Intelligent Affect: Rational Decision Making for Socially Aligned Agents

UNIVERSITY OF WATERLOO

Nabiha Asghar and Jesse Hoey

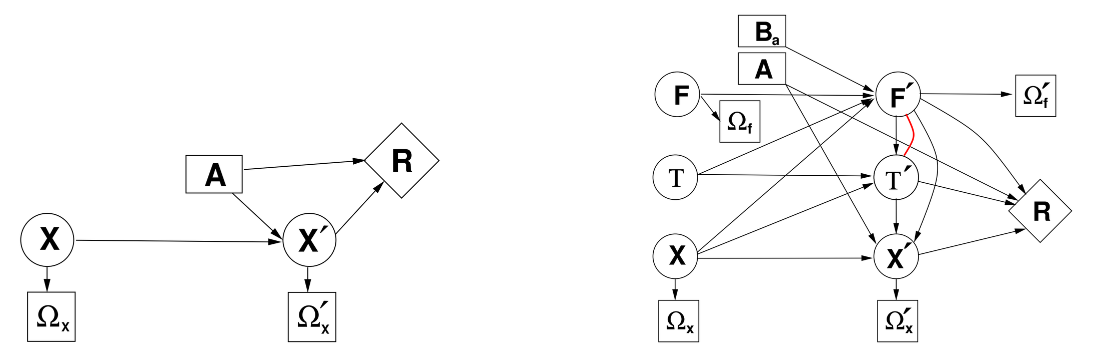
David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada

WATERLOO CHERITON SCHOOL OF COMPUTER SCIENCE

Introduction

- Affect Control Theory (ACT)[2]:
  - sociological model of human interaction
  - humans have shared cultural *sentiments* about identities, behaviours, and interaction dynamics
  - cultural consistency "gestalt" is a keystone of intelligence
  - used to make predictions of other's behaviours,
  - and to guide action choices for an agent,
- ACT proposes affective prescriptions for action:
  - results in affective ecosystem of roles and behaviours,
  - an equilibrium that yields a social order [1],
  - "System 1" thinking [6].
- Bayesian Affect Control Theory (*BayesAct*)[5]:
  - sentiments are probability distributions
  - propositional (non-affective) states
  - explicit utility function
- Planning in *BayesAct* away from the cultural norm:
  - ACT prescriptions as affective heuristic:
    - guides search for beneficial, yet affectively appropriate, actions
    - bounds required resources and implicitly solves social dilemmas
  - planning yields individually beneficial solutions
  - that are still (somewhat) culturally acceptable
  - allows for manipulation (deception and altercasting)
  - "System 2" thinking [6]

