Machine-Learning for Human-Machine Interaction

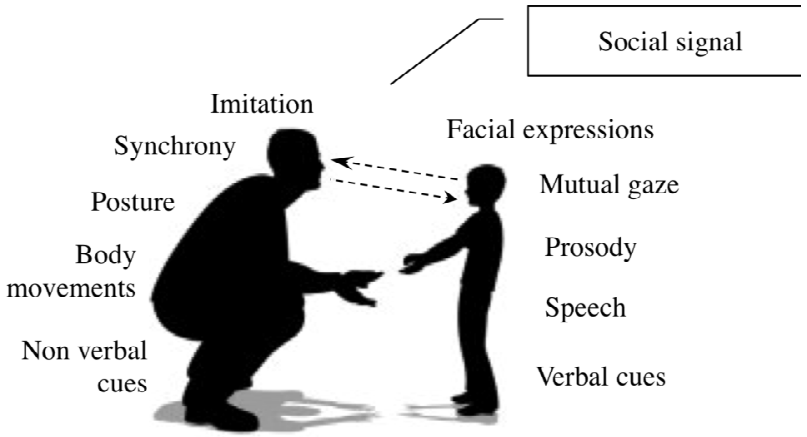
Mohamed CHETOUANI
Professeur des Universités

Institut des Systèmes Intelligents et de Robotique (ISIR)
Sorbonne Université

mohamed.chetouani@sorbonne-universite.fr

From Impersonal to Interpersonal Interaction

- Human-Machine Interaction has traditionally been inspired by Interpersonal Interaction
- Interpersonal interaction involves the exchange of verbal and non-verbal messages
- As for Human-Human Interactions, Interpersonal Human-Machine Interactions could be ranged along a continuum: from impersonal at one end to highly personal at the other.

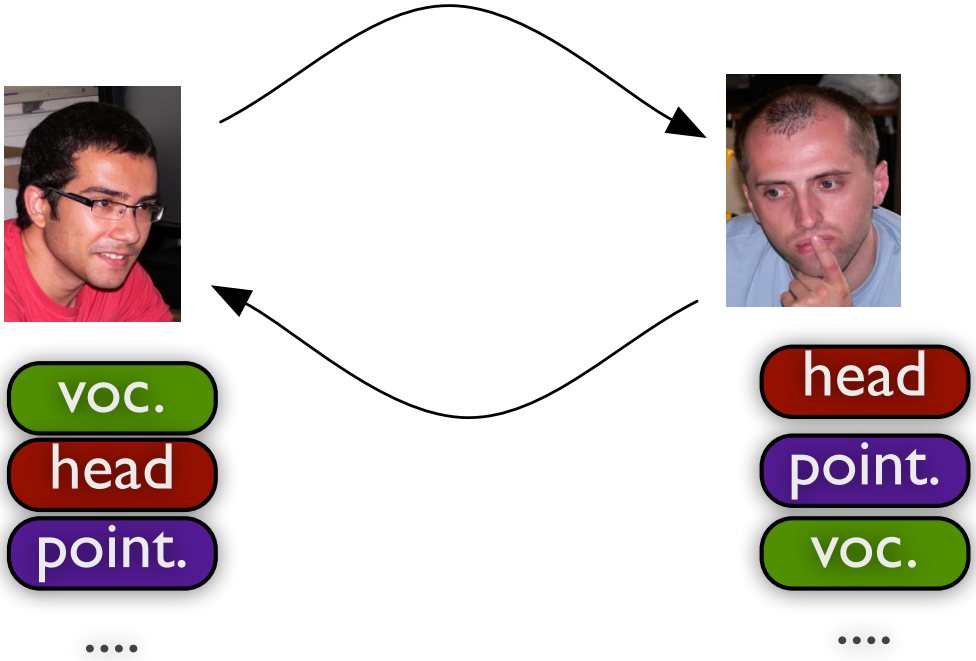


Impersonal

Interpersonal

Interpersonal interaction is a highly dynamic process

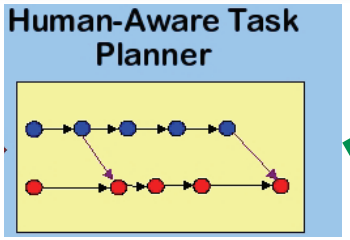
- Behavioral dynamics: non-verbal signals (e.g. gesture)
- Individual dynamics: multimodal signals (e.g. gesture + speech)
- Interpersonal dynamics: social signals (e.g. gazing in response to pointing of the partner)
- The «Telegraphist model» of communication (Shannon) is usually considered in Human-Computer Interaction
  - Emit / Receive / Respond (Answer)



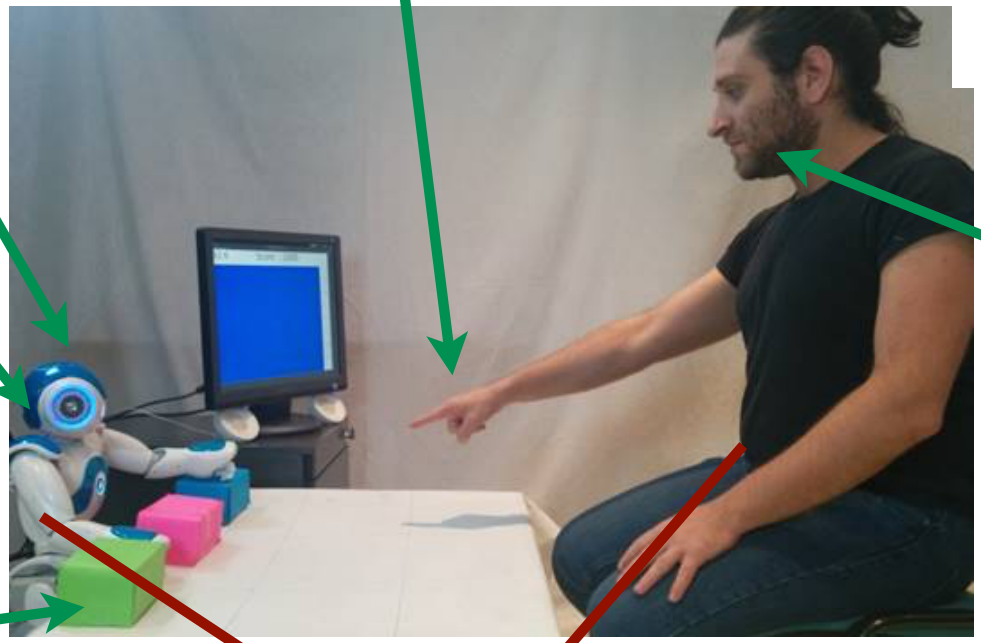
- **While Interpersonal Interaction in Humans involves «connected individuals»:**
  - **Interdependent individuals**
  - **Inherently relational (e.g. role)**
  - **Transactional (a person serves simultaneously as speaker and listener)**

Non-verbal behaviors in Human-Machine Interaction

Task Learning, Planning, Decision



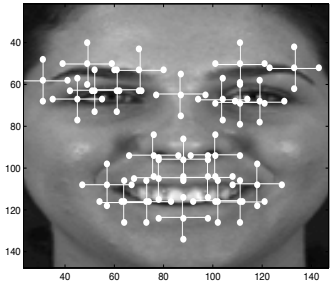
Action recognition



Automatic Speech Recognition and Natural Language Processing

« Look at this box »

Affective computing



Behavior synthesis

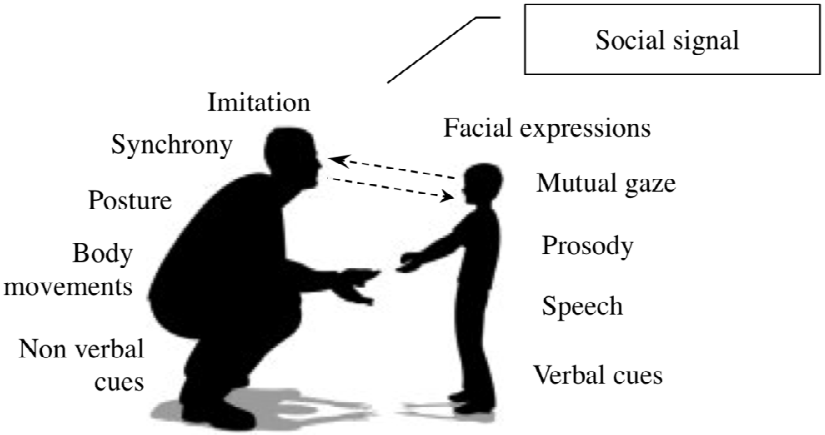
Object recognition

Social signal processing

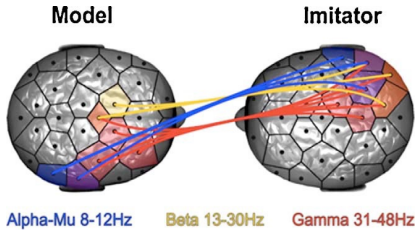
Modeling, Analysis and Synthesis of Machine-detectable traces of psychological and social phenomena (e.g., mimicry, engagement, conflict, interest...)

