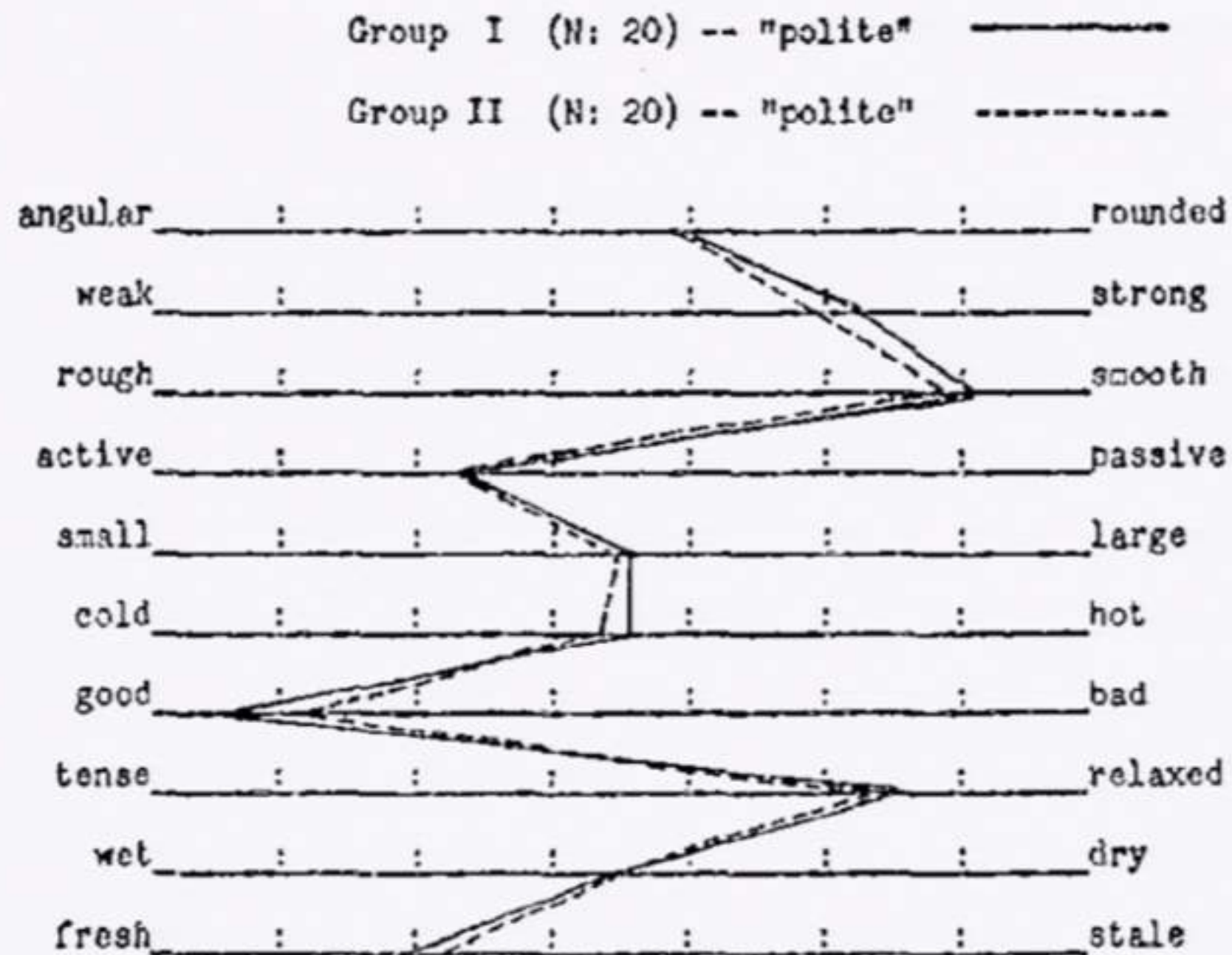
A cognitive machine disrupted by emotion

Affect Control Theory



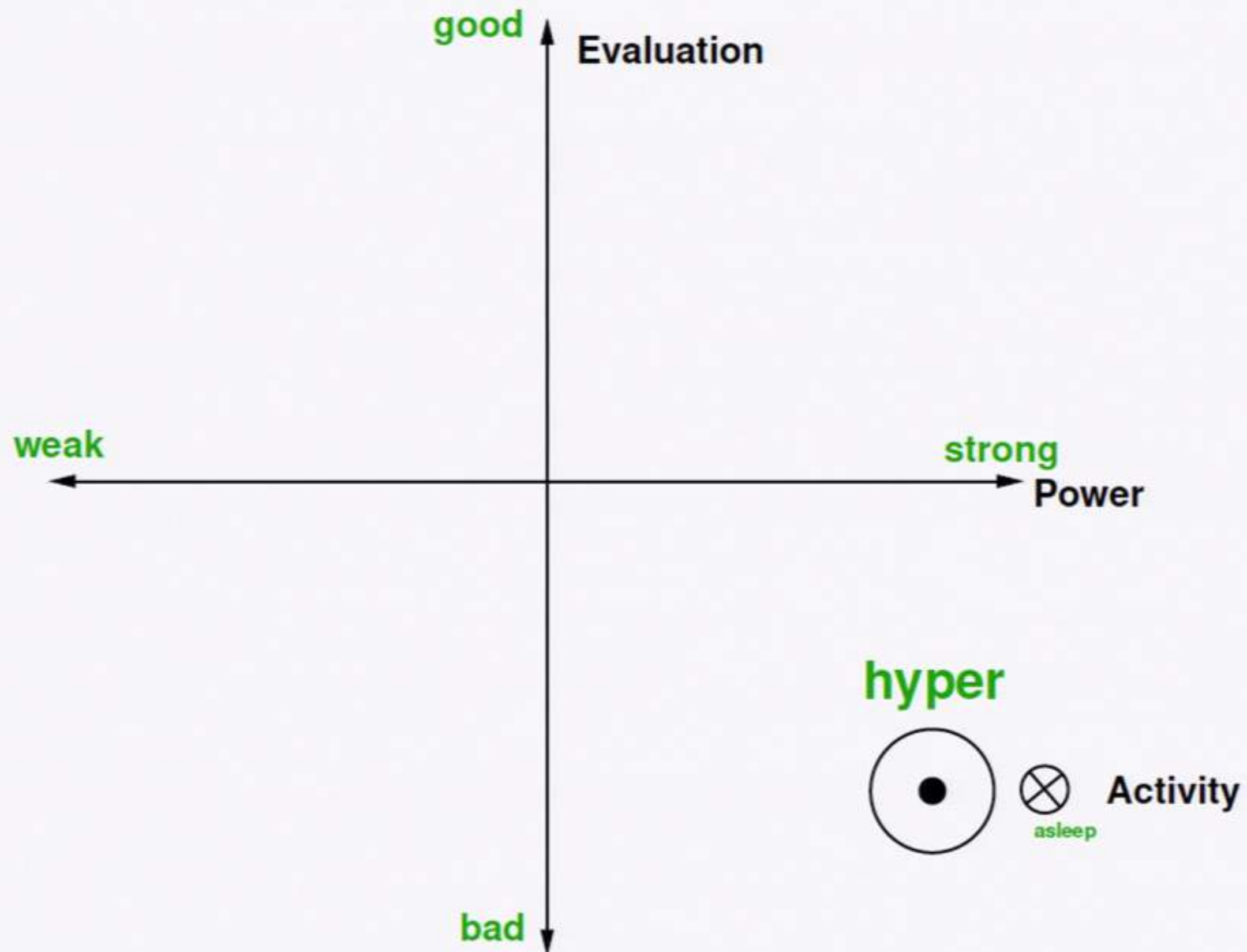
An emotional machine disrupted by cognition

Osgood's Semantic Differential

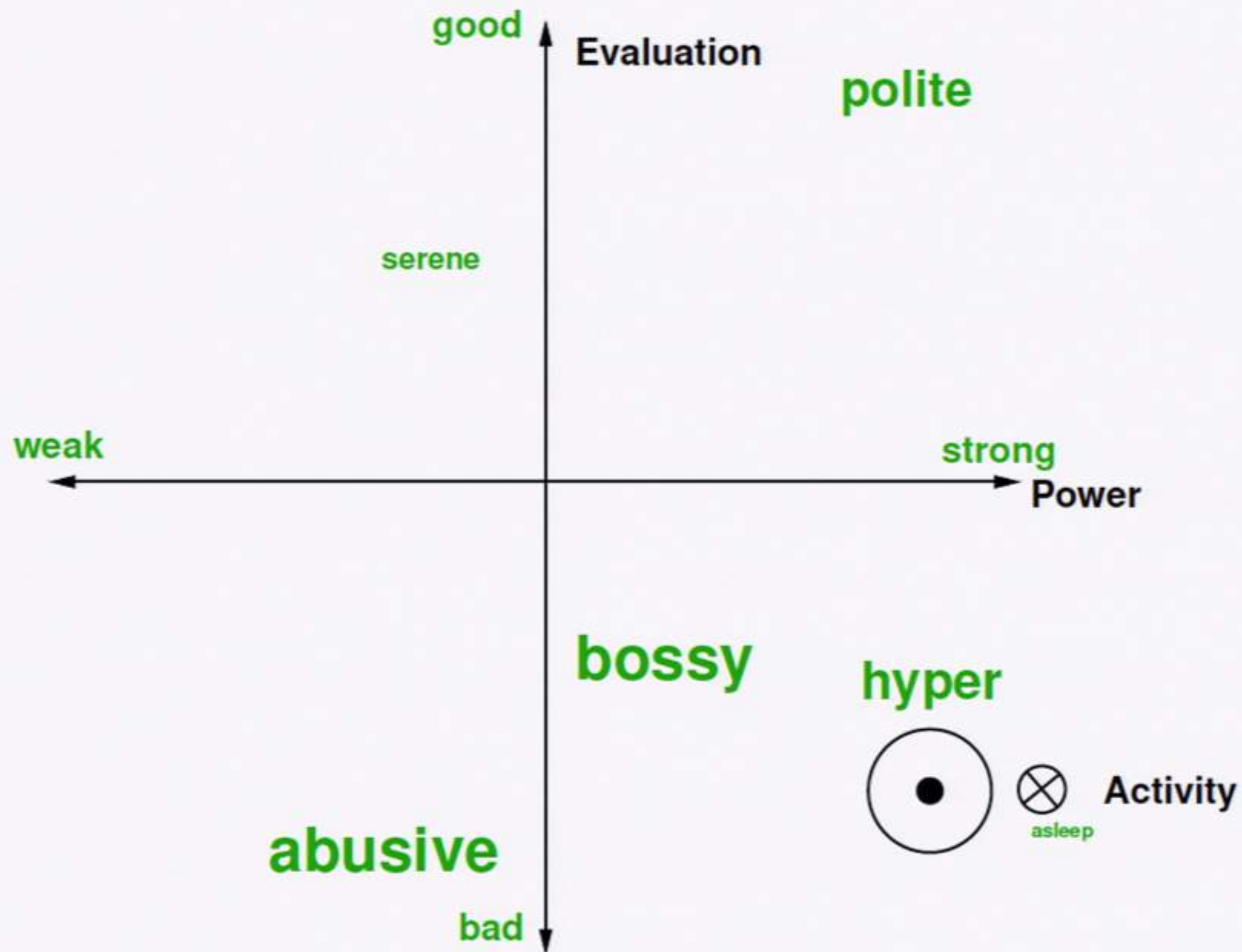


Charles E Osgood. The nature and measurement of meaning. *Psychological bulletin*, 49(3):197, 1952.

