Computational emotions
for social agents

Christian Becker-Asano

Foundations of Artificial Intelligence

Department of Computer Science

Albert-Ludwigs-Universität Freiburg

Overview

- **What are emotions?**
  - Introduction
  - Feedback theories
  - Basic emotions
  - Dimensional theories
  - Appraisal theories
- Computational modeling with WASABI
- Social emotions and DEL
- Some open questions

Five components:

- The component of **cognitive appraisal** or **evaluation** of stimuli and situations
- The **physiological component** of activation and **arousal**
- The component of **motor expression**
- The **motivation**al component, including behavior intentions or behavioral readiness
- The component of **subjective feeling** state

Appraisal theories

Feedback theories
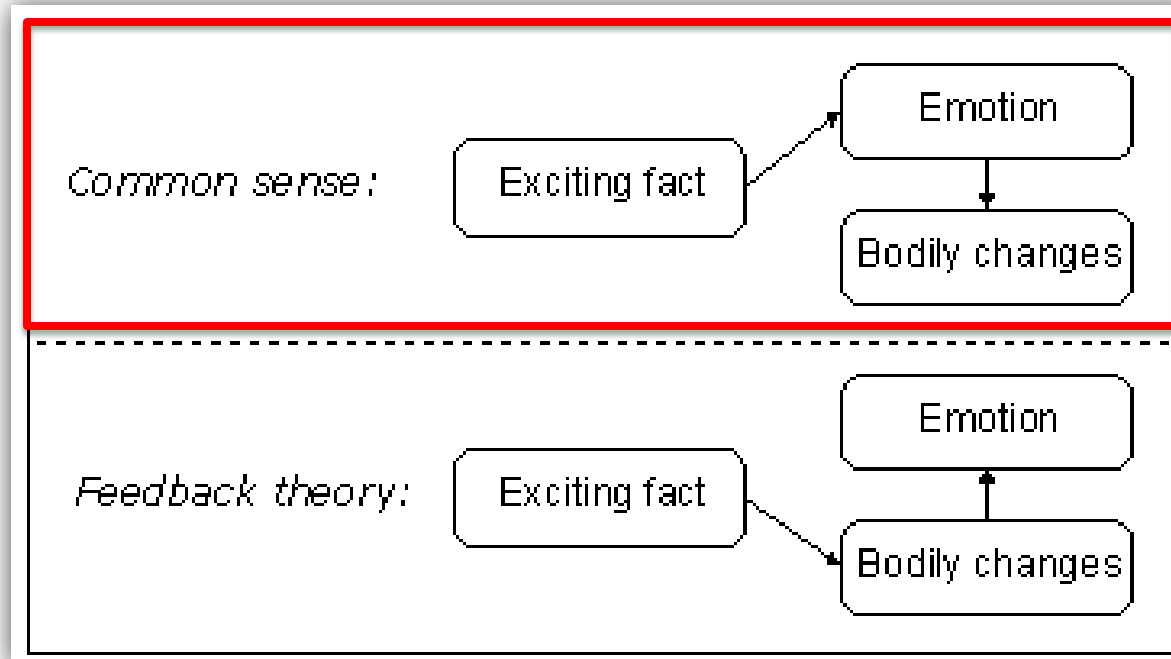
Basic emotions

Dimensional theories

Basic idea:

"We don't cry because of feeling sad, but we feel sad because we cry."

Common sense: Exciting fact → Emotion → Bodily changes

Feedback theory: Exciting fact → Bodily changes → Emotion

James' "standard emotions":

- "surprise, curiosity, rapture, fear, anger, lust, greed, and the like"

- purely result from the perception of bodily changes

- which directly follow the perception of an exiting fact

- in form of reflexes, so-called "nervous anticipations"

Five components:

- The component of **cognitive appraisal** or **evaluation** of stimuli and situations
- The **physiological component** of activation and **arousal**
- The component of **motor expression**
- The **motivation**al component, including behavior intentions or behavioral readiness
- The component of **subjective feeling** state

Appraisal theories

Feedback theories

Basic emotions

Dimensional theories

What are emotions?
Basic emotions

Example: Ekman et al. (1969) → 6,
             Ekman (1999) → +11 = 17 (cf. Prinz 2004)

Happy
Surprised
Fearful

Angry
Disgusted
Sad

(Ekman 1999, „Facial expressions", p. 304,
refering to an intercultural study in 1969)

Characteristics (Dist. → Distinctive):
1. Dist. universal signals
2. Dist. physiology
3. Automatic appraisal, tuned to:
4. Dist. universals in antecedent events
5. Dist. appearance developmentally
[…]
9. Unbidden occurence
10. Dist. thoughts, memories, images
11. Dist. subjective experience
(Ekman 1999, „Basic emotions", p. 56)

Five components:

- The component of **cognitive appraisal** or **evaluation** of stimuli and situations
- The **physiological component** of activation and **arousal**
- The component of **motor expression**
- The **motivation**al component, including behavior intentions or behavioral readiness
- The component of **subjective feeling** state

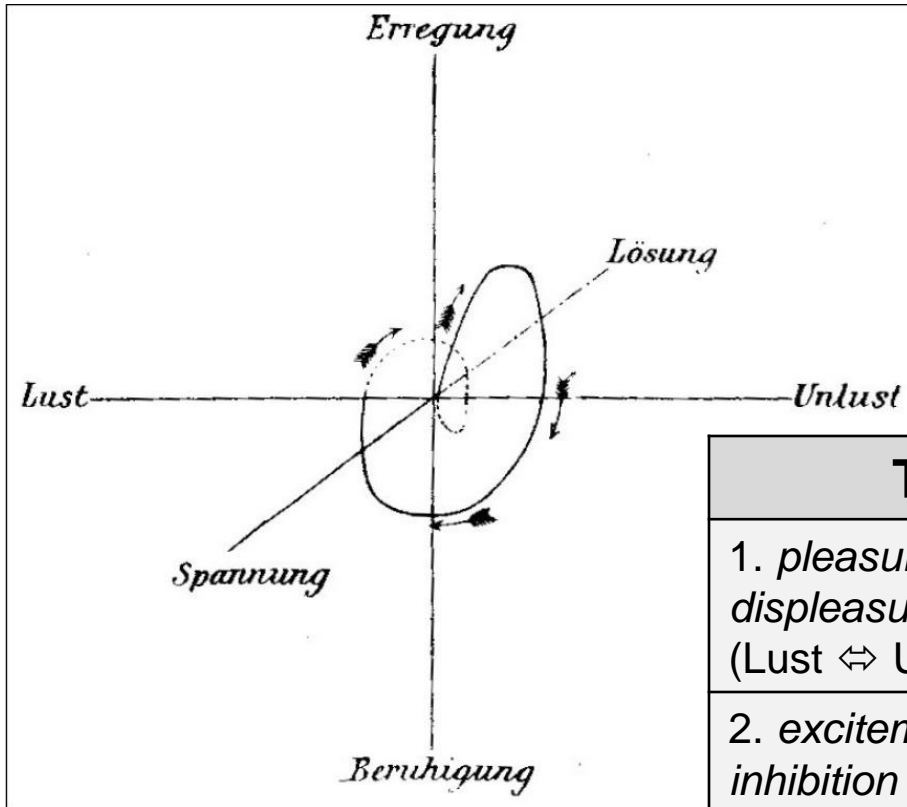Appraisal theories

Feedback theories

Basic emotions

Dimensional theories

What are emotions?
Wundt's 3D emotion space

A concrete event results in a "certain, continuous course of feeling" and in principle describes a trajectory that "represents the feeling state in any given moment."

