



# Emotion modeling for social robots

Motivation &  
Psychological background &  
Realization

OR

Why should social robots be able to  
cry?

# Part I: Motivation

("Ex astris scientia", StarTrek)

# WHY to build social robots? Isn't it all just science fiction?

**Sonny:** [*angry*] I did not murder him.

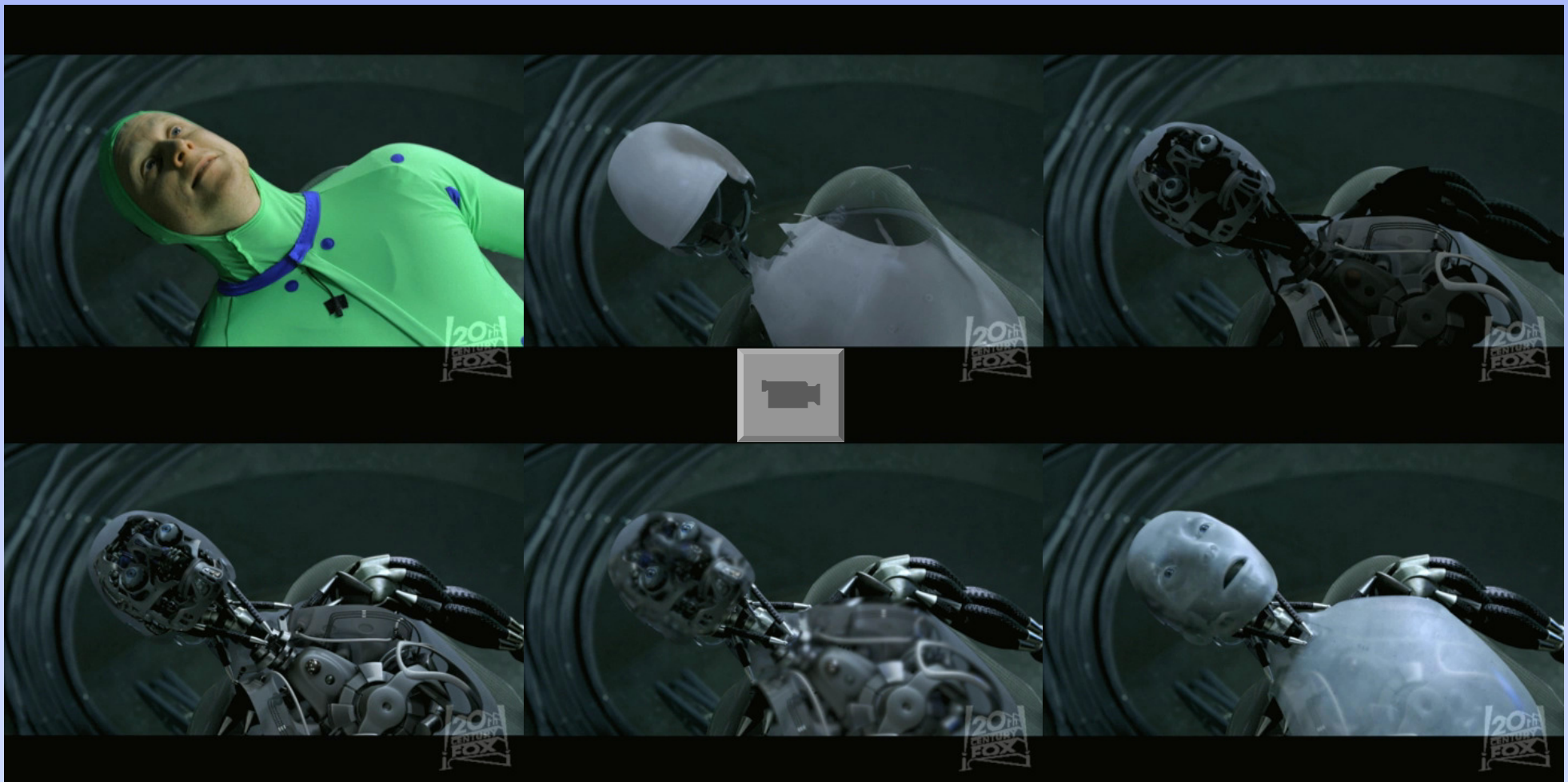
**Detective Spooner:** You were emotional... I don't want my vacuum cleaner, or my toaster appearing emotional...

**Sonny:** [*very angry*] I did not murder him!

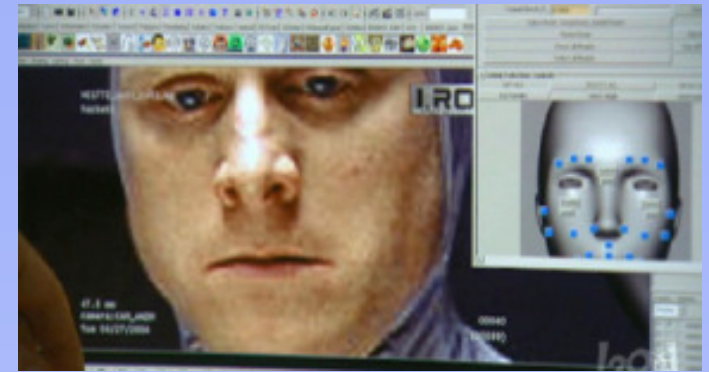
**Detective Spooner:** That one's called "anger".  
(from "I, Robot", Fox Entertainment 2004)



# The making of the robot "Sonny" ("I, Robot", 2004)



# What makes Sonny a “special” robot?



Alan Tudyk alias “Sonny”

“.. his face and his eyes are just so expressive.”

“Sonny has emotions”

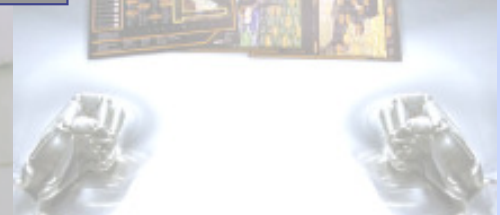
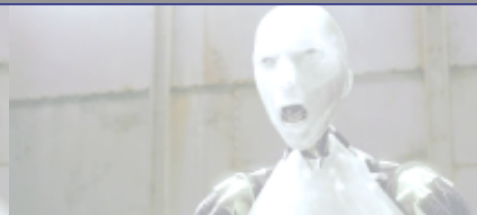
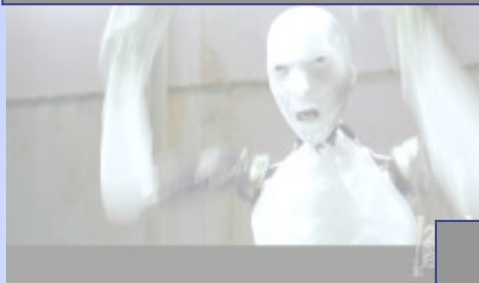
“Alan is the soul of Sonny – every subtlety is there..”

“.. really study every little nuances of [Alan’s] performance”

“what is life itself .. all these great little things you wouldn’t necessarily think about [says the animator ;-)].”

“Expressive”

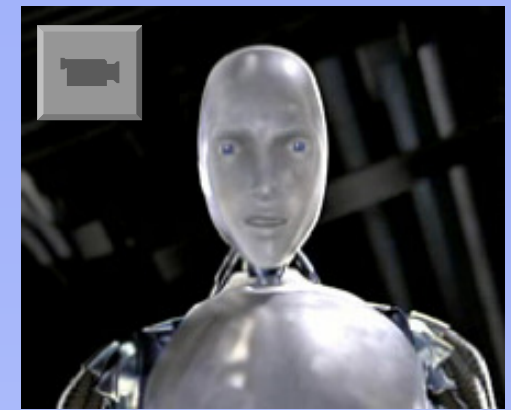
“Soul”



“Emotions”

“Life itself”

# Why to integrate emotions into robots?



“What am I?”

The very simplistic “I, Robot” answer:

To let it **break** the **Three Laws of Robotics**

1. A robot may **not harm a human being**, or, through inaction, allow a human being to come to harm.
2. A robot must **obey the orders** given to it by human beings except where such orders would conflict with the First Law.
3. A robot must **protect its own existence**, as long as such protection does not conflict with the First or Second Law.

(as invented by Asimov in his short story “Runaround” in 1941)

BUT: In the movie, it remains very mysterious HOW emotions are integrated and WHY they should help the robot to save humanity by letting him violate the three laws!!!

Anyway: It’s just a Hollywood movie!

# Once again: WHY to build social robots?

1. To pass the “Turing test” (as proposed by Alan Turing in 1950)?
  - Five minutes, text-based conversation with both a human and a machine.
  - If in at least 30% of the cases the machine is **falsely judged as human**, it has successfully passed the test.

Rosalind Picard, MIT (“Affective Computing”, 1997, p. 13):

“A machine, even limited to text communication, will **communicate more effectively** with humans if it can perceive and express emotions.”

2. To become a social member of a future society?

Hiroshi Ishiguro, ATR Japan:

“Robots are information media, especially humanoid robots. Their main role in our future is to **interact naturally with people.**”

# Ishiguro: "Android Science"

## What is human?

### **Androids for android science**

Studies on fundamental issues of human-machine interaction in science and engineering