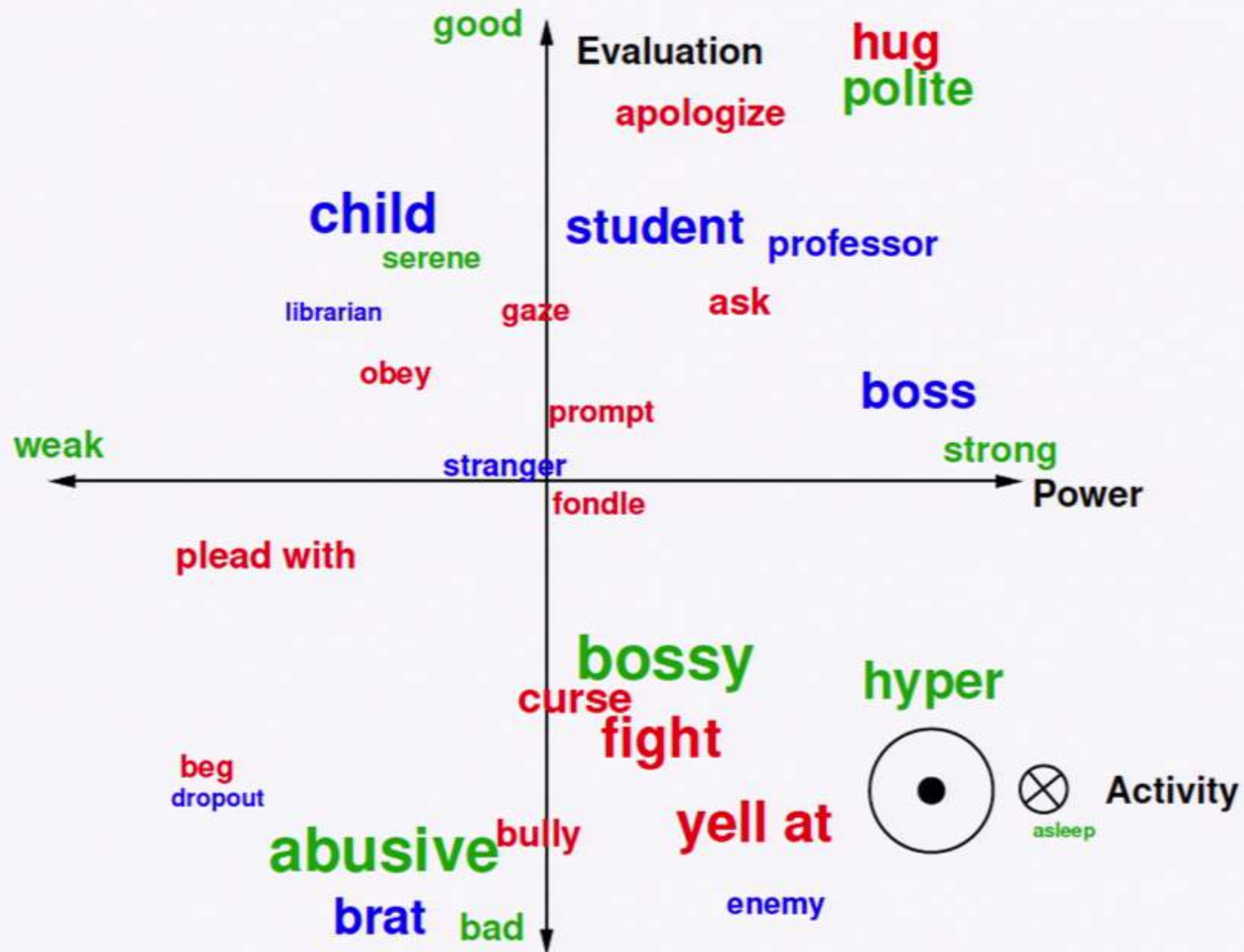
Fundamental Sentiments



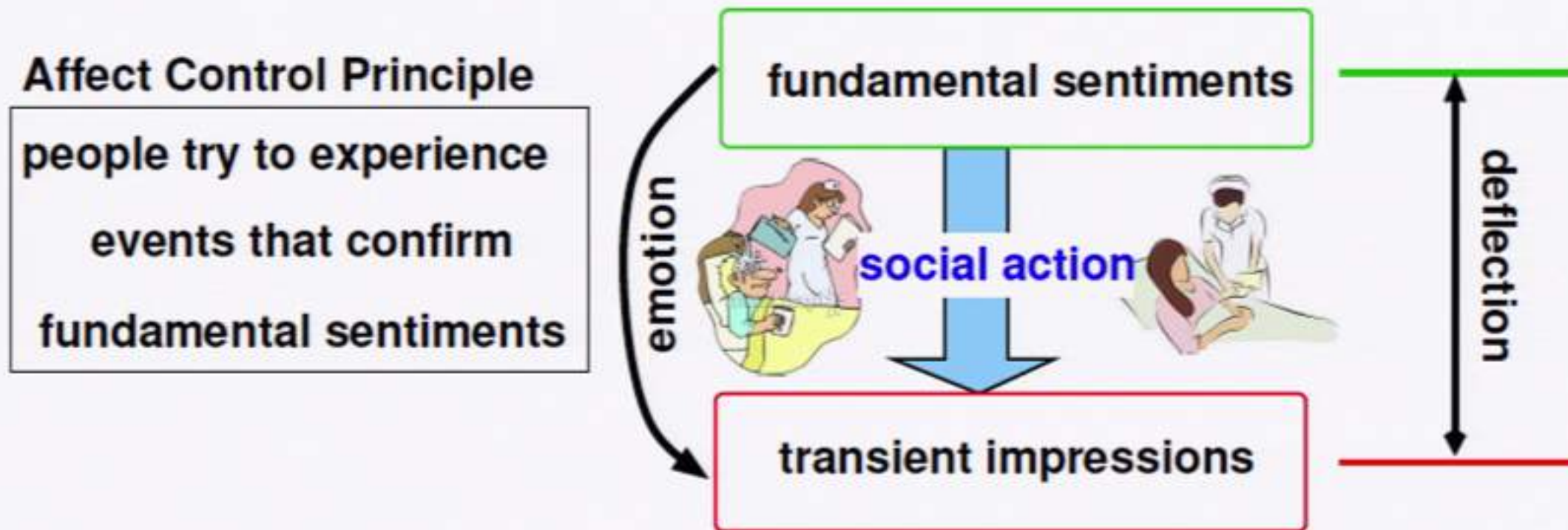
Fundamental Sentiments



Fundamental Sentiments

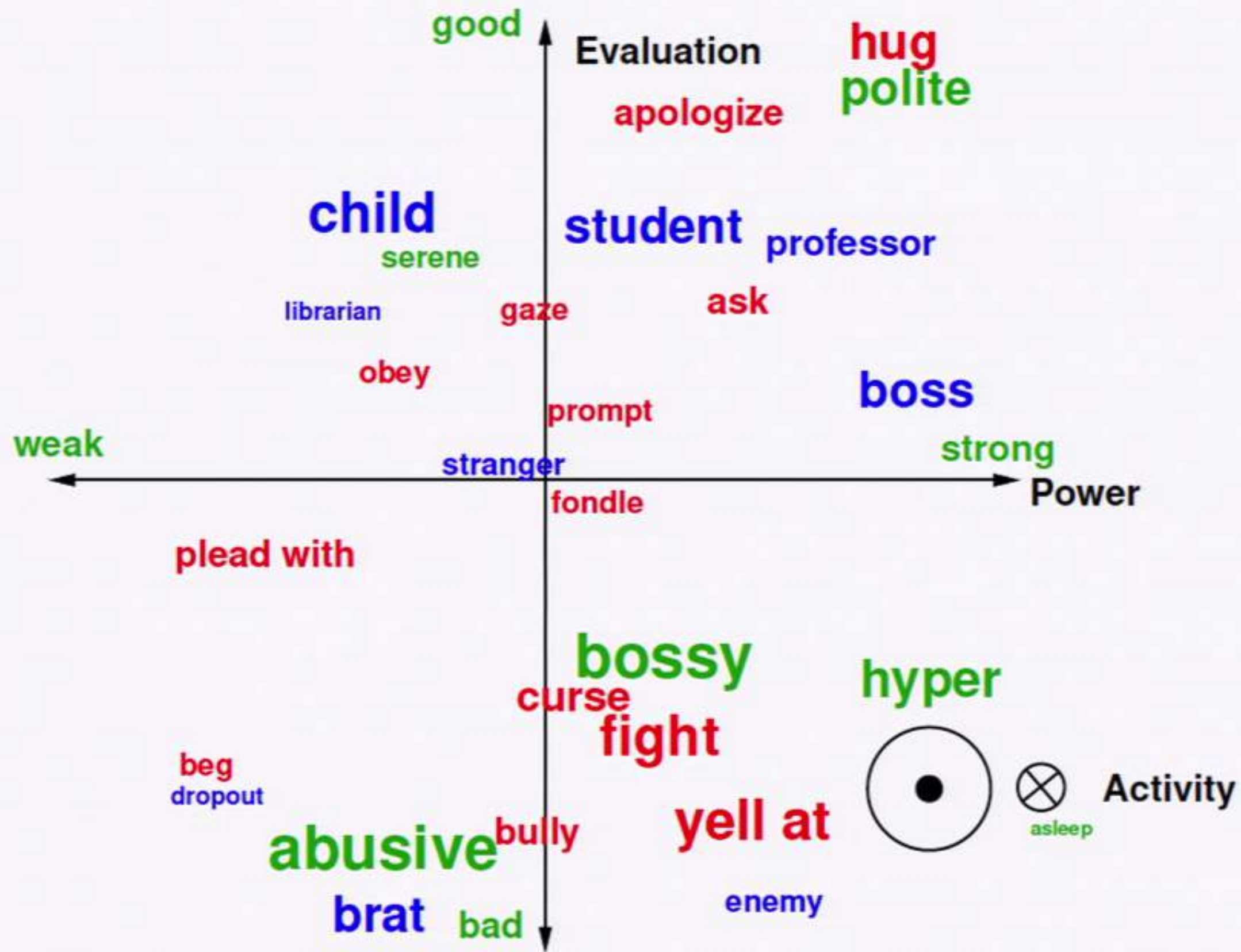


Affect Control Theory

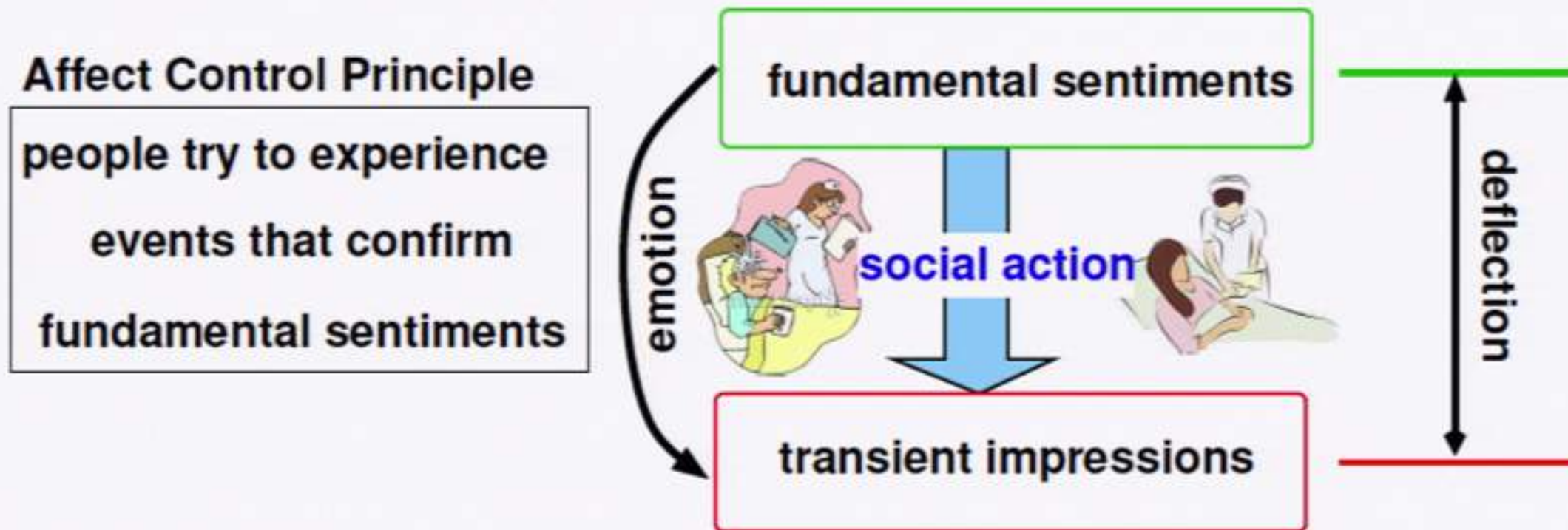


- Shared sentiments
- Shared **emotional dynamics**
- Shared **consistency** → **Cooperation**
- David Heise. **Expressive Order: Confirming Sentiments in Social Actions**, Springer, 2007

Fundamental Sentiments



Affect Control Theory

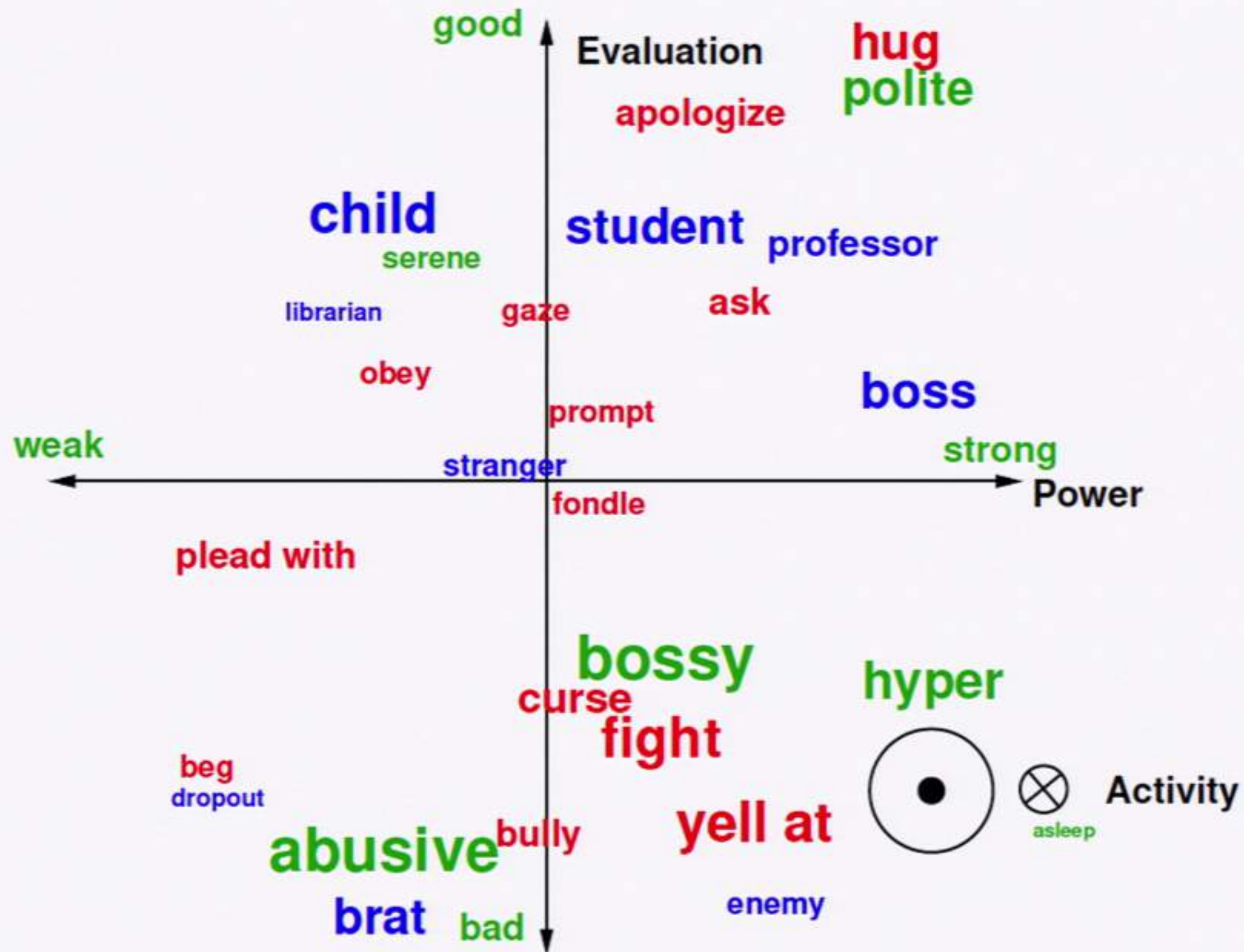


- Shared sentiments
- Shared **emotional dynamics**
- Shared **consistency** → **Cooperation**
- David Heise. **Expressive Order: Confirming Sentiments in Social Actions**, Springer, 2007

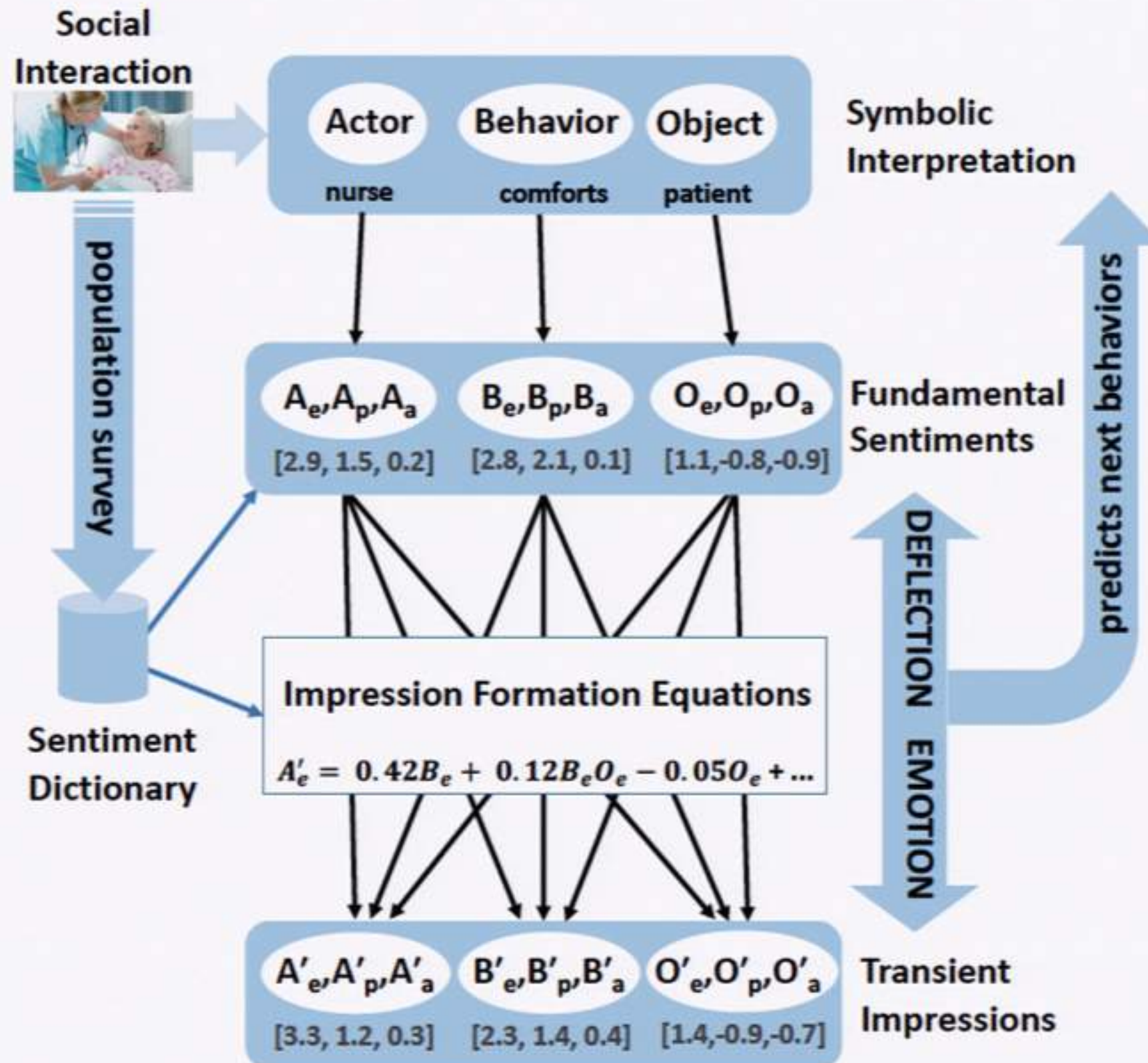
Affect Control Theory

- Actor-Behaviour-Object
- fundamental sentiments: $\mathbf{F} \in [-4.3, 4.3]^9$
- transient impressions: $\mathbf{T} \in [-4.3, 4.3]^9$
- prediction $\mathbf{T}_{t+1} = \mathbf{M}^{\mathcal{G}}(\mathbf{F}_t, \mathbf{T}_t)$ **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 $\epsilon \propto \mathbf{f} - \boldsymbol{\tau}$