Similar to Zajonc's "affective primacy idea" (Zajonc 1980)

| Three elementary feelings	
1. *pleasure* ⇔ *displeasure* (Lust ⇔ Unlust)	Quality or hedonic valence of emotional experience
2. *excitement* ⇔ *inhibition* (Erregung ⇔ Beruhigung)	Level of (physiological) arousal or (neurological) activation accompanying an emotional experience
3. *tension* ⇔ *relaxation* (Spannung ⇔ Lösung)	Temporal aspect of the emotion eliciting event

What are emotions?
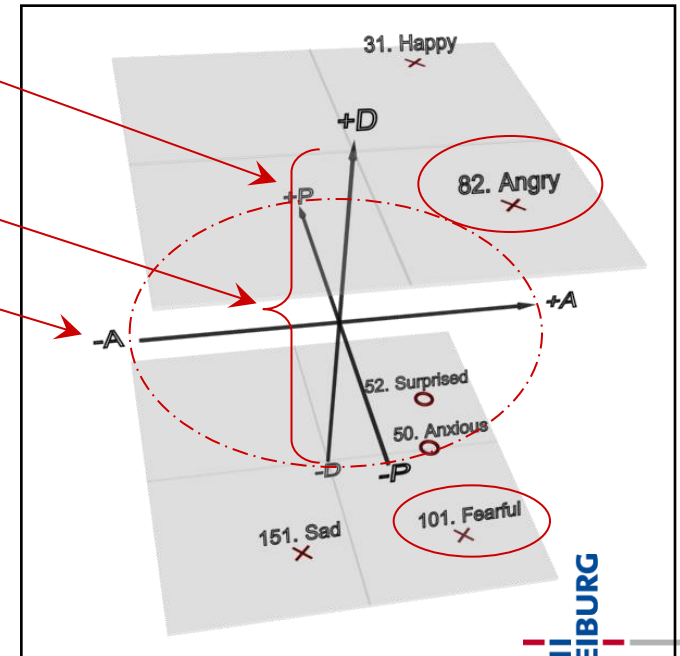# Core Affect & PAD space
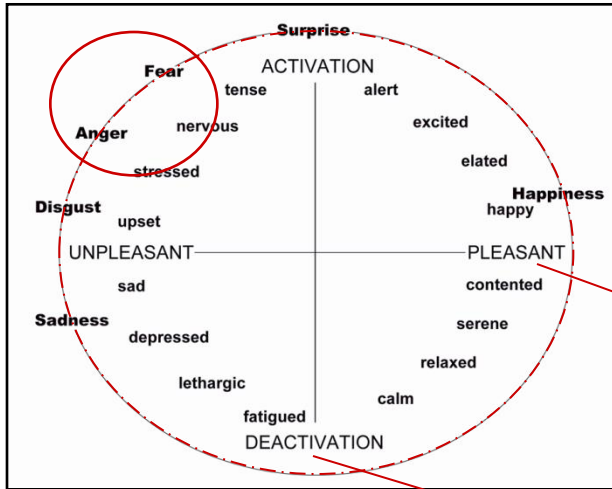
Assumption underlying "Core Affect":

→ Emotions not identifiable by distinct categories from the start

→ "Circumplex model of Core Affect" (Pleasentness & Activation)

Problem: "Fear" and "Anger" close together!
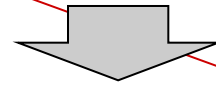
PAD space

Term	Pleasure		Arousal		**Dominance**	
	Mean	SD	Mean	SD	Mean	SD
31. Happy	.81	.21	.51	.26	.46	.38
50. Anxious	.01*	.45	.59	.31	-.15*	.32
52. Surprised	.40	.30	.67	.27	-.13*	.38
82. **Angry**	-.51	.20	.59	.33	**.25**	.39
101. **Fearful**	-.64	.20	.60	.32	**-.43**	.30
151. Sad	-.63	.23	-.27	.34	-.33	.22

(Russel & Feldmann Barrett 1999; Russel & Mehrabian 1977)

- **Pleasure/Valence:**
  - Basic evaluation of „good" or „bad"
  - based on subjective feeling or cognitive appraisal

- **Arousal/Activation/Excitement:**
  - Level of physiological arousal or neural activation
  - Not equal to an emotion's intensity!

- **Dominance/Control/Power:**
  - Reflecting a person's level of control and
  - social power in a situation

(Russell & Mehrabian 1977; Gehm & Scherer 1988)

Five components:

- The component of **cognitive appraisal** or **evaluation** of stimuli and situations
- The **physiological component** of activation and **arousal**
- The component of **motor expression**
- The **motivation**al component, including behavior intentions or behavioral readiness
- The component of **subjective feeling** state