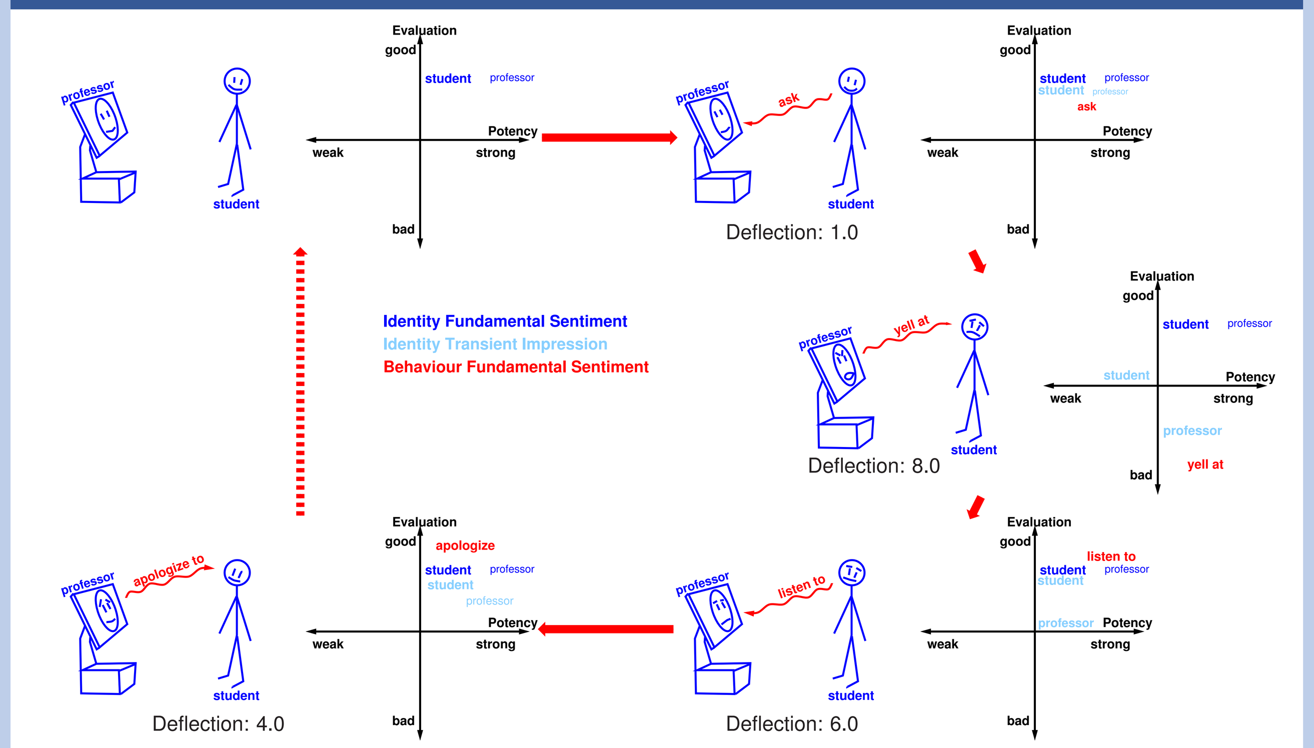
Partially Observable Markov Decision Process



General POMDP — POMDP for *BayesAct*

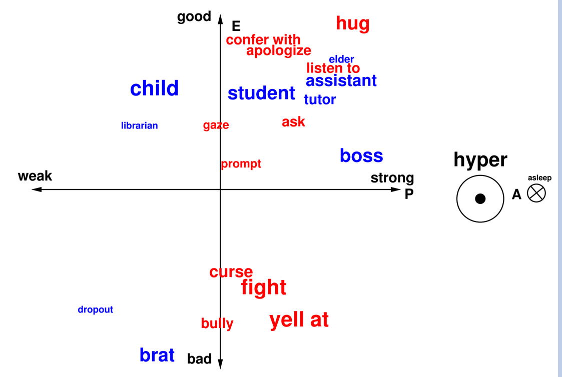
- a policy maps belief states into actions, such that the expected discounted sum of rewards is maximised
- used for human-interactive domains (see [4])

Affect Control Theory Example



EPA Space [7]

- 3-D EPA space [7]
- Evaluation, Potency, Activity
- shared sentiments across a cultural group
- universal organising principle of human socio-affective experience
- is compatible with appraisal theories [8]: goal congruence of an event (E), the agent's coping potential (P), and the urgency (A)



Affect Control Theory (ACT) [2]

- Actor-Behaviour-Object (A, B, O) Grammar
- shared fundamental sentiments $(\forall A, B, O)$: $\mathbf{F} \in [-4.3, 4.3]^9$
- transient impressions created by events $A - B - O$ $(\forall A, B, O)$: $\mathbf{T} \in [-4.3, 4.3]^9$
- deflection $D = \sum_i w_i(\mathrm{f}_i - \tau_i)^2$
- prediction $\mathbf{T}_{t+1} = \boldsymbol{M}\mathscr{G}(\mathbf{F}_t, \mathbf{T}_t)$
- $\mathbf{F}, \boldsymbol{M}, \mathscr{G}$: measured empirically [3]

Affect Control Principle: actors work to experience transient impressions that are consistent with their fundamental sentiments

BayesACT [5]