Intepersonal Human-Human Interaction

- Human communication dynamics (Delaherche et al. 2012a):
  - Inter-personal interaction: mutual and dynamic influence of partners
  - Key concepts in psycho-pathology, social robotics and Human-Machine Interaction



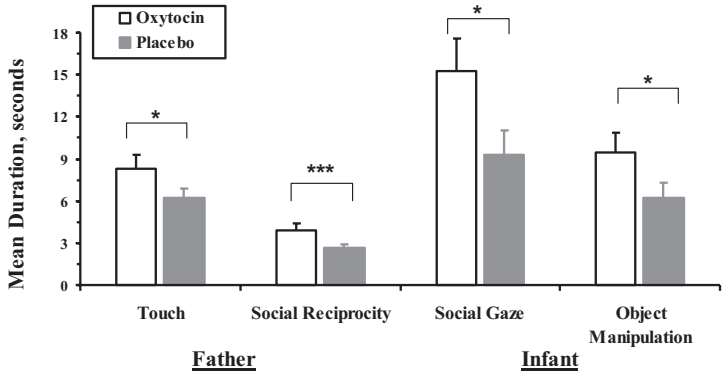
Cerebral basis



Dumas et al., 2011

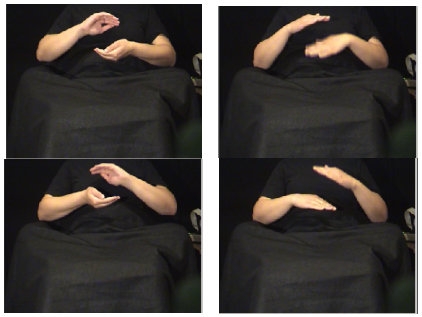
Still face experiments

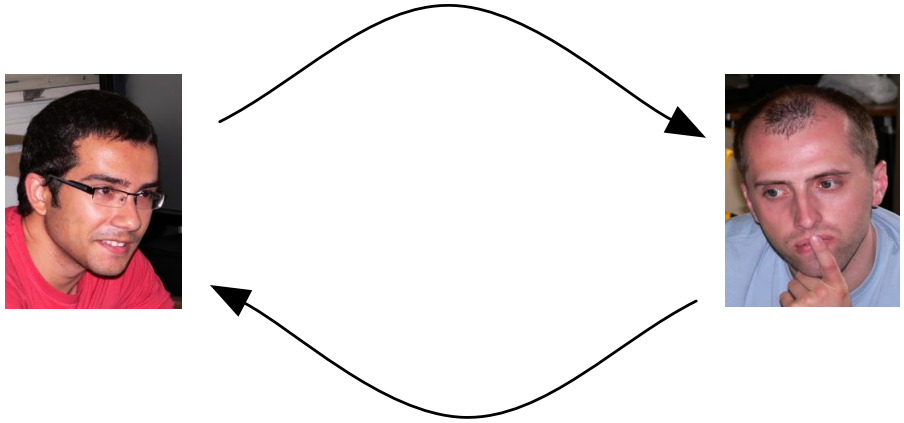
Physiology

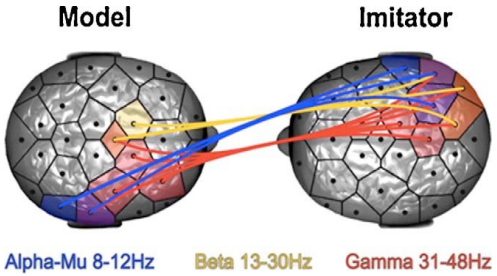


Weisman et al., 2012

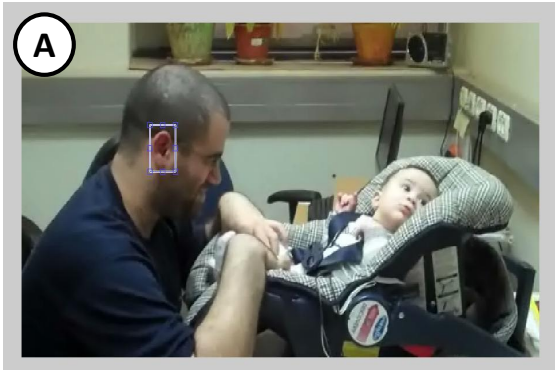
But these processes involve more than behaviors...

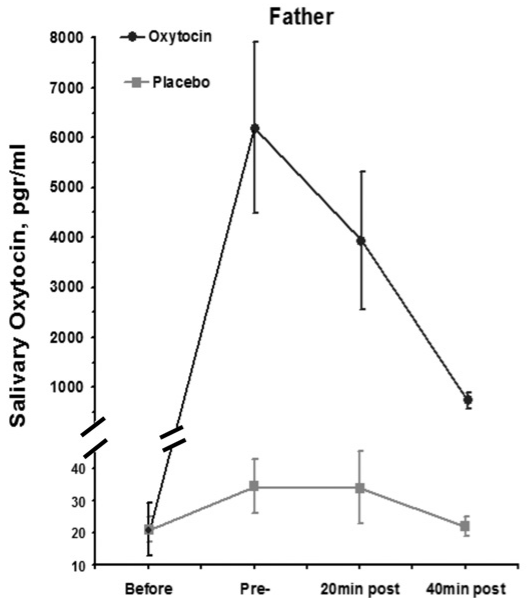


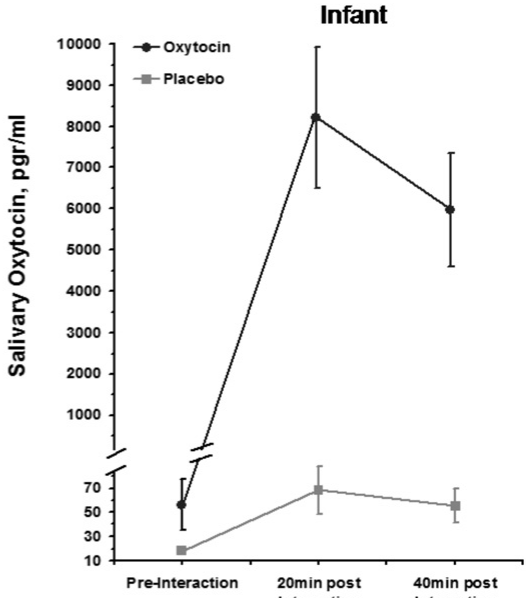




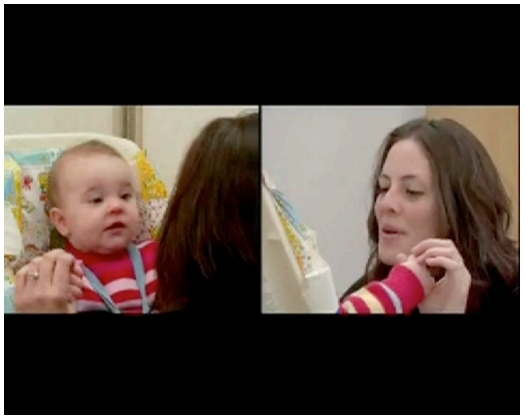
- **While Interpersonal Interaction in Humans involves «connected individuals»:**
  - **« Two body neuroscience »**
  - **« Biological synchrony »**