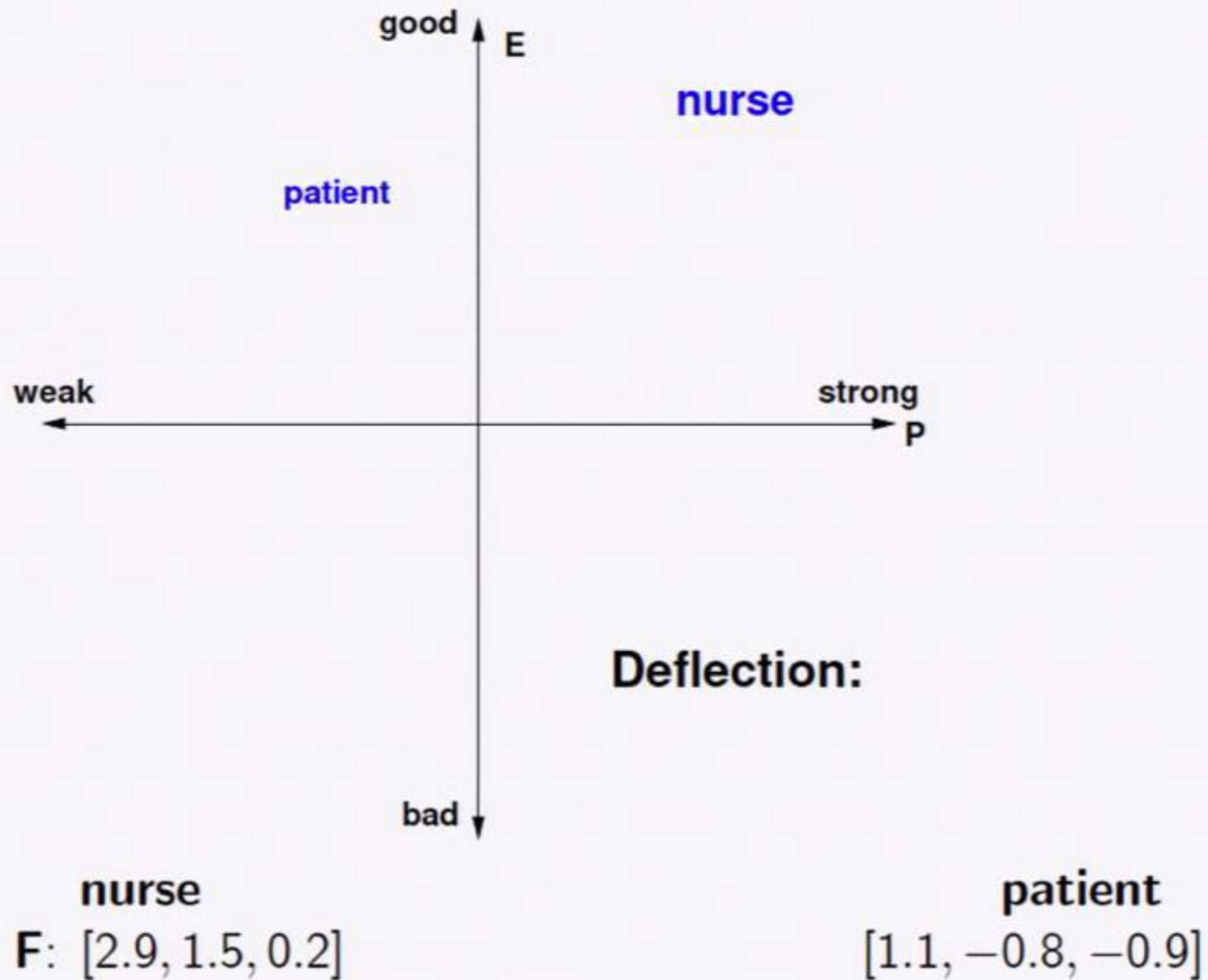
Fundamental Sentiments



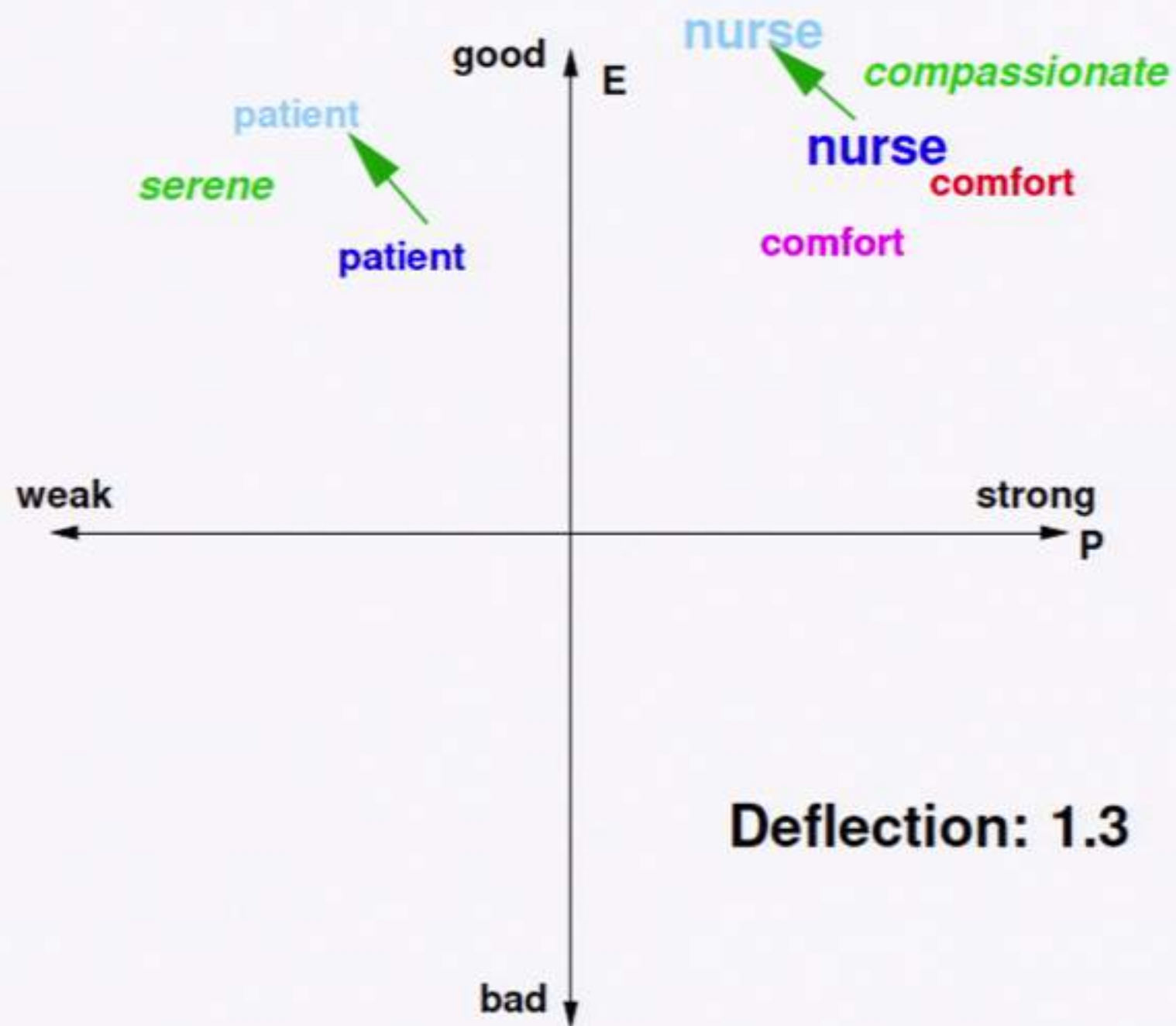
Affect Control Theory



Affect Control Theory

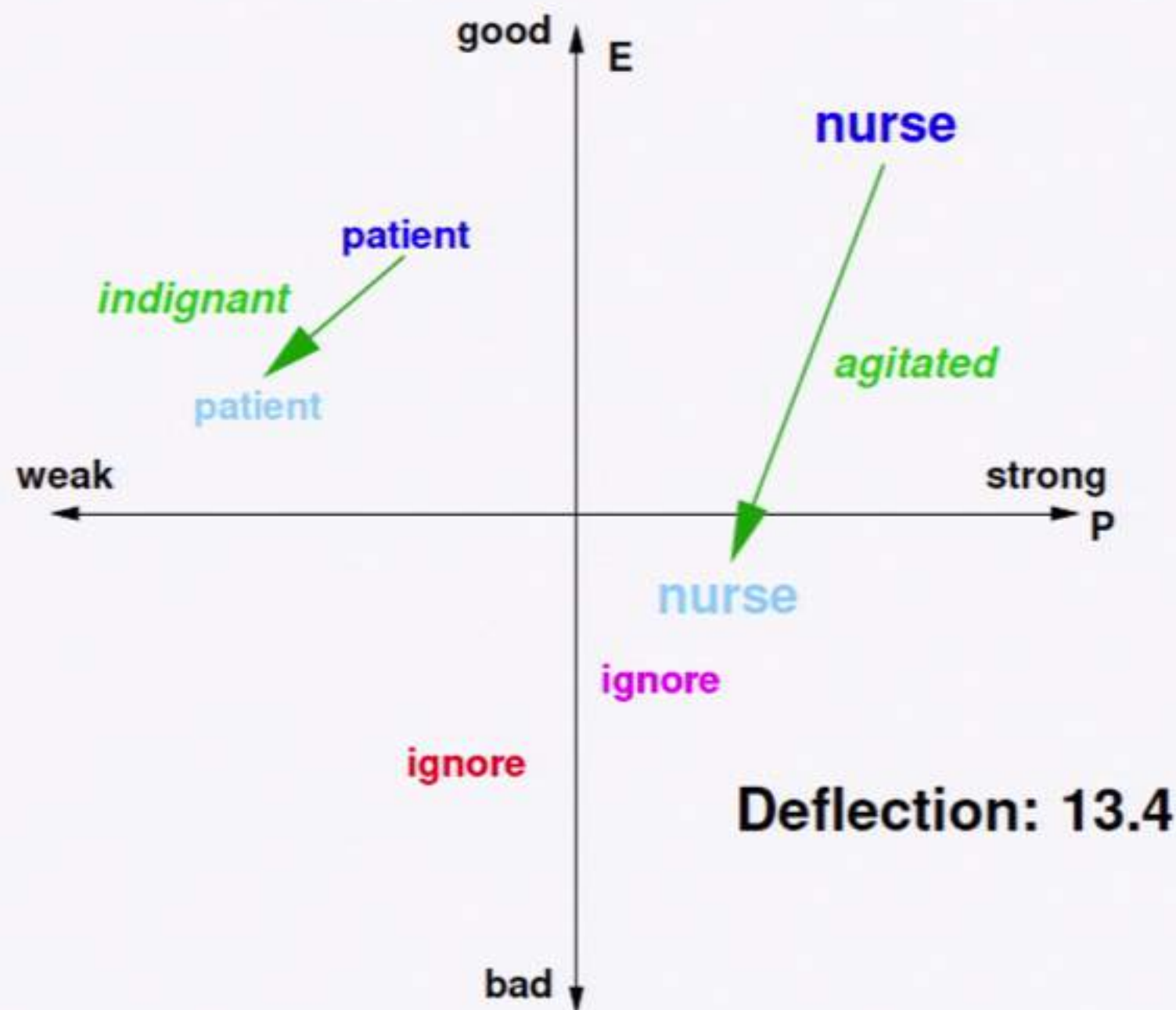


Affect Control Theory



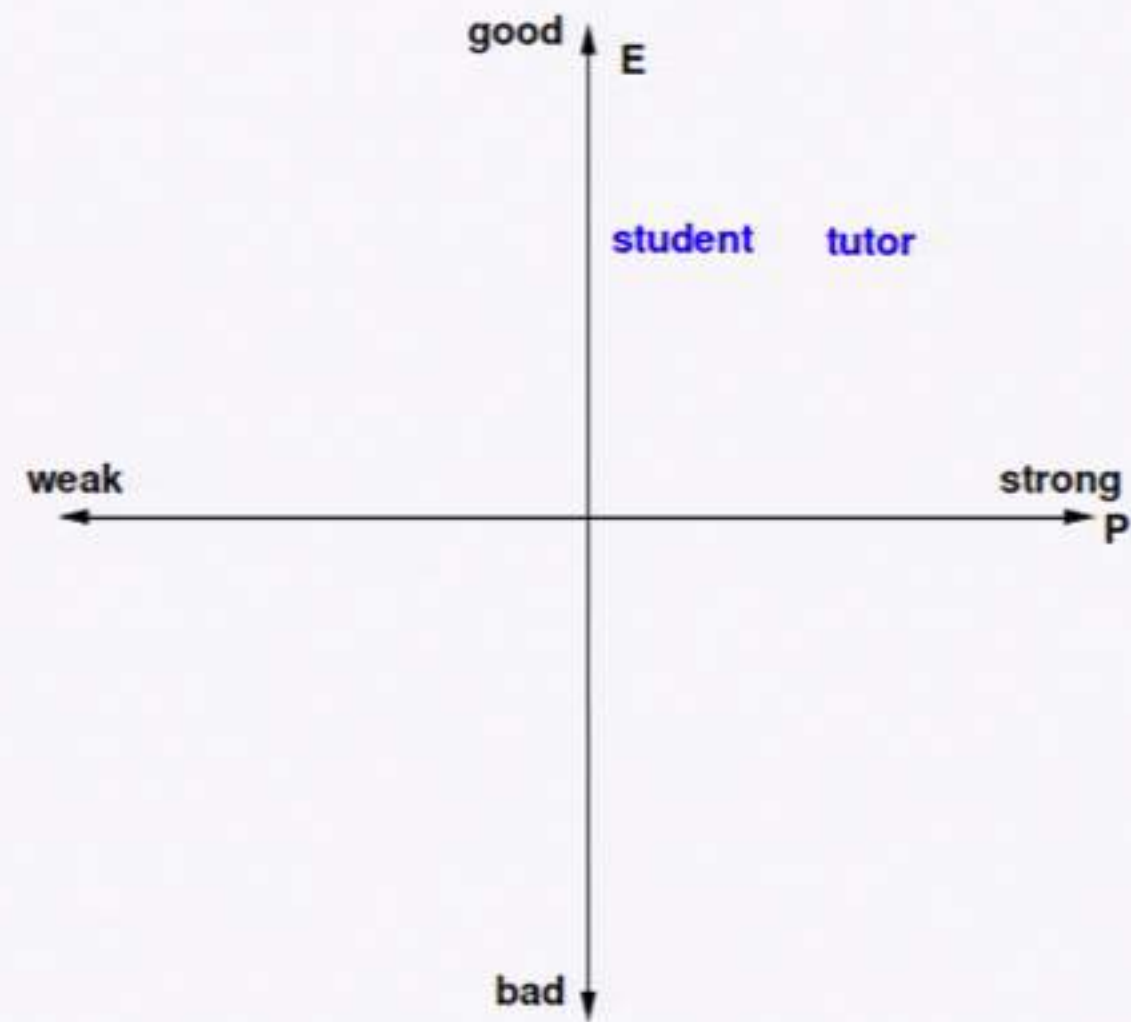
nurse	comforts	patient
F: [2.9, 1.5, 0.2]	[2.8, 2.1, 0.1]	[0.9, -0.7, -1.1]
T: [3.3, 1.2, 0.3]	[2.3, 1.4, 0.4]	[1.4, -0.9, -0.7]

Affect Control Theory



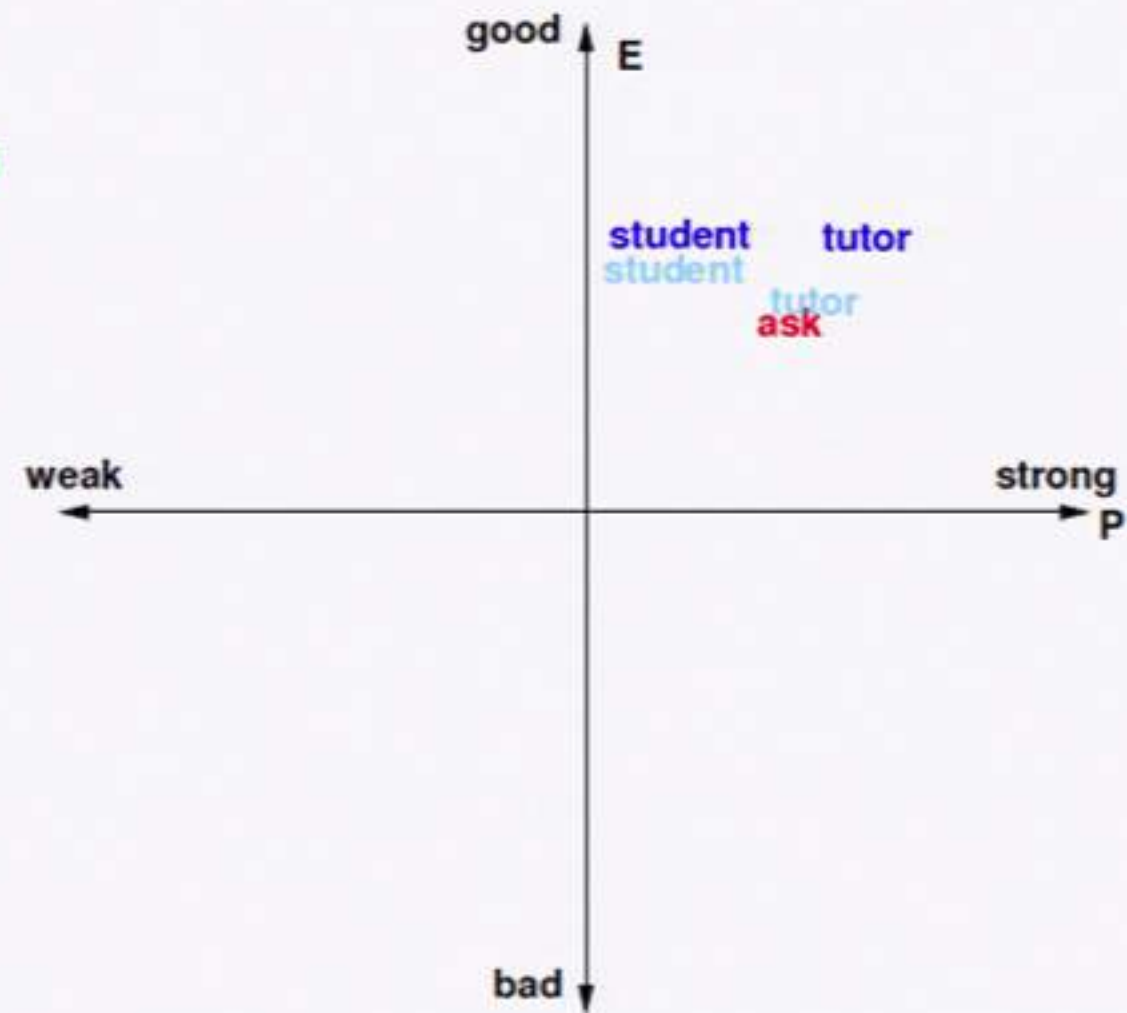
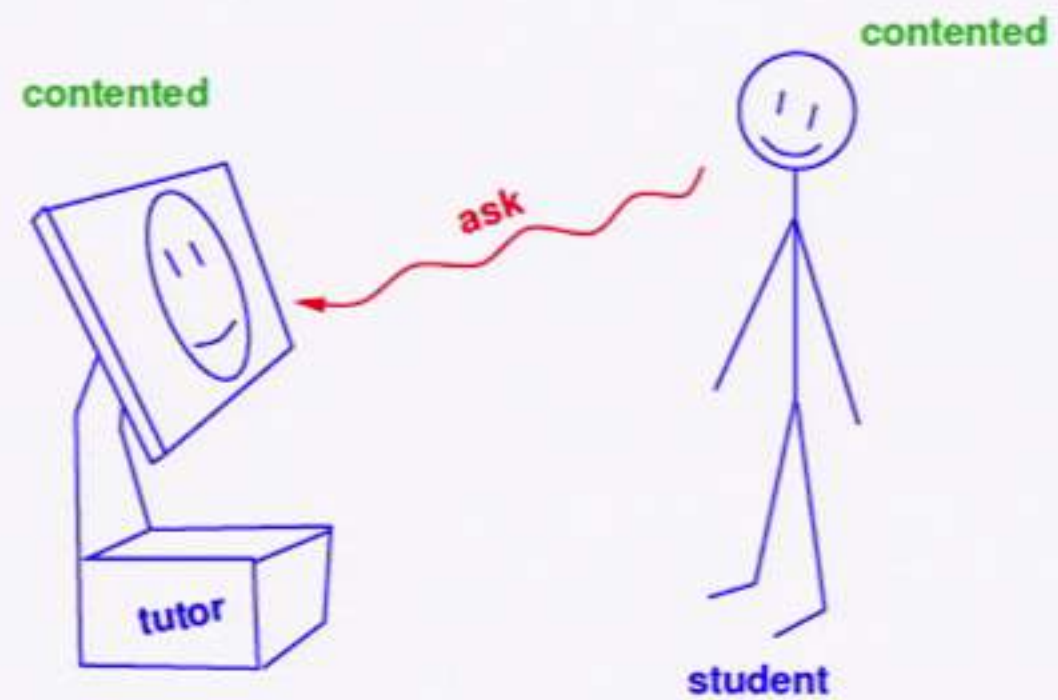
nurse	ignores	patient
F: [2.9, 1.5, 0.2]	[-1.9, -0.3, -0.9]	[0.9, -0.7, -1.1]
T: [-0.5, 0.9, 0.3]	[-1.2, 0.4, -0.4]	[0.4, -1.4, -0.8]

ACT Examples



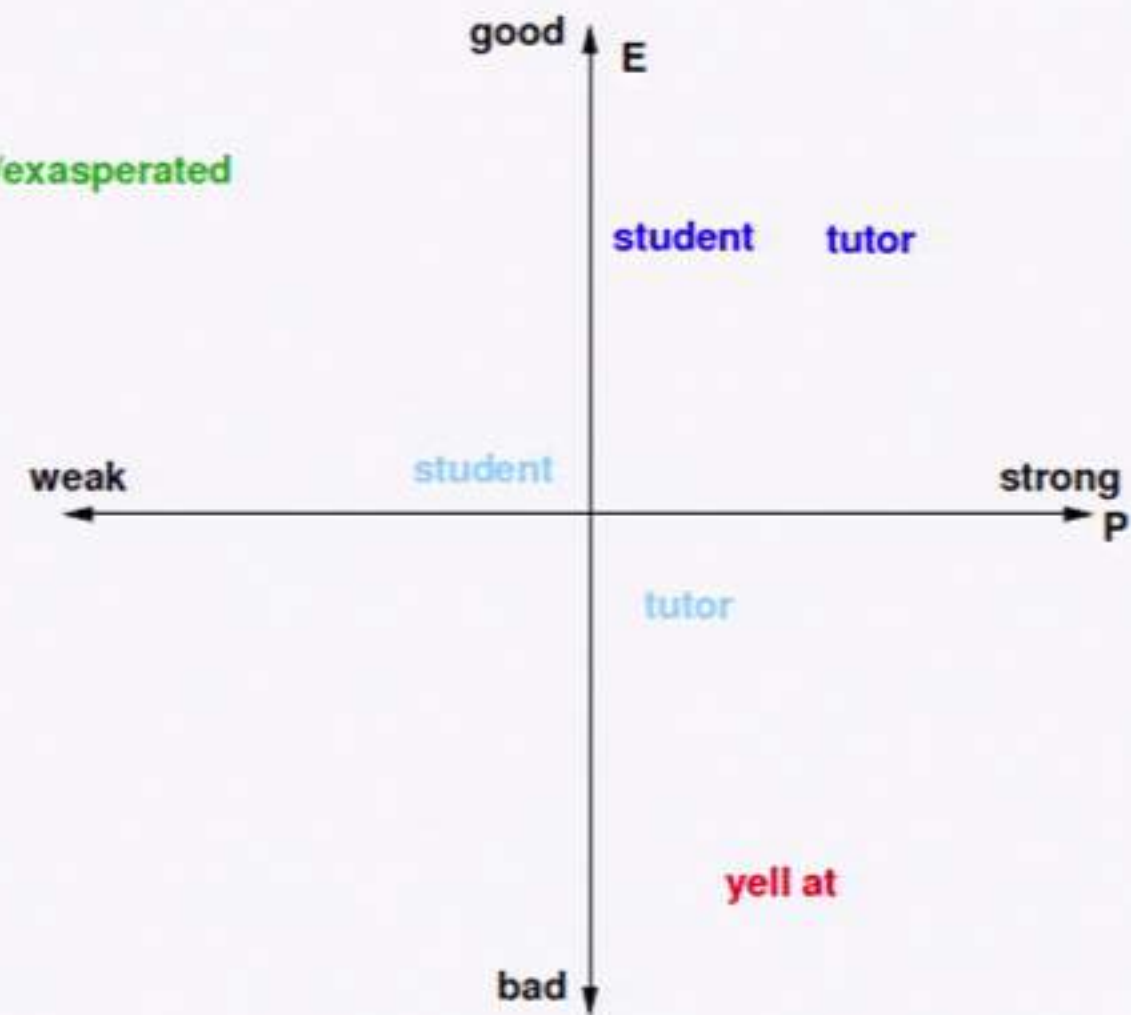
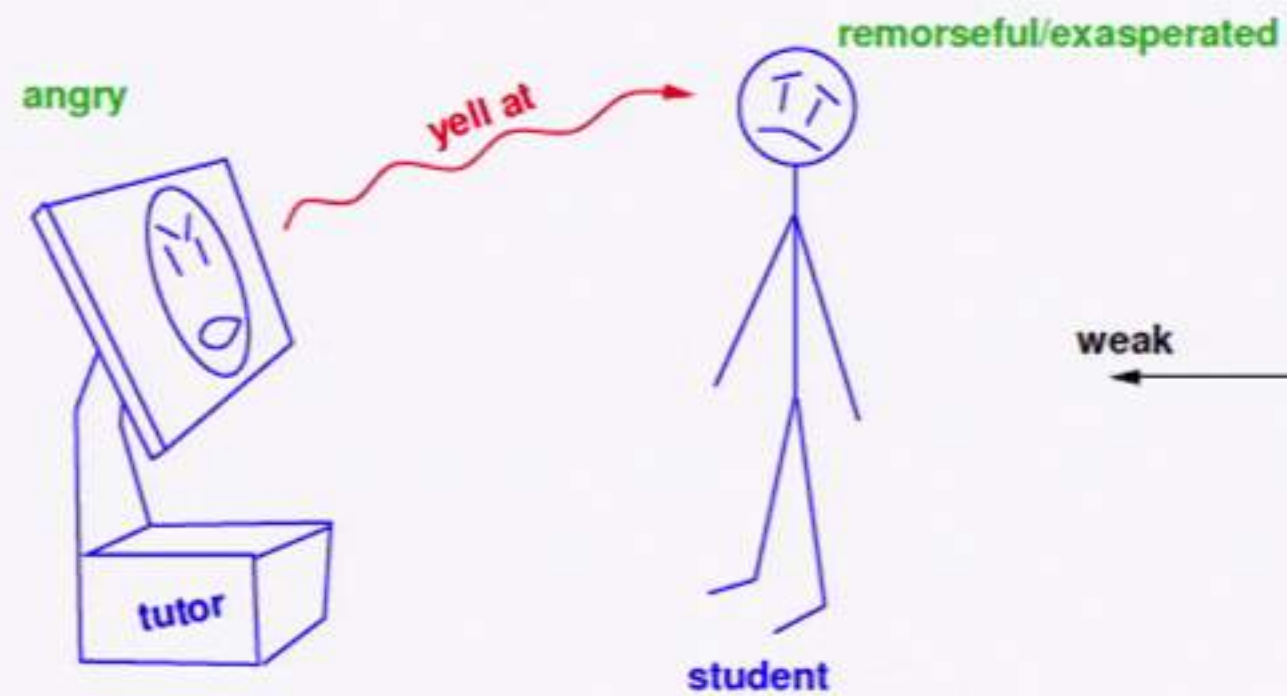
Deflection:

ACT Examples



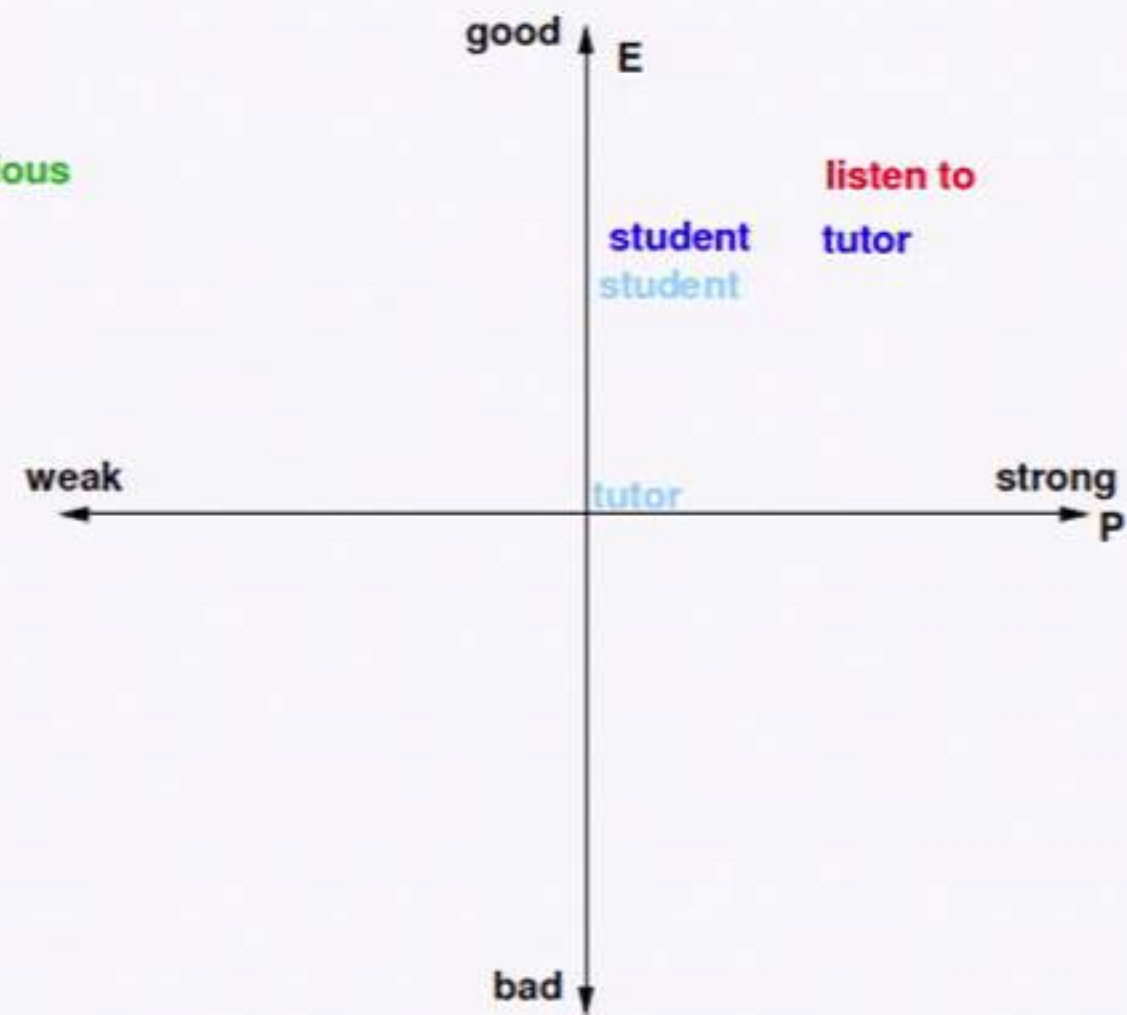
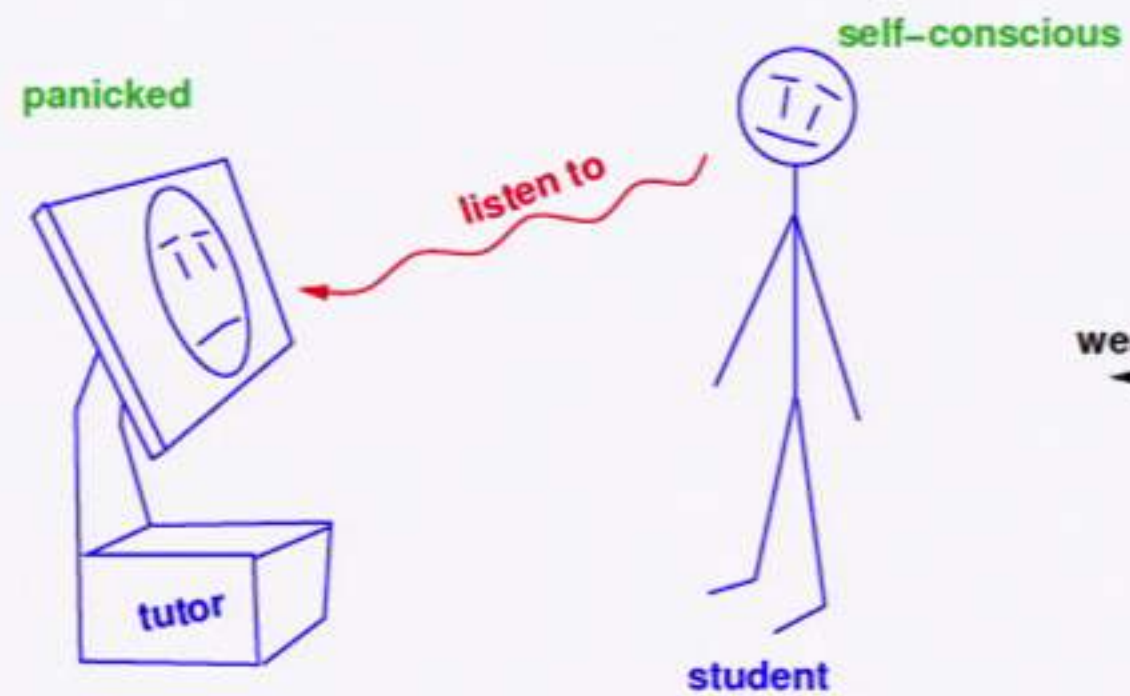
Deflection: 1.0

ACT Examples



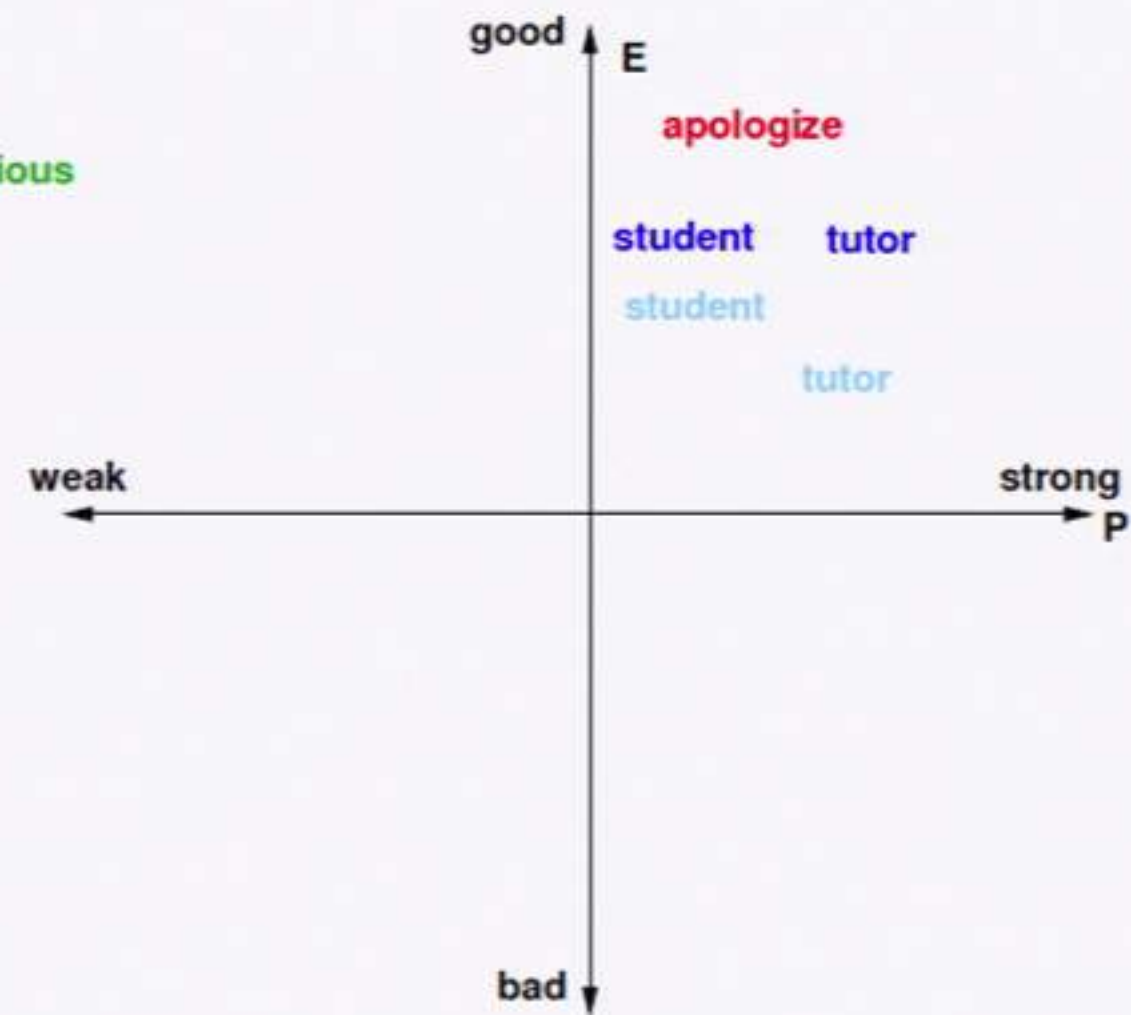
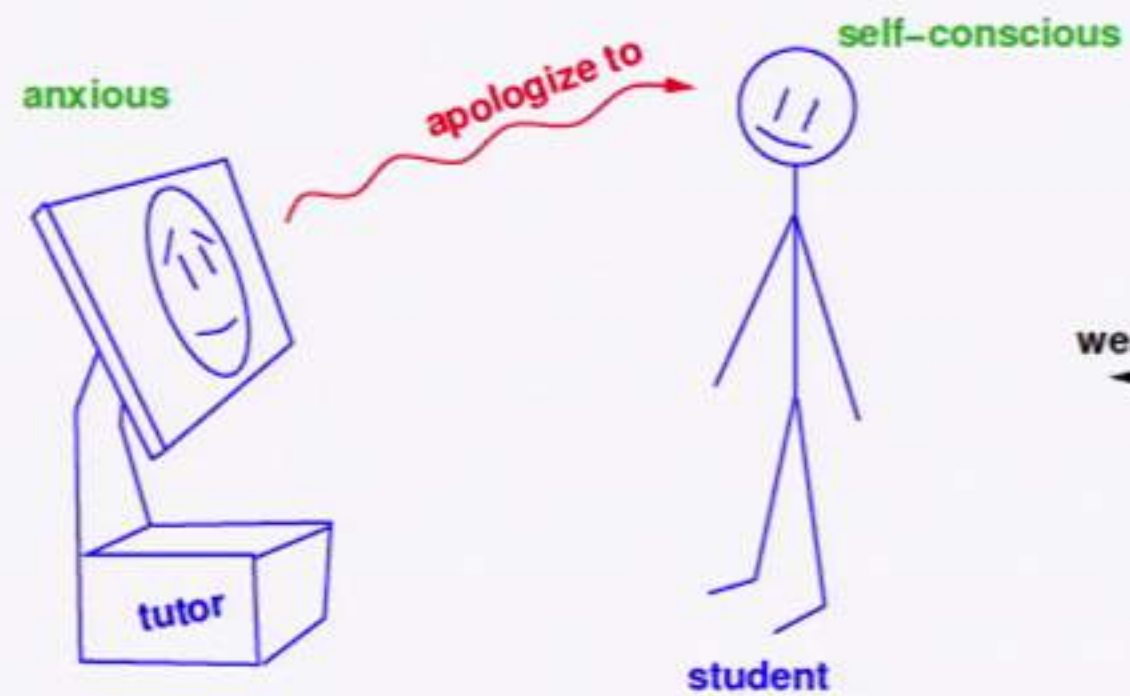
Deflection: 8.0