- fundamental sentiments $\mathbf{F} = \{\mathrm{F}_{ij}\}$ where $\mathrm{F}_{ij}, i \in \{a, b, c\}, j \in \{e, p, a\}$
- transient impressions $\mathbf{T} = \{\mathrm{T}_{ij}\}$
- application states $\mathbf{X}$
- actions: affective ($\mathbf{b_a}$) and cognitive (a)
- POMDP with $Pr(\mathbf{s}'|\mathbf{s}, \mathbf{b_a}, a) =$ $Pr(\tau'|\tau, \mathbf{f}', \mathbf{x})Pr(\mathbf{f}'|\mathbf{f}, \tau, \mathbf{x}, \mathbf{b_a})Pr(\mathbf{x}'|\mathbf{x}, \mathbf{f}', \tau', a)$
- transient dynamics $Pr(\tau'|\tau, \mathbf{f}', \mathbf{x}) = \delta(\tau' - \boldsymbol{M}\mathscr{G}(\mathbf{f}', \tau, \mathbf{x}))$
- affect control potential $\varphi(\mathbf{f}', \tau') \propto e^{-(\mathbf{f}'-\tau')^T\Sigma^{-1}(\mathbf{f}'-\tau')}$
- reward function $R(\mathbf{f}, \tau, \mathbf{x}) = R_x(\mathbf{x}) + R_s(\mathbf{f}, \tau)$ combines application goals and deflection minimizing goal
- application dynamics $Pr(\mathbf{x}'|\mathbf{x}, \mathbf{f}', \tau', a)$
- observation functions $Pr(\omega_f|\mathbf{f}), Pr(\omega_x|\mathbf{x})$

generalisation of the affect control principle:

$$\psi(\mathbf{f}', \tau, \mathbf{x}) = (\mathbf{f}' - \boldsymbol{M}(\mathbf{x})\mathscr{G}(\mathbf{f}', \tau, \mathbf{x}))^T\Sigma^{-1}(\mathbf{f}' - \boldsymbol{M}(\mathbf{x})\mathscr{G}(\mathbf{f}', \tau, \mathbf{x}))$$

affective "inertia":

$$\xi(\mathbf{f}', \mathbf{f}, \mathbf{b_a}, \mathbf{x}) \equiv (\mathbf{f}' - \langle\mathbf{f}, \mathbf{b_a}\rangle)^T\Sigma_f^{-1}(\mathbf{x})(\mathbf{f}' - \langle\mathbf{f}, \mathbf{b_a}\rangle)$$

fundamental dynamics:

$$Pr(\mathbf{f}'|\mathbf{f}, \tau, \mathbf{x}, \mathbf{b_a}, \varphi) \propto e^{-\psi(\mathbf{f}', \tau, \mathbf{x}) - \xi(\mathbf{f}', \mathbf{f}, \mathbf{b_a}, \mathbf{x})}$$

Bayesact Instances

Two key elements for each application domain

- Normative Action Bias (NAB)
  Gives the affective prescription (norm)
  $\pi^\dagger(\mathbf{f}_b') = \int_{\mathbf{f}_a', \mathbf{f}_c'} \int_{\mathbf{s}} Pr(\mathbf{f}'|\mathbf{f}, \tau, \mathbf{x})b(\mathbf{s})$
  $\mathbf{b_a} = \arg\max_{\mathbf{f}_b'} \pi^\dagger(\mathbf{f}_b')$
- Social Coordination Bias (SCB)
  Defines $Pr(\mathbf{x}'|\mathbf{f}', \tau', \mathbf{x}, a)$
  Describes how we expect the state to change in a given relationship (identities)

The Normative Action Bias gives a mechanism for relaying identities, while the Social Coordination Bias allows agents to predict actions *given* identities.