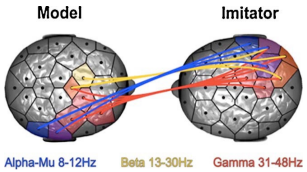


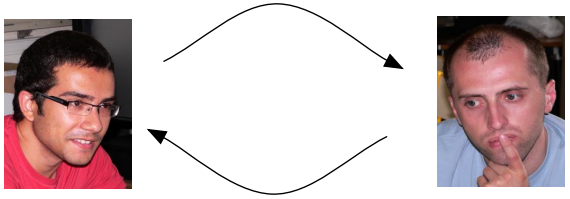


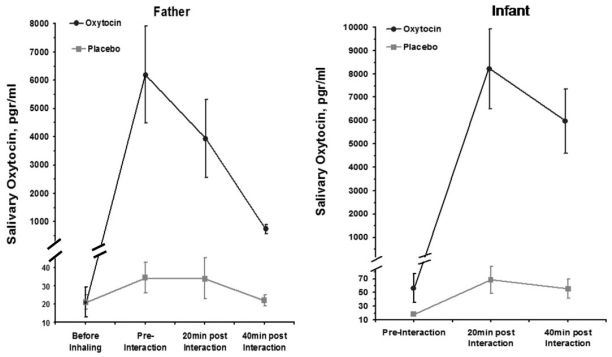
Timing issues



Still face experiments





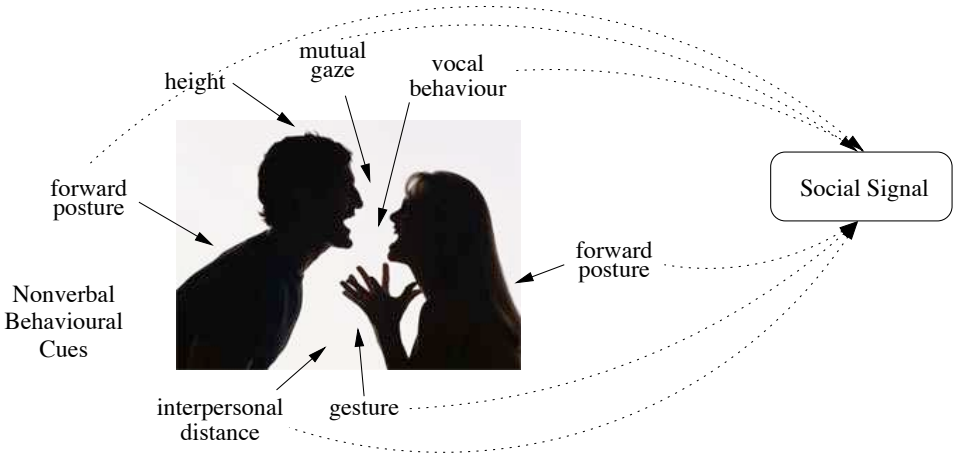


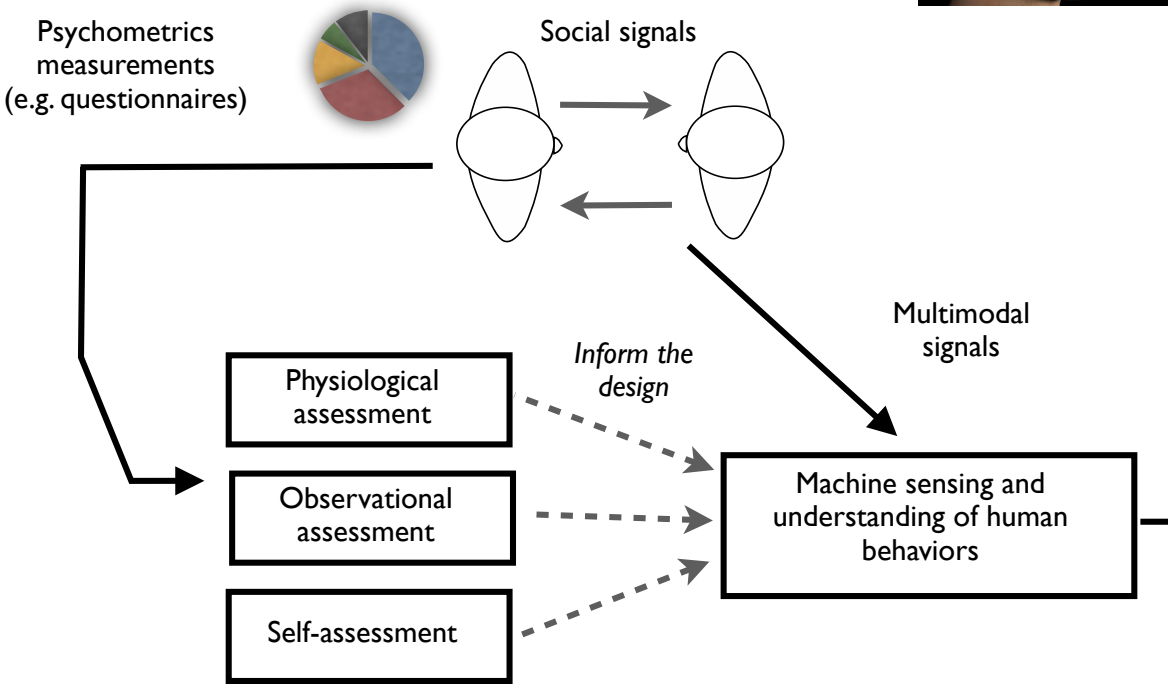
Neural level | Behavioral level | Hormonal level

miliseconds | seconds | minutes

Time-scales

Social Signal Processing, Social Robotics, Affective Computing





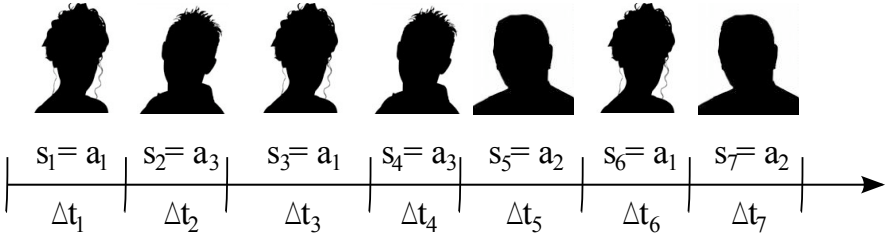
About Social Intelligence

Mapping non-verbal cues into social traits (emotions, role…) and vice-and-versa

Some examples

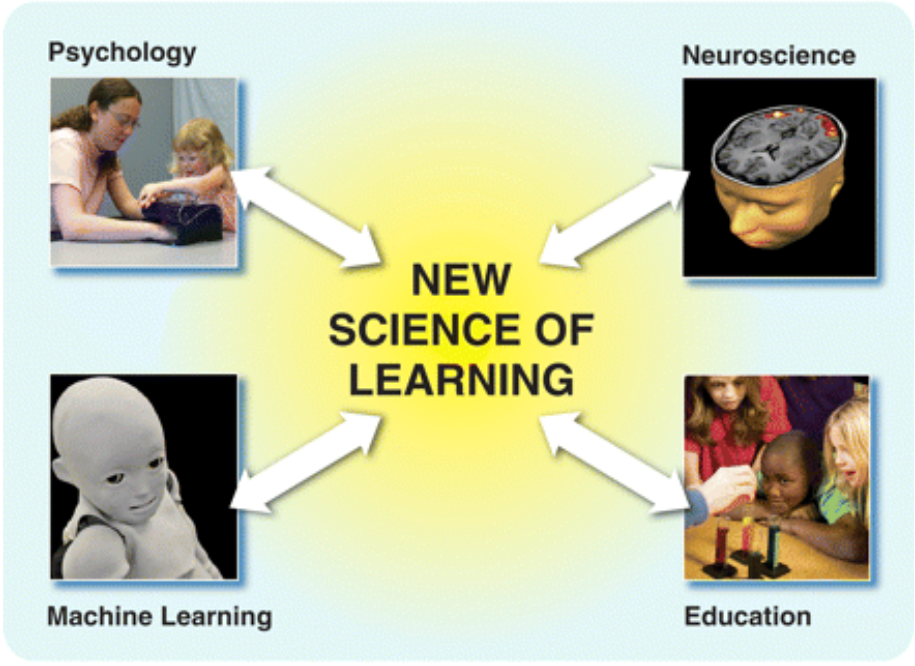
Reality mining (Pentland, 2005)