Appraisal theories

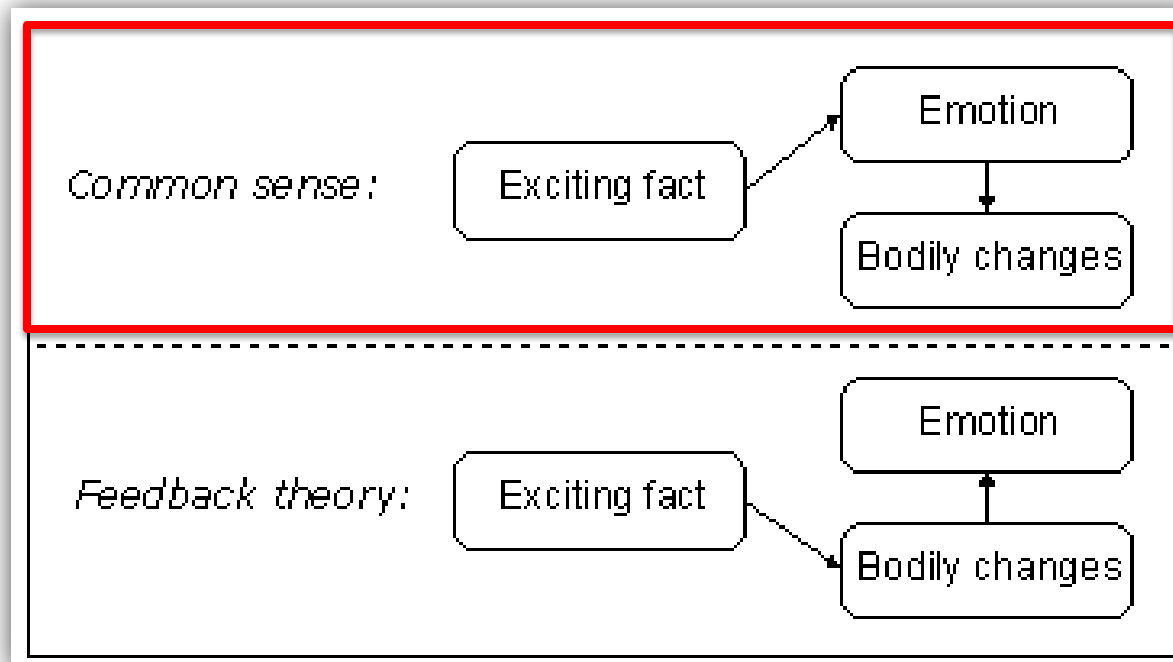**Feedback theories**

Basic emotions

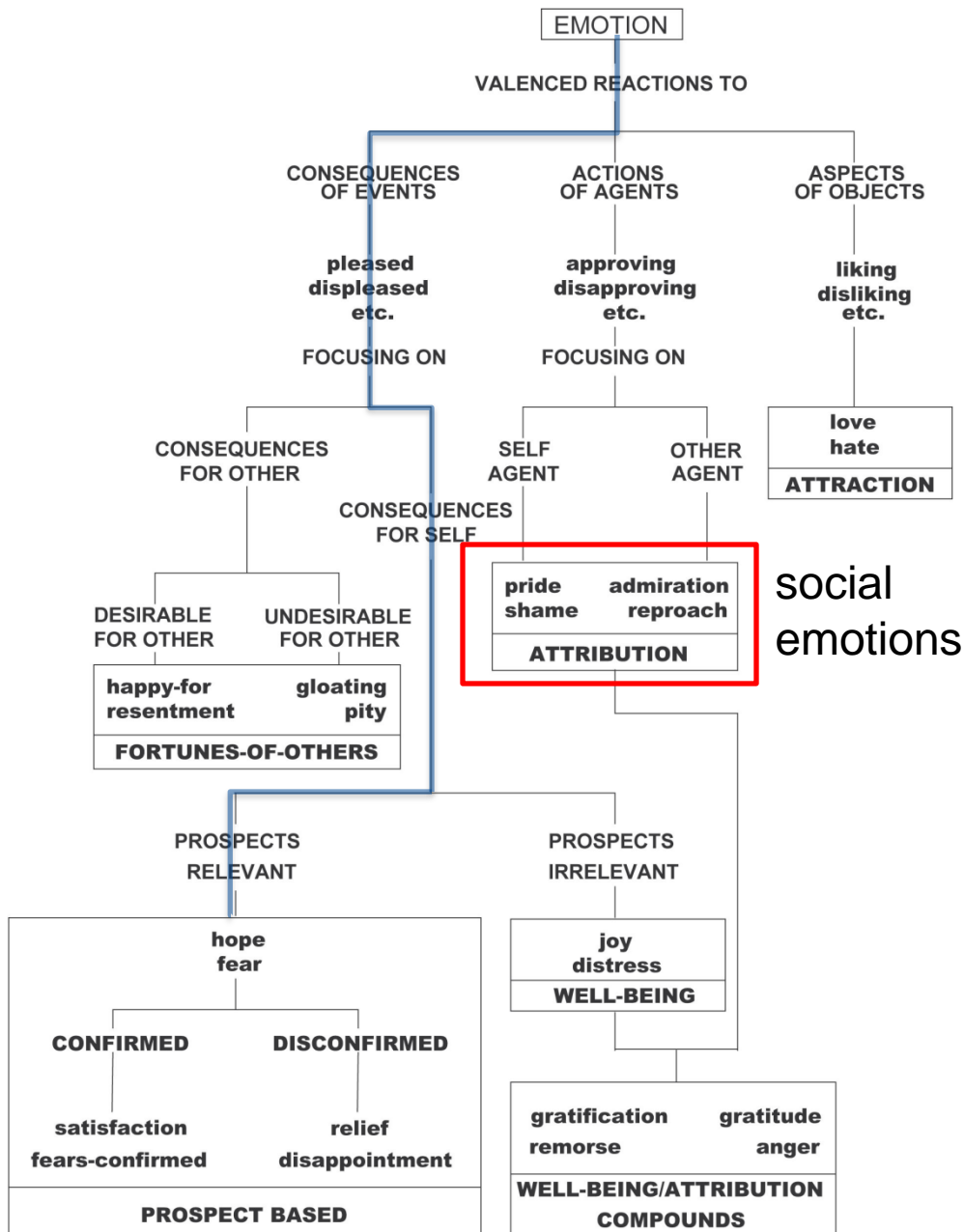**Dimensional theories**

What are emotions?
Appraisal theories

In common sense an emotion is a reaction to some event after its implication for the self has been assessed by an individual.

The OCC model

EMOTION

VALENCED REACTIONS TO

CONSEQUENCES OF EVENTS — ACTIONS OF AGENTS — ASPECTS OF OBJECTS

pleased displeased etc. — approving disapproving etc. — liking disliking etc.

FOCUSING ON — FOCUSING ON

CONSEQUENCES FOR OTHER

CONSEQUENCES FOR SELF

SELF AGENT — OTHER AGENT

love hate
ATTRACTION

DESIRABLE FOR OTHER — UNDESIRABLE FOR OTHER

pride admiration
shame reproach
ATTRIBUTION

social emotions

happy-for gloating
resentment pity
FORTUNES-OF-OTHERS

PROSPECTS RELEVANT — PROSPECTS IRRELEVANT

hope fear

joy distress
WELL-BEING

CONFIRMED — DISCONFIRMED

satisfaction relief
fears-confirmed disappointment
PROSPECT BASED

gratification gratitude
remorse anger
WELL-BEING/ATTRIBUTION COMPOUNDS

o "The cognitive structure of emotions"

o A *semantics-based* theory

o 22 emotions in 6 groups

Consequences of:
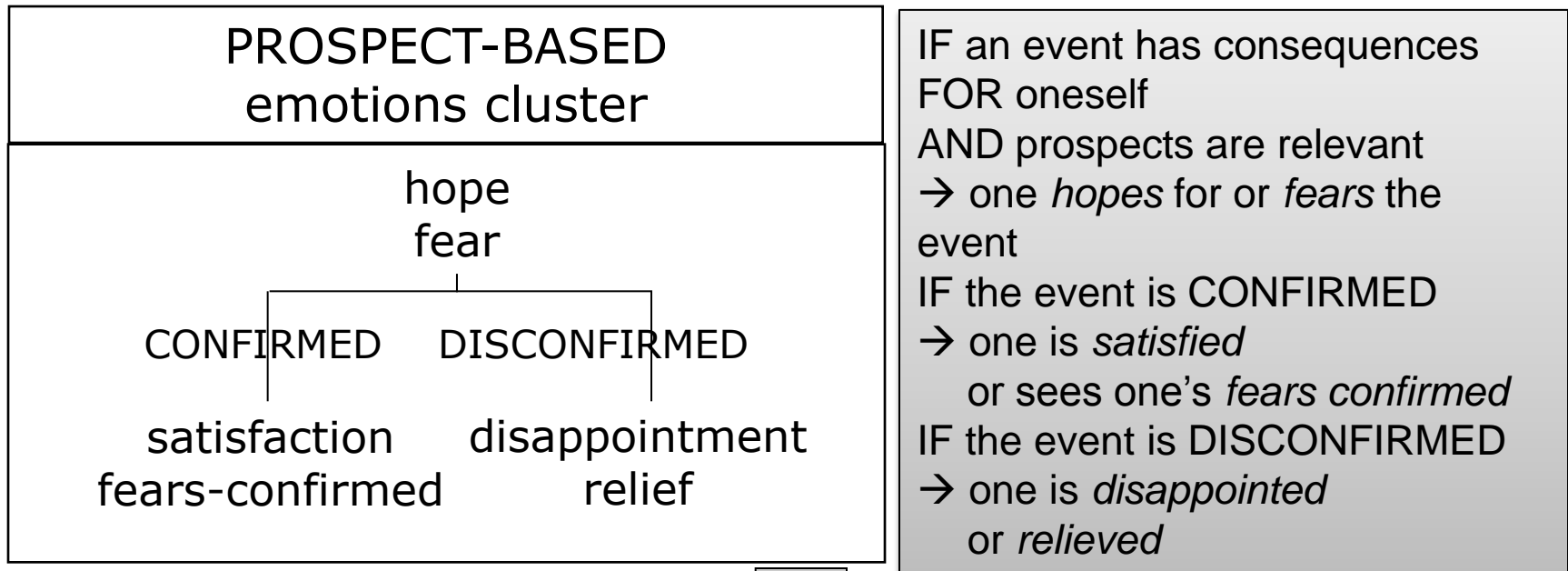
1. *Events*
2. *Agents*
3. *Objects*

1. Events → Consequ. for:

a) *Self*
b) *Other*

…

UNI FREIBURG

What are emotions?
OCC's prospect-based emotions

PROSPECT-BASED emotions cluster
hope fear
CONFIRMED   DISCONFIRMED
satisfaction   disappointment fears-confirmed   relief

IF an event has consequences
FOR oneself
AND prospects are relevant
→ one *hopes* for or *fears* the event
IF the event is CONFIRMED
→ one is *satisfied*
   or sees one's *fears confirmed*
IF the event is DISCONFIRMED
→ one is *disappointed*
   or *relieved*