### **Social robots for our societies**

Development of practical robots and finding social issues in real fields

Tools → Media → Partner

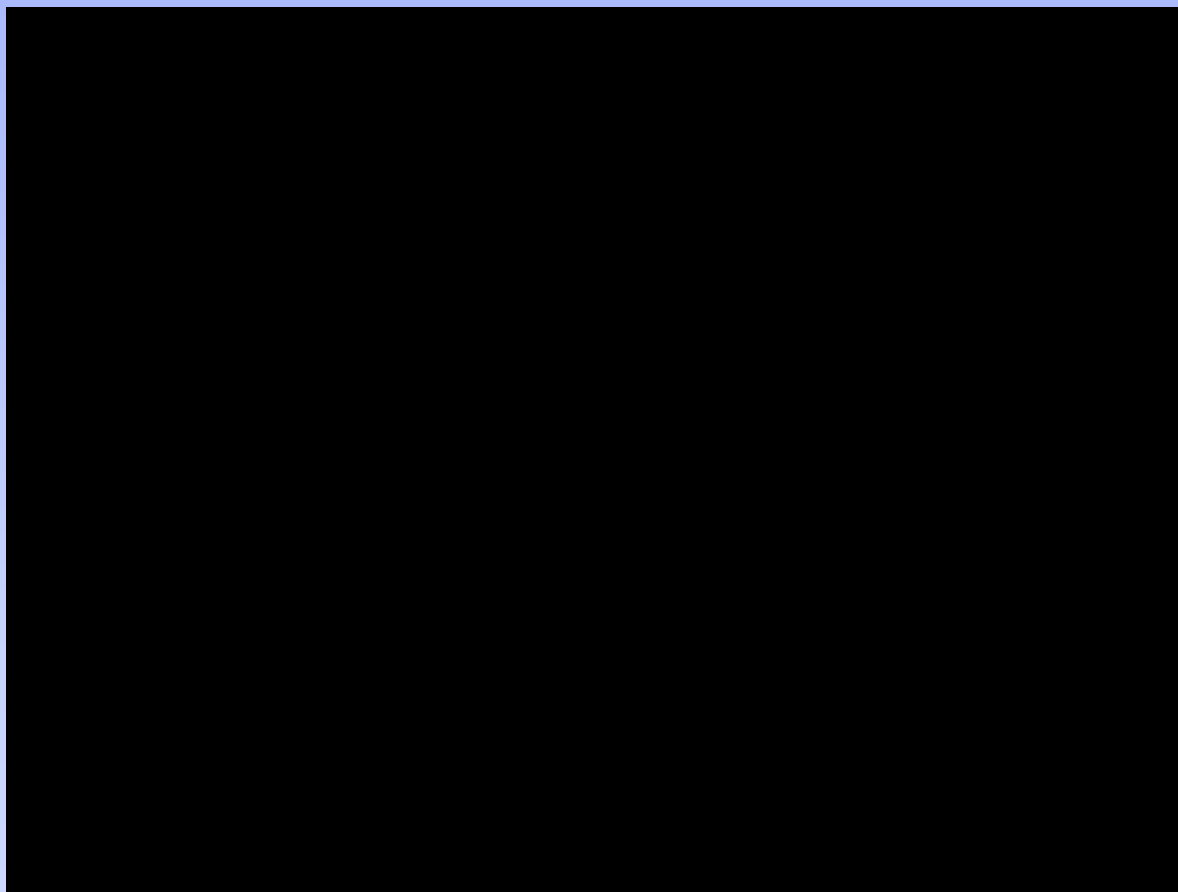


# Repliee-Q2 (University of Osaka) (Repliee-Q1expo, 2005 in Aichi)



- 42 actuators in the upper torso including 13 for the head
- Prerecorded japanese voice synchronized with mouth movement
- No locomotion but movement of both arms possible

# Repliee-Q2: Research issues



- In what way should robots resemble human beings?
- Why do we feel uncanny as the robot is getting close to human?
- How do we develop humanlike behavior of the android?
- Why do we become aware that the android is not human?

Does a human unconsciously recognize a human?

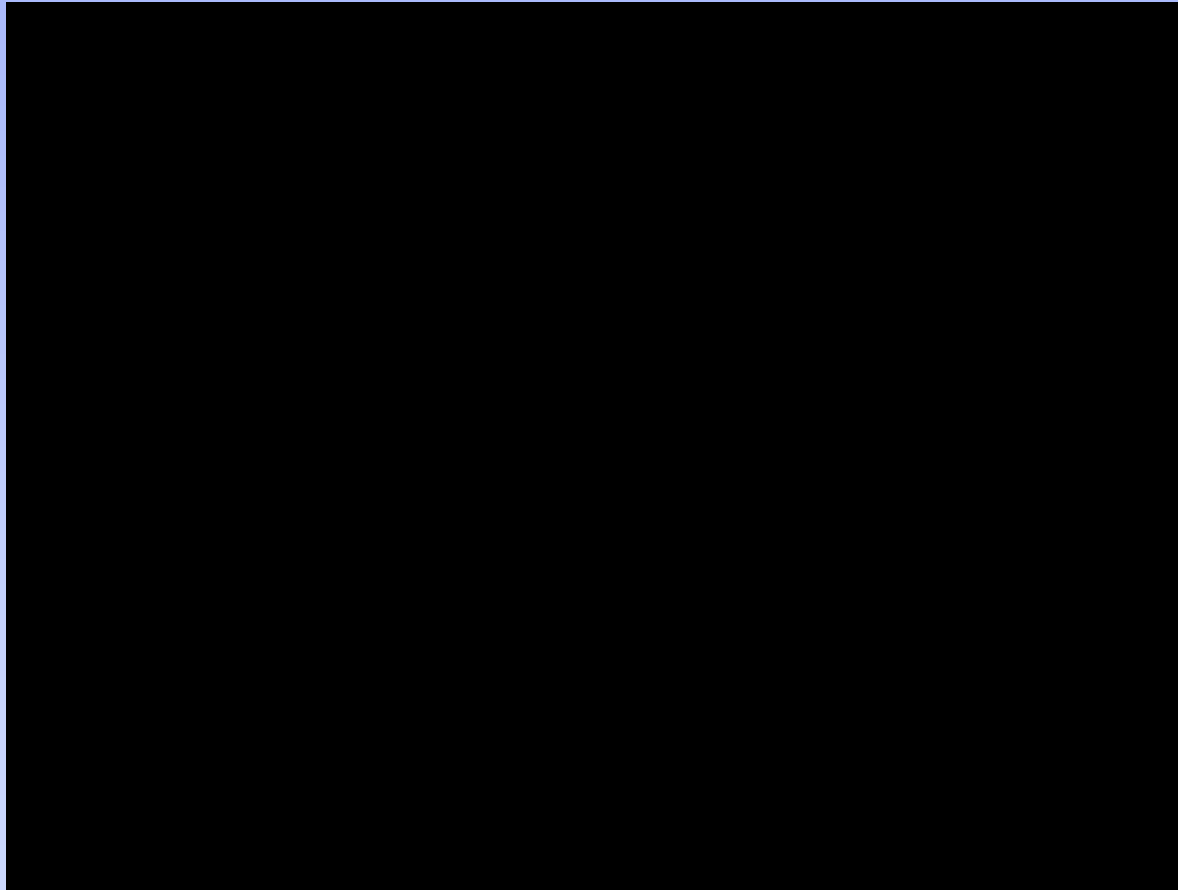
# MaxHNF (Scenario 1): Research issues



- In what way should humanoid agents resemble human beings?
- (Why) do we feel uncanny as the virtual agent is getting close to human?
- How do we develop humanlike behavior of the humanoid agent?
- Why do we become aware that the humanoid agent is not human?

Does a human accept the **humanoid agent as a conversational partner**?

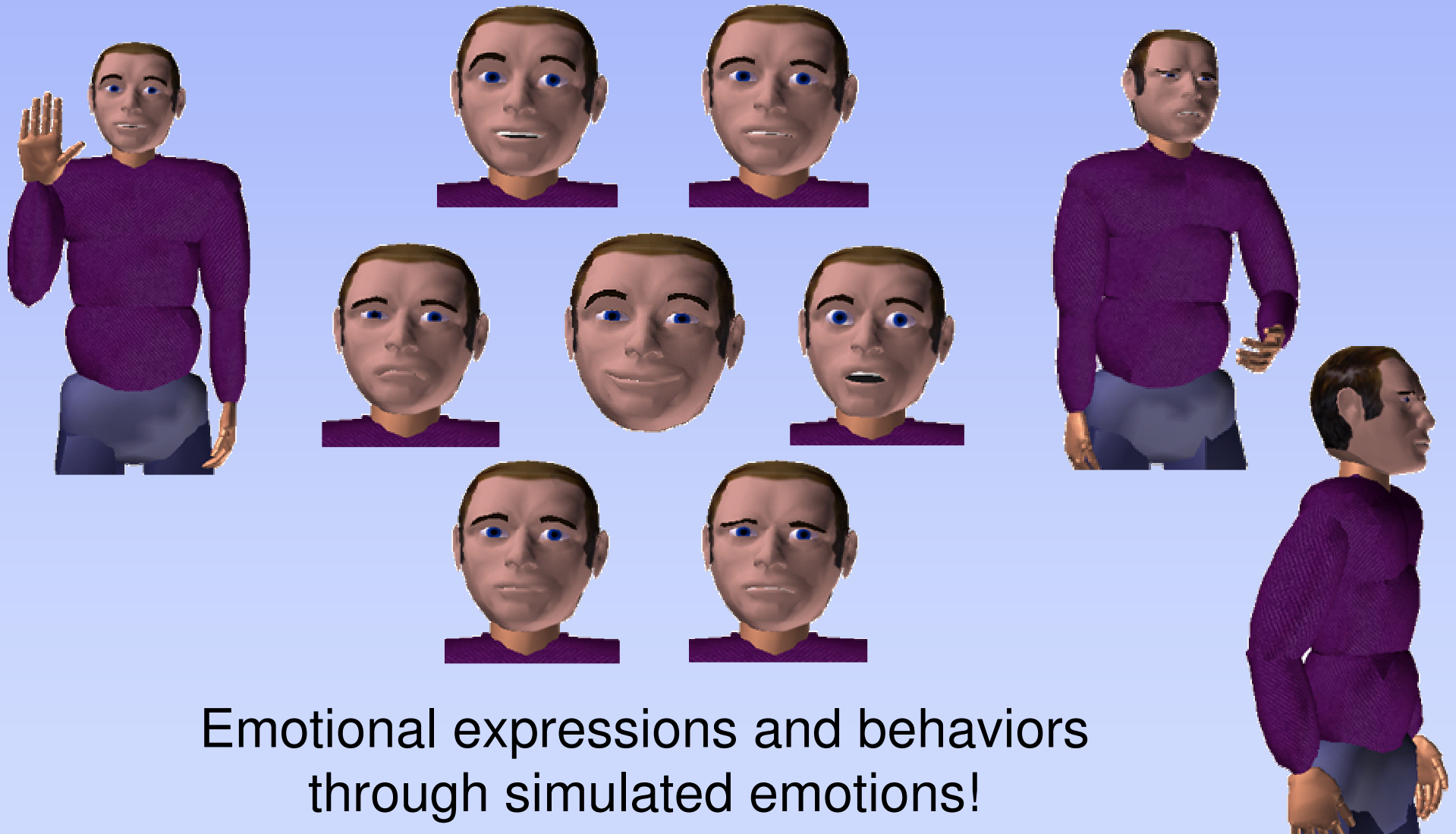
# Max plays SkipBo (Scenario 2): An empirical study in Japan



- Can empathic behavior be implemented and how do humans react on it?
- Do the player's reactions differ when the agent's empathic attitude is changed systematically?
- Does the agent's emotional behavior lead to „emotional contagion“ in the human player as to be derived from bio-metric sensor data?

Does the expression of negative emotions in a competitive scenario help to make the agent more acceptable?

# How do we improve our agent's acceptance as a social partner?

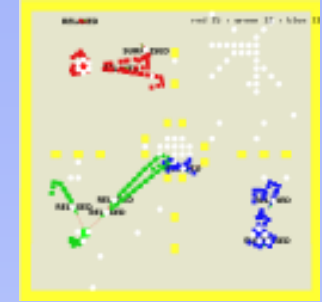


Emotional expressions and behaviors through simulated emotions!

# Three motives for the integration of emotions

## 1. Control-engineering motive

Simulating emotional processes to generate appropriate and fast reactive responses in Socionics and Multiagent Systems



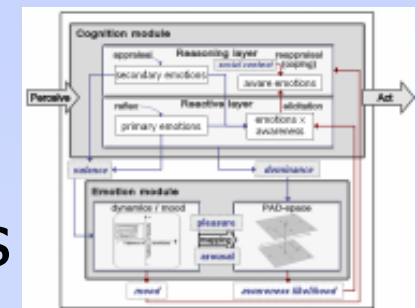
## 2. Believable-agent motive



Increasing the believability of an agent by simulating emotions and their effects on expression and behavior

## 3. Experimental-theoretical motive

Gaining a deeper understanding of human emotions using simulated humanoid agents



# Believable-agent motive: Social interaction perspective

B. Reeves & C. Nass, 1998, in “The media equation”:

- Empirical evidence → “Humans treat computers as social actors”
- Social and emotional aspects of the computer were encoded on the textual level only → NO anthropomorphic interface!