Role recognition during meetings (Vinciarelli, 2007)



Building Autonomous Sensitive Artificial Listeners (Schröder et al., 2011)

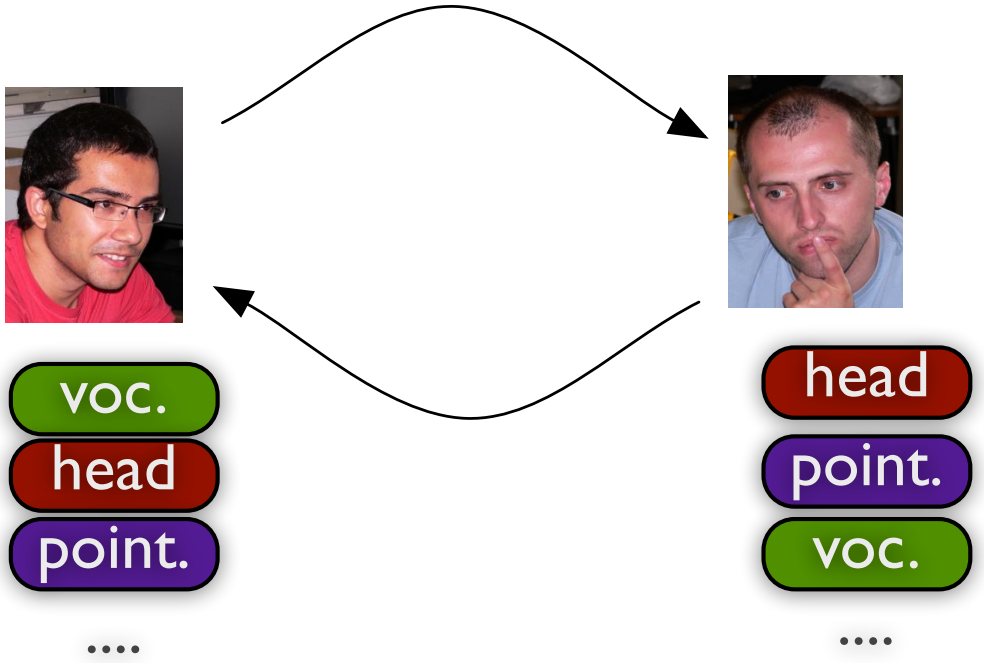
Social learning (Meltzoff et al., 2009)



Verbal and Non-verbal Communication

- Verbal communication: communication which occurs according to a linguistic code known to the participants using one or several methods (oral, gestural, tactile or graphical)
  - Elements that codify meaning are called "signs"

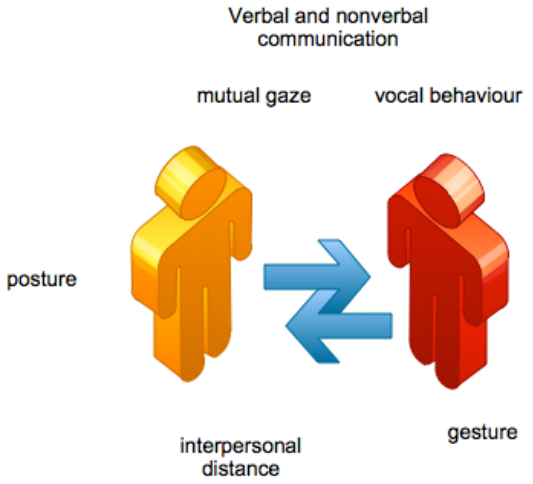
- Non-verbal communication: communication without « words »
  - Facial expressions, gaze, touching, tone of voice…



Why using non-verbal communication?

- Words have limitations: explain the shape, directions…
- Nonverbal signals are powerful: expression of inner feelings
- Nonverbal message are likely to be more genuine: nonverbal behaviors cannot be controlled as easily as spoken words
- Help to send complex messages: A speaker can modify and add information to the verbal message through simple nonverbal signals

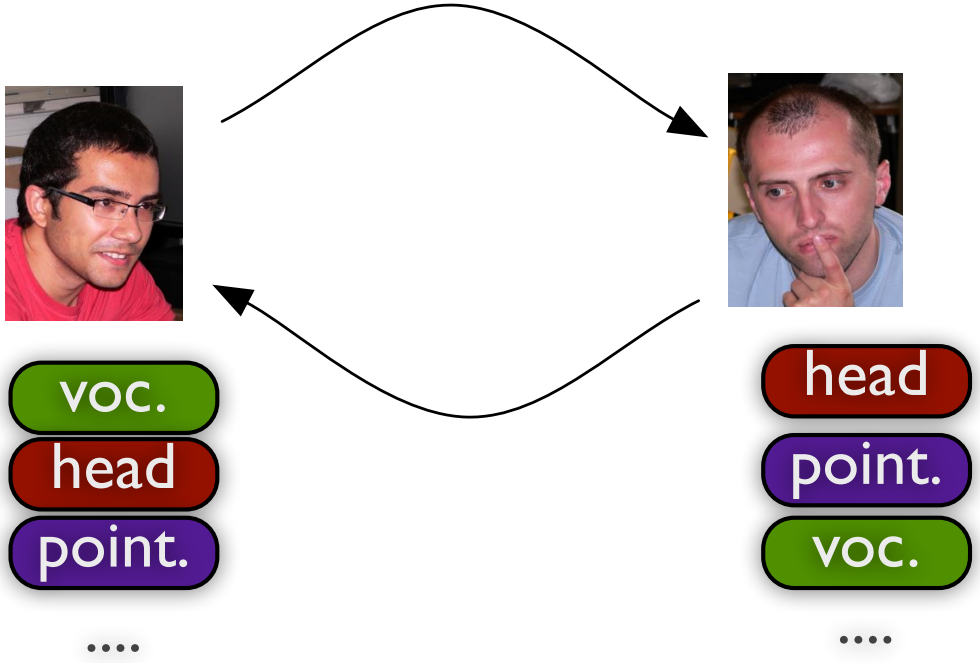
- Verbal communication: semantic content of the message
- Non-verbal communication: gesture, gaze, tonal expression



Non-verbal communication

- Nonverbal communication is learned shortly after birth and practiced and refined throughout a person's life

- Nonverbal behavior is a continuous source of signals which convey information about feelings, mental state, personality, and other traits.



Meltzoff & Moore : Imitation of facial and manual gestures by human neonates. Science (1977)

Non-verbal communication

Nonverbal relevant patterns:

- Face and eyes behavior: facial expressions, gaze exchanges
- Vocal behavior: vocal outbursts, turn-taking, silences and pauses
- Gestures and postures: head movements and body orientation with respect to others
- Physical appearance (somatotype and clothes)
- Use of space and environment: seating arrangements and interpersonal distances

Vinciarelli et al., Social signal procesing: a survey of an emerging domain, Image and Vison Computing (2009)

Face and eye behaviors

Face is involved in several activities:

- Identity
- Speech production
- Communication of affective state, intentions => Facial expressions
- Personality, attractiveness, age and gender information

Face and eye behaviors

Facial expressions

Ekman proposed 6 basic emotions

- Disgust
- Happiness
- Sadness
- Anger
- Fear
- Surprise

Esposito, Cognitive Computation (2009)

Fig. 4. Basic emotions. Prototypic facial expressions of six basic emotions (disgust, happiness, sadness, anger, fear, and surprise).

Vocal behaviour

Vocal non-verbal behavior includes all spoken cues that surround the message and influence its actual meaning.