Many implementations of the OCC-model exist, but they are only moderately successful

OCC-model best suited to reason about emotions

The non-cognitive emergence of emotions is neglected or at least not covered by the OCC model

What are emotions?
The emotional brain

Fear conditioning and the Amygdala
→ "emotional circuits" of the brain
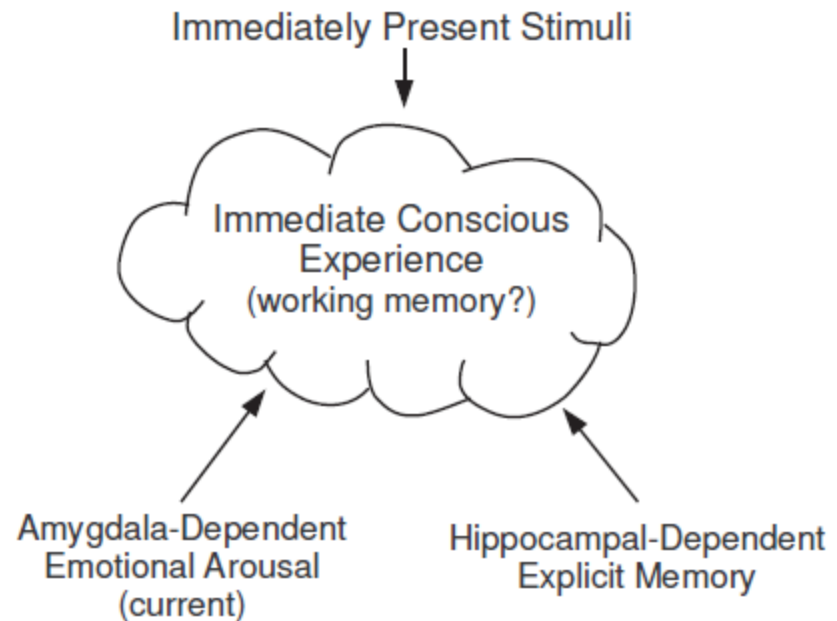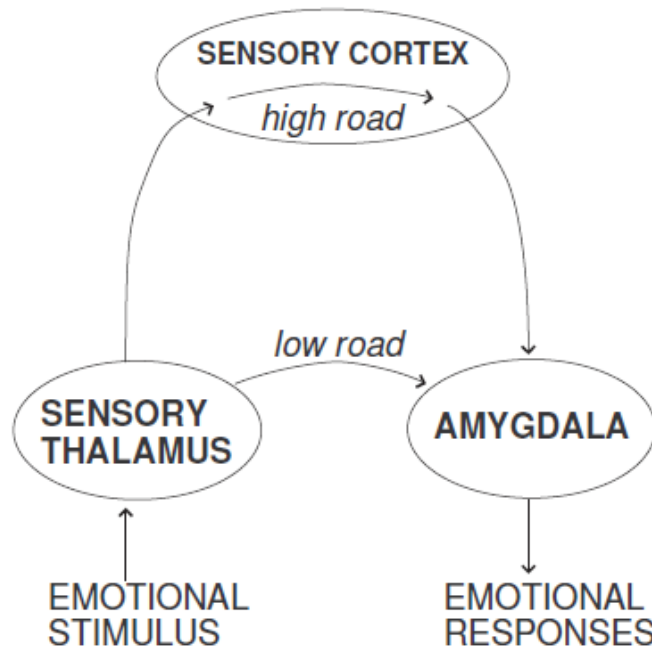
"Sensations are given meaning through memory." (LeDoux 2007)

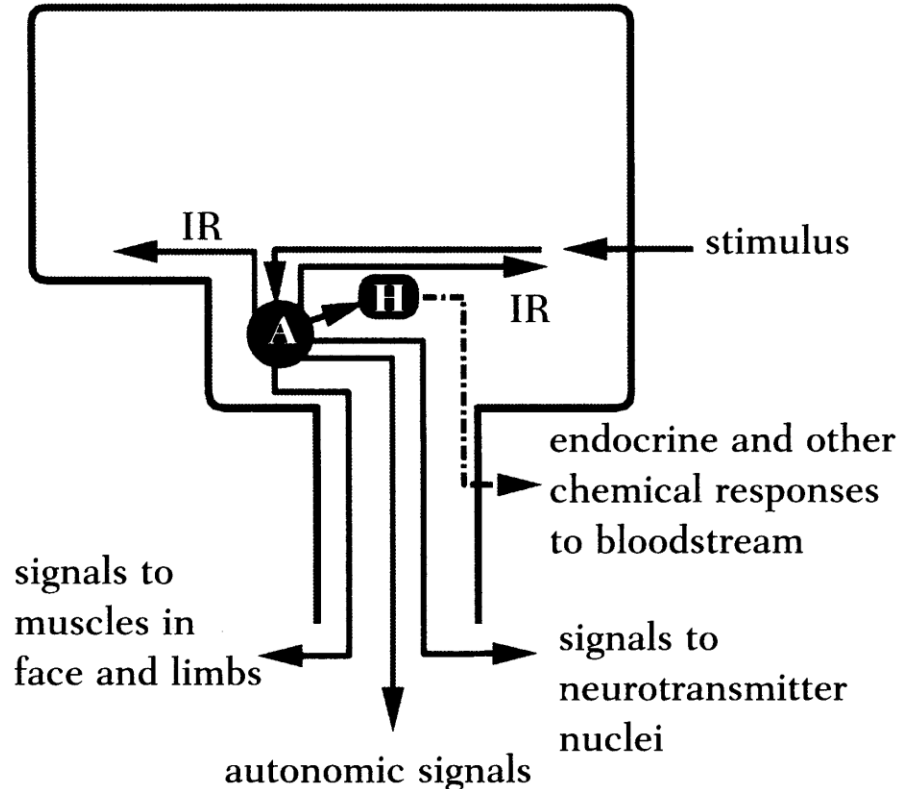## Primary emotions (fear, anger, joy, ...):

- fast, hard-wired stimulus response patterns

- trigger fight-or-flight behaviors
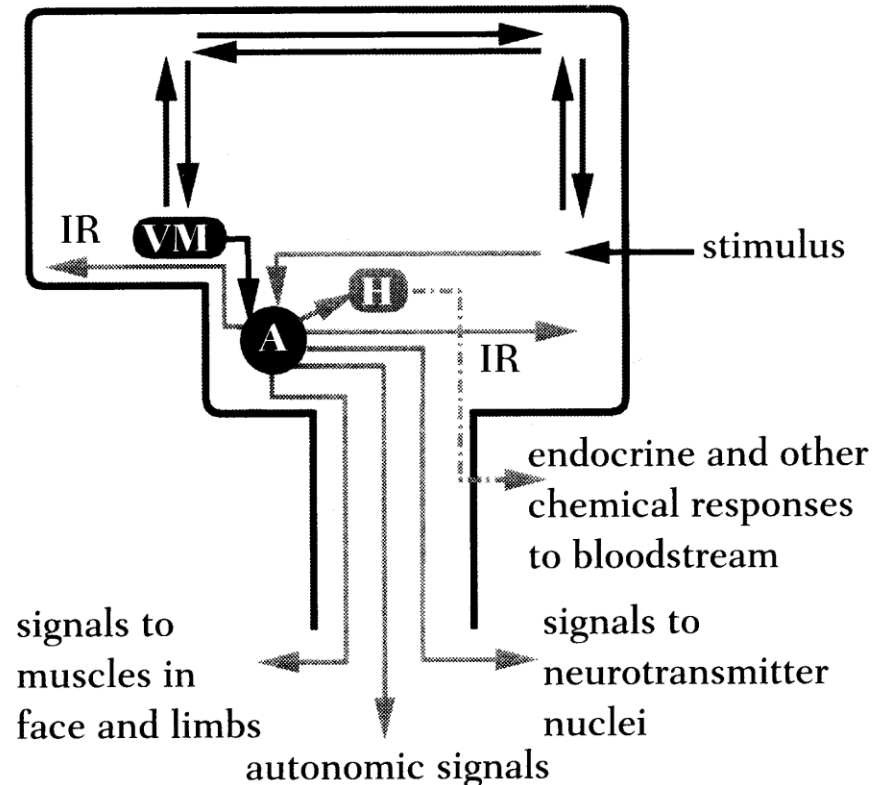
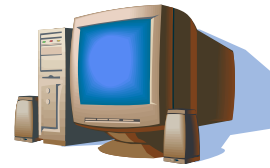- ontogenetically earlier types of emotion