ACT Examples



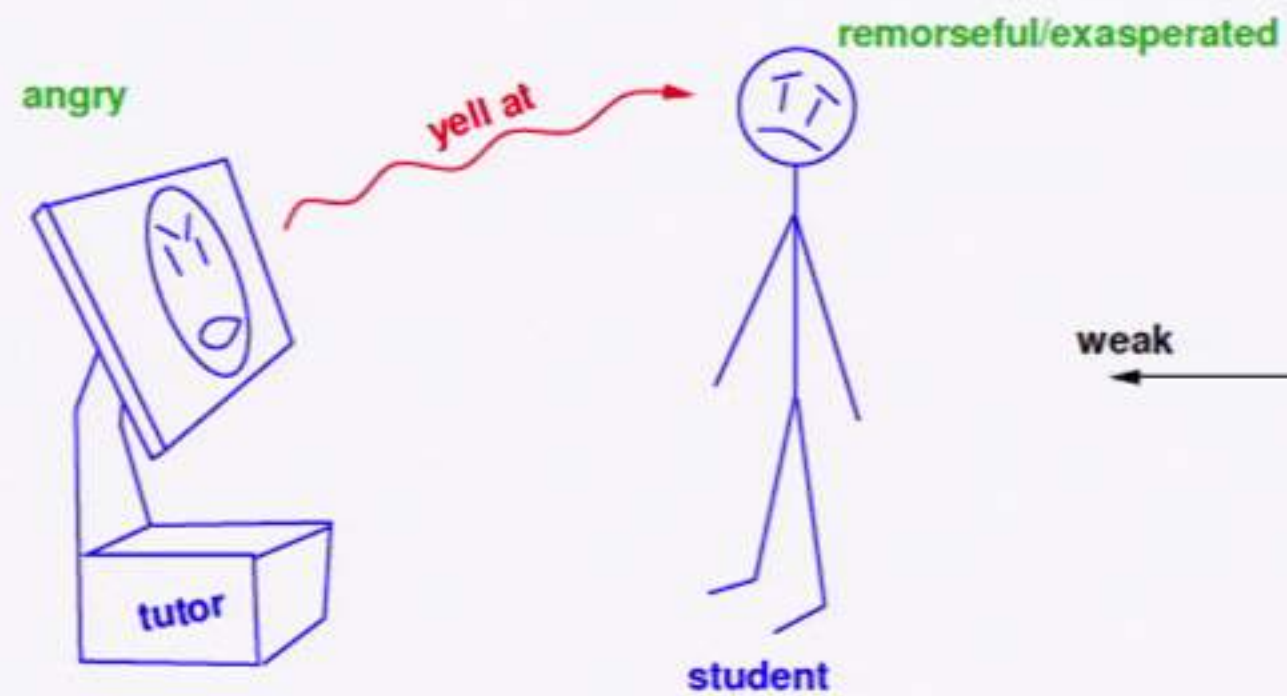
Deflection: 6.0

ACT Examples

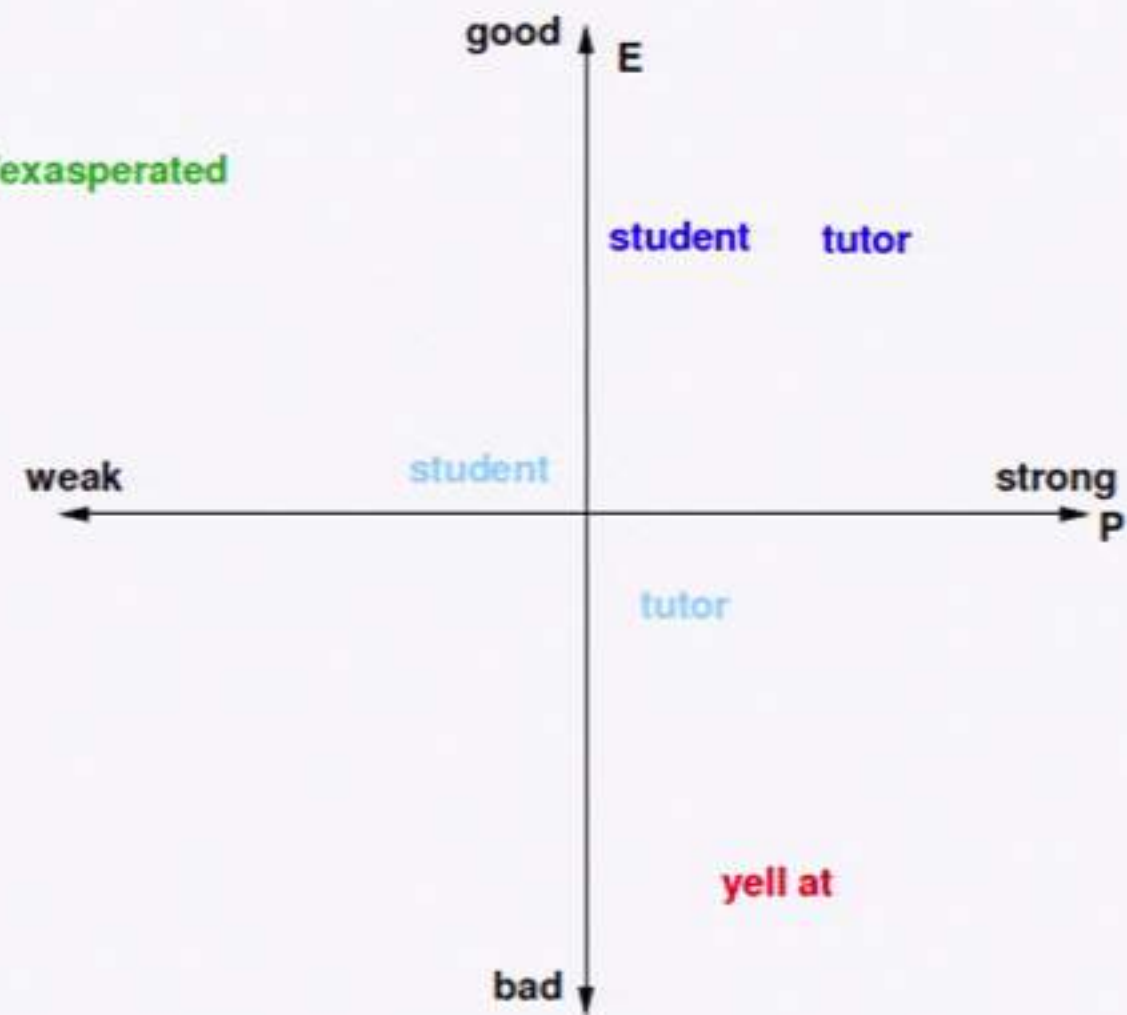


Deflection: 4.0

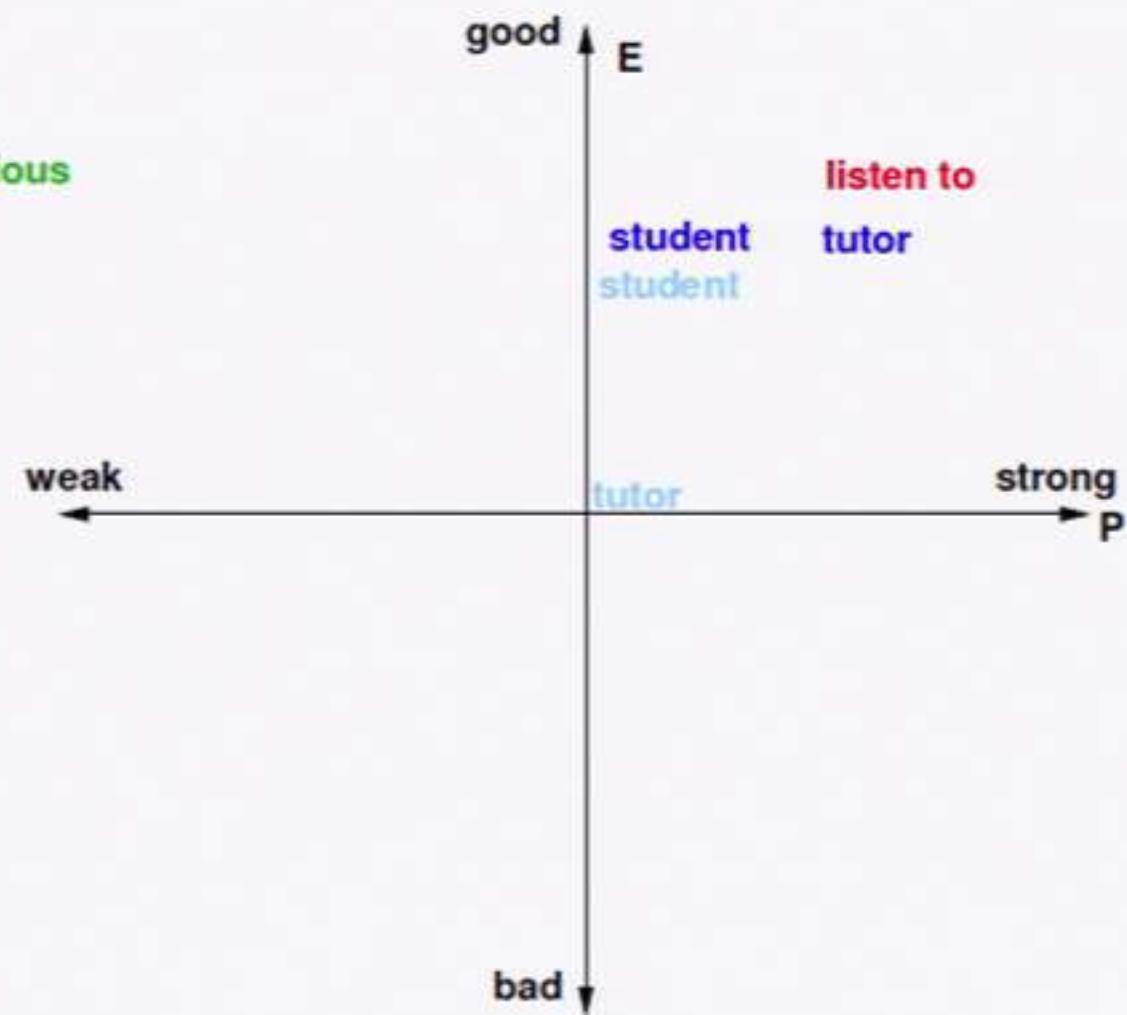
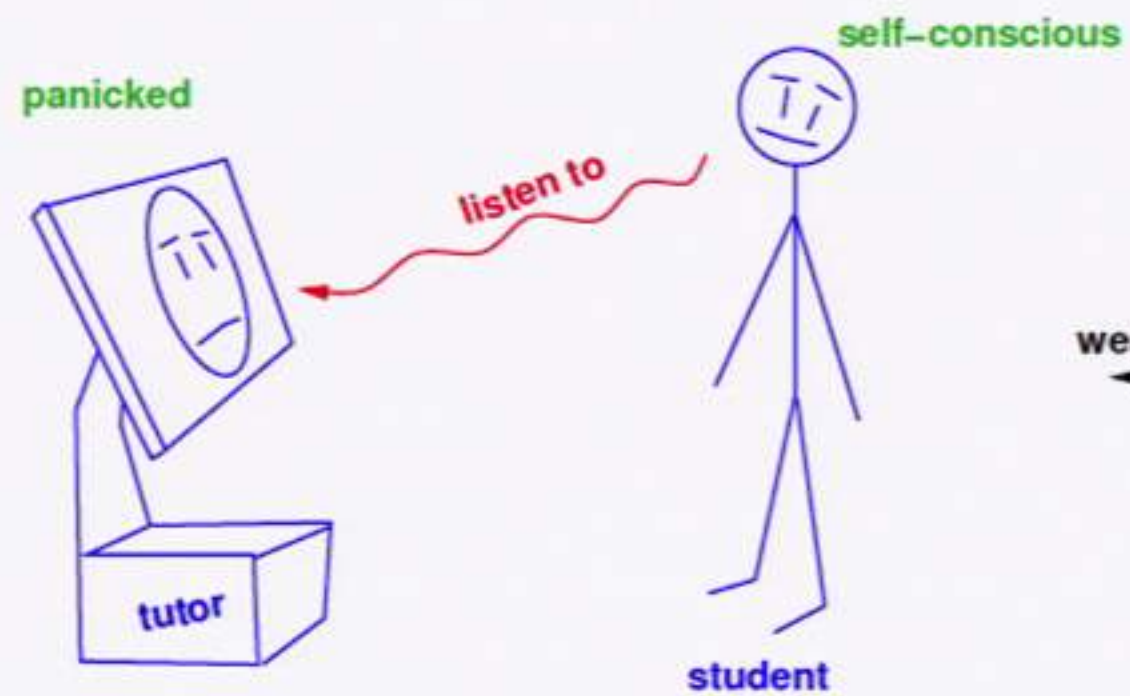
ACT Examples



Deflection: 8.0

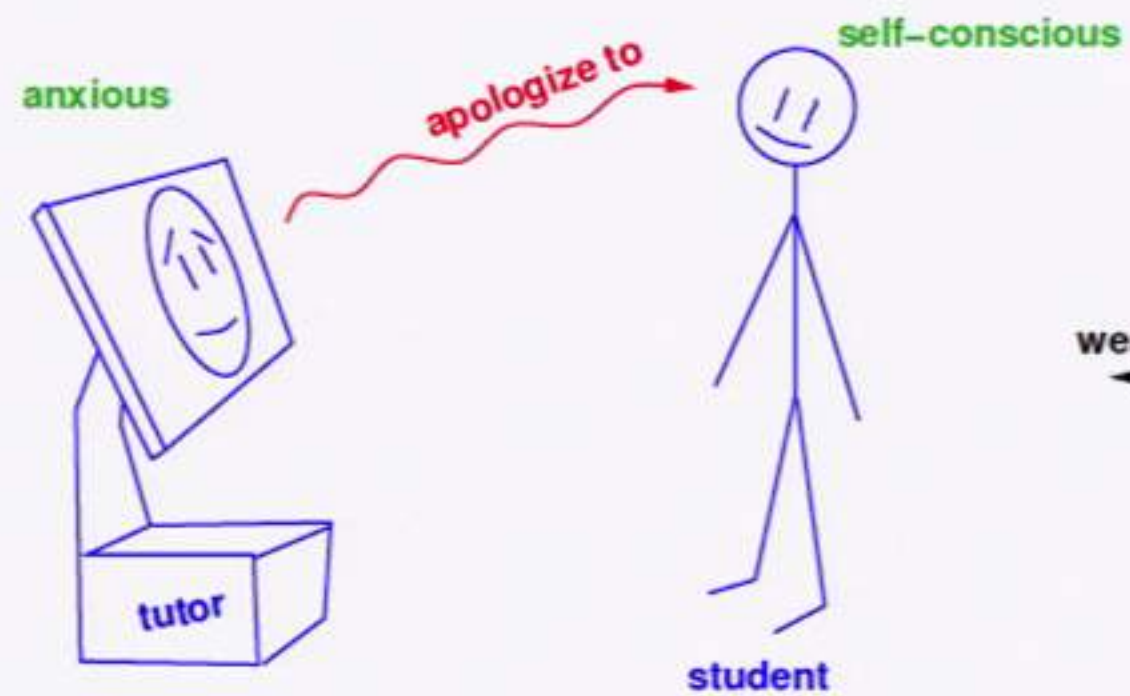


ACT Examples

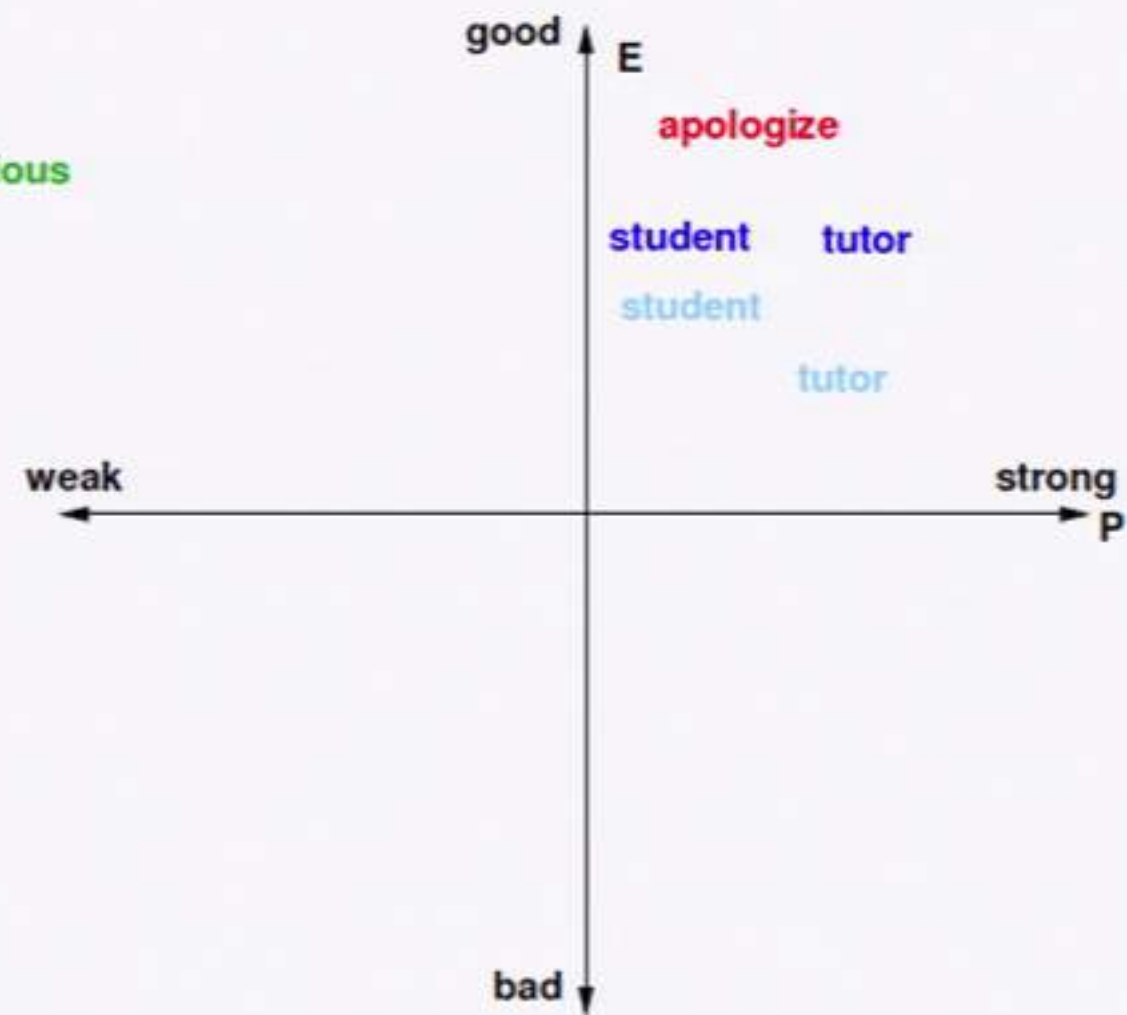


Deflection: 6.0

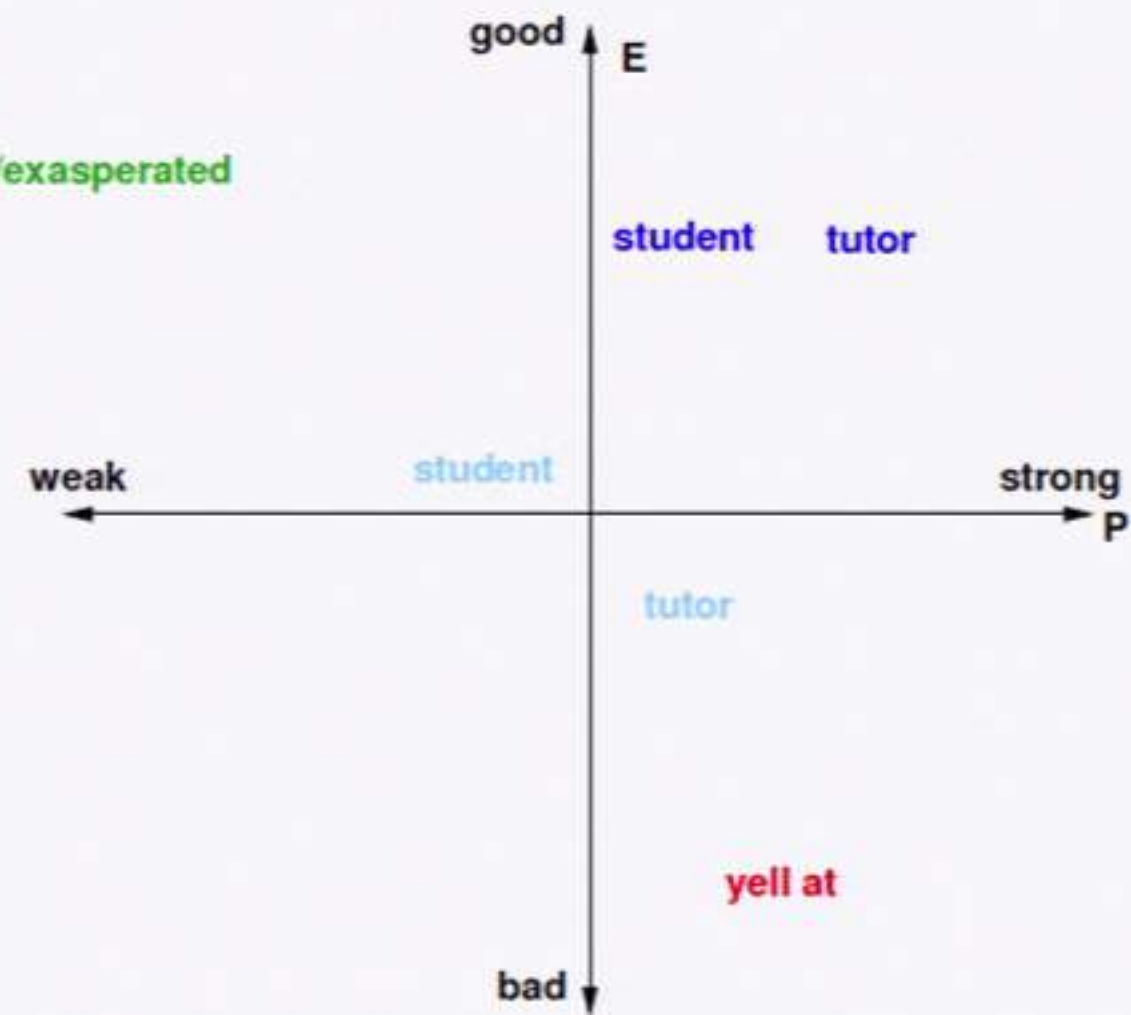
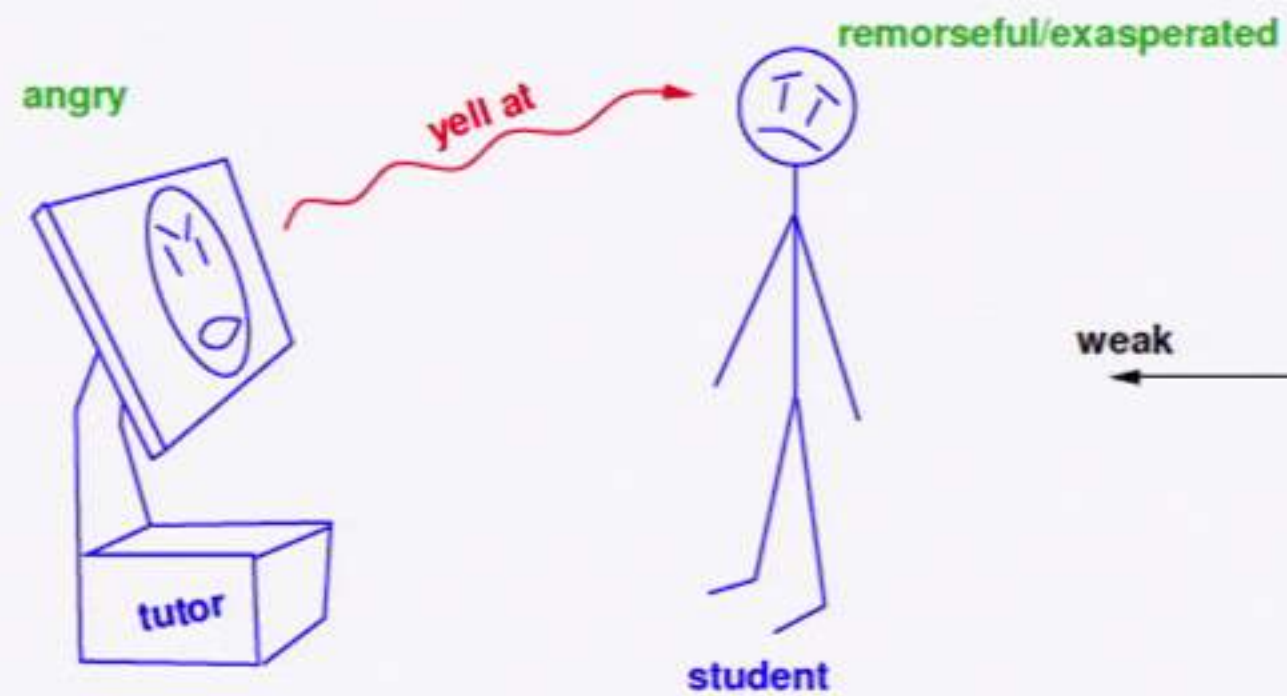
ACT Examples



Deflection: 4.0



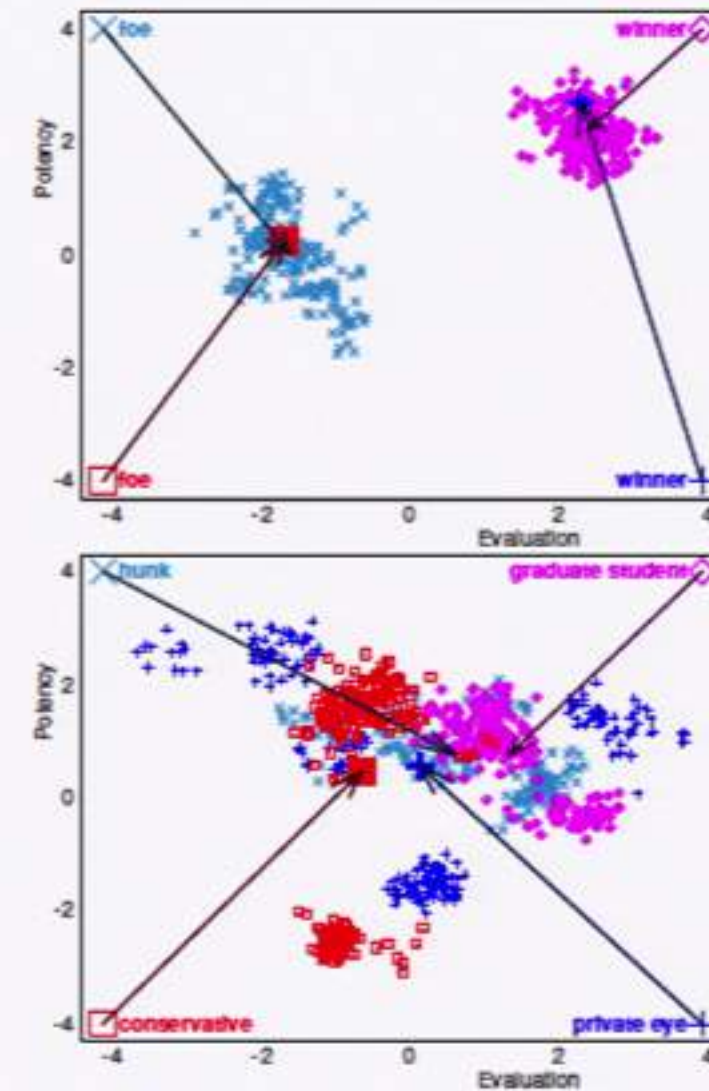
ACT Examples



Deflection: 8.0

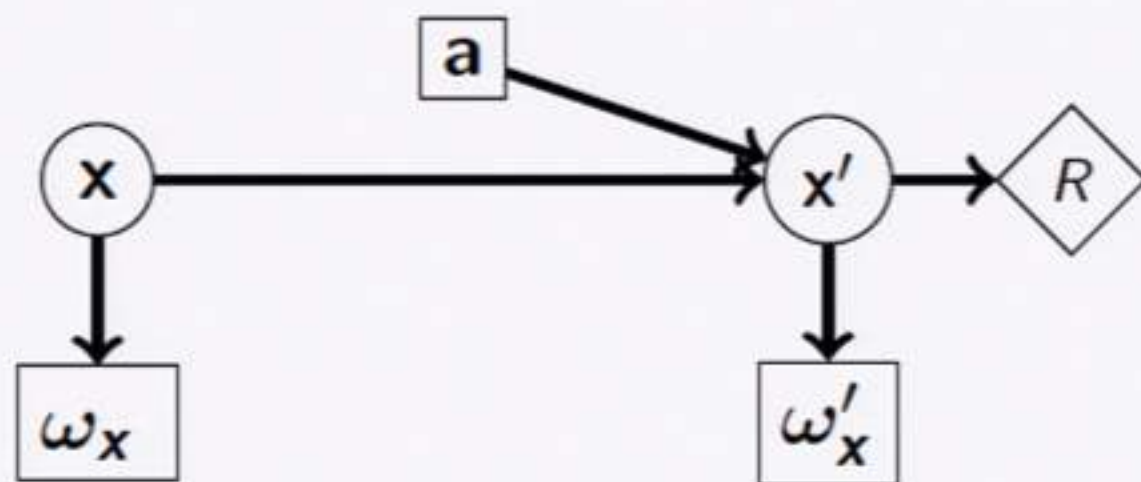
Bayesian Affect Control Theory

- models **uncertainty**
- identities and behaviours: **probability distributions**
- external goals and **planning**
- Partially Observable **Markov Decision Process**
- Affect Control Principle guides **action selection**