- Voice quality
- Filled pauses
- Non-linguistic vocalizations
- Silences
- Turn-taking

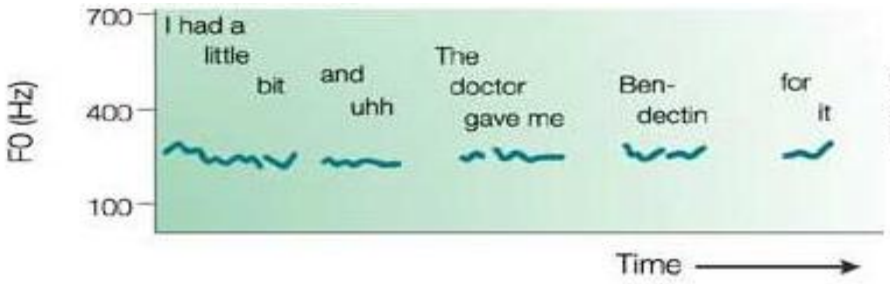
Vocal behaviour

Voice quality:

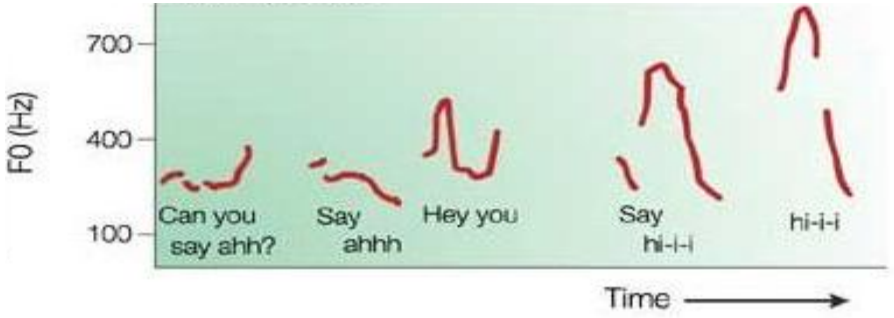
- Prosodic features: pitch, loudness, syllable length
- Acoustic features: fundamental frequency (f0), energy, duration

- How something is said:

Adult-directed speech



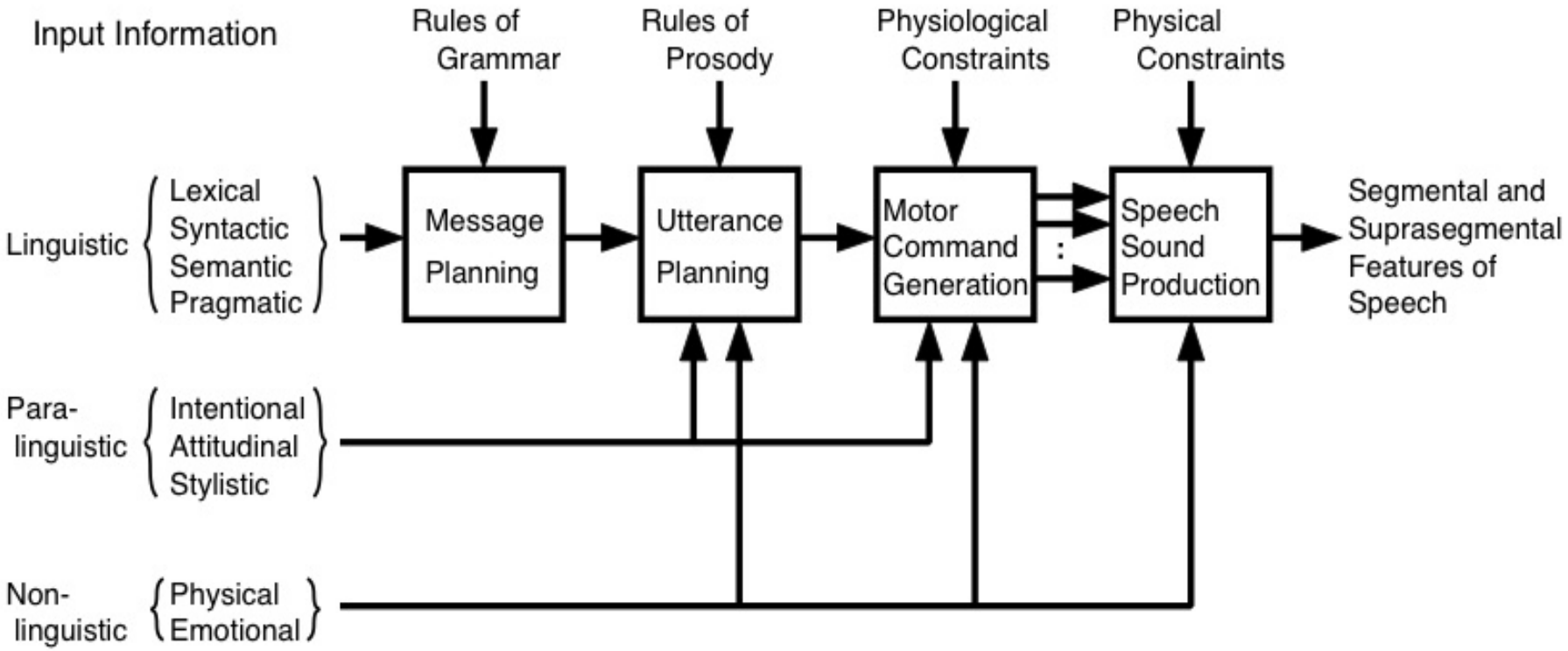
Infant-directed speech



Vocal behaviour

Prosody conveys several socially relevant cues:

- Emotions, perception of dominance, extroversion, speaker fluency
- Structure of the message (question, pitch accents, change rhythm…)



Vocal behaviour

Filled pauses:

- Non-words: "ehm", "ah-ah", "hum"
- Used in spontaneous speech
- May replace words: To answer a question "ehm"
- Disfluencies: embarrassment
- Back-channeling: fillers are used to accompany someone else speaking

Non-linguistic vocalizations (vocal outbursts)

- Non-verbal sounds
- Laughing, sobbing, crying, groaning…
- May accompany words

Vocal behaviour

Silences:

- Non-speech event
- Hesitation silence, psycholinguistic silence and interactive silence
- Sign of respect
- Ignoring persons, attract attention

Turn-taking

- Regulation of the conversation (coordination)
- Employs gaze, voice quality, fillers, silences…
- Synchronization, fluency, overlapping speech (disputes)

Gestures and postures

- Posture and body/limb change with emotion expressed
- Head inclination, posture shifting often accompany social affective states like shame and embarrassment.

- 90% of body gestures are associates with speech

McNeil, Hand and Mind: what gestures reveal about thought. University of Chicago (1996)

Jiang Xu, Patrick J. Gannon, Karen Emmorey, Jason F. Smith, Allen R. Braun Symbolic gestures and spoken language are processed by a common neural system Proceedings of the National Academy of

Gestures and postures

- To communicate specific meaning (thumbs up)
- To punctuate discourse
- To greet

Unconscious gestures:

- Self-manipulations (ear touching…)
- Manipulation of small objects (pens…)
- Self-protection gestures (moving legs…)

Gestures and postures

https://www.youtube.com/watch?v=1p7sOsDHLiw

Gestures and postures

Posture:

- Used to indicate attitudes, status, affective moods
- Self-confidence, energy, fatigue, boredom
- Conveys gross or overall affect while specific emotions are communicated by other signals (facial expressions, speech…).

- Face-to-face vs. parallel body orientation

- Congruence vs. Incongruence: symmetric postures

Physical appearance

- Height, body shape, hair color…
- Clothes, ornaments, make up…

- Attractiveness

- "Halo effect": first traits we recognize in other people

- Do not necessarily correspond to the reality
- See robot capacities vs. embodiment