Secondary emotions
(hope, relief, shame, ...):

- lead to cognitively
  elaborated, deliberative
  behaviors

- are based on memories
  and expectations

- "social emotions"
  developed during
  infancy
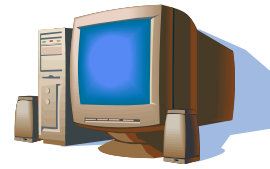
- "utilize the machinery
  of primary emotions"

WASABI

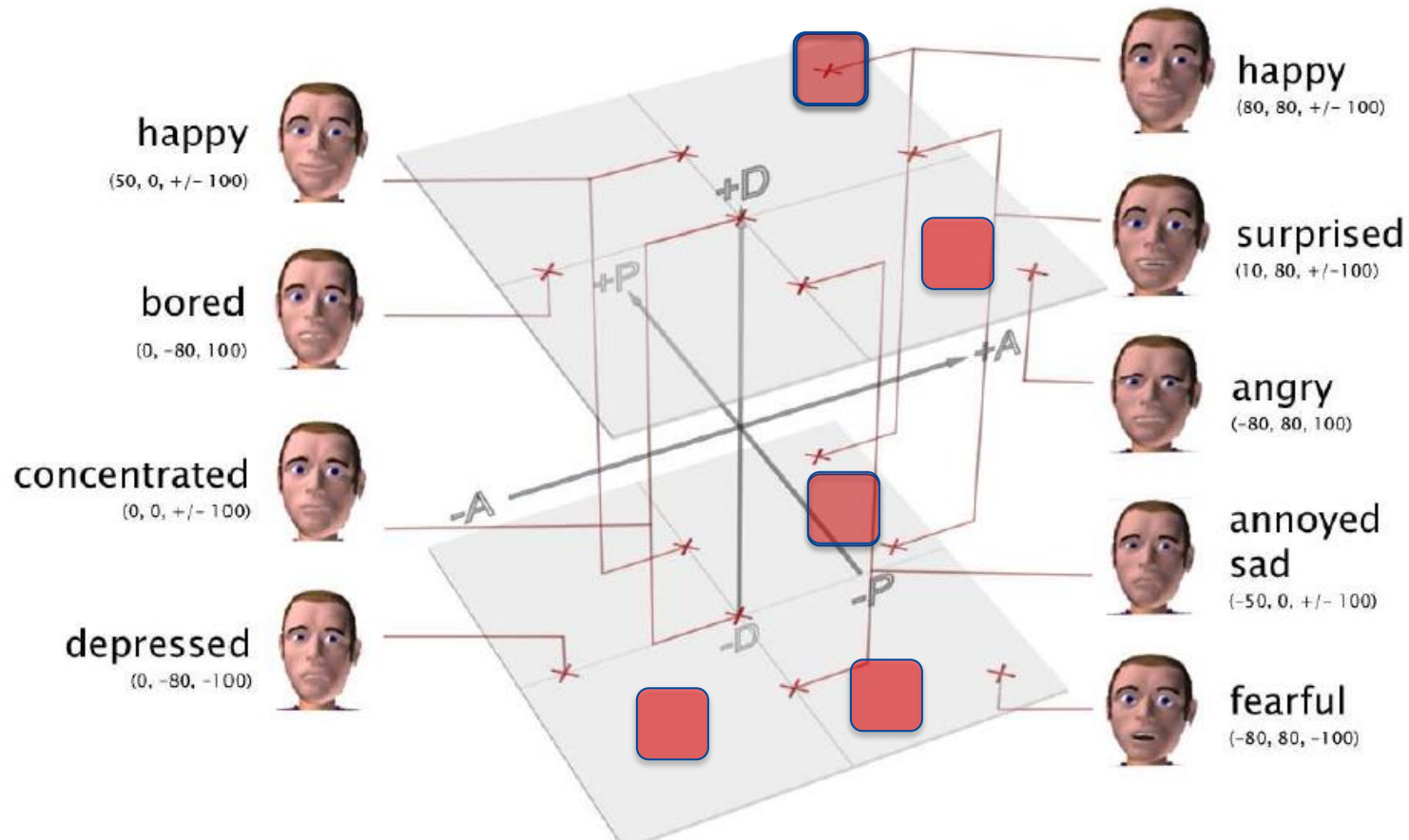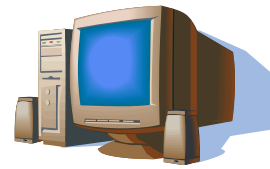## [W]ASABI [A]ffect [S]imulation for [A]gents with [B]elievable [I]nteractivity

- "Believable Characters" (in literature, film, etc.) need not necessarily be "honest and reliable"

- but must permit the audience the "suspension of disbelief"

- "Believable Agents"
  = Believable Characters + Interactivity

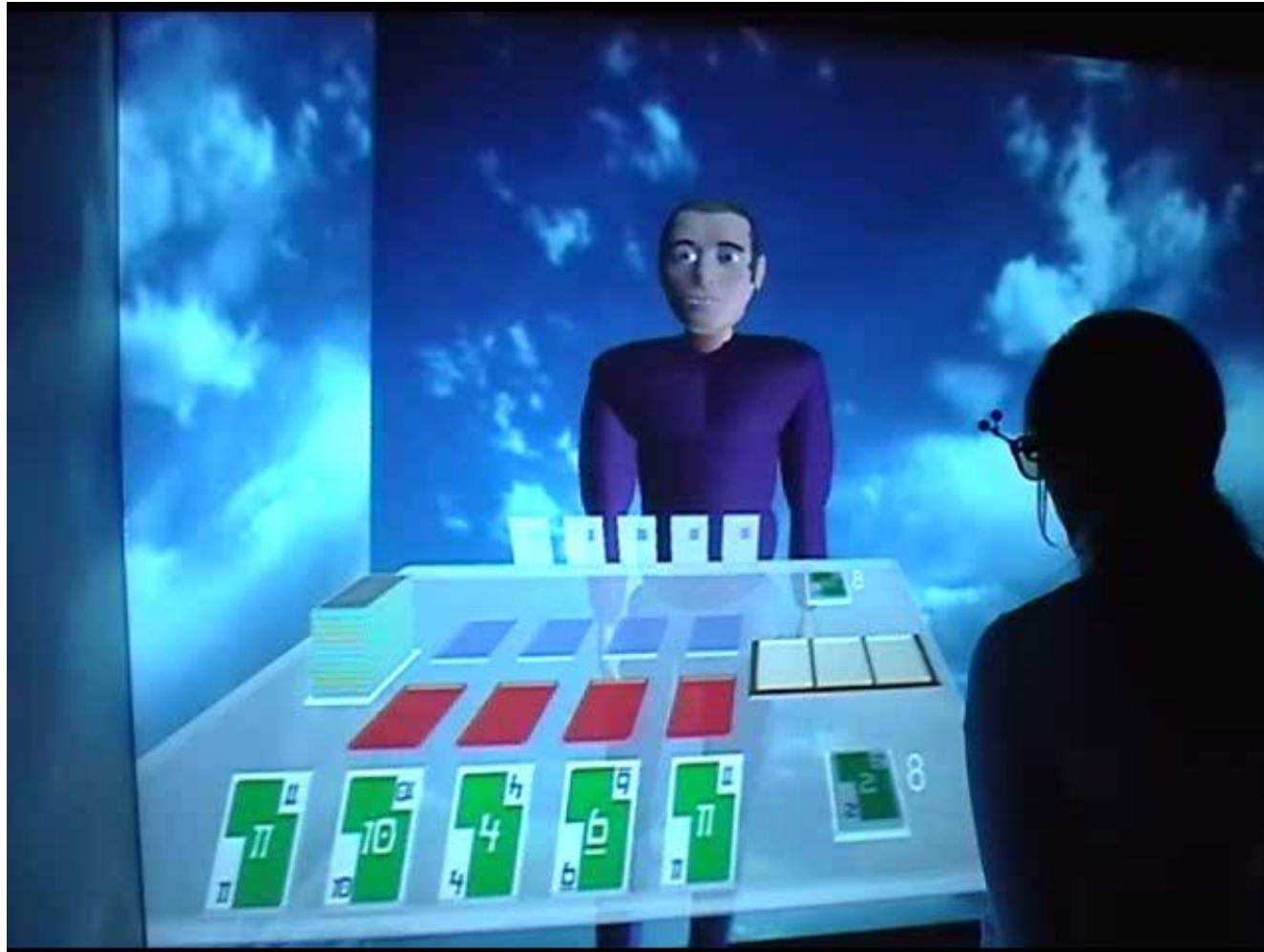# WASABI architecture
the virtual human MAX