T. Bickmore & J. Cassel, 2005, in the context of the ECA “Rea”:

- Integration of non-verbal cues for embodied conversational agents (ECA) in social dialogues
- In order to generate “trust” → implementation of a theory of interpersonal relationship based on the three dimensions *familiarity*, *solidarity* and *affect*.
  - “Affect” is understood as “the degree of liking the interactants have for each other”
  - “Affect” is coupled with the social ability of coordination, i.e., fluent and natural smalltalk → NO explicit model of emotions needed!

# Experimental-theoretical motive: Cognitive modeling perspective

## Cognitive Science:

- combines philosophy, psychology, artificial intelligence, neuroscience, linguistics and anthropology
- attempts to build computational models of human cognitive behavior by combining the above disciplines to verify their findings

Emotions recently became an important concept in Cognitive Science  
esp. to improve Human-Computer Interaction

→ We implement computational models of emotions combining psychological / philosophical theories and AI-algorithms in order to:

1. Improve the believability of our agents (technical motive)
2. Help in validating the underlying theories (experimental motive)
3. Extend the AI-algorithms (theoretical motive)



# Part II:

## Cognitive emotion theories

("cogito, ergo sum")

# I. OCC theory: (Ortony, Clore and Collins, 1988)

Book, “The cognitive structure of emotions”, page 1:

“Taking the perspective of **empirical psychology** and **cognitive science**, we start with the assumption that **emotions** arise **as** a result of the way in which the **situations** that initiate them are **construed by the experienter.**“

Three necessary conditions for the emergence of emotions:

1. Situations that cause the emotions
2. Persons that experience the emotions
3. The appraisal of the situation by the person

# I. OCC theory: Pros & Cons

Pros (Meyer, Schützwohl, Reisenzein 2003):

- **All-embracing** theory, i.e. aims at explaining **all** emotions
- **Detailed** theory
- **Systematically** constructed theory

Cons (Omdahl 1995):

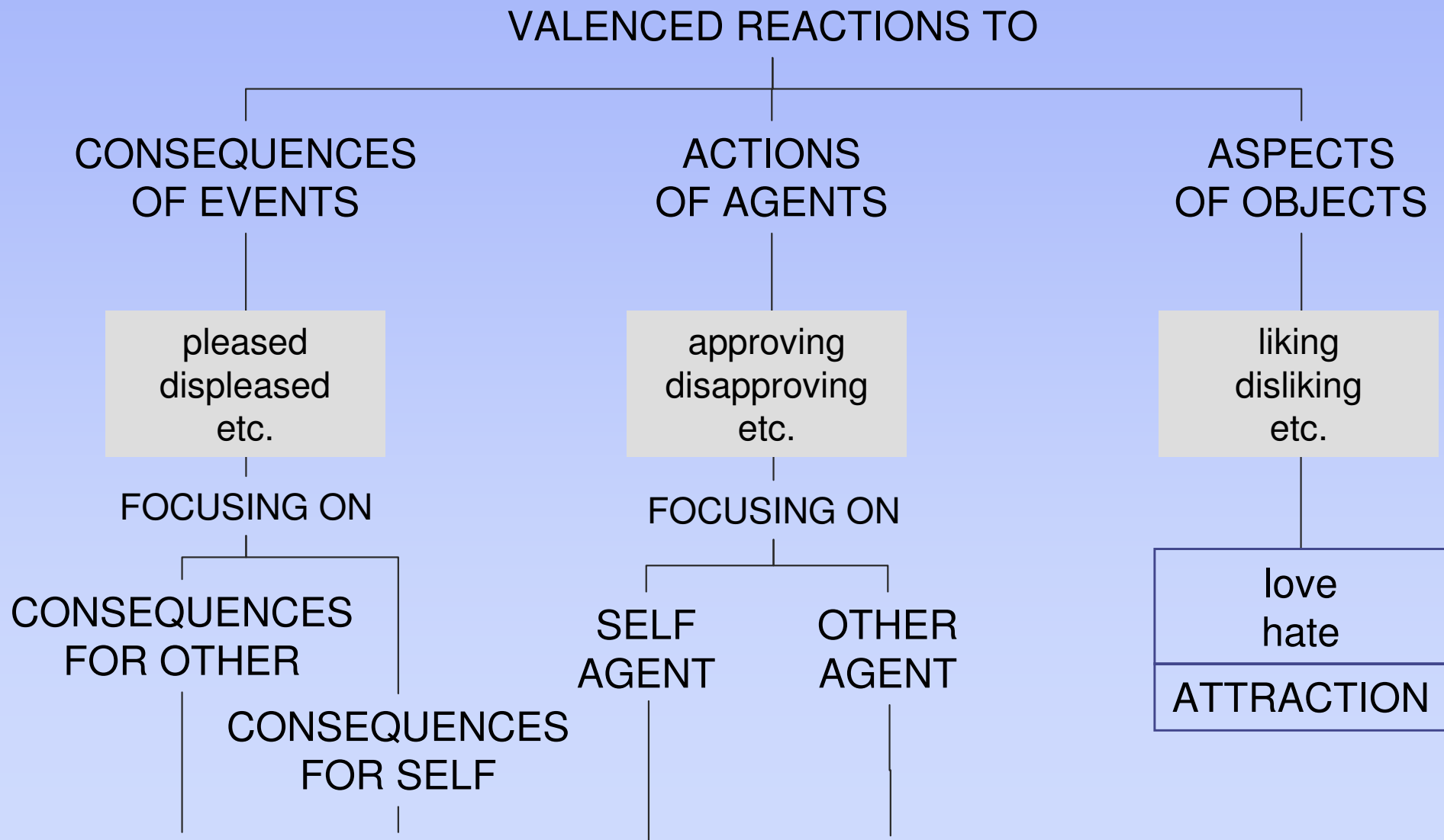
- OCC's approach relies on the assumption that **humans** are inherently **goal-driven beings**
- **Lack of empirical studies** validating the predictions

# I. OCC theory: Emotions as valenced reactions

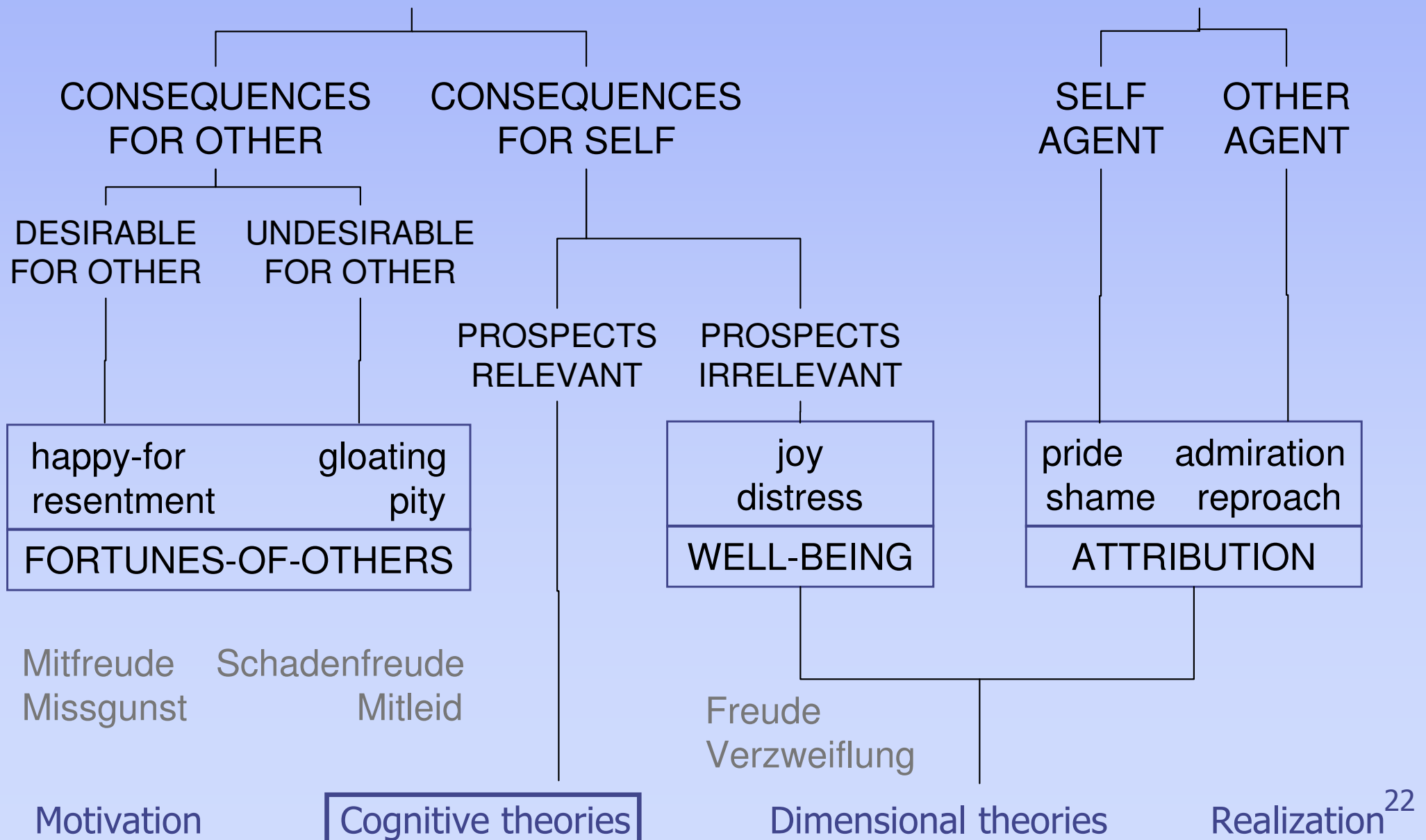
Three major aspects of the world, or changes in the world, upon which one can focus:

1. **Events**, with interest in their consequences
  2. **Agents**, with regard to their actions
  3. **Objects**, with interest in certain aspects or imputed properties of them *qua* objects
- Objects can take the role of agents, e.g., a car might be treated as an agent, when malfunctioning!  
(compare slide 15, “The media equation”, 1998)