POMCP-C

POMCP: Monte-Carlo Tree Search method for solving large POMDPs[9]

POMCP-C:

- extends POMCP to work with continuous action and observation spaces
- yields affective prescriptions quickly using NAB: $\pi^\dagger$
- computes rational actions more slowly

Procedure SEARCH(B^*, h)
repeat
  if $h = \emptyset$ **then**
    $s \sim B^*$
  else
    $s \sim B(h)$
  end
  SIMULATE($s, h, 0$)
until *TIMEOUT*()
return $\arg\max_b V(hb)$

Function DiscretizeObs(o, h)
if $\exists_{\mathbf{o_j} \in \mathbf{o}(h)} : |o - \bar{\mathbf{o}}_\mathbf{j}| < \delta_o$ **then**
  $\mathbf{o_j} \leftarrow \mathbf{o_j} \cup \{o\}$
  return $\bar{\mathbf{o}}_\mathbf{j}$
else
  $\mathbf{o}(h) \leftarrow \mathbf{o}(h) \cup \{\{o\}\}$
  return o
end

Procedure PruneTree(h, a, o)
$i^* \leftarrow \arg\min_i |a - a_i(h)|$
$j^* \leftarrow \arg\min_j |o - \bar{\mathbf{o}}_j|$
$T \leftarrow T(hi^*j^*)$

Procedure SIMULATE(s, h, d)
if $\gamma^d < \epsilon$ **then**
  return *0*
end
if $N_A(h) < N_A^{max}$ **then**
  $a \sim \pi^\dagger(s)$
  if $\mathbf{a}(h) = \emptyset \vee \forall_{a_j \in \mathbf{a}(h)}|a - a_j| > \delta_a$ **then**
    $i \leftarrow N_A(h)$
    $T(hi) \leftarrow (N_{init}(hi), V_{init}(hi), \emptyset)$
    $N_A(h) \leftarrow N_A(h) + 1$
    $\mathbf{a}(h) \leftarrow \mathbf{a}(h) \cup \{a\}$
    return ROLLOUT*(s,h,d)*
  end
end
$i \leftarrow \arg\max_{j=1...N_A(h)} V(hj) + c\sqrt{\frac{\log N(h)}{N(hj)}}$
$(s', o, r) \sim \mathcal{G}(s, a_i(h))$
$o^\dagger \leftarrow$ DiscretizeObs(o, h)
$R \leftarrow r + \gamma$.SIMULATE($s', ha_i(h)o^\dagger, d + 1$)
$B(h) \leftarrow B(h) \cup \{s\}$
$N(h) \leftarrow N(h) + 1$
$N(hi) \leftarrow N(hi) + 1$
$V(hi) \leftarrow V(hi) + \frac{R - V(hi)}{N(hi)}$
return R

POMCP-C is *anytime*, epitomizing *fast* vs. *slow* thinking, blending *affective* and *cognitive* reasoning.

References

[1] Erving Goffman. *Behavior in Public Places*. The Free Press, New York, 1963.

[2] David R. Heise. *Expressive Order: Confirming Sentiments in Social Actions*. Springer, 2007.

[3] David R. Heise. *Surveying Cultures: Discovering Shared Conceptions and Sentiments*. Wiley, 2010.

[4] Jesse Hoey, Craig Boutilier, Pascal Poupart, Patrick Olivier, Andrew Monk, and Alex Mihailidis. People, sensors, decisions: Customizable and adaptive technologies for assistance in healthcare. *ACM TIIS*, 2(4), Jan. 2012.

[5] Jesse Hoey, Tobias Schröder, and Areej Alhothali. Bayesian affect control theory. In *Proc. ACII* 2013.

[6] Daniel Kahneman. *Thinking, Fast and Slow*. Doubleday, 2011.

[7] Charles E. Osgood, William H. May, and Murray S. Miron. *Cross-Cultural Universals of Affective Meaning*. University of Illinois Press, 1975.