„HNFMax" in Paderborn (2004)

Nine primary emotions
in PAD space

happy
(50, 0, +/- 100)

bored
(0, -80, 100)

concentrated
(0, 0, +/- 100)

depressed
(0, -80, -100)

+D

+P

+A

-A

-P

-D

happy
(80, 80, +/- 100)

surprised
(10, 80, +/-100)

angry
(-80, 80, 100)

annoyed
sad
(-50, 0, +/- 100)

fearful
(-80, 80, -100)
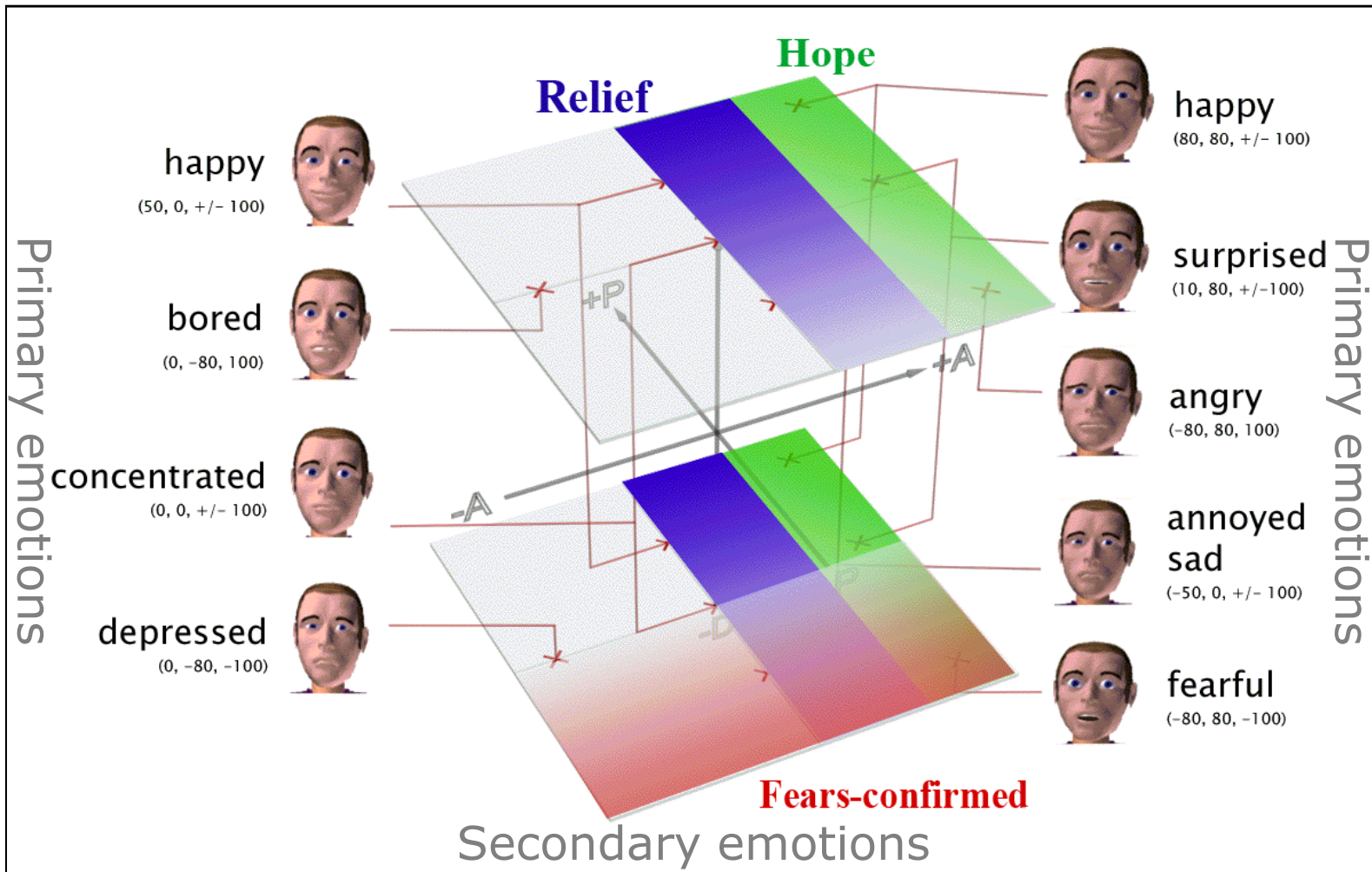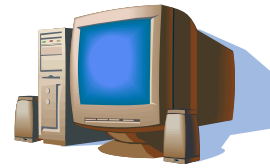
As of 2005 only primary emotions

WASABI architecture: SkipBo MAX → Secondary emotions
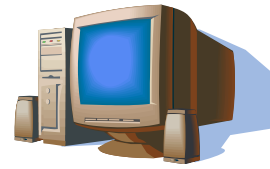
„CAVEMax" in Bielefeld (2008)

Primary emotions

happy
(50, 0, +/− 100)

bored
(0, −80, 100)

concentrated
(0, 0, +/− 100)

depressed
(0, −80, −100)

Relief

Hope

+P

+A

−A

−D

Fears-confirmed

Secondary emotions

Primary emotions

happy
(80, 80, +/− 100)

surprised
(10, 80, +/−100)

angry
(−80, 80, 100)

annoyed
sad
(−50, 0, +/− 100)

fearful
(−80, 80, −100)

Secondary emotion
example
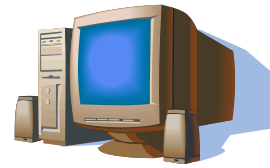
Cognition: The prospect of an undesirable event was confirmed

happy
(50, 0, +/- 100)

bored
(0, -80, 100)

concentrated
(0, 0, +/- 100)

depressed
(0, -80, -100)

happy
(80, 80, +/- 100)

surprised
(10, 80, +/-100)

angry
(-80, 80, 100)

annoyed sad
(-50, 0, +/- 100)

fearful
(-80, 80, -100)