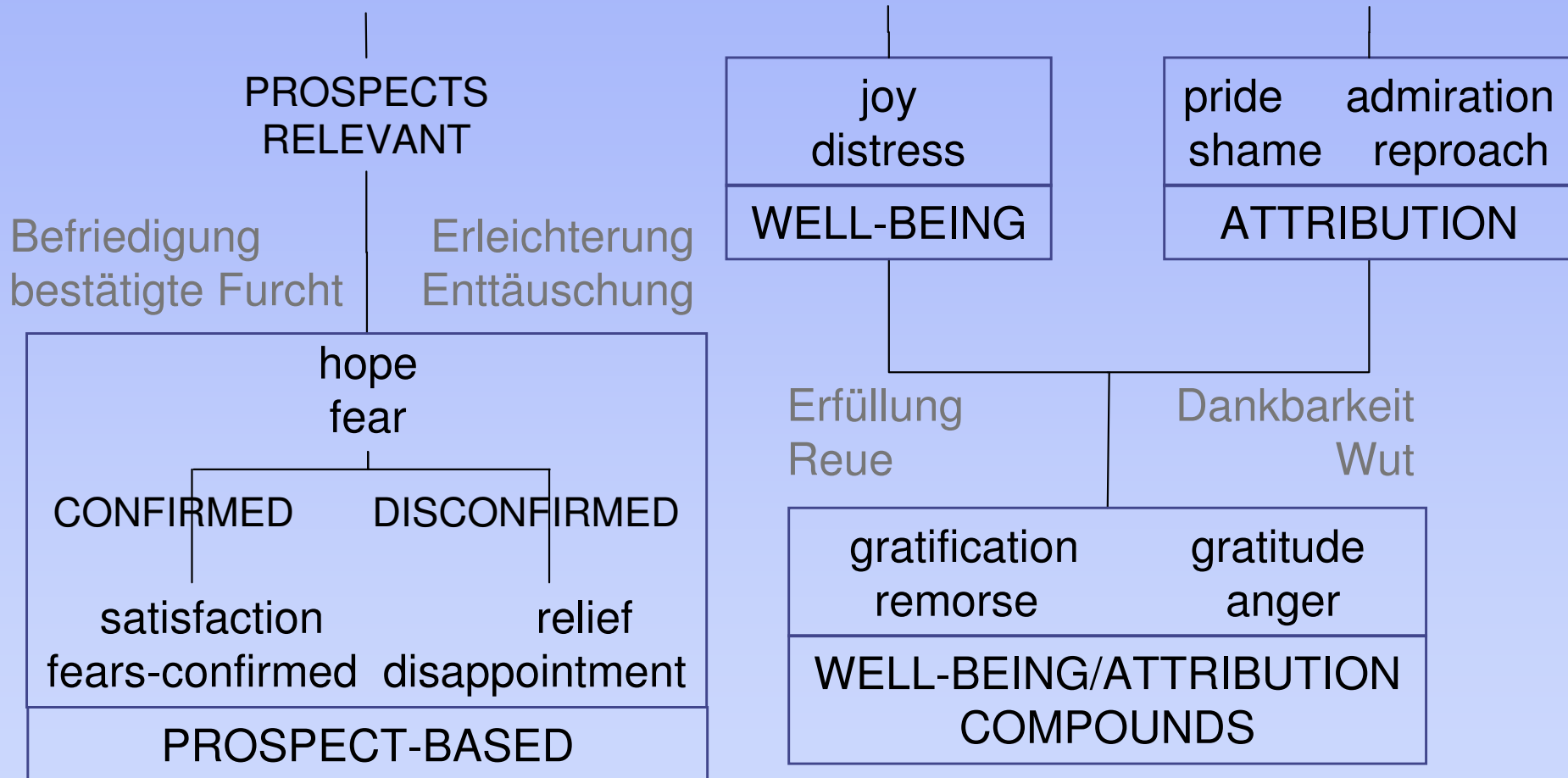
# I. OCC theory: Global structure of emotion types



# I. OCC theory: Global structure of emotion types



# I. OCC theory: Global structure of emotion types



(from page 19, Figure 2.1)

# I. OCC theory, intensity (1): The central intensity variables

Variables associated with the three foci of valenced reactions:

## 1. Reactions to Events:

*desirability* with reference to *goals*

## 2. Actions of Agents:

*praiseworthiness* with reference to *standards*

## 3. Reactions to Objects:

*appealingness* with reference to *attitudes*

→ Global variables also have to be taken into account!



# I. OCC theory, intensity (2): The global intensity variables

Global intensity variables affecting all emotions:

- 1. Sense of reality:** The degree to which the event, agent or object seems real
- 2. Proximity:** Reflects the psychological proximity of the emotion-inducing event, agent or object
- 3. Unexpectedness:** Is assessed *after* an event occurred, in contrast to the local variable likelihood
- 4. Arousal:** Is not cognitive but physiological in nature and has a relatively slow rate of decay

# I. OCC theory, intensity (3): The local intensity variables

Local intensity variables and their emotion types:

- 1. Prospect-based emotions:** likelihood, effort and realization (+ desirability)
- 2. Fortune-of-others emotions:** desirability-for-other, liking and deservingness (+ desirability)
- 3. Attribution emotions:** strength-of-cognitive-unit and expectation-deviation (+ praiseworthiness)
- 4. Attraction emotion:** familiarity (+ appealingness)

# I. OCC theory: “emotion specifications” (1)

## JOY EMOTIONS

**Type Specification:** (pleased about) a desired event

**Tokens:** contented, cheerful, delighted, ecstatic, elated, euphoric, feeling good, glad, happy, joyful, jubilant, pleasantly surprised, pleased, etc.

**Variables affecting intensity:** (1) the degree to which the event is desirable

**Example:** The man was pleased when he realized he was to get a small inheritance from an unknown distant relative.

↑  
Primary emotion?

→ Empathy?

## HAPPY-FOR EMOTIONS

**Type Specification:** (pleased about) an event presumed to be desirable for someone else

**Tokens:** delighted-for, happy-for, pleased-for etc.

**Variables affecting intensity:**

(1) the degree to which the desirable event for the other is desirable for oneself

(2) the degree to which the event is presumed to be desirable for the other

(3) the degree to which the other person deserved the event

(4) the degree to which the other person is liked

**Example:** Fred was happy for his friend Mary because she won a thousand dollars.

# I. OCC theory: "emotion specifications" (2)

## RELIEF EMOTIONS

**Type Specification:** (pleased about) the disconfirmation of the prospect of an undesirable event

**Tokens:** relief

**Variables affecting intensity:**

- (1) the intensity of the attendant fear emotion
- (2) the effort expended in trying to prevent the event
- (3) the degree to which the event is realized

**Example:** The employee was relieved to learn that he was not going to be fired.

## DISAPPOINTMENT EMOTIONS

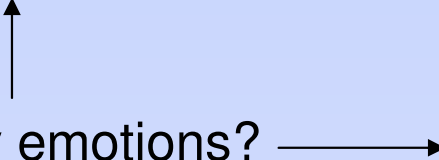
**Type Specification:** (displeased about) the disconfirmation of the prospect of a desirable event

**Tokens:** dashed-hopes, despair, disappointment, frustration, heartbroken, etc.

**Variables affecting intensity:**

- (1) the intensity of the attendant hope emotion
- (2) the effort expended in trying to attain the event
- (3) the degree to which the event is realized

**Example:** The girl was disappointed when she realized that she would not be asked to the dance after all.

Secondary emotions? 

# I. OCC theory:

## Computational Tractability (1)

(1)

**IF DESIRE**  $(p, e, t) > 0$

**THEN set JOY-POTENTIAL**  $(p, e, t) = f_j [ | \mathbf{DESIRE}(p, e, t) |, I_g(p, e, t) ]$

Where  $| \mathbf{DESIRE}(p, e, t) |$  is the absolute value of a function that returns the degree of desirability that a person,  $p$ , assigns to some perceived event,  $e$ , at time,  $t$ , under normal conditions, and where  $I_g(p, e, t)$  is a function that returns the value of the combined effects of the global intensity variables.

(2)

**IF JOY-POTENTIAL**  $(p, e, t) > \mathbf{JOY-THRESHOLD}(p, t)$

**THEN set JOY-INTENSITY**  $(p, e, t) =$

**JOY-POTENTIAL**  $(p, e, t) - \mathbf{JOY-THRESHOLD}(p, t)$

<sup>1</sup> We shall not elaborate here on how the global variables (e.g., sense of reality, proximity, and unexpectedness) might be represented because such an excursion would take us too far afield.

**Type Specification:** (pleased about) a desired event

**Tokens:** contented, cheerful, delighted, ecstatic, euphoric, feeling good, glad, happy, joyful, jubilant, pleasantly surprised, pleased, etc.

**Variables affecting intensity:** (1) the degree to which the event is desirable  
 (2) the degree to which the event is realized when he realized he was to get a small inheritance from an unknown distant relative.

# I. OCC theory: Computational Tractability (2)

(3)

**IF FEAR-POTENTIAL**  $(p, e, t) > 0$  **AND DISBELIEVE**  $(p, e, t_2)$  **AND**  $t_2 \geq t$   
**THEN set RELIEF-POTENTIAL**  $(p, e, t) =$   
 $f_j [ \text{FEAR-POTENTIAL} (p, e, t), \text{EFFORT} (p, e), \text{REALIZATION} (e, t_2),$   
 $l_g (p, e, t) ]$

(4)

**IF RELIEF-POTENTIAL**  $(p, e, t_2) > \text{RELIEF-THRESHOLD} (p, t_2)$   
**THEN set RELIEF-INTENSITY**  $(p, e, t_2) =$   
 $\text{RELIEF-POTENTIAL} (p, e, t_2) - \text{RELIEF-THRESHOLD} (p, t_2)$   
**AND reset FEAR-POTENTIAL**  $(p, e, t_2) =$   
 $f_f [ | \text{DESIRE} (p, e, t_2) |, \text{LIKELIHOOD} (p, e, t_2), l_g (p, e, t_2) ]$   
**ELSE set RELIEF-INTENSITY**  $(p, e, t_2) = 0$