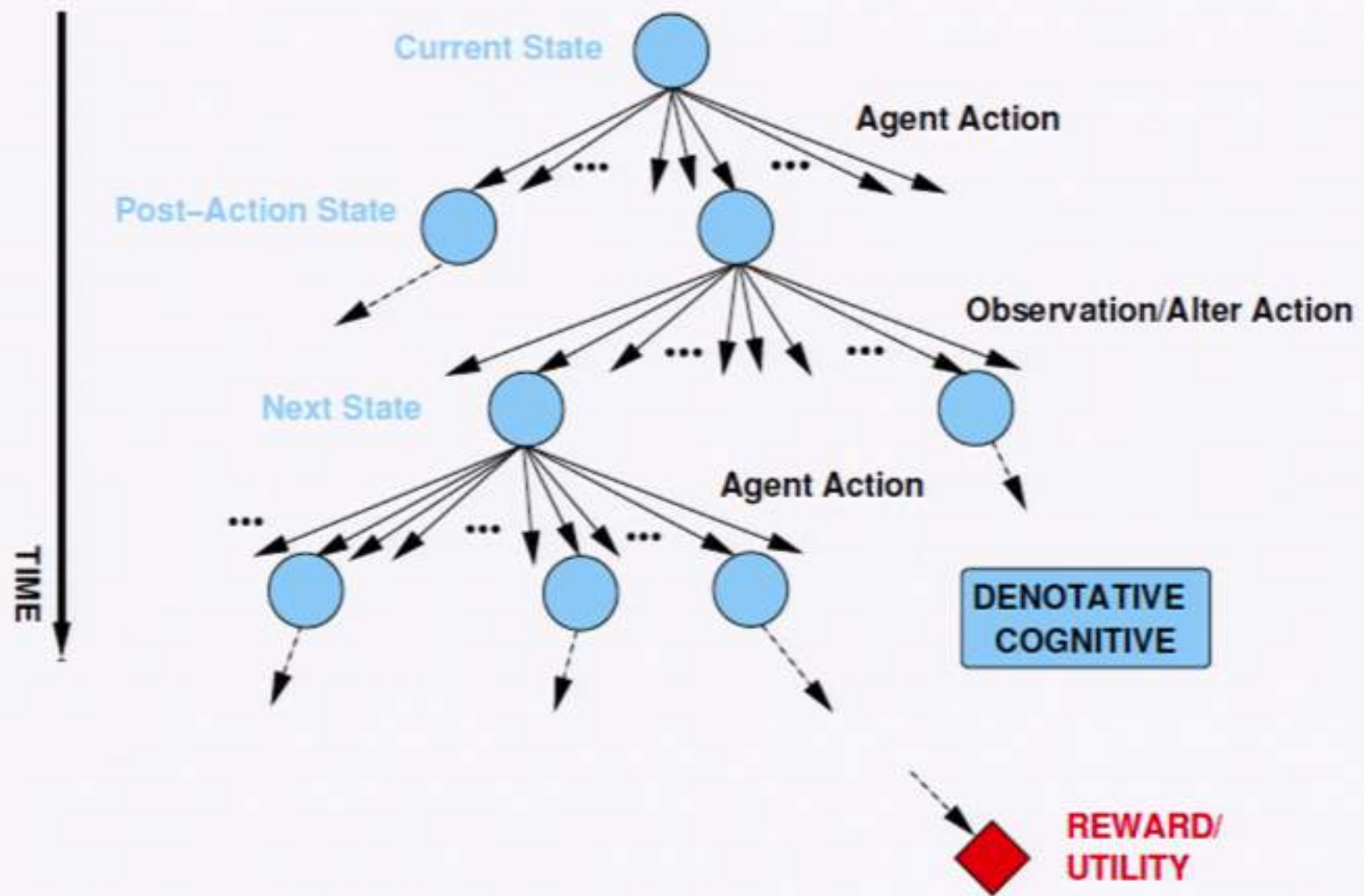
- Jesse Hoey, Tobias Schröder and Areej Alhothali. Affect control processes: Intelligent affective interaction using a partially observable Markov decision process. *Artificial Intelligence*, 230, 2016.
- Tobias Schröder, Jesse Hoey and Kimberly B. Rogers. Modeling Dynamic Identities and Uncertainty in Social Interactions: Bayesian Affect Control Theory. *American Sociological Review*, 81, 4, 2016.

Partially Observable Markov Decision Process

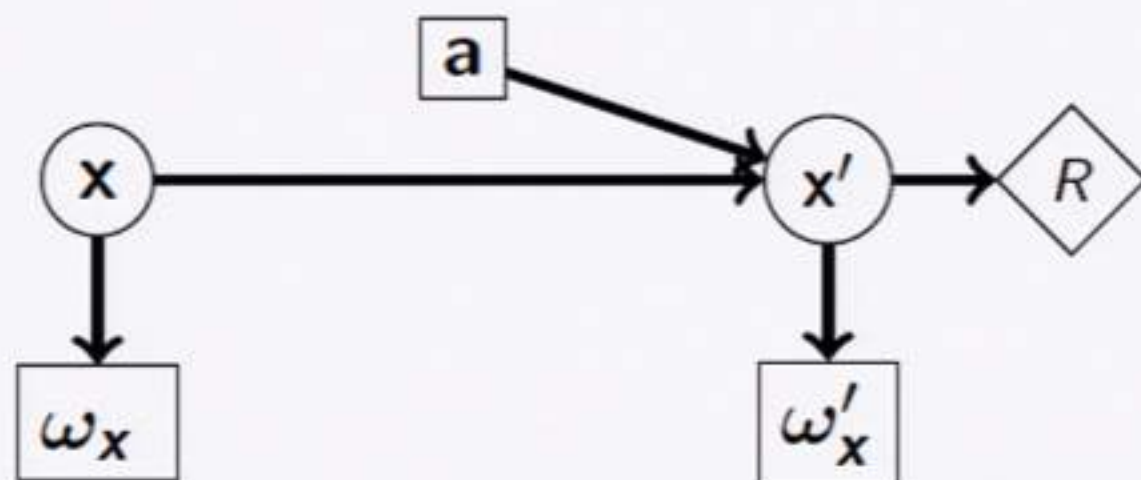


- $Pr(x'|x, a)$
- $Pr(\omega_x|x)$
- $R(x')$

Artificial Intelligence: Decision Theoretic

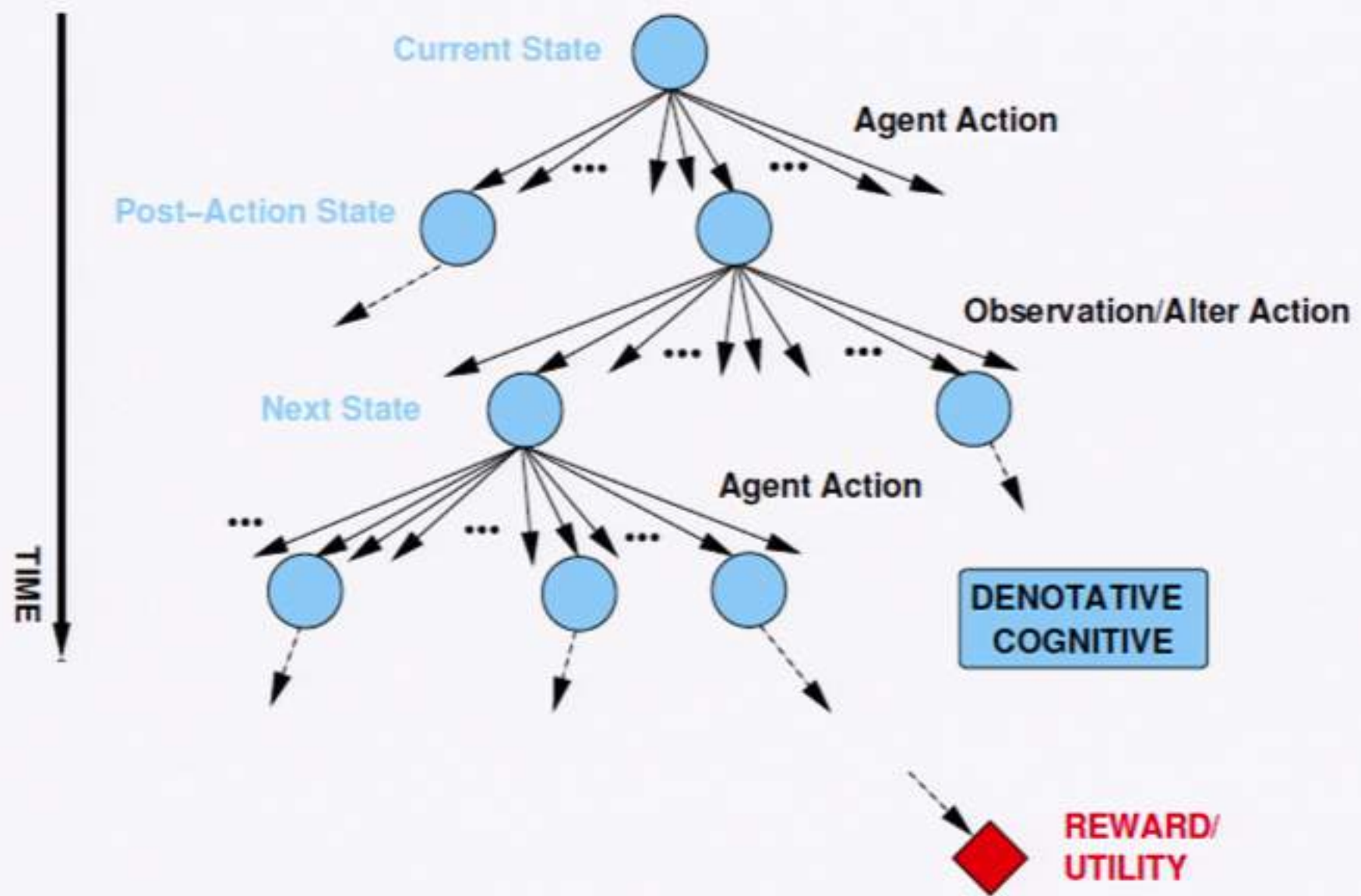


Partially Observable Markov Decision Process

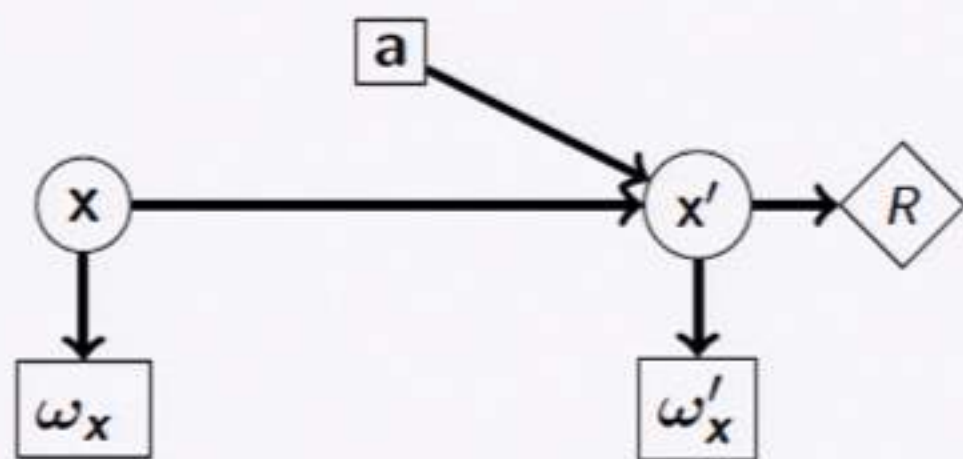


- $Pr(x'|x, a)$
- $Pr(\omega_x|x)$
- $R(x')$

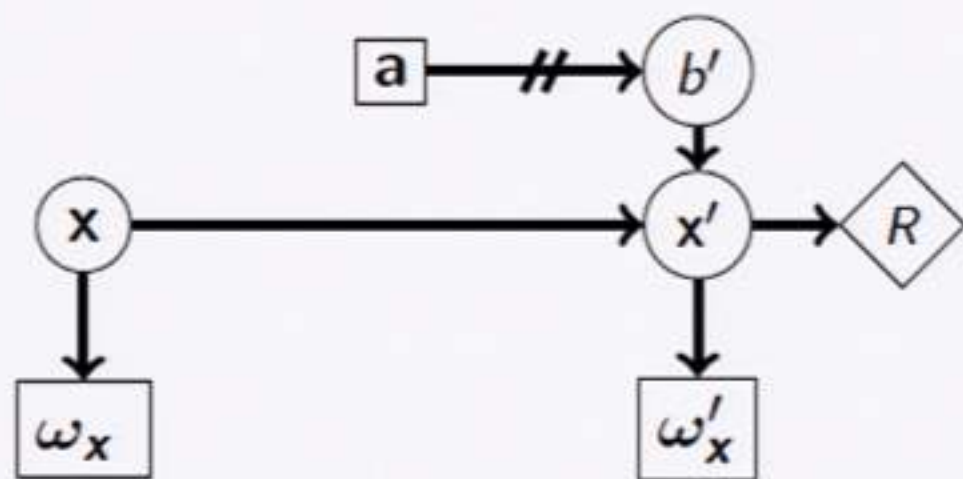
Artificial Intelligence: Decision Theoretic



BayesACT

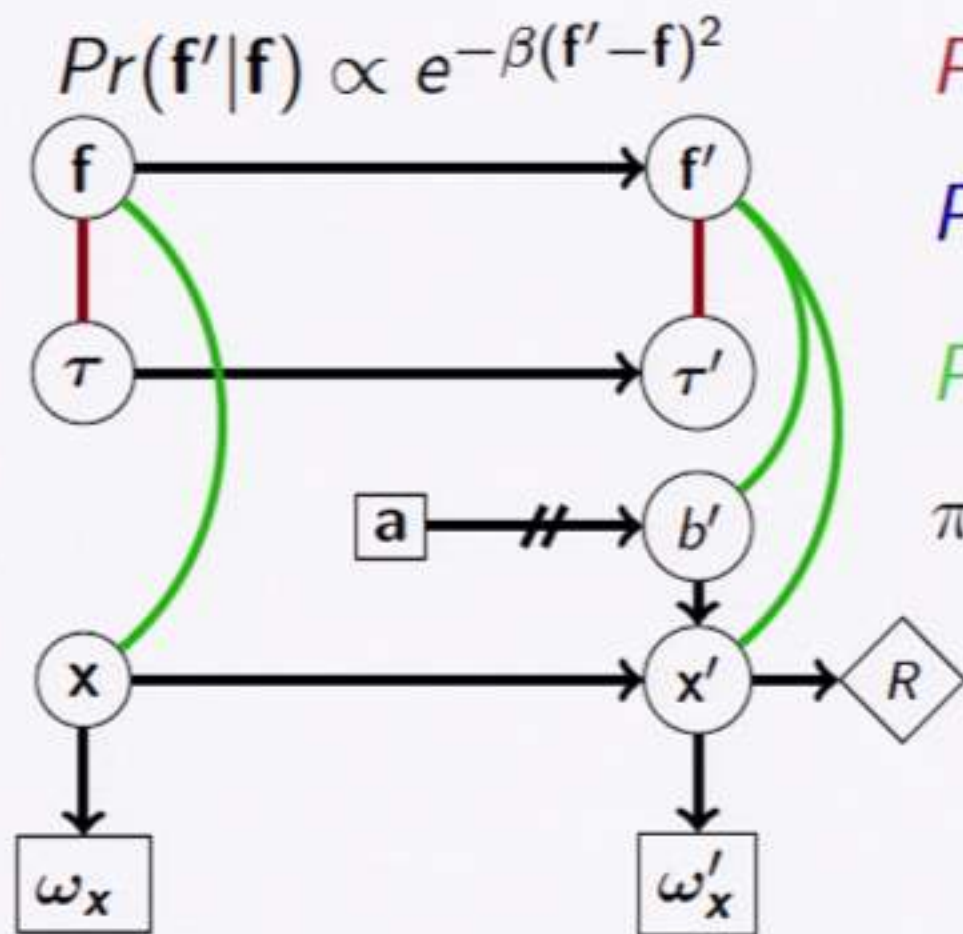


- $Pr(x'|x, a)$
- $Pr(\omega_x|x)$
- $R(x')$



- $Pr(x'|x, a)$
- $Pr(\omega_x|x)$
- $R(x')$

BayesACT



$$Pr(\mathbf{f}'|\mathbf{f}) \propto e^{-\beta(\mathbf{f}'-\mathbf{f})^2}$$

$$Pr(\mathbf{f}', \tau') \propto e^{-\alpha(\mathbf{f}'-\tau')^2}$$

$$Pr(\tau'|\tau, \mathbf{f}') = \delta(\tau' - M\mathcal{G}(\mathbf{f}', \tau))$$