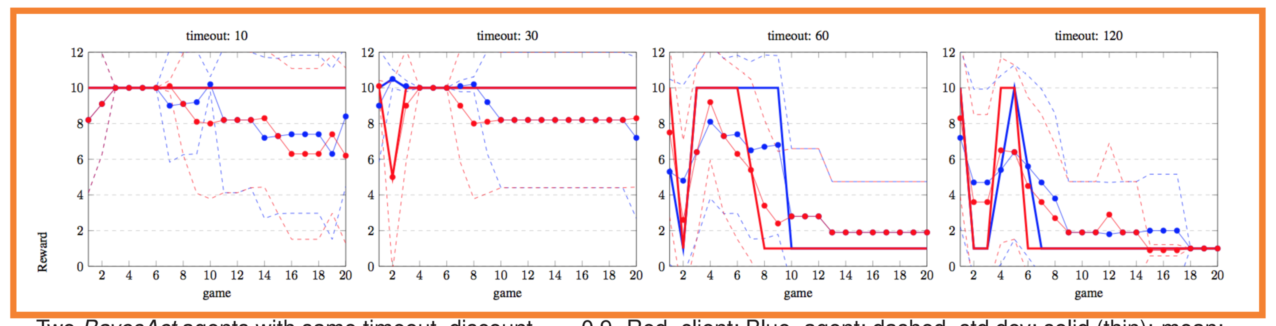
[8] Kimberly B. Rogers, Tobias Schröder, and Christian von Scheve. Dissecting the sociality of emotion: A multi-level approach. *Emotion Review*, 6(2):124–133, 2014.

[9] David Silver and Joel Veness. Monte-Carlo planning in large POMDPs. In *NIPS* 2010.

Prisoner's Dilemma

agent	client	optimal behaviour	closest labels	distance from collborate	distance from abandon
friend	friend	1.98, 1.09, 0.96	treat toast	0.4	23.9
friend	scrooge	0.46, 1.14, -0.27	reform lend money to	1.7	10.5
scrooge	friend	-0.26, -0.81, -0.77	curry favor look away	8.5	4.2
scrooge	scrooge	-0.91, -0.80, -0.01	borrow money chastise	9.6	2.7

-Deflection minimising behaviours for identities: friend (EPA:{2.75, 1.88, 1.38}), scrooge (EPA:{-2.15, -0.21, -0.54}).
-Affective distance from *collaborate* (EPA:{1.44, 1.11, 0.61}) and *abandon* (EPA:{-2.28, -0.48, -0.84}) show what action will be taken in the game based only on affective prescriptions



-Two *BayesAct* agents with same timeout, discount $\gamma = 0.9$. Red=client; Blue=agent; dashed=std.dev; solid (thin): mean; solid (thick): median. As timeout increases, more defections give less reward for both agents.
-When $\gamma = 0.99$, agents always cooperate.
-With more discounting ($\gamma = 0.9$), more time buys more breadth of search (the *agent* gets to explore more short-term options), and finds more of them that look appealing (it can get away with a defection for a short while).
-With less discounting ($\gamma = 0.99$, not shown), more time buys more depth, and results in better long-term decisions.

game #	post-play sentiments (*agent*) $\mathbf{f}_a$	$\mathbf{f}_c$	$\mathbf{f}_b$	deflection	identities agent	identities client	emotions agent	emotions client	actions agent	actions client
1	-1.36,-0.01,-0.35	2.32,1.61,1.27	2.62,1.58,1.73	4.44	failure	newlywed	easygoing	idealistic	cooperate	cooperate
2	-0.66,0.04,-0.05	1.77,1.27,1.06	2.23,1.00,1.76	3.70	parolee	husband	easygoing	self-conscious	cooperate	cooperate
3	-0.23,-0.08,0.20	1.02,0.93,0.84	2.49,0.97,1.87	7.19	stepmother	purchaser	female	immoral	cooperate	defect
4	-0.12,-0.33,0.33	0.27,0.62,0.62	2.37,0.48,1.34	4.99	stuffed_shirt	roommate	dependent	unfair	cooperate	defect
5	-0.26,-0.47,0.32	-0.26,0.26,0.42	-0.59,0.41,-0.23	3.27	divorcée	gun_moll	dependent	selfish	defect	defect
6	-0.37,-0.66,0.26	-0.61,0.00,0.28	-0.10,-0.41,-0.27	2.29	divorcée	hussy	disapproving	selfish	defect	defect