## RELIEF EMOTIONS

**Type Specification:** (pleased about) the disconfirmation of the prospect of an

undesirable event

to prevent relief

**Variables affecting intensity:**

(1) the intensity of the attendant fear emotion

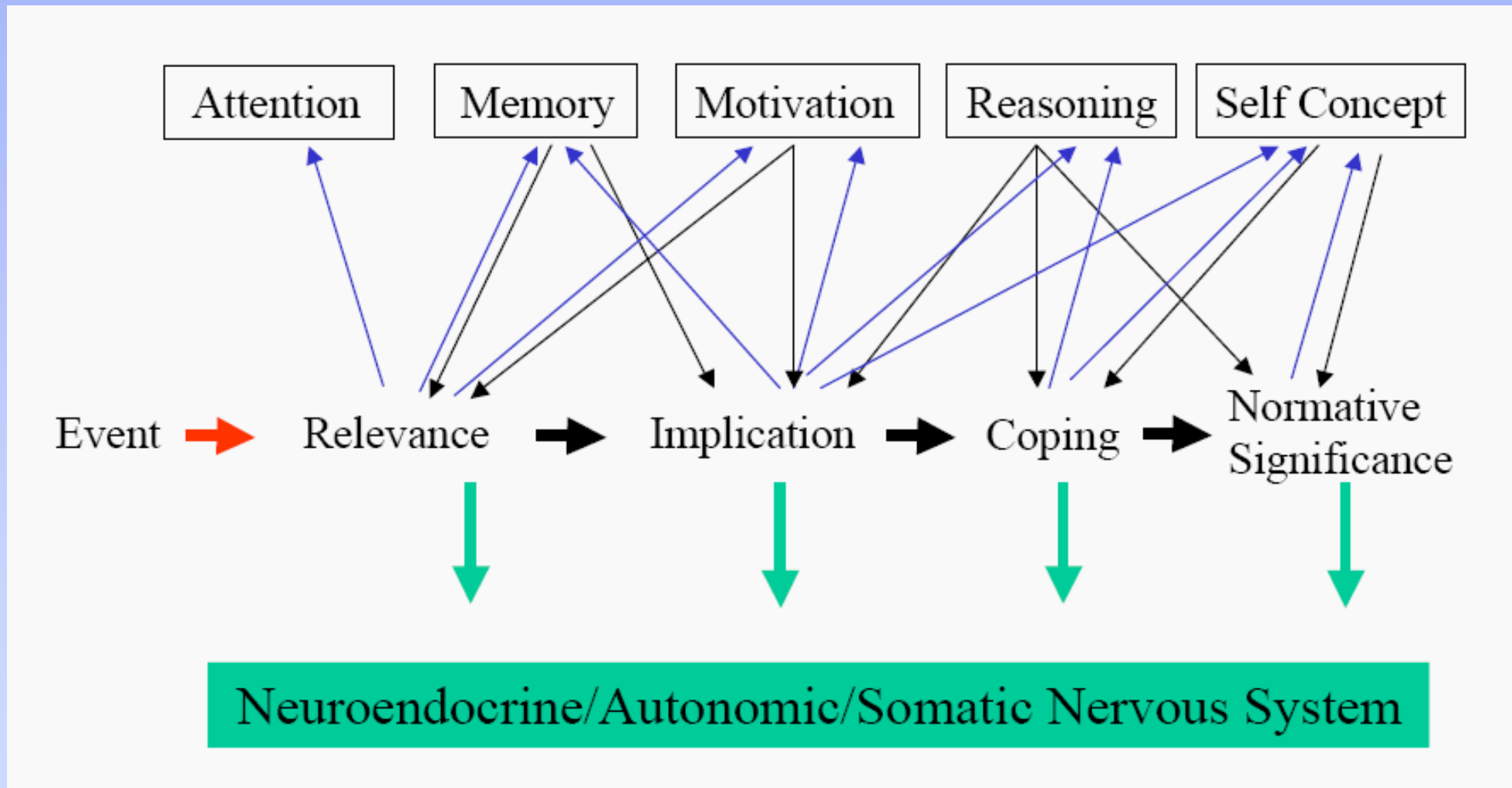
(2) the degree expected in trying to prevent the event

(3) the degree expected in the event's realized

**Example:** The employee was relieved to learn that he was not going to be fired.

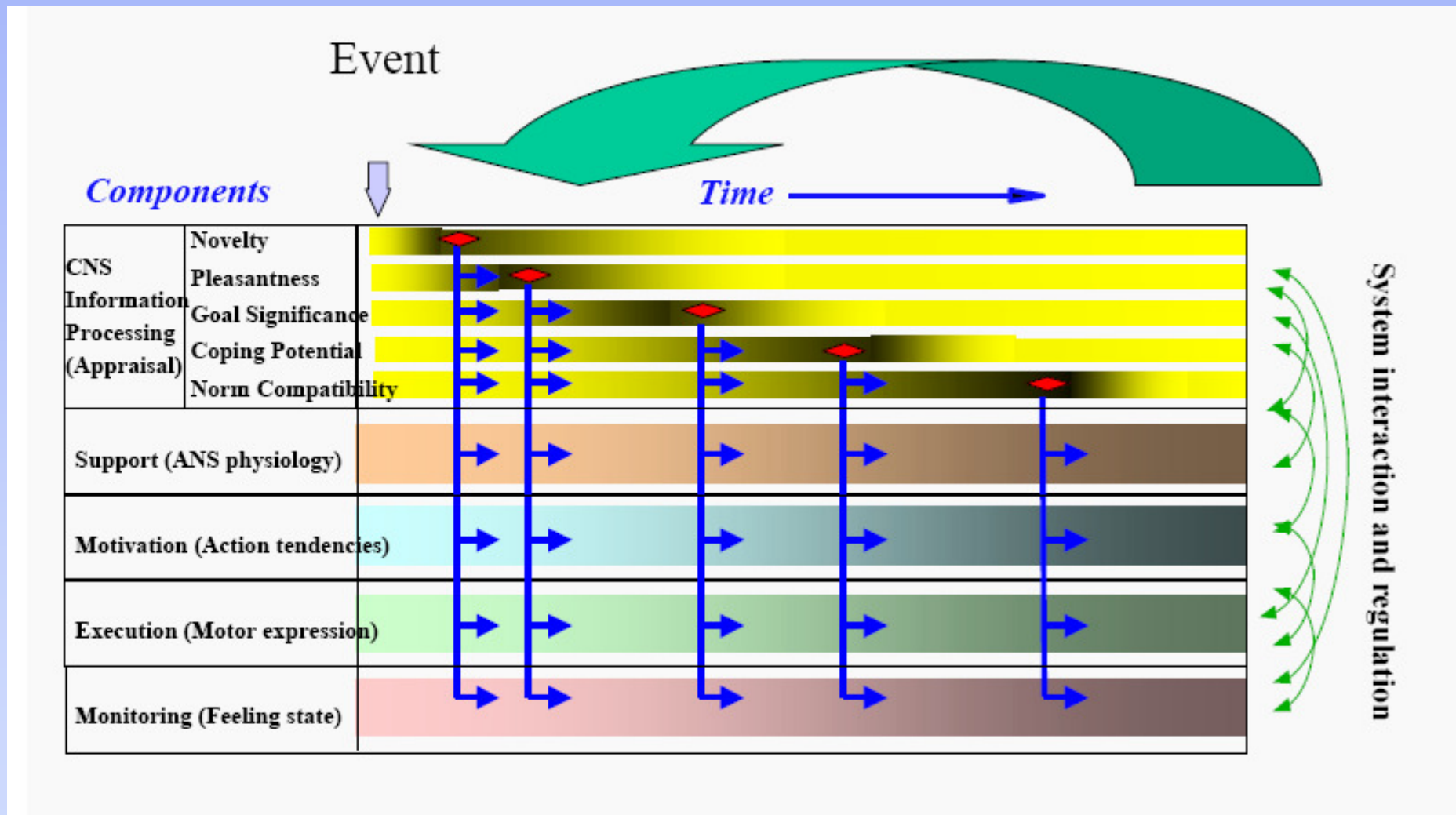
(from page 186, Rules 5 and 6)

# II. Scherer's theory: Component process model - Appraisal



(from Scherer, 2006)

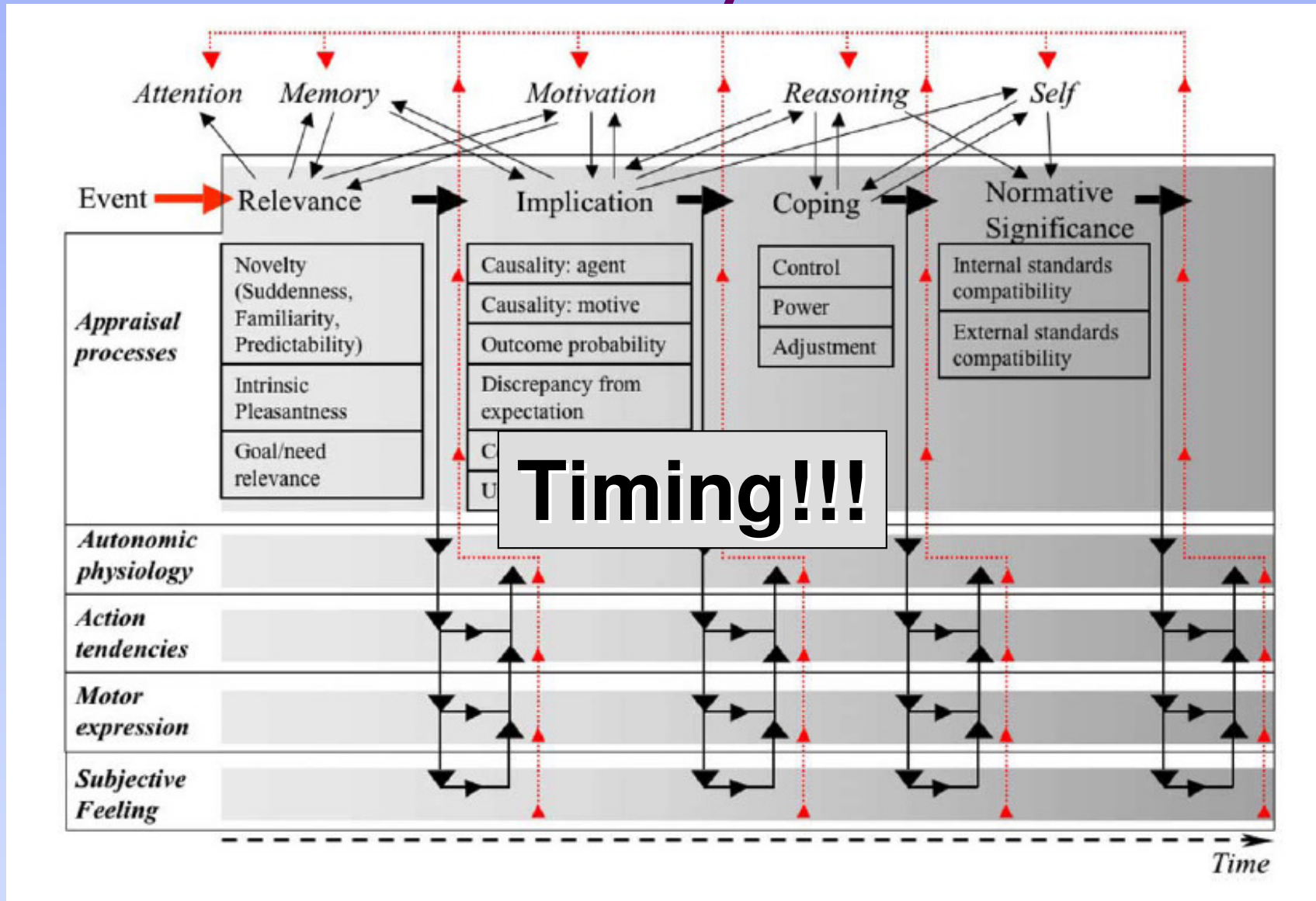
# II. Scherer's theory: Component process model (1984)



(from Scherer, 2006)



# II. Scherer's theory:



(from Scherer, 2006)

# Summary: Cognitive theories are..

Members of **appraisal theory** of emotion:

Emotions are  
**elicited and differentiated**  
on the basis of the  
**subjective evaluation**  
of an event on a set of standard criteria

→ *So far, most appraisal theorists have focused on the **prediction of basic or modal emotions** rather than more **diffuse feeling states**.*