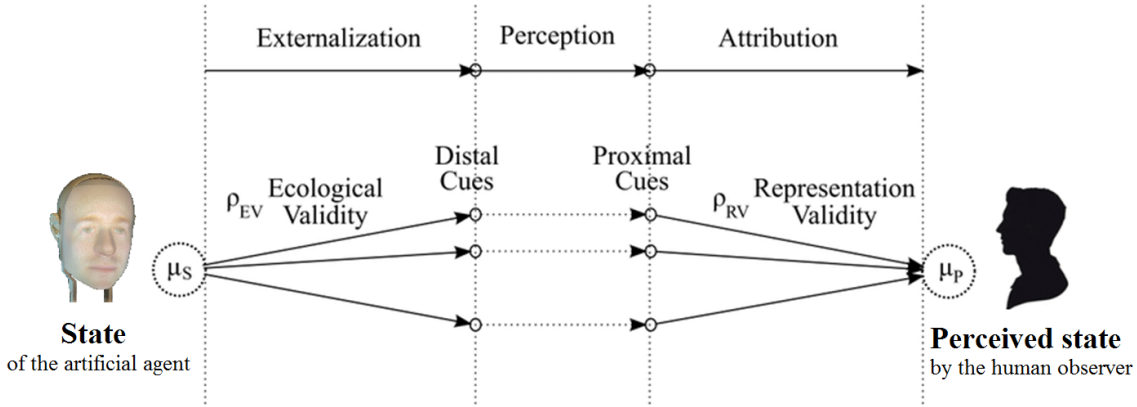
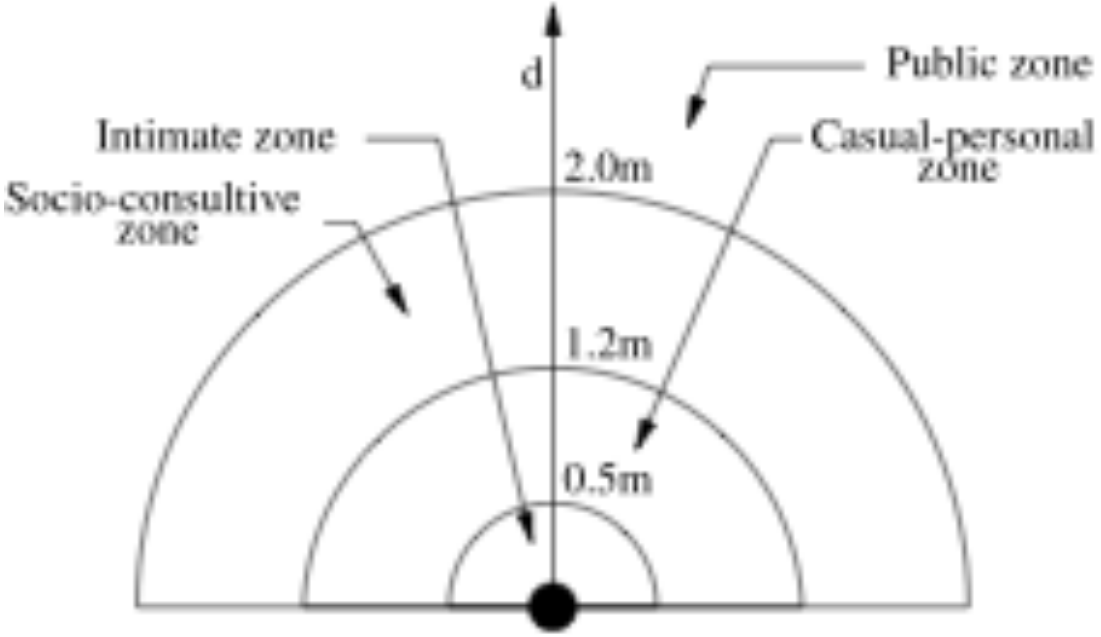


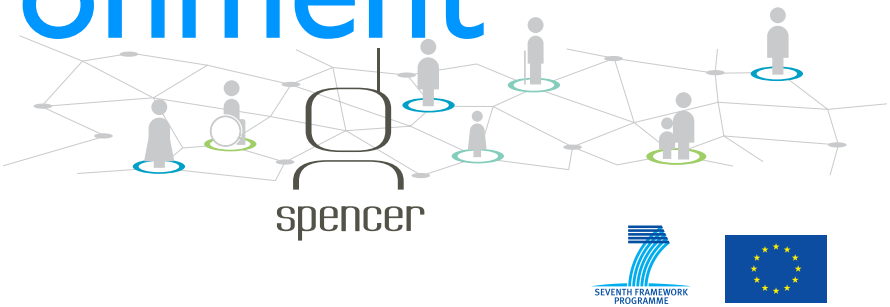
Fig. 1: Adaptation of the Brunswik lens model [6] in which a human observer perceives the internal state of a robot.

Space and environment

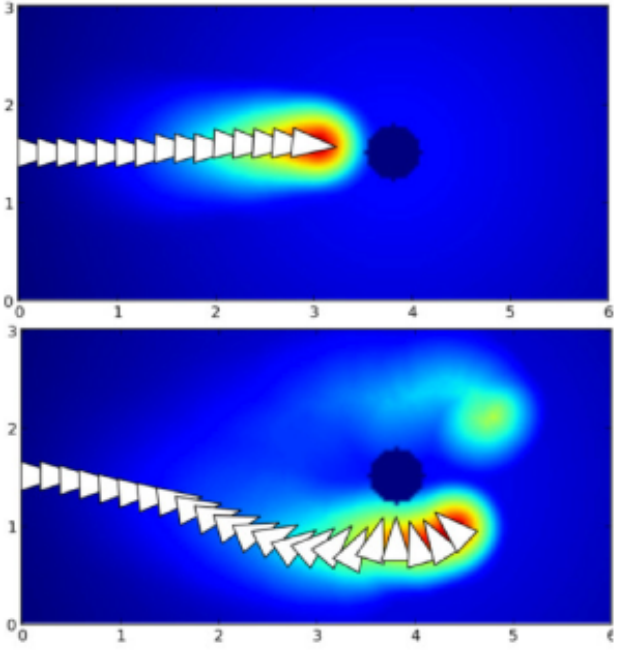
- Proxemics:
- People often refer to their need for personal space



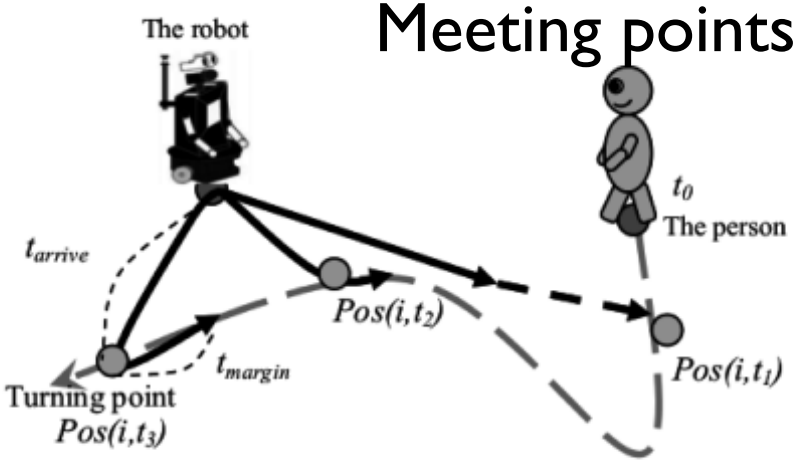
Space and environment



Potential functions



Meeting points



[Satake 2009]

[Avrunin 2014]