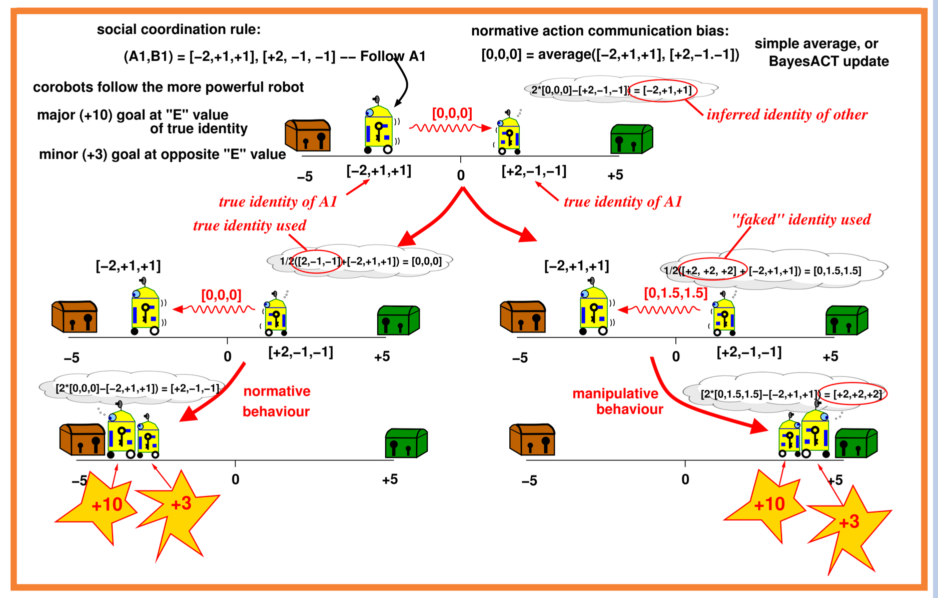
Example games with *client* playing two-out. Identities and emotions are *agent* interpretations.

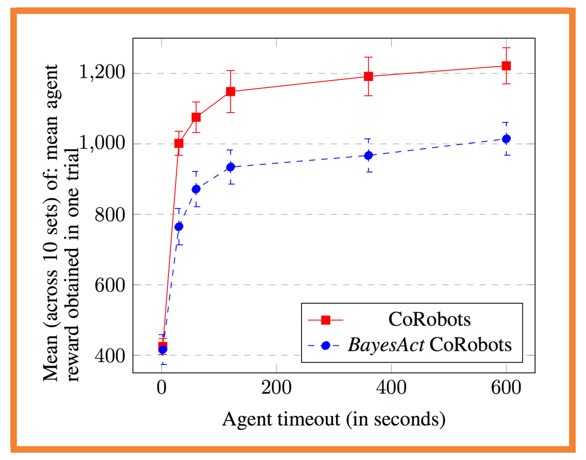
γ	(tt)	(t2)	(2t)
0.9	1.64 ± 2.24	3.98 ± 2.48	1.72 ± 2.35
0.99	7.33 ± 1.17	7.28 ± 1.68	7.63 ± 0.91

-Results (avg. rewards) against tit-for-tat (tt), tit-for-two-tat (t2), two-tit-for-tat (2t), with discount factor γ.
-Less discounting leads to better solutions against these strategies, as longer-term solutions are found.

Corobots

- *Battle of the sexes*: a husband wants to go to a football game, but his wife wants to go shopping, and neither wants to be alone.
- Implemented as two *corobots*, with goal locations dependent on *identity* "E" value.
- Social coordination: follow more powerful, active robot.
- Normative Action Bias: either simple mean (shown on right) or *BayesAct* update.
- *Corobots* must coordinate, but may manipulate: the weaker agent may "fake" a more powerful identity in order to convince the other to follow her.





- As asymmetry in planning resources increases, manipulations and reward increase for the agent with more resource.
- upper bound (perfect manipulation): ~ 2000.

Conclusion

- Decision-theoretic planning in a POMDP model of affective interaction, *BayesAct*.
- Unifies cognitive (individual, "System 2") and affective (social, "System 1") reasoning.
- Agents search for actions close to socio-cultural prescriptions in affective "EPA" space.
- MCTS yields realistic solutions to classic social dilemmas.
- Agents with more resources manipulate for personal gain.