"I was already afraid of that!"
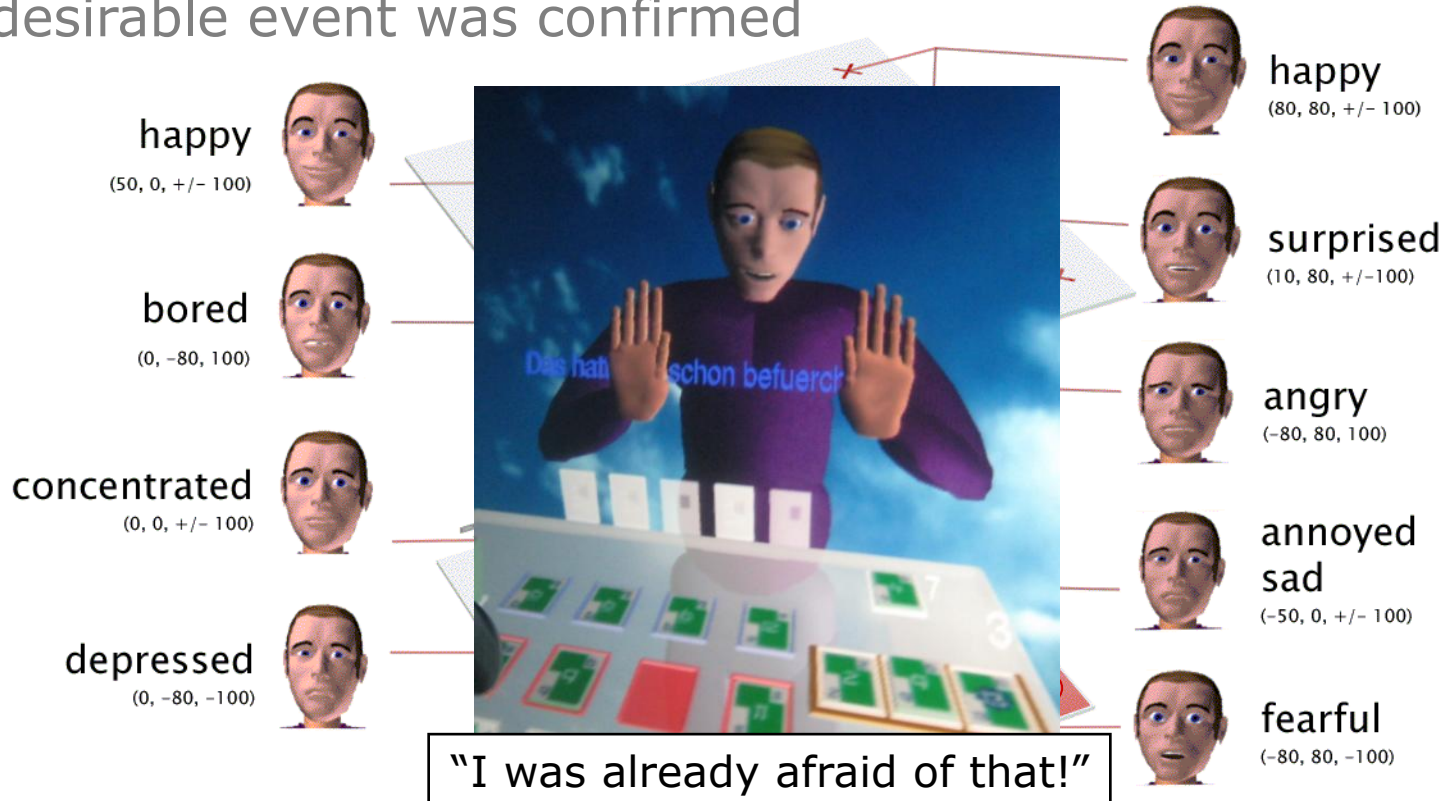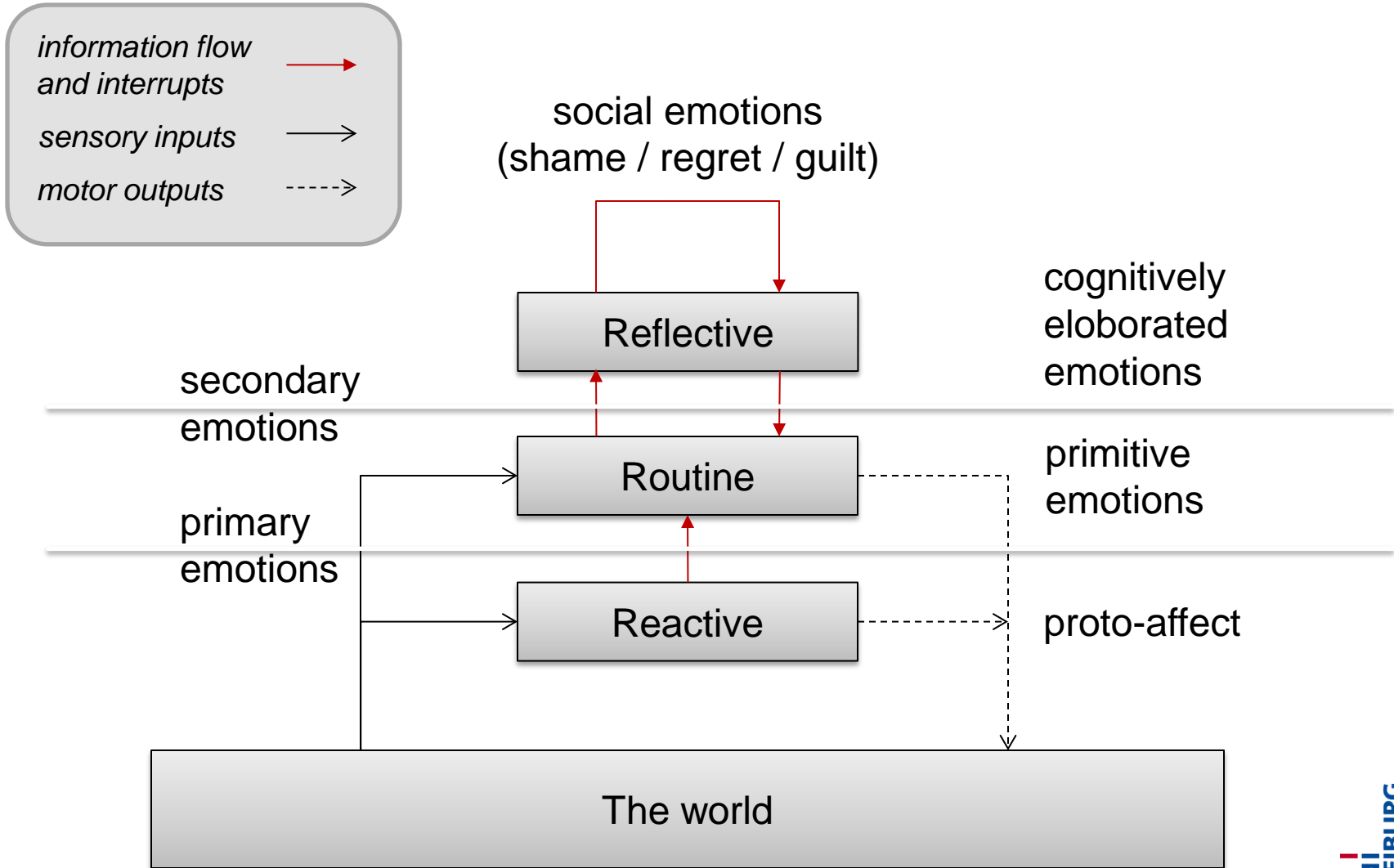
Fears-confirmed

→ trigger **Fears-confirmed**
   awareness likelihood = (0.3 * fearful, 0.2 * sad, 0.6 * **Fears-confirmed**)