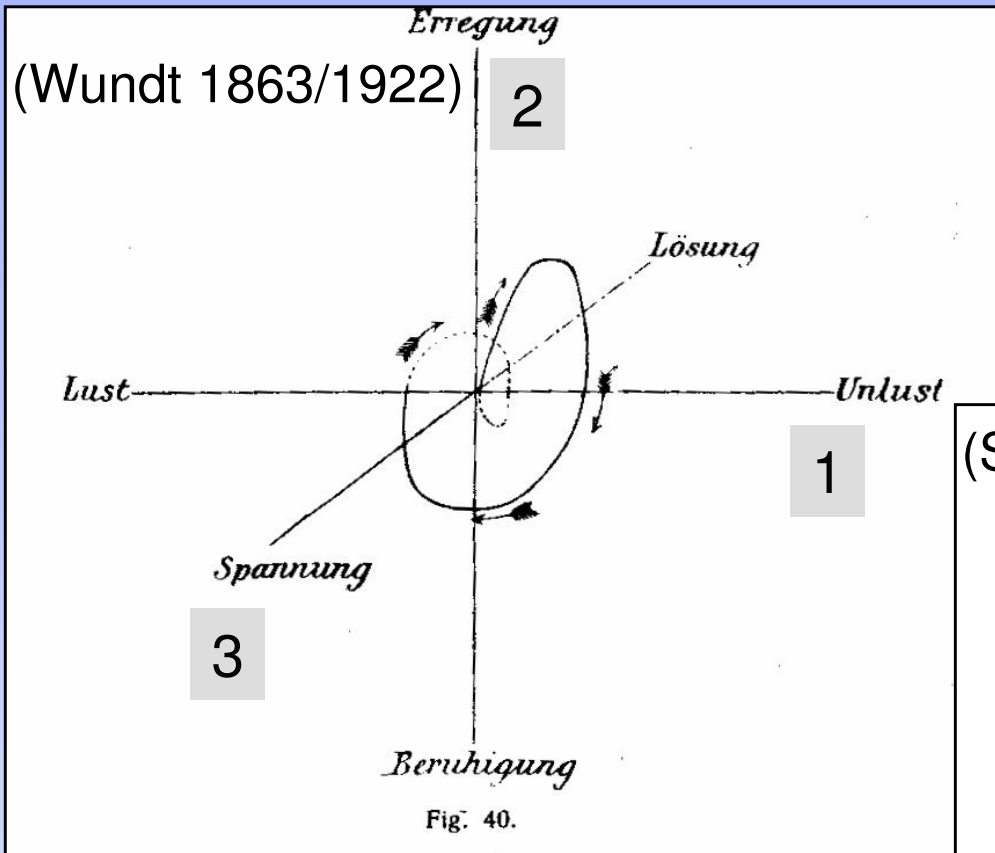
*(Scherer 2006)*

# Part III:

## Dimensional emotion theories

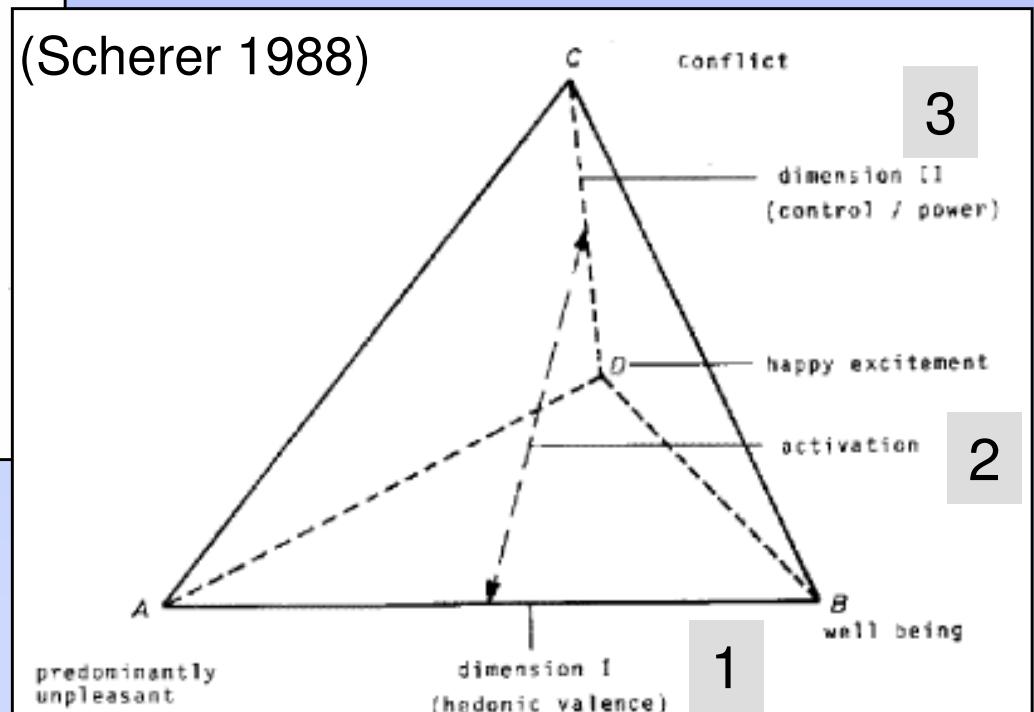
("sentio, ergo sum")

# What is a "diffuse feeling state": → Dimensional theories



Mostly three dimensions:

1. Pleasure (Lust – Unlust)
2. Arousal (Erregung – Beruhigung)
3. Dominance (Spannung – Lösung)



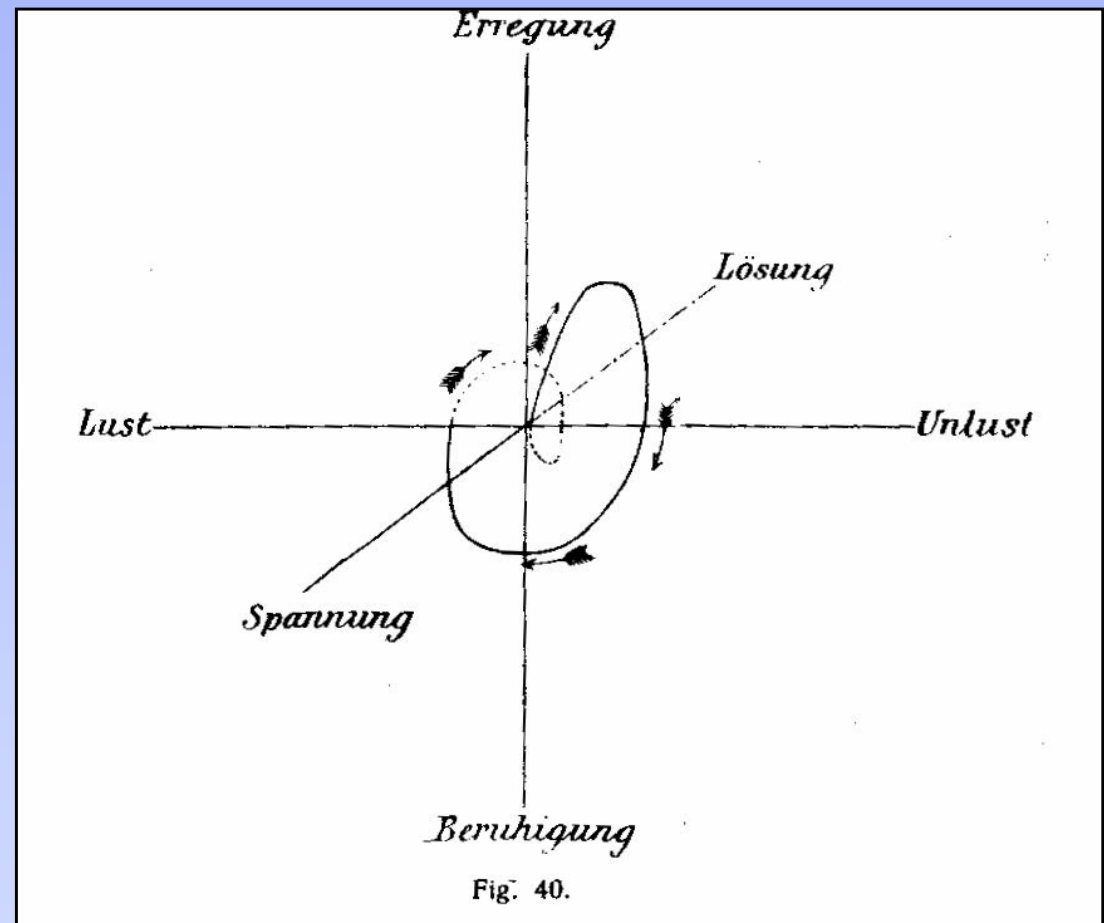
- What is the difference between emotions and feelings?
- How many emotions/feelings?

# Dimensional theories: Wundt (1863) "Gefühlsverlauf"

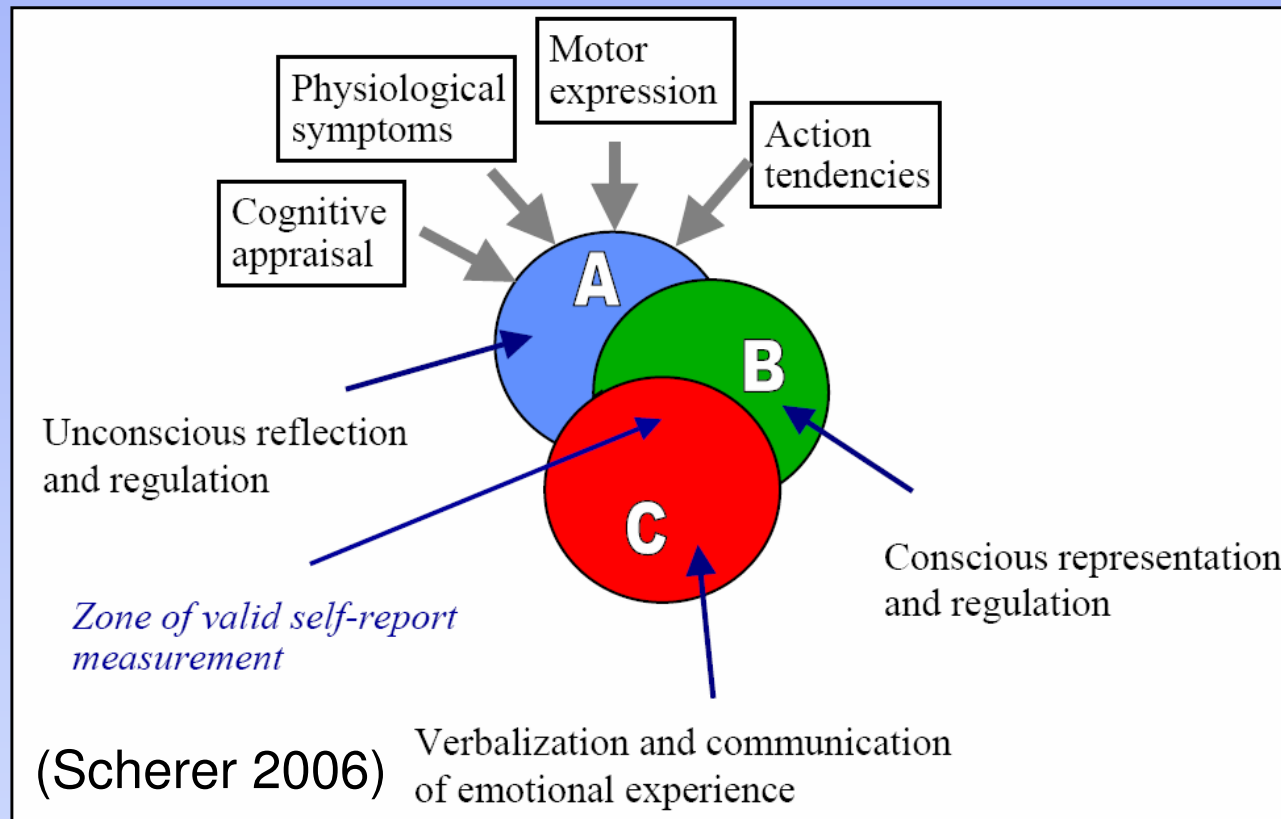
„Indem ein einzelner Punkt nur ein **momentanes Gefühl** bezeichnet, wird aber **irgendein konkretes Geschehen immer** in einem bestimmten [...] **Gefühlsverlauf** bestehen [...]“ (p. 245)

→ Dynamics of emotions

→ Unlimited number of feelings possible!?