Space and environment

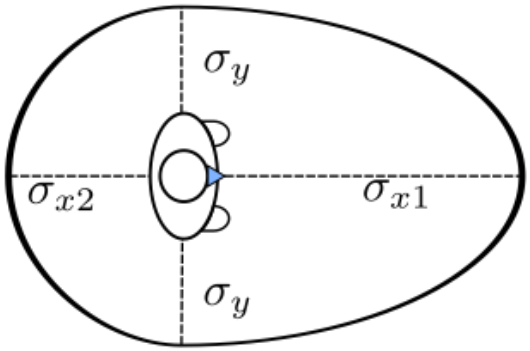
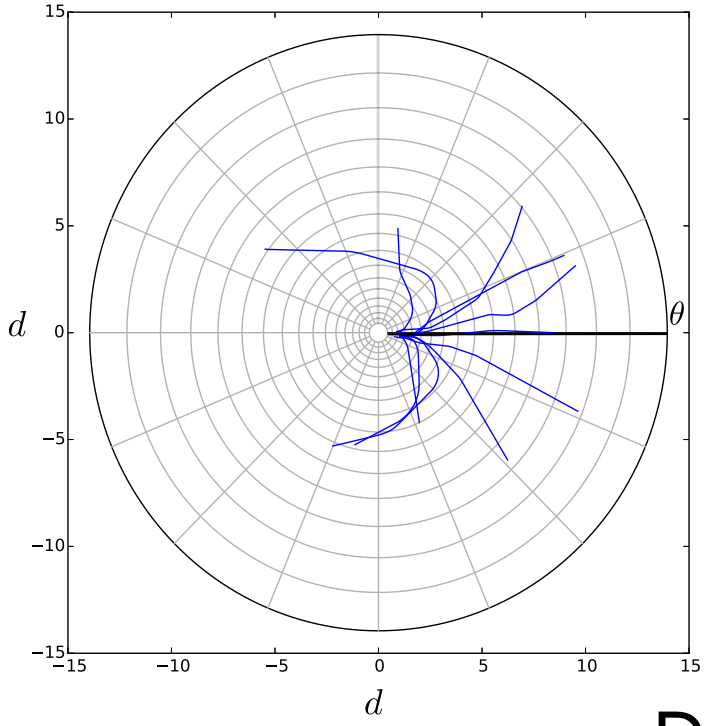
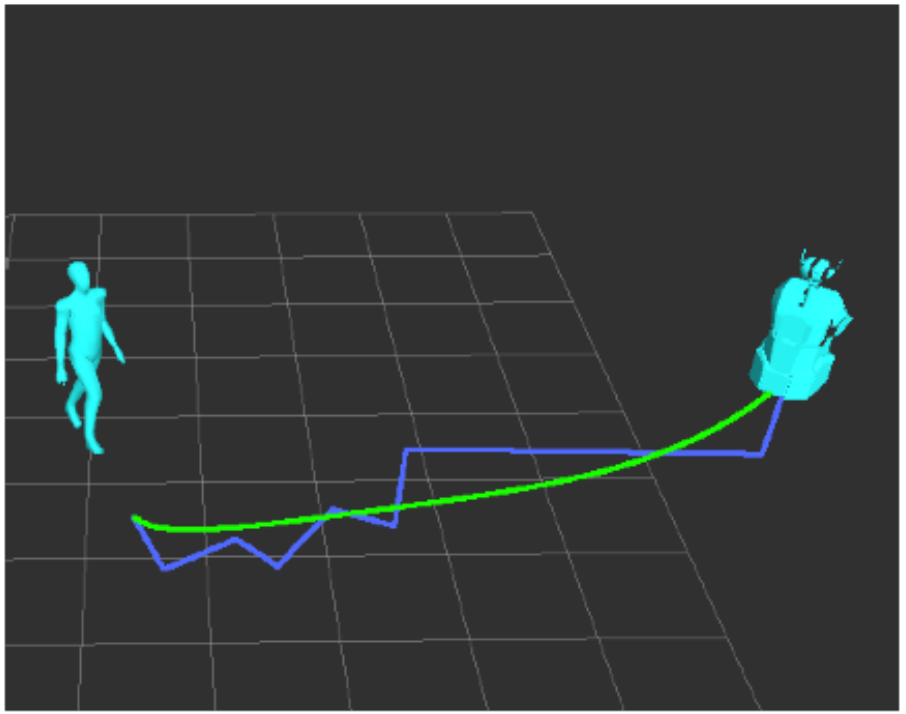


Fig. 1. Human representation with parameters σ_x and σ_y, which distances start from the center of the person.



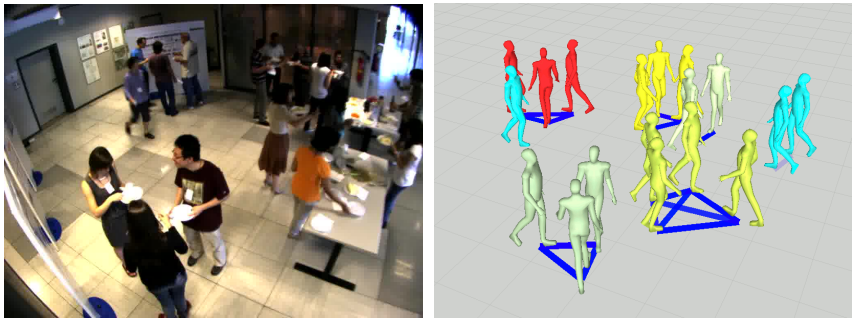


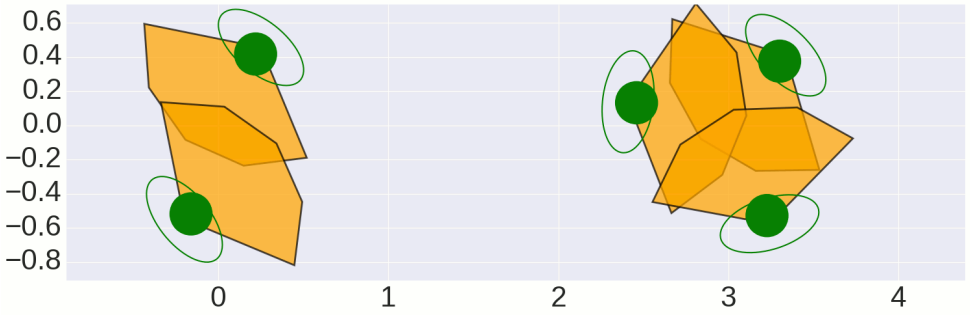
Inverse Reinforcement Learning:
Demonstrations + Learning Cost functions

O. Islas et al. (2016).Robots Learning How and Where to Approach People IEEE RO-MAN 2016

Space and environment

- Dyadic to group modeling
  - What it is a group?





