

SHAME: WHEN EMOTION AND REASONING ARE LINKED

EUMAS 2013
12-13 December, Toulouse

Carole Adam

UJF-LIG/MAGMA, Grenoble (France)
Carole.Adam@imag.fr

Dominique Longin

CNRS-IRIT/LILaC, Toulouse (France)
Dominique.Longin@irit.fr

Why a study of shame?

- **Complex** emotion, with vague / **ambiguous** / contradicting definitions
 - Interesting (and challenging) to formalise and disambiguate
- Realism and expressivity for single agents
 - Expression of shame in appropriate situations
 - Applications: virtual worlds for training or entertainment (games, serious games)
- Essential to account for during interactions with other agents/humans
 - Understand shame felt by another agent
 - Reason about its causes, react appropriately
 - Application: intelligent tutors (strategies to remove some inhibitions, etc)
- Basis of many human social behaviours (similarly to guilt)
 - Model shame to understand social behaviours
 - Application: multi-agent based social simulation
 - Model the dynamic role of shame in human groups to reproduce it in MAS
 - Application: shame-based cooperation and cohesion in MAS

What definition of shame?

- Castelfranchi and Poggi's cognitive analysis
- Agent *i* feels shame about a fact *F* in front of an agent *j* if the following four conditions hold:
 - (1) agent *i* believes that *j* believes that *F* is true
 - (2) agent *i* believes that *j* believes that if *F* is true then agent *i* is negatively appraised w.r.t. a certain criterion *C*
 - (3) agent *i* believes that *i* and *j* commonly believe that the criterion *C* is a shared normative standard for them both
 - (4) agent *i* is not indifferent to *j*'s opinion of him w.r.t. *C*. (i.e. agent *i* prefers *j* to have a positive opinion of him with respect to *C*, i.e. to believe that he has property *C*)

Definition of shame, illustrated

- Castelfranchi and Poggi's example
- A doctor is ashamed in front of their patient for not knowing a new medicine, making him a bad doctor in this patient's eye if:
 - (1) the doctor believes that his patient believes that he does not know about this new medicine
 - (2) the doctor believes that according to his patient, ignorance of this new medicine makes him a bad doctor
 - (3) the doctor and his patient commonly believe that it is a normative standard to be a good doctor
 - (4) the doctor prefers his patient to think that he is a good doctor

Formal framework

- $MBel_G\varphi$ is read: “ φ is a mutual belief of the agents in group G .”
 - $MBel_{\{i\}}\varphi \equiv Bel_i\varphi$
 - K45 logic for Bel_i
- $Goal_i\varphi$ is read “agent i prefers that φ is true”
 - KD logic
- $NS_i\varphi$ is read “ φ is a normative standard of agent i ”
 - $NS_G\varphi \equiv \bigwedge_{i \in G} NS_i\varphi$
 - KD logic

Definition of shame, formalised

- Castelfranchi and Poggi's four conditions, expressed in our formal framework
- $\text{Shame}_i (G, \varphi, p_i) =$
 - (1) $\text{Bel}_i \text{MBel}_G \varphi \wedge$
 - (2) $\text{Bel}_i \text{MBel}_G (\varphi \rightarrow \neg p_i) \wedge$
 - (3) $\text{Bel}_i \text{MBel}_{G \cup \{i\}} \text{NS}_{G \cup \{i\}} p_i \wedge$
 - (4) $\text{Goal}_i \text{Bel}_j p_i$
- Special cases
 - i is in G : i is ashamed in front of all agents in G , **including** itself
 - i is not in G : i is ashamed only in front of external agents
 - G is reduced to $\{i\}$: i is ashamed only in front of itself

Example of shame, formalised

- A doctor (agent **d**) is ashamed in front of their patient (agent **p**, so group **G** is reduced to **{p}**) for not knowing a new medicine (fact **ignoreMedic** is true), making him a bad doctor (**d** does not have property **goodDoc_d**) in this patient's eye
 - (1) the doctor believes that his patient believes that he does not know about this new medicine
 - **Bel_d Bel_p ignoreMedic ∧**
 - (2) the doctor believes that according to his patient, ignorance of this new medicine makes him a bad doctor
 - **Bel_d Bel_p (ignoreMedic => ~goodDoc_d) ∧**
 - (3) the doctor and his patient commonly believe that it is a normative standard to be a good doctor
 - **Bel_d Mbel_{d,p} NS_{d,p} goodDoc_d ∧**
 - (4) the doctor prefers his patient to think that he is a good doctor
 - **Goal_d Bel_p goodDoc_d**

Shame

and the dynamics of announcements

- $[\varphi!]\psi$ is read “After the announcement of φ , ψ is true”
 - KD logic
- Public announcements allow to express some principles like:
 - The doctor is ashamed before herself for not knowing a new medicine: $Shame\downarrow d (\{d\}, \neg kdNewMedecineP, goodDoctor\downarrow d)$
 - As she believes her patient does not know that she does not know this new medicine, she is not ashamed before her patient: $Bel\downarrow d (\neg Bel\downarrow p kdNewMedecineP \wedge \neg Bel\downarrow p \neg kdNewMedecineP) \wedge \neg Shame\downarrow d (\{d\}, \neg kdNewMedecineP, goodDoctor\downarrow d)$
 - But after the announcement that d does not know the new medicine, the doctor will believe that her patient believes that she does not know this new medicine, and she will be ashamed before herself and her patient

$$[\neg kdNewMedecineP](Bel\downarrow d Bel\downarrow p \neg kdNewMedecineP \wedge Shame\downarrow d (\{d, p\}, \neg kdNewMedecineP, goodDoctor\downarrow d))$$

Conclusion and future works

- Focus of current work:
 - Formalisation of a concept of shame
 - Handling of shame by an individual agent
- Future work on individual agents:
 - Action tendencies associated with shame
 - Coping strategies aiming at reducing or suppressing shame
 - Shame and self-deception?
- Future work on the social role of shame:
 - How does shame regulate an agent's social behaviour?
 - How does shame serve to maintain cohesion in a MAS?