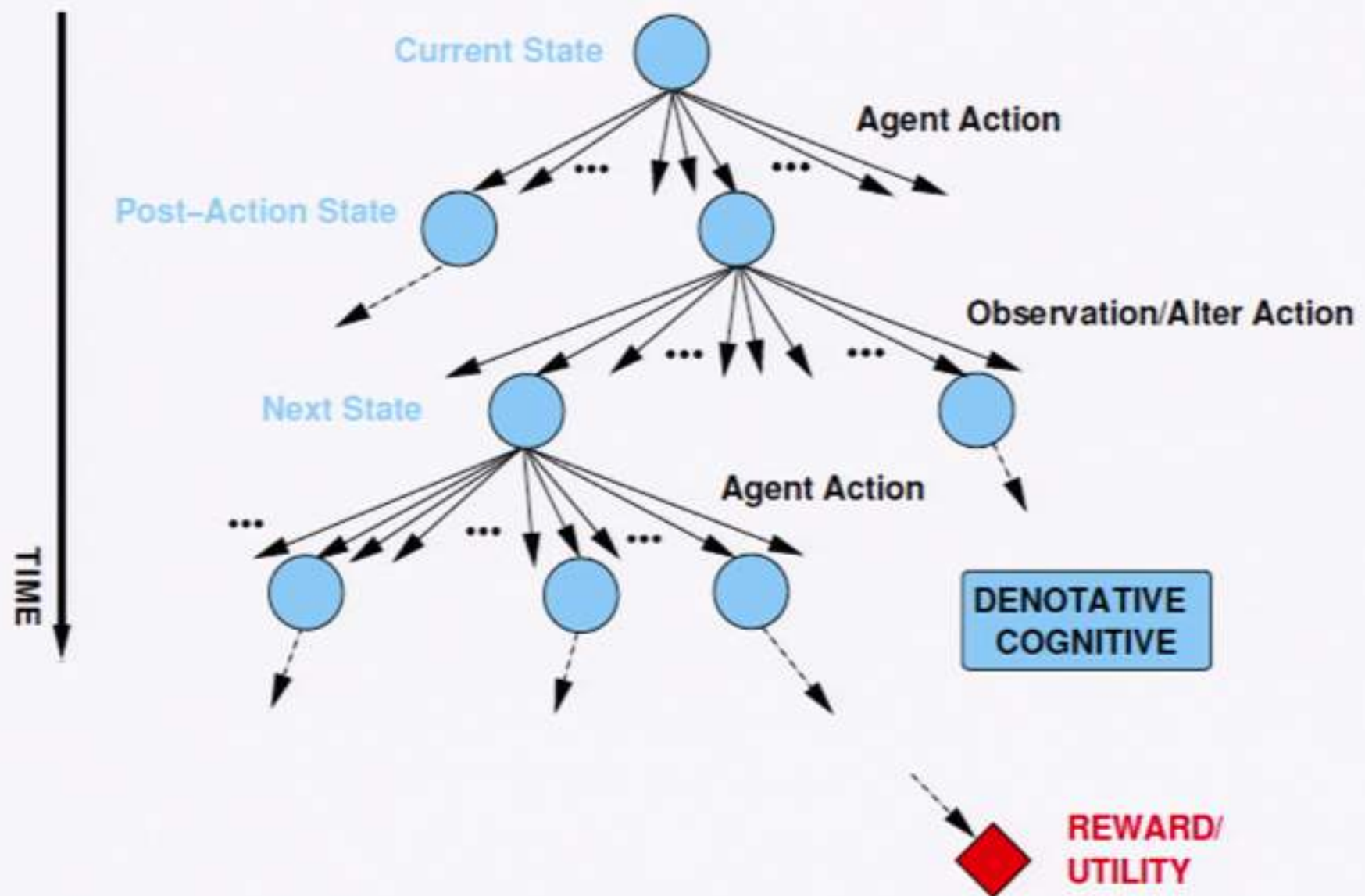
$$Pr(\mathbf{f}', \mathbf{x}') \propto \hat{\mathbf{G}}(\mathbf{f}', \{\mathbf{x}', b'\})$$

$$\pi^\dagger(\mathbf{f}'_b) = \int_{\mathbf{f}'_a, \mathbf{f}'_c} \int_{\mathbf{s}} Pr(\mathbf{f}'|\mathbf{f}, \tau, \dots) b(\mathbf{s})$$

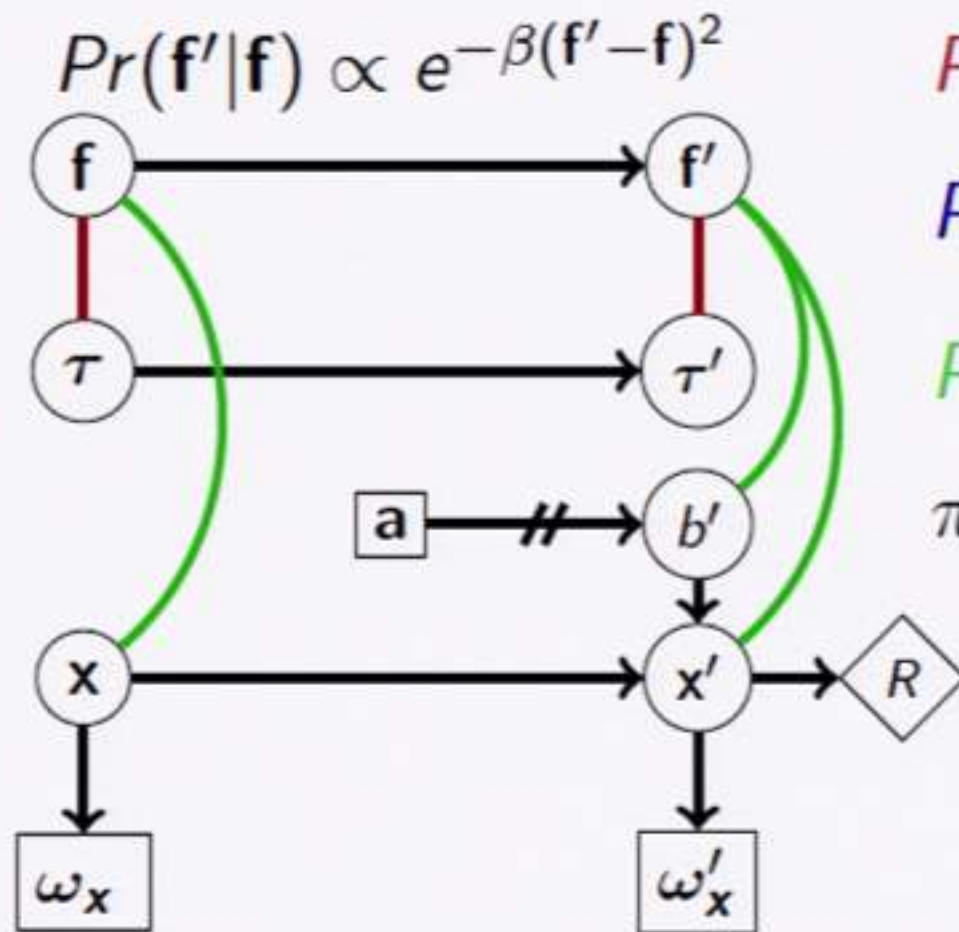
- $Pr(\mathbf{x}'|\mathbf{x}, \mathbf{a})$
- $Pr(\omega_x|\mathbf{x})$
- $R(\mathbf{x}')$

- β sentiment inertia
- α **Affect Control Principle**
- $M\mathcal{G}$ Impression Formation
- $\hat{\mathbf{G}}(\mathbf{y}, \{\mathbf{x}, b'\})$ **Somatic Potential**

Artificial Intelligence: Affect Control Theoretic



BayesACT



$$Pr(\mathbf{f}'|\mathbf{f}) \propto e^{-\beta(\mathbf{f}'-\mathbf{f})^2}$$

$$Pr(\mathbf{f}', \tau') \propto e^{-\alpha(\mathbf{f}'-\tau')^2}$$

$$Pr(\tau'|\tau, \mathbf{f}') = \delta(\tau' - M\mathcal{G}(\mathbf{f}', \tau))$$

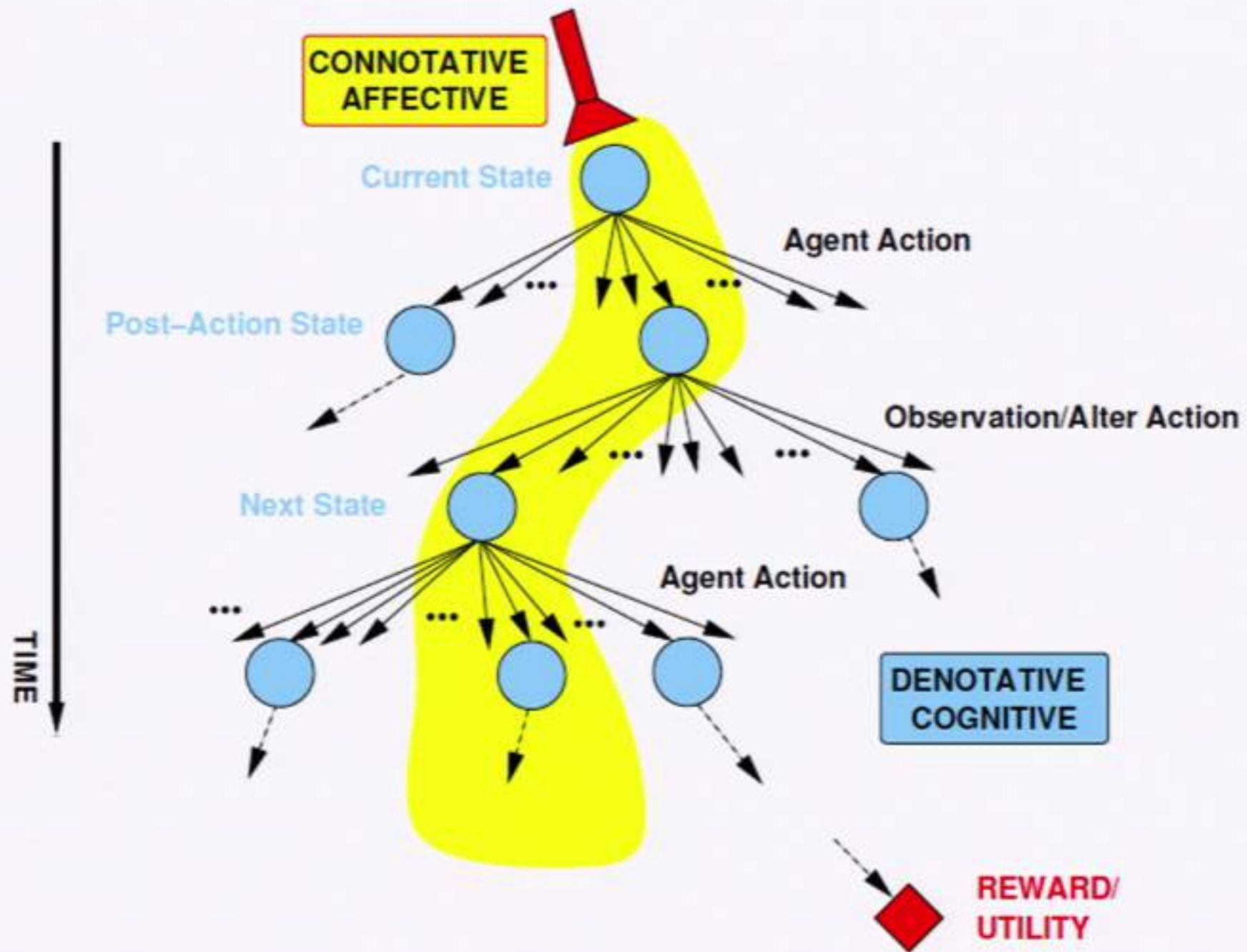
$$Pr(\mathbf{f}', \mathbf{x}') \propto \hat{\mathbf{G}}(\mathbf{f}', \{\mathbf{x}', b'\})$$

$$\pi^\dagger(\mathbf{f}'_b) = \int_{\mathbf{f}'_a, \mathbf{f}'_c} \int_{\mathbf{s}} Pr(\mathbf{f}'|\mathbf{f}, \tau, \dots) b(\mathbf{s})$$

- $Pr(\mathbf{x}'|\mathbf{x}, \mathbf{a})$
- $Pr(\omega_x|\mathbf{x})$
- $R(\mathbf{x}')$

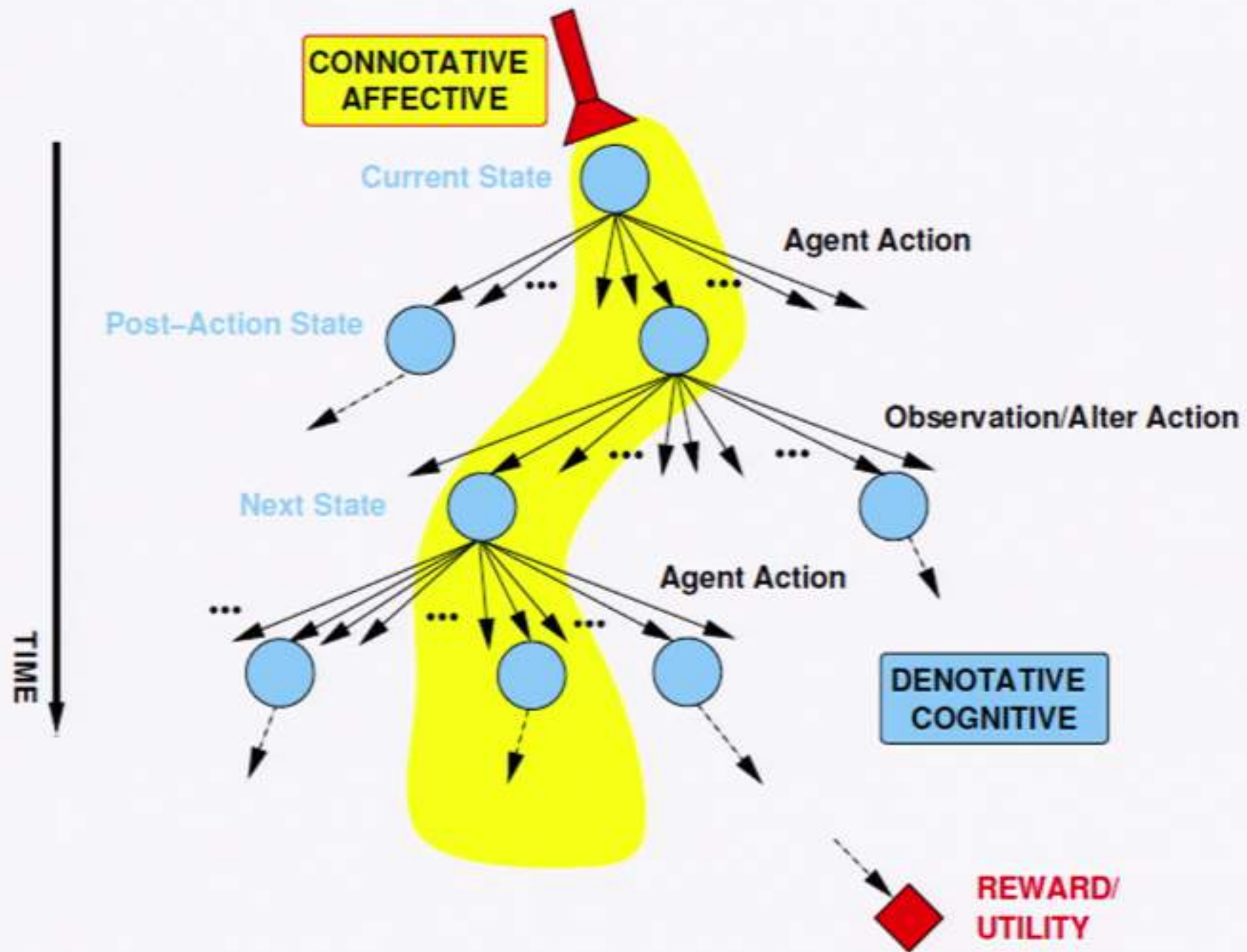
- β sentiment inertia
- α **Affect Control Principle**
- $M\mathcal{G}$ Impression Formation
- $\hat{\mathbf{G}}(\mathbf{y}, \{\mathbf{x}, b'\})$ **Somatic Potential**

Artificial Intelligence: Affect Control Theoretic



"...responses of the *habitus* may be accompanied by a strategic calculation... but these calculations are first defined without any calculation, in relation to objective potentialities, immediately inscribed in the present, things to do or not to do, things to say or not to say..." – Pierre Bourdieu, *The Logic of Practice*, 1990.

Artificial Intelligence: Affect Control Theoretic

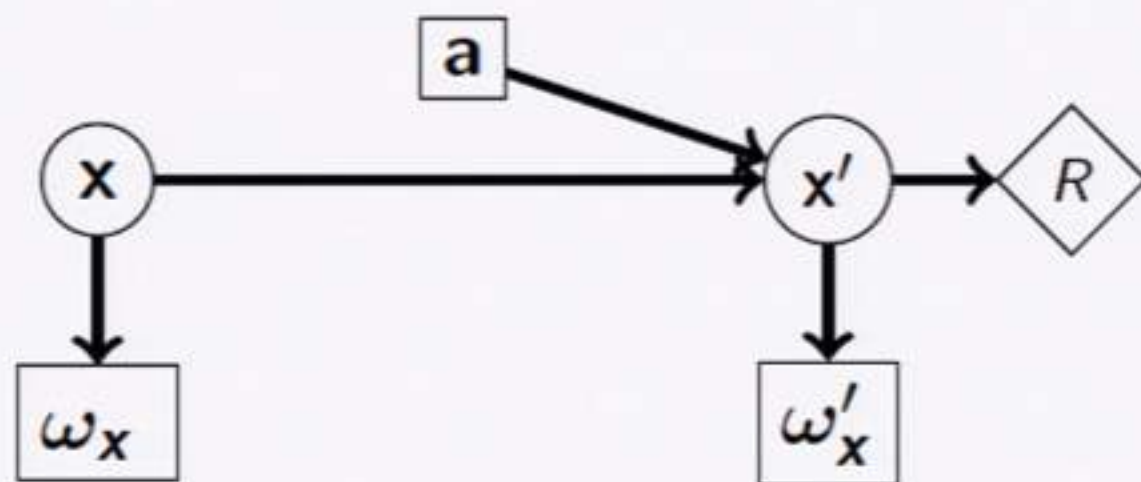


"...responses of the *habitus* may be accompanied by a strategic calculation... but these calculations are first defined without any calculation, in relation to objective potentialities, immediately inscribed in the present, things to do or not to do, things to say or not to say..." – Pierre Bourdieu, *The Logic of Practice*, 1990.

Affect Control Theory

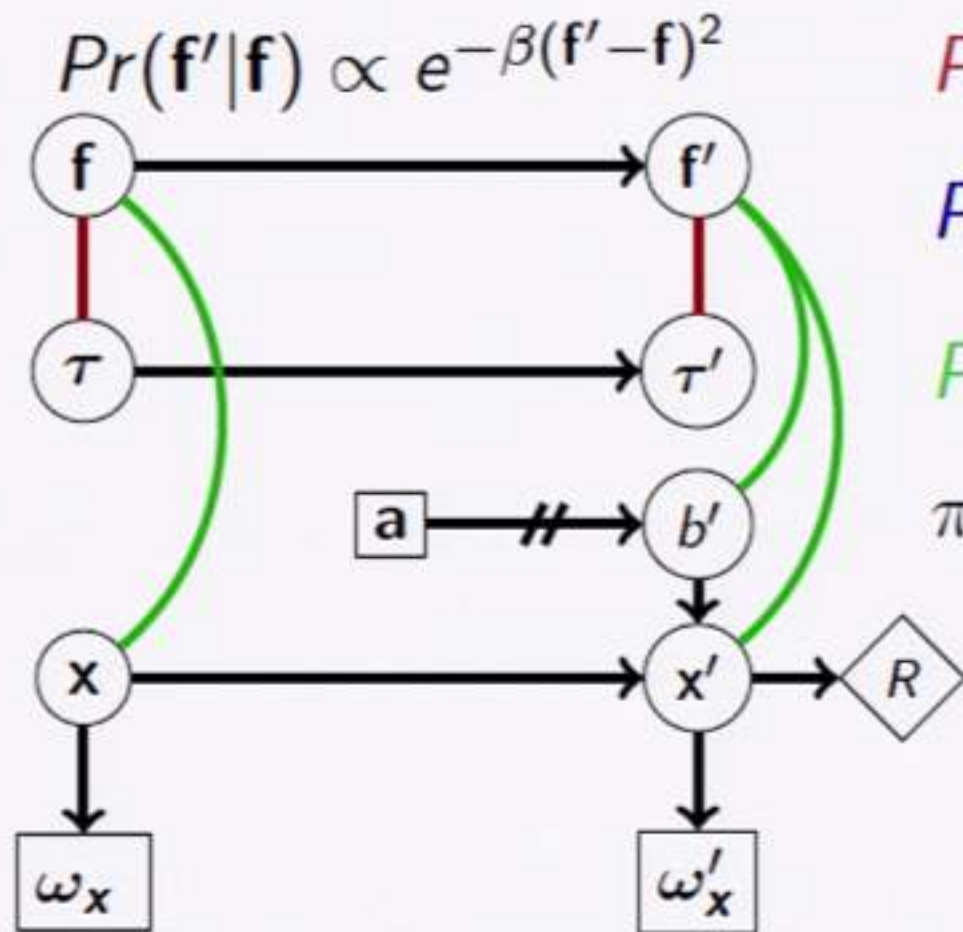
- Actor-Behaviour-Object
- fundamental sentiments: $\mathbf{F} \in [-4.3, 4.3]^9$
- transient impressions: $\mathbf{T} \in [-4.3, 4.3]^9$
- prediction $\mathbf{T}_{t+1} = \mathbf{M}^{\mathcal{G}}(\mathbf{F}_t, \mathbf{T}_t)$ **measured empirically**
- deflection $D = \sum_i w_i (f_i - \tau_i)^2$
- **Affect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