The WASABI architecture

Perceive

Appraisal module

Act

beliefs, desires, intentions

goals, plans, expectations

Reactive appraisal
- intrinsic pleasantness
- expectation deviation

Cognitive appraisal
- goal conduciveness
- Dominance assessment

Cognitive reappraisal
- causal (mis)attribution
- coping

<<trigger>>

primary emotions

<<send>>

emotional impulses

<<change>>

Dominance

<<trigger>>

secondary emotions

aware emotions

<<send>>

Emotion dynamics

✕

emotion

mood

Pleasure Arousal

PAD space

Integration/categorization module

UNI FREIBURG

Tertiary / social emotions
Affect & Proto-affect (Ortony 2005)

information flow and interrupts →

sensory inputs →

motor outputs ----→

social emotions
(shame / regret / guilt)

Reflective

cognitively
eloborated
emotions

secondary
emotions

Routine

primitive
emotions

primary
emotions

Reactive

proto-affect

The world

Social emotions
Dynamic (epistemic) logic

Following **Turrini, Meyer, & Castelfranchi (2010)** from here:

„Coping with shame and sense of guilt: a Dynamic Logic Account"

According to Ortony, Clore, & Collins (OCC, 1988):

„In order to feel shame one must have violated a standard one takes to be important, as moral standards are. Such violations are held to be inexcusable. This is not necessary for a person who is feeling guilty.(...) In fact, we do not think that there is a distinct emotion of feeling guilty. Rather, we view feelings of guilt as mixtures of distinct emotions such as shame and regret, perhaps accompanied by certain cognitive states, such as the belief that one was, at least technically, responsible." (p. 142-143)

- „mixture of emotions"?
- „technical responsibility"?
- violations only in case of shame inexcusable?

Turrini et al. (2010, p. 403) comment:

„If we find the distinction between shame and sense of guilt and all other related feelings as meaningful at all, we need to have clear-cut definitions that relate those feelings to agents' mental states and to precisely understand their functioning."

→ extending logical framework „KARO" (BDI)

→ taking the „Multi Agent nature" of emotions such as shame into account (not discussed in this talk, though)

Running example (Torrini et al. 2010):

January 1998

„I did not have sexual relations with that woman, Miss Lewinsky. I never told anybody to lie, not a single time - never. These allegations are false. And I need to go back to work for the American people."
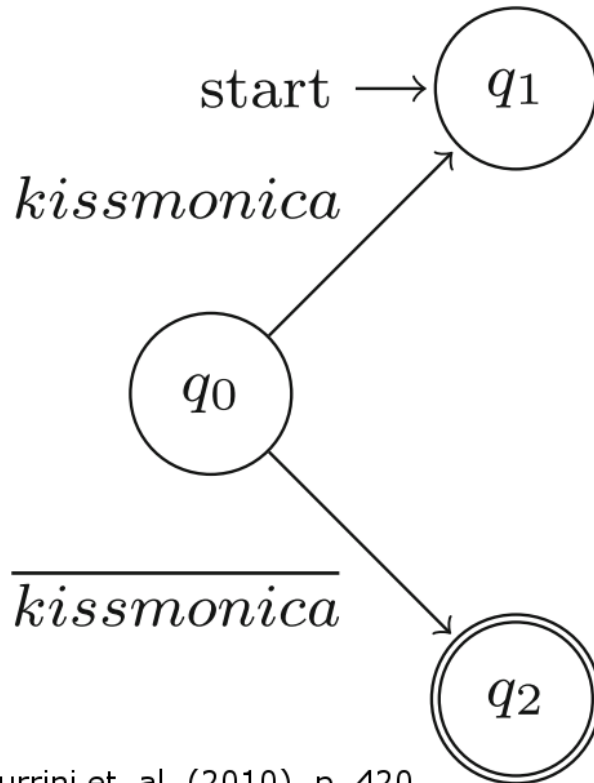
August 1998

„I did have a relationship with Miss Lewinsky that was not appropriate. In fact, it was wrong. It constituted a critical lapse in judgment and a personal failure on my part for which I am solely and completely responsible."

Epistemic state of agents feeling guilty

Epistemic state of agents feeling ashamed

Turrini et. al. (2010), p. 420

Turrini et. al. (2010), p. 420

Syntax (Torrini et al. 2010, p. 409):

$$\phi ::= p \,|\, L(i) \,|\, H(i) \,|\, guilty(i, a, j) \,|\, ashamed(i, a, j) \,|\, V_i \,|\, \neg\phi \,|\, \phi \wedge \psi \,|$$
$$\mathbf{Sig}_{i,j} \,|\, \mathbf{B}_i\phi \,|\, \mathbf{D}_i\phi \,|\, [\xi]\phi \,|\, [\xi]^{-1}\phi$$

$p \in \Pi_0$, set of atomic propositions.

$i, j \in Agt, \alpha \in Act, \xi \in Evt$, with Agt agents, Act actions, Evt events.

$L(i), H(i)$ indicate low/high self esteem personality type

$guilty(i, a, j)$ and $ashamed(i, a, j)$ means agent i feels guilty/ashamed for action a relative to other agent j.

V_i are special atoms to „describe also emotional states and agent types [e.g., violations of norms!]"

…

$\mathbf{Sig}_{i,j}$ encodes that for agent i another agent j is a „significant other"

<u>Abbreviations</u>:

$$\boldsymbol{P}\phi := \left\langle \left(\vee_{i \in Agt} \vee_{a \in Act} (i : a) \right) \right\rangle^{-1} \phi$$

„ϕ was just true before the latest action"

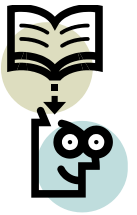$$\mathbf{Done}_j(b) := \langle (j : b) \rangle^{-1} (p \vee \neg p)$$

„agent j did b"

$$\mathbf{Done}_j(\overline{b}) := \vee_{c \in Agt, c \neq b} \langle (j : b) \rangle^{-1} (p \vee \neg p)$$

„agent j did not do b"

Social emotions
Coping with shame and guilt

Sense of guilt:

$$\mathbf{B}_i\big(\mathbf{Sig}_{i,j} \wedge V_j \wedge \mathbf{Done}_i(a) \wedge \mathbf{P}\big(\langle i\colon \bar{a}\rangle \neg V_j\big)\big) \to guilty(i,a,j)$$

1. high self esteem agent reaction:

$$H(i) \wedge \mathbf{B}_i\big(V_j\big) \wedge guilty(i,a,j)$$
$$\to [deliberate_i](\langle i\colon \mathbf{eliminate}(j)\rangle(p \vee \neg p) \vee \mathbf{B}_i(\neg V_i)$$

2. low self esteem agent reaction:

$$L(i) \wedge \mathbf{B}_i\big(V_j\big) \wedge guilty(i,a,j) \to [deliberate_i]\mathbf{B}_i(\mathbf{P}[i\colon \bar{a}]V_j))$$

Shame:

$$\mathbf{B}_i\big(\mathbf{Sig}_{i,j} \wedge V_j \wedge \mathbf{Done}_i(a) \wedge \mathbf{P}\big(\langle i\colon \bar{a}\rangle V_j\big)\big) \to shame(i,a,j)$$
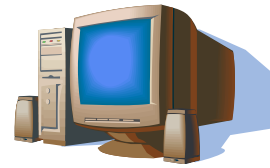
1. high self esteem agent reaction:

$$H(i) \wedge shame(i,a,j) \to [deliberate_i]\langle i\colon \mathbf{eliminate}(j)\rangle(p \vee \neg p)$$

Why not also „$\vee\, \mathbf{B}_i(\neg V_j)$" here?

Unity3D demo

And how about planning?

Social agents..

- planning to avoid feeling ashamed?
  - Not based on Turrini et al.'s work, because post-hoc reasoning, i.e. coping, not helpful here
- simulating another's emotional state?
  - Partially done based on WASABI
    (PhD thesis by Hana Boukricha, 2013)

Open challenges (selection):

- Integrating (DEL-based) TOM on reflective level
- Combining bottom-up dynamics simulation with top-down reasoning and planning capabilities

Thank you for your attention