

# \* Honte ou culpabilité ?

*(That is the question)*

Carole Adam

UJF--LIG/MAGMA

carole.adam@imag.fr

Dominique Longin

CNRS--IRIT/LILaC

Dominique.Longin@irit.fr

WACAI - 16 Novembre 2012 - Grenoble

- \* Intérêt intrinsèque  
(émotions complexes, difficiles à cerner)
- \* Intérêt par rapport à une typologie des émotions  
(une propriété caractéristique de l'une peut en caractériser une autre par son absence)
- \* Utiles dans :
  - \* Les ACA (comprendre pour agir ou réagir)
  - \* La théorie des jeux (prise de décision en situation stratégique)

\* Pourquoi étudier ces émotions ?

- \* Émotions très étudiées en psychologie
  - \* Voir notamment les travaux de Tangney
- \* Souvent confondues ou peu différenciées
  - \* Par les individus eux-mêmes (résultats expérimentaux, cf. [Tangney & Dearing, 2002] par exemple)
  - \* Par les spécialistes eux-mêmes (cf. par exemple [OCC, 1988])

**\* Pourquoi les étudier ensemble ?**

- \* Valence négative
- \* Émotions sociales
  - \* « Médiateurs cognitifs du comportement social de l'individu » [Castelfranchi & Poggi, 1990]
  - \* + vrai pour la honte (caractéristique) que pour la culpabilité (non nécessaire)
- \* Relèvent de la violation de normes/valeurs morales internalisées ([Elster, 1999] [OCC, 1988] [Lazarus, 1990])
- \* Une même situation peut mener à de la honte ou de la culpabilité (Princesse de Clève & Duc de Nemours *versus* Mathilde de la Molle & Julien Sorel, [Elster, 1999])

\* **Quand la honte  $\approx$   
la culpabilité**

\* [H.B. Lewis, 1975] :

\* Culpabilité : focus sur les actions de l'agent  
(j'ai fait telle action [ce qui a conduit à cette situation négative])

\* Honte : focus sur l'agent lui-même jugé comme un tout (j'ai fait telle action [ce qui a conduit à cette situation négative])

\* Responsabilité causale = caractéristique de la culpabilité mais pas de la honte

\* Relation avec les autres (honte devant/face à quelqu'un possible, mais pas la culpabilité)

**\* Quand la honte  $\neq$   
la culpabilité**

- \* Formalisation de la structure cognitive de l'émotion (quels états mentaux doit posséder un agent pour éprouver une émotion donnée)
  - \* Pas de composante, physiologique, physique, ou liée au ressenti ou aux tendances à l'action
- \* Une émotion est toujours à propos de quelque chose (Intentionnalité)
  - \* Honte et regret = direction d'ajustement des états mentaux au monde (provient d'une croyance sur le fait que l'objet de l'émotion est vrai)

\* Cadre de la formalisation

- \*  $Bel_i\varphi$  : l'agent  $i$  croit que  $\varphi$  est vrai dans le monde (savoir subjectif)
- \*  $Value_i\varphi$  :  $\varphi$  est une valeur (morale) de l'agent  $i$
- \*  $X\varphi$  :  $\varphi$  sera vrai l'instant suivant
- \*  $X^{-1}\varphi$  :  $\varphi$  était vrai l'instant précédent
- \*  $Fait_i\varphi$  : l'agent  $i$  fait en sorte que  $\varphi$  soit vrai (quel que soit ce que les autres agents aient choisi de faire)
- \*  $Emp_i\varphi$  : l'agent  $i$  peut empêcher le fait que  $\varphi$  soit vrai (quel que soit ce que les autres agents aient choisi de faire)

\* **Langage formel  
utilisé**

- \* L'agent  $i$  est responsable (causalement) du fait que  $\varphi$  soit vrai ssi
  - \* Il croit que  $\varphi$  est vrai ET
  - \* Il aurait pu éviter que  $\varphi$  soit vrai OU il a fait en sorte que  $\varphi$  soit vrai
  
- \* Resp. indirecte :  $Resp_i\varphi \equiv Bel_i\varphi \wedge X^{-1}Emp_iX\varphi$
- \* Resp. directe :  $Resp'_i\varphi \equiv Bel_i\varphi \wedge X^{-1}Fait_iX\varphi$
  
- \*  $Guilt_i\varphi \equiv Bel_i(Value_i\neg\varphi \wedge Resp_i\varphi)$

## \* Modélisation de la culpabilité

\*  $MBel_G \varphi$  :  $\varphi$  est une croyance mutuelle entre les agents du groupe  $G$

Ex. :  $MBel_{\{i,j\}} \varphi \rightarrow Bel_i \varphi \wedge Bel_j \varphi \wedge Bel_i Bel_j \varphi \wedge Bel_j Bel_i \varphi \wedge \dots$

\*  $Value^{++}_i \varphi$  :  $\varphi$  est une valeur particulièrement importante (au sens de [OCC, 1988]) de l'agent  $i$  susceptible de lui faire perdre la face

\*  $Value^{++}_G \varphi \equiv \bigwedge_{i \in G} Value^{++}_i \varphi$

\* Opérateurs  
additionnels

- \* Avoir honte à propos de  $\varphi$  devant un groupe  $G$  signifie [Castelfranchi & Poggi, 1990] :
  1. Qu'il y a croyance mutuelle de  $G$  que  $\varphi$  est vrai
  2. Qu'il y a croyance mutuelle de  $G$  que le fait que  $\varphi$  soit vrai a pour conséquence que nous ne vérifions pas une certaine propriété
  3. Qu'il y a croyance mutuelle sur le fait que tous les agent de  $G$  pensent que l'absence de cette propriété est susceptible de nous faire perdre la face
- \* Note : dans 1. et 2.  $G$  peut se réduire à  $G \setminus \{i\}$

## \* Propriétés formelles

\* Soit

\*  $G \in 2^{AGT}$  tel que  $G \neq \emptyset$  et soit  $i \in AGT$   
( $i$  peut appartenir ou non à  $G$ )

\*  $\varphi$  une formule représentant l'objet de la honte de l'agent  $i$

\*  $\psi_i$  une formule représentant une propriété particulière de l'agent  $i$

\*  $Shame_i(G, \varphi, \psi_i)$  ssi :

1.  $Bel_i M Bel_G \varphi \wedge$

2.  $Bel_i M Bel_G (\varphi \rightarrow \neg \psi_i) \wedge$

3.  $Bel_i M Bel_{G \cup \{i\}} Value_{G \cup \{i\}}^{++} \psi_i$

\* **Formalisation de la  
honte**

- \*  $\varphi$  représente le fait que Dominique ne fait pas partie des 50% des chercheurs de la session du vendredi matin de WACAI 2012 ayant eu un *best award paper*
- \*  $\psi_d$  représente le fait que Dominique a la propriété d'être un bon chercheur
- \*  $G$  est l'ensemble des participants à WACAI 2012

## \* Exemple simple

- \* Intégration dans un formalisme dynamique permettant de représenter des raisonnements du type :
  - \* L'agent  $i$  sait qu'après avoir accompli l'action  $\alpha$  l'agent  $j$  éprouvera de la honte
  - \* Auquel cas l'agent  $i$  (ne) va (pas) exécuter l'action  $\alpha$
- \* Étudier les liens entre la honte et d'autres émotions (par ex. la désapprobation morale)

\* Perspectives

! Merci \*

\* Merci !