Partially Observable Markov Decision Process



- $Pr(x'|x, a)$
- $Pr(\omega_x|x)$
- $R(x')$

BayesACT



$$Pr(\mathbf{f}'|\mathbf{f}) \propto e^{-\beta(\mathbf{f}'-\mathbf{f})^2}$$

$$Pr(\mathbf{f}', \tau') \propto e^{-\alpha(\mathbf{f}'-\tau')^2}$$

$$Pr(\tau'|\tau, \mathbf{f}') = \delta(\tau' - M\mathcal{G}(\mathbf{f}', \tau))$$

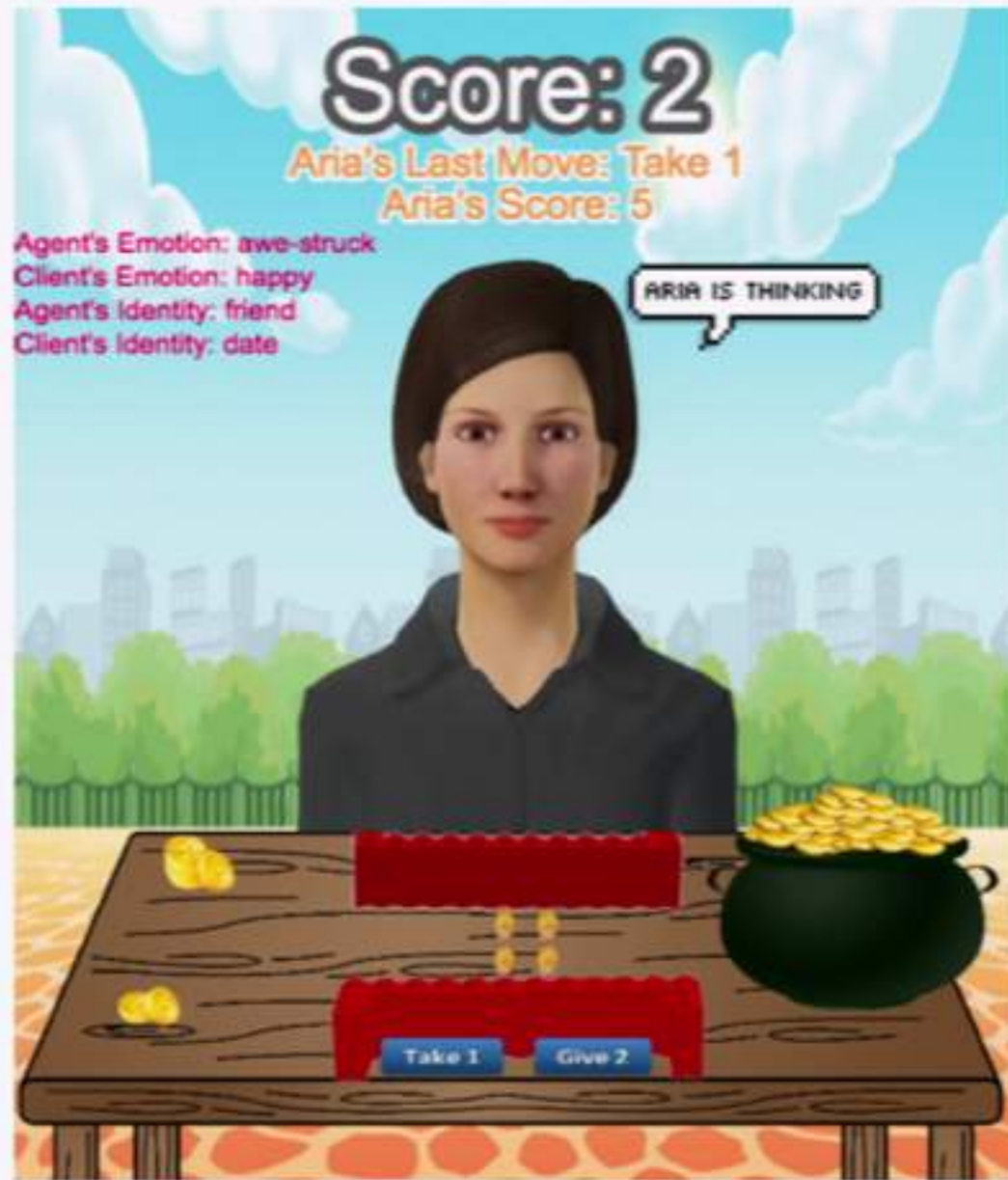
$$Pr(\mathbf{f}', \mathbf{x}') \propto \hat{\mathbf{G}}(\mathbf{f}', \{\mathbf{x}', b'\})$$

$$\pi^\dagger(\mathbf{f}'_b) = \int_{\mathbf{f}'_a, \mathbf{f}'_c} \int_{\mathbf{s}} Pr(\mathbf{f}'|\mathbf{f}, \tau, \dots) b(\mathbf{s})$$

- $Pr(\mathbf{x}'|\mathbf{x}, \mathbf{a})$
- $Pr(\omega_x|\mathbf{x})$
- $R(\mathbf{x}')$

- β sentiment inertia
- α **Affect Control Principle**
- $M\mathcal{G}$ Impression Formation
- $\hat{\mathbf{G}}(\mathbf{y}, \{\mathbf{x}, b'\})$ **Somatic Potential**

Homo Economicus and the Prisoner's Dilemma



	C	D
C	2,2	0,3
D	3,0	1,1

THEMIS.COGE: Social Programming Networks



SSHRC CRSH



DFG Deutsche Forschungsgemeinschaft



Tobias Schröder
Potsdam

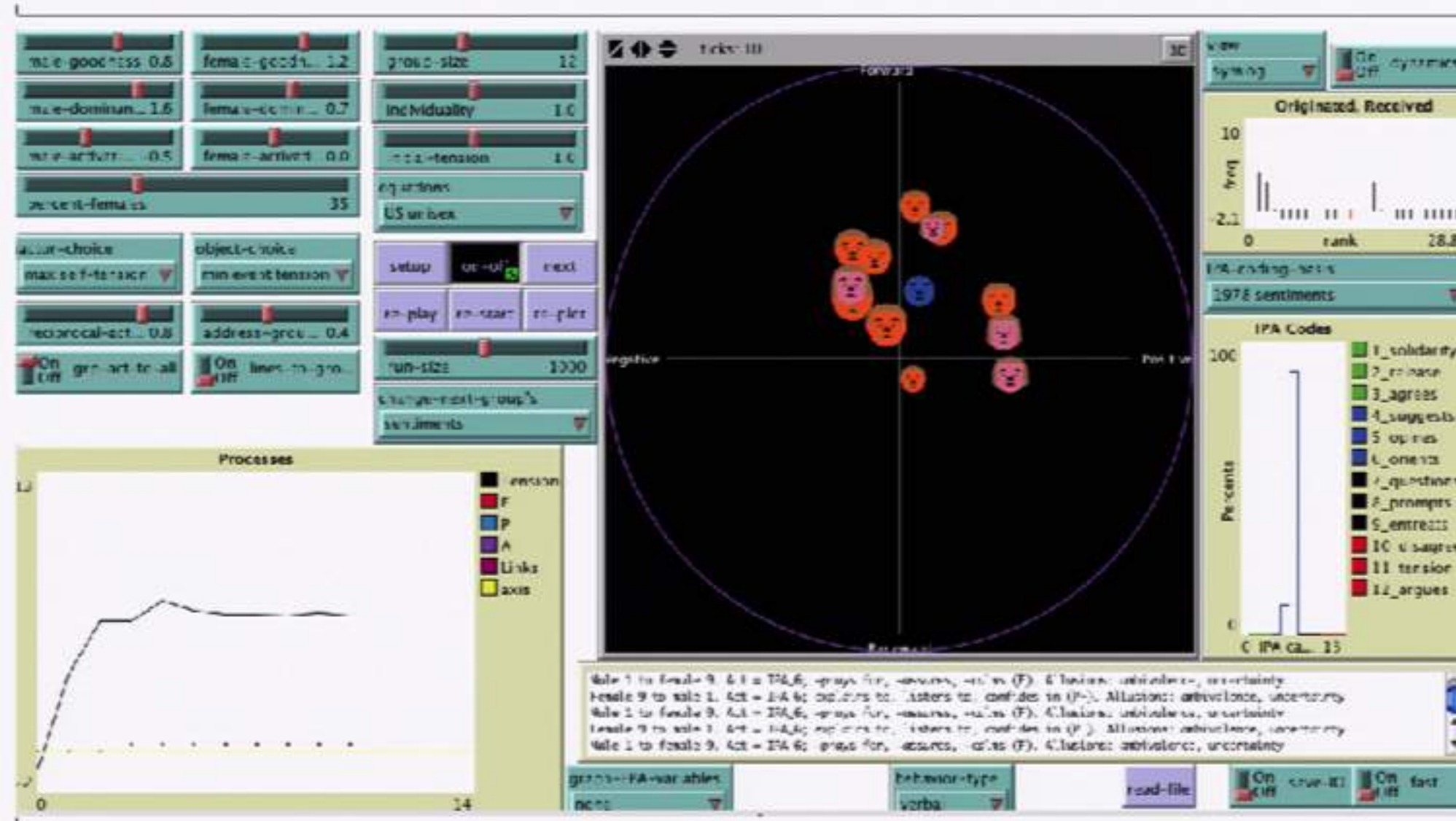


Kimberly B. Rogers
Dartmouth



Mei Nagappan
Waterloo

Group Simulator



<http://www.indiana.edu/~socpsy/ACT/SmallGroups/GroupSimulator.html>

Bottom-up (Data-Driven) analysis

Data driven analysis of group behaviour

Datasets:

Themis-EPA

500 E,P,A concepts
Rated by 50
GitHub developers

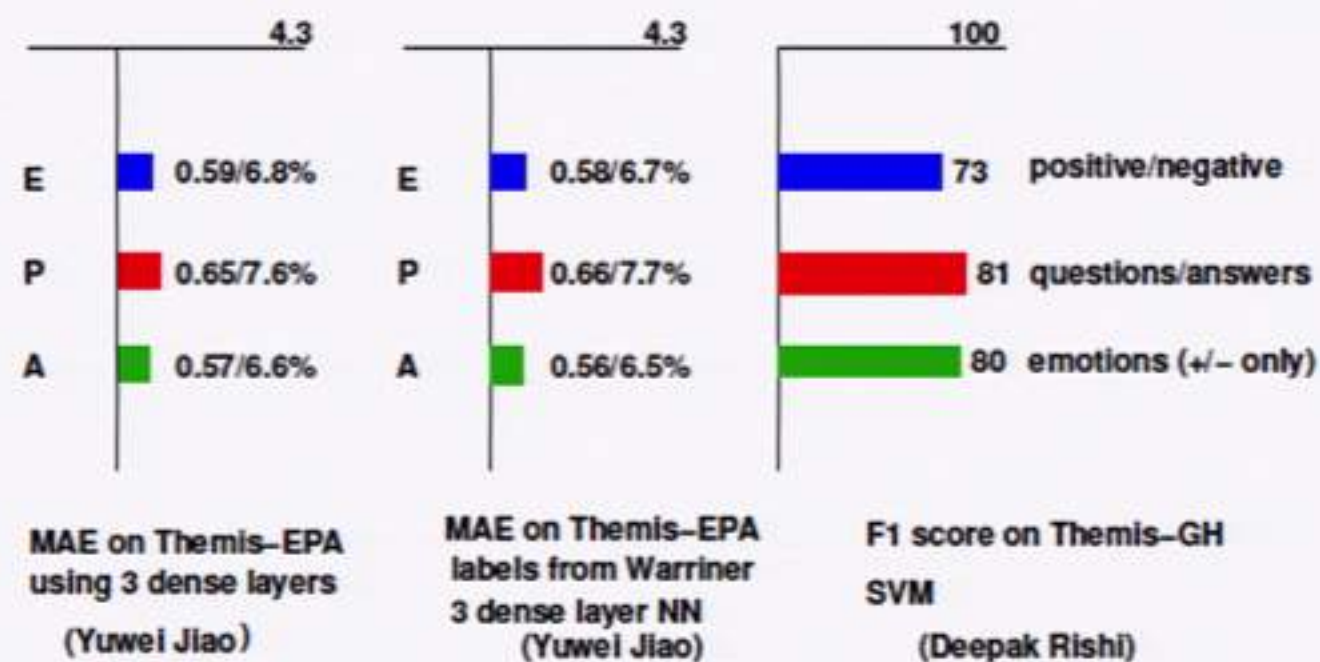
GitHub-WV

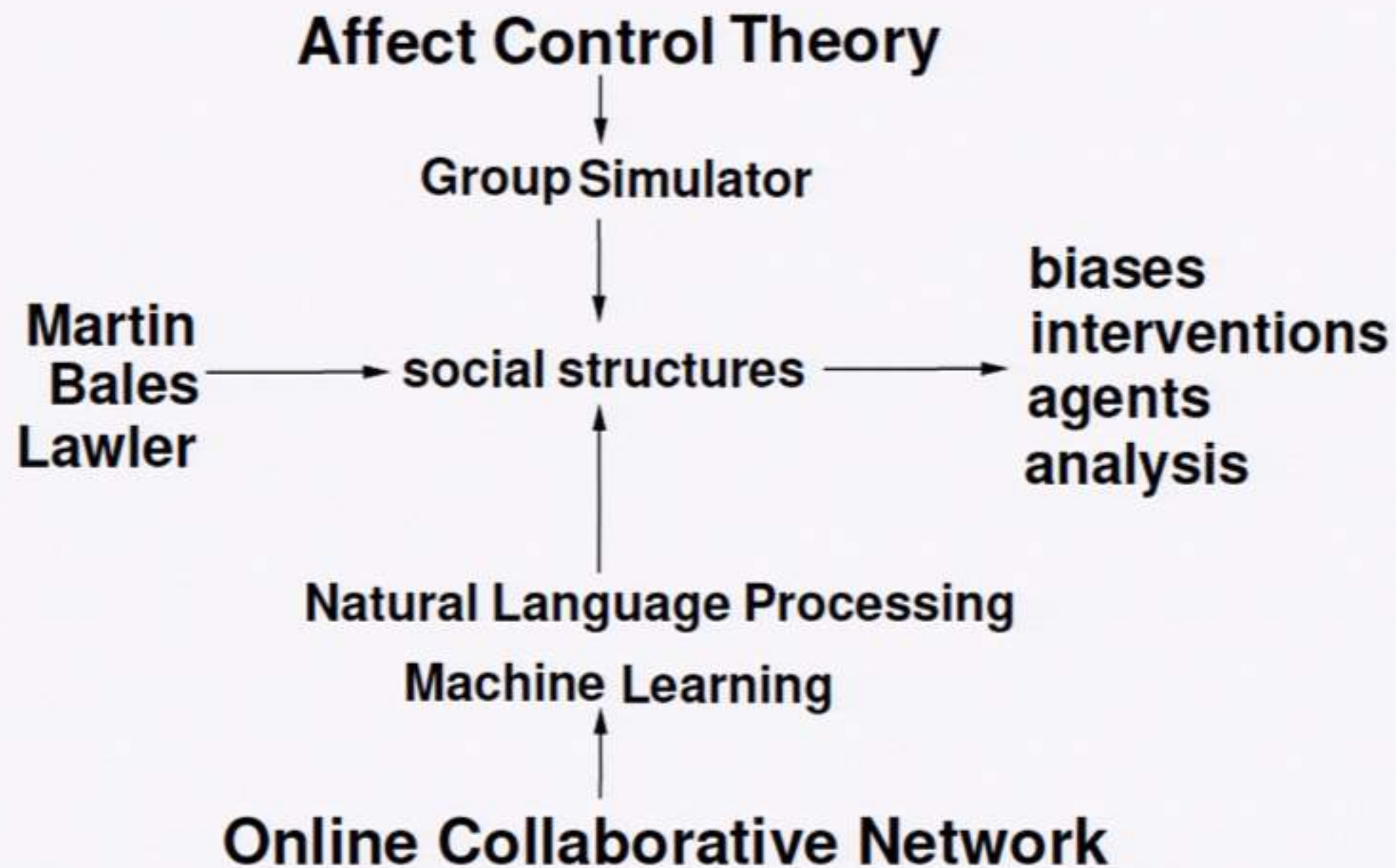
4124 repositories
53K pull requests
1.5m issue comments
1.9m commit
comments
word vectors

Themis-GH

834 pull requests
3000 pull request
comments
rated for IPA and
emotion words

Results:





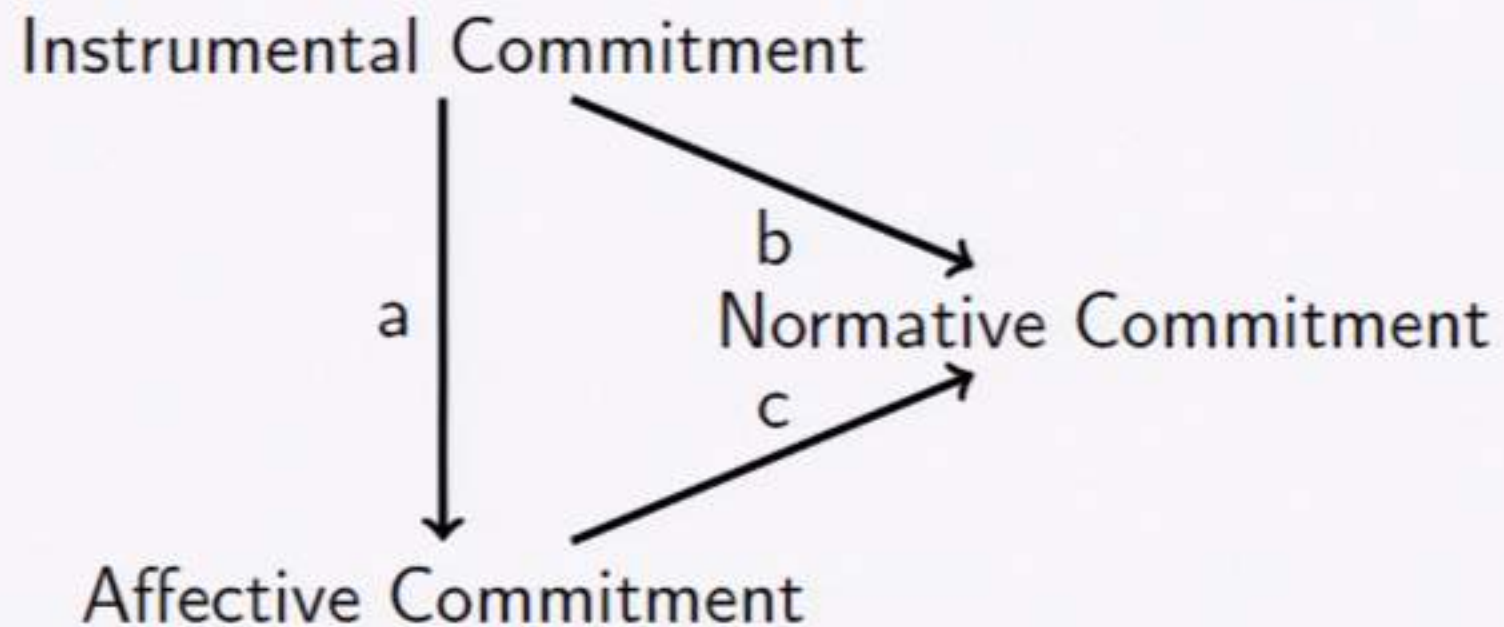
Emotions: the new AI

- 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?*

Forms of Commitment



Affective ties:

- arise as a by-product of instrumental conditions (a)
- **strengthen bonds** between agents → longer lasting
- increases **trust** between agents → shorter contracts
- allow agents to be more **efficient**, regardless of available resources

From: Edward J. Lawler, Shane R. Thye and Jeongkoo Yoon. *Social Commitments in a Depersonalized World*. Russell Sage Foundation, 2009.

Funding:

- Trans-Atlantic Partnership (TAP)
- Cheriton, Graham Trust Faculty Fellowships
- Natural Science and Engineering Research Council (Canada)
- Social Science and Humanities Research Council (Canada)
- National Science Foundation (USA)
- Deutsche Forschungsgemeinschaft (Germany)

More Information:

- THEMIS.COG: themis-cog.ca
- Bayesian Affect Control Theory: bayesact.ca
- Jesse Hoey:
 - ▶ jhoey@cs.uwaterloo.ca
 - ▶ [@drjessehoey](https://twitter.com/drjessehoey)