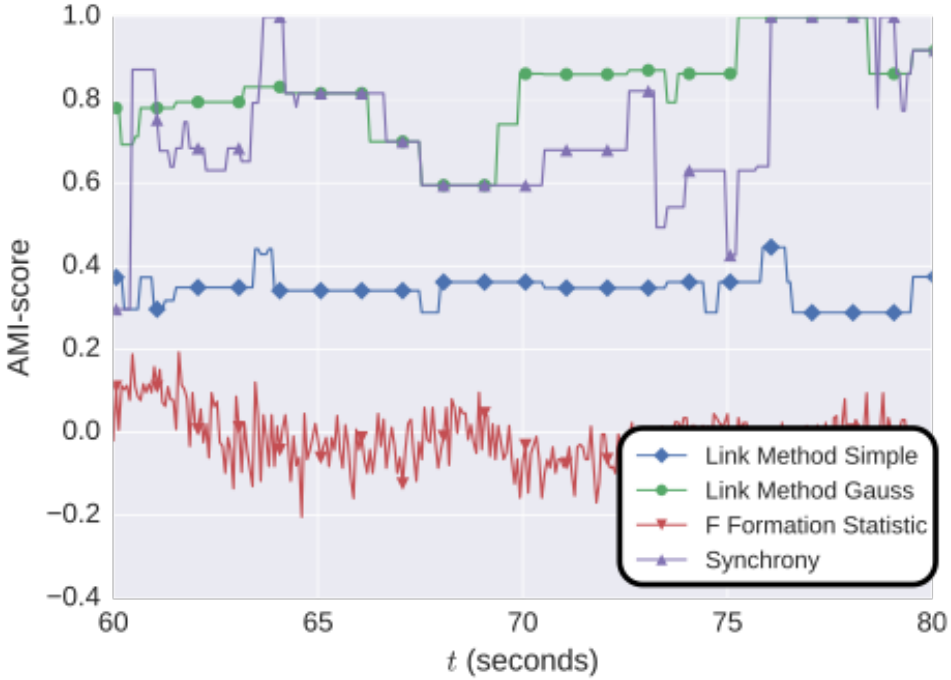


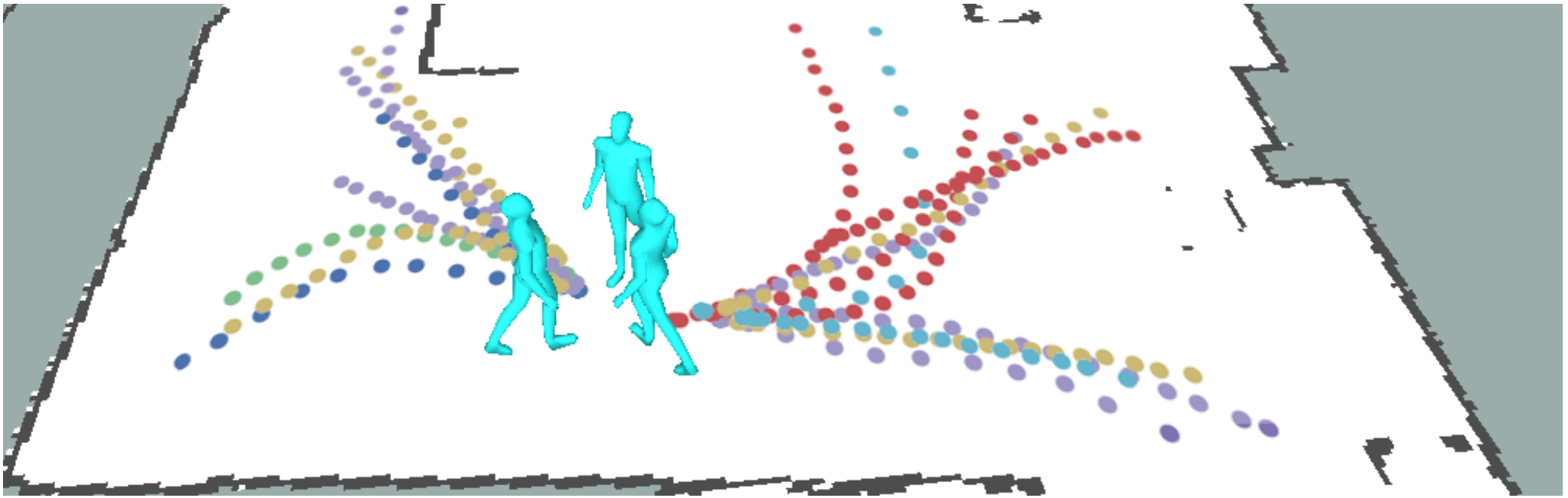
Fig. 6: AMI shape during a time segment of 20 s extracted from the video sequence SALSA S1.

O. Islas, G. Varni, M. Andries, R. Chatila, M. Chetouani (2016). Modeling the dynamics of individual behaviors for group detection in crowds using low-level features. IEEE RO-MAN 2016

J. Allen et al.:Comparing Social Science and Computer Science Workflow Processes for Studying Group Interactions, Small Group Resarch *(2017)*

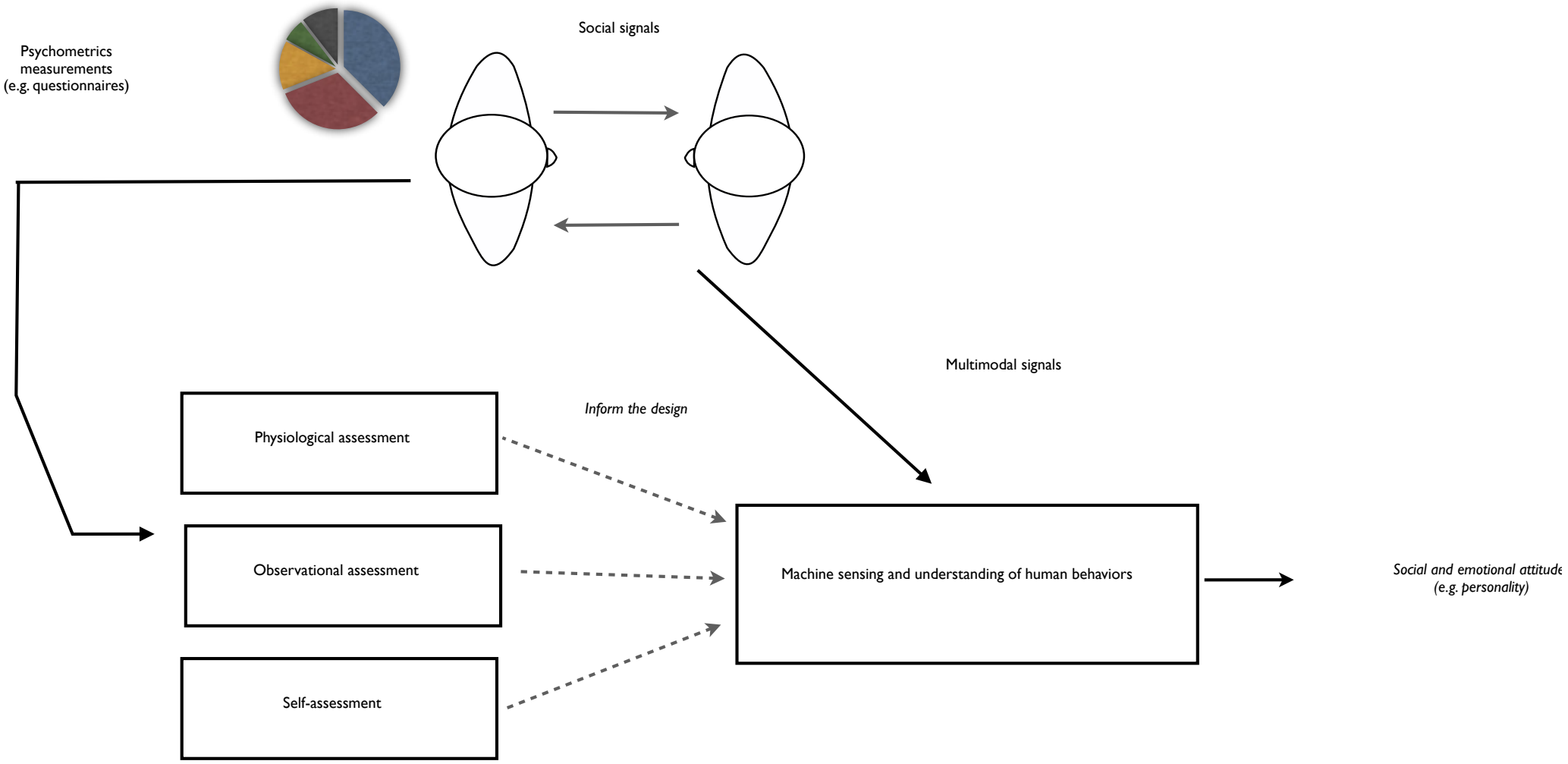
Space and environment

Learning to approach a group

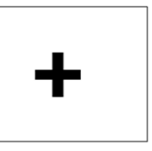


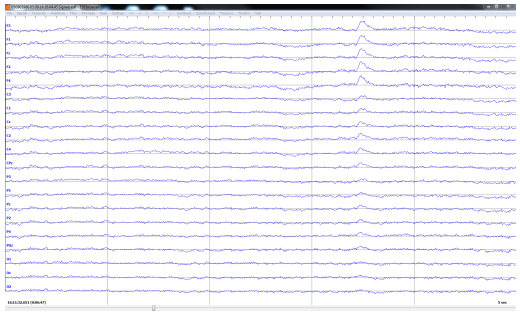
O. Islas, G. Varni, M. Andries, R. Chatila, M. Chetouani (2016). Modeling the dynamics of individual behaviors for group detection in crowds using low-level features. IEEE RO-MAN 2016

Affective Computing



Affective Computing

Black cross on white screen	Stimulus (image block or video)	White screen	Self-annotation
			
3 s	12,5 s (image block) or 15 s (video)	10 s	



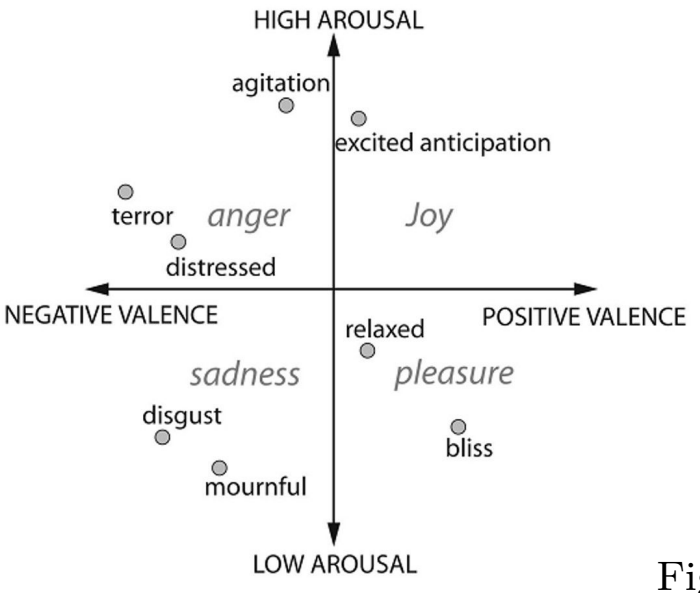


FIGURE 1.1 – The valence-arousal space