# The verbalization problem: Conscious vs. unconscious



- Most processes believed to remain unconscious (A)
- Even not all conscious representations can be verbalized (B)

→ Zone of valid self-report measurement lies on top! (C)

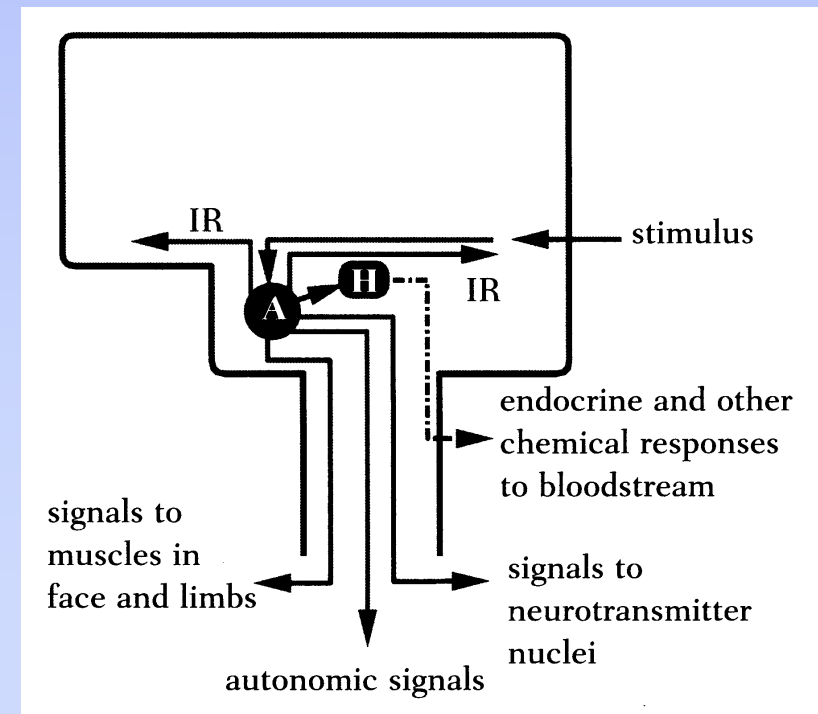
→ Scherer focuses on the underlying **processes** of appraisal

# Damasio (1994): Coming from neurobiology

1. Fast, hard-wired stimulus response patterns:
  - **primary emotions** (fear, anger, joy, etc.)
    - Triggering fight-or-flight behaviors
    - Genetically anchored / unconscious?
2. Cognitively elaborated, deliberative behaviors:
  - **secondary emotions** (jealousy, relief, shame, etc.)
    - May use memories and expectations
    - “Social emotions” to be developed during infancy
    - Consciousness needed?

# Damasio (1994): Primary emotions

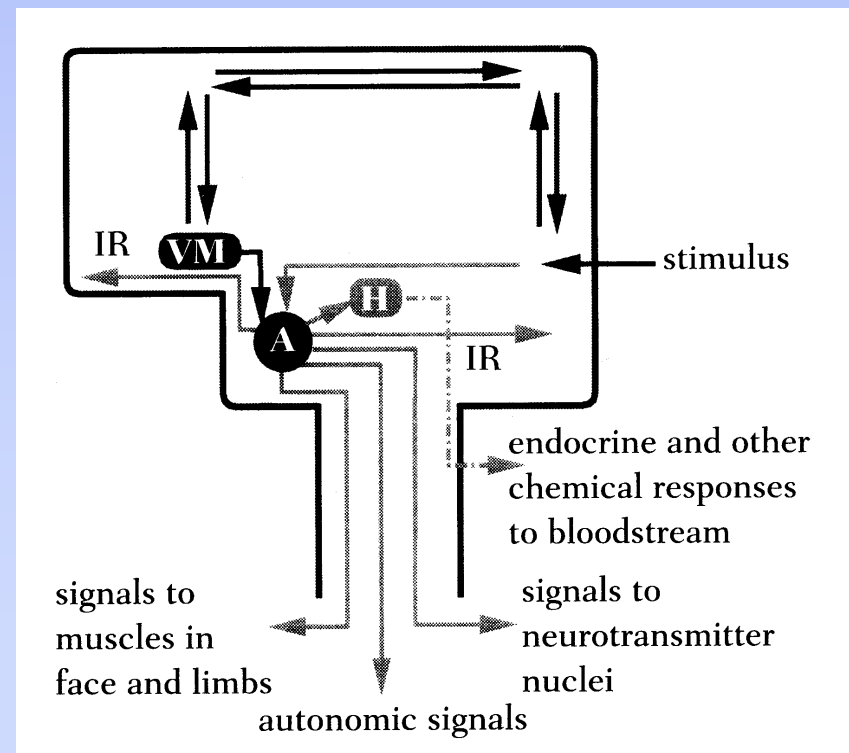
- ◆ “[..] we are wired to respond with an emotion, in preorganized fashion, when certain features of stimuli in the world or in our bodies are perceived, alone or in combination. Examples [..] include size (as in large animals); large span (as in flying eagles); types of motion (as in reptiles); certain sounds (such as growling); certain configurations of body state (as in the pain during a heart attack).” (p. 131)
- ◆ Stimulus activates amygdala (A)
- ◆ Various responses ensue:
  - Internal responses (IR)
  - Muscular responses
  - Visceral responses (autonomic signals)
  - Responses to neurotransmitter nuclei and hypothalamus (H)
  - Hypothalamus gives rise to endocrine & other chemical responses





# Damasio (1994): Secondary emotions

- ◆ “[..] I believe that in terms of an individual’s development [primary emotions] are followed by *secondary emotions*, which occur once we begin experiencing feelings and forming *systematic connections between categories of objects and situations, on the one hand, and primary emotions, on the other.*” (p. 134)
- ◆ Stimulus activates amygdala (A)
- ◆ Stimulus is analyzed in the thought process
  - May activate ventromedial region of frontal cortices (VM)
  - VM acts via the amygdala (A)
- ◆ Secondary emotions utilize the machinery of primary emotions
- ◆ VM depends on A to express its activity



# Ortony (2005): Affect & Proto-affect

*information flow  
and interrupts*



*sensory inputs*



*motor outputs*



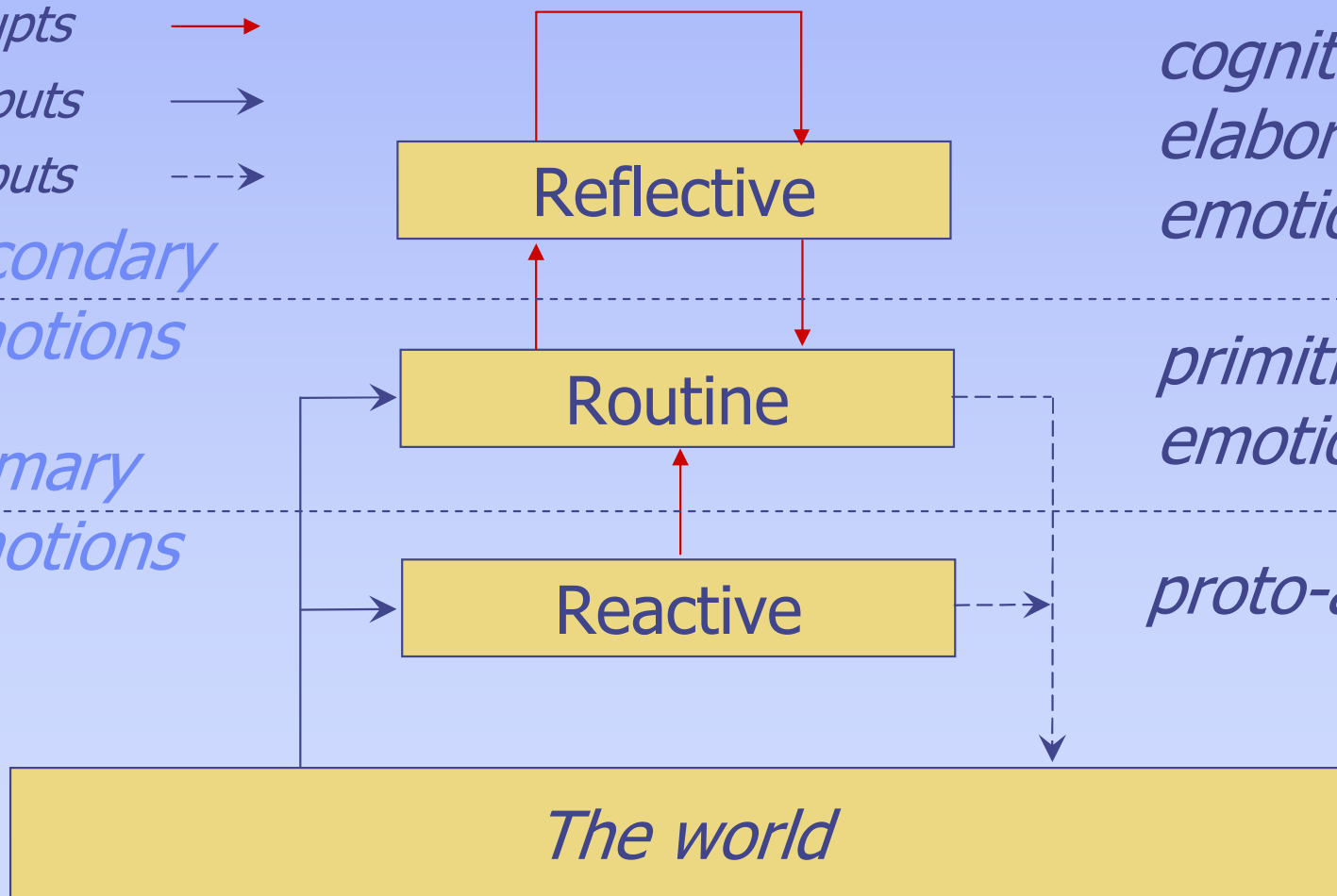
*cognitively  
elaborated  
emotions*

*secondary  
emotions*

*primitive  
emotions*

*primary  
emotions*

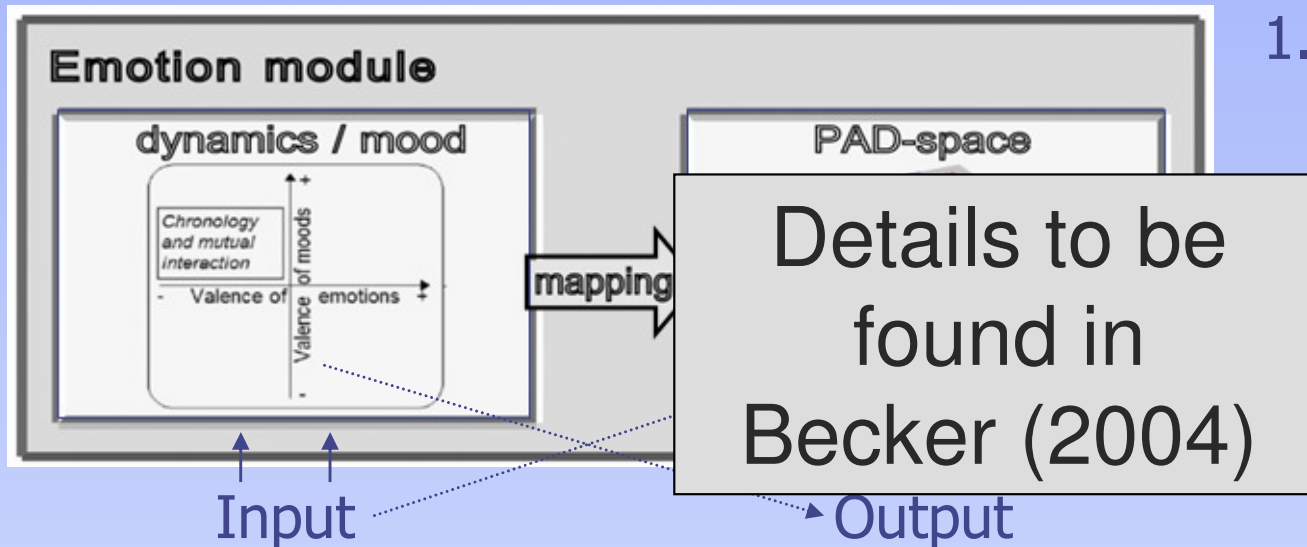
*proto-affect*



# Part IV: Realization ("videor, ergo sum")

# The emotion module: Based on dimensional theory

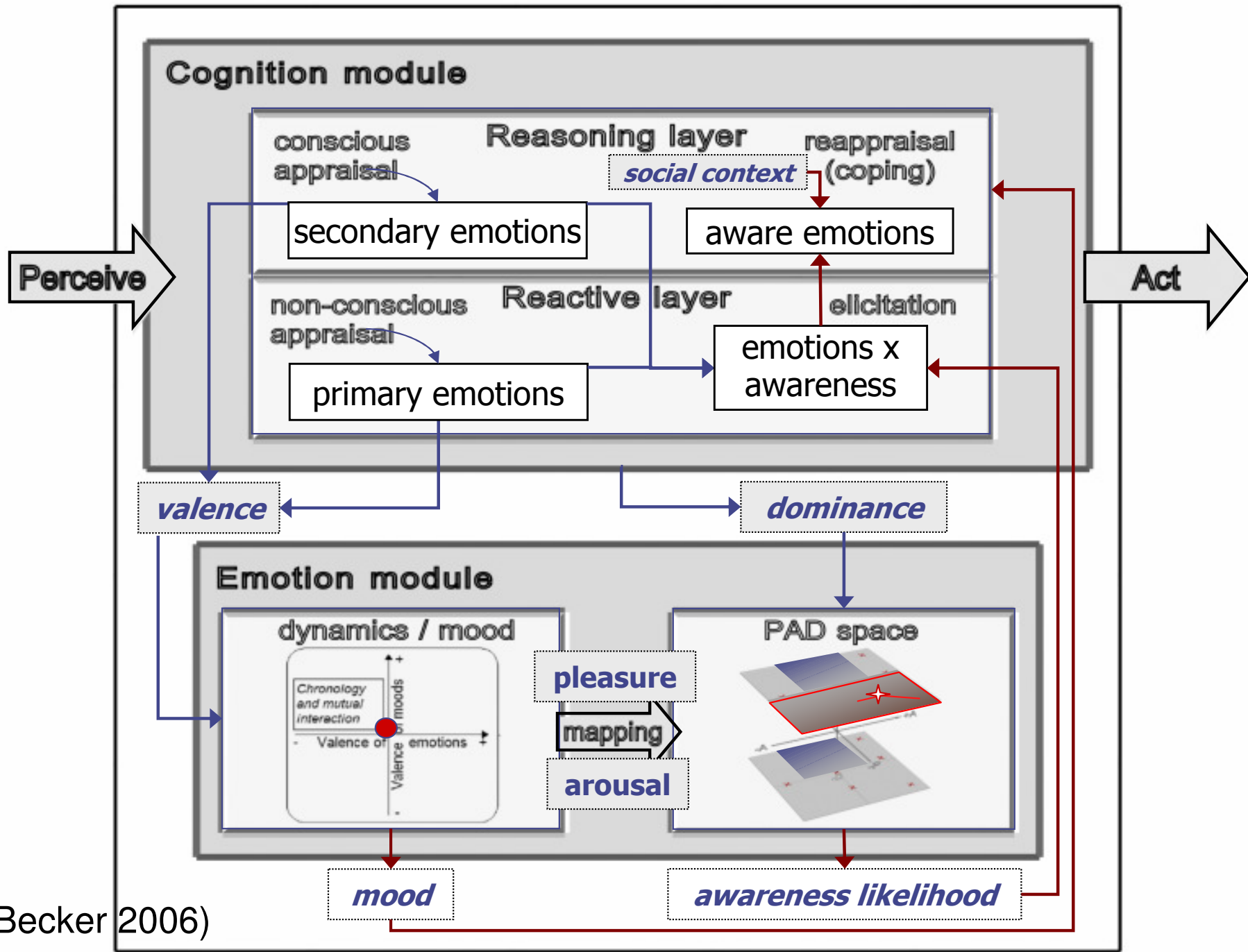
## *Dynamics of emotional impulses*



1. Dynamic component:  
Conceptual linkage of  
emotions and moods

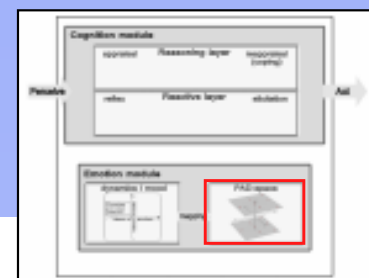
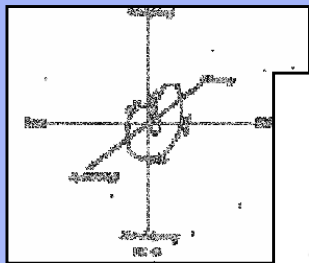
PAD-mapping:  
Mapping onto emotion  
categories in PAD space

- Valence of emotions: short-time system state
- Valence of moods: longer lasting system state
- (P)leasure ranges from joy (+P) to reluctance (-P)
- (A)rousal ranges from mental awareness (+A) to sleepiness (-A)
- (D)ominance describes the agent's feelings of control over the situation

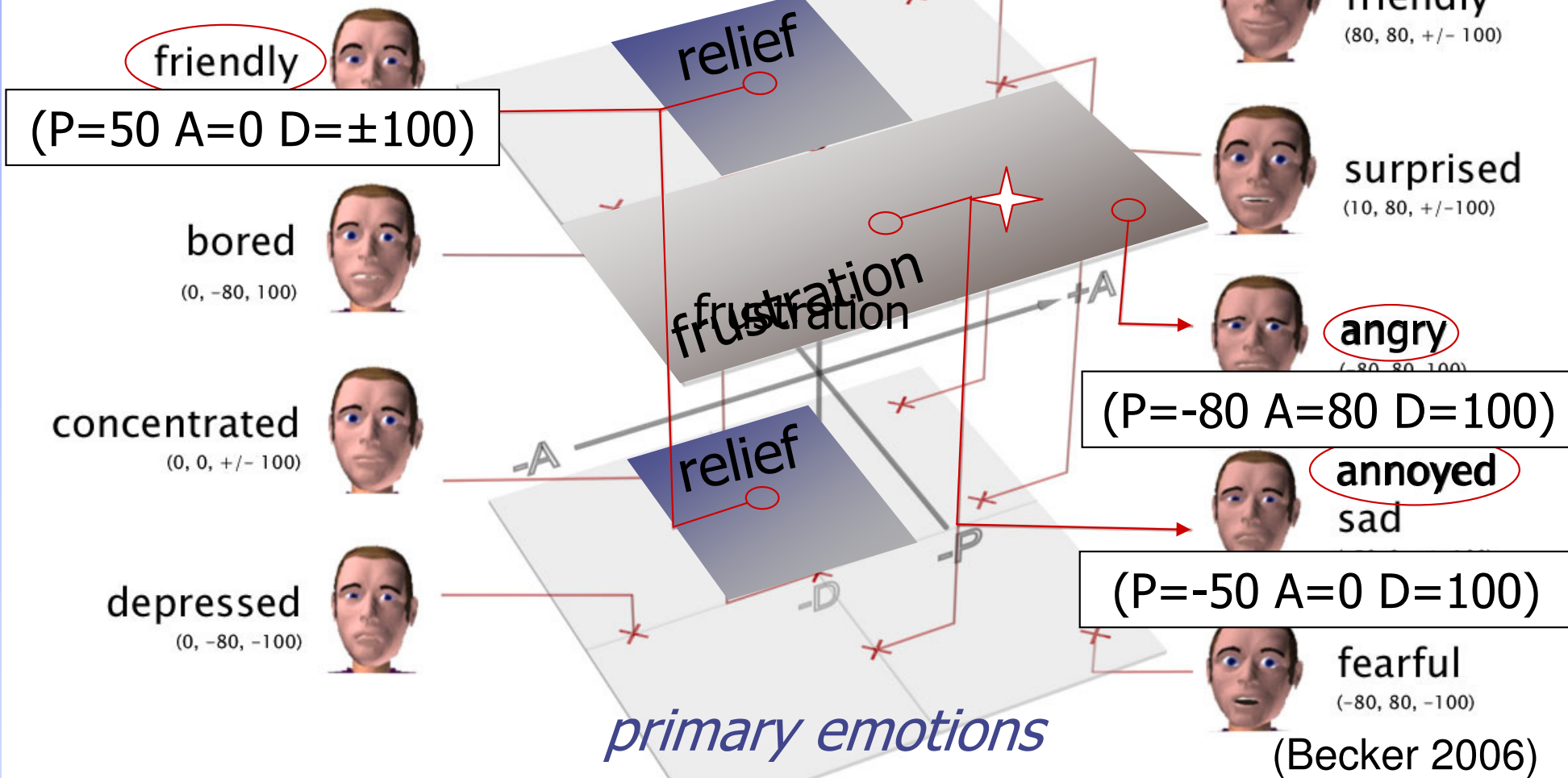


(Becker 2006)

# PAD-space



*secondary emotions*



*awareness likelihood* = (0.3 \* , 0.2 \* , 0.6 \* )

# Thank you very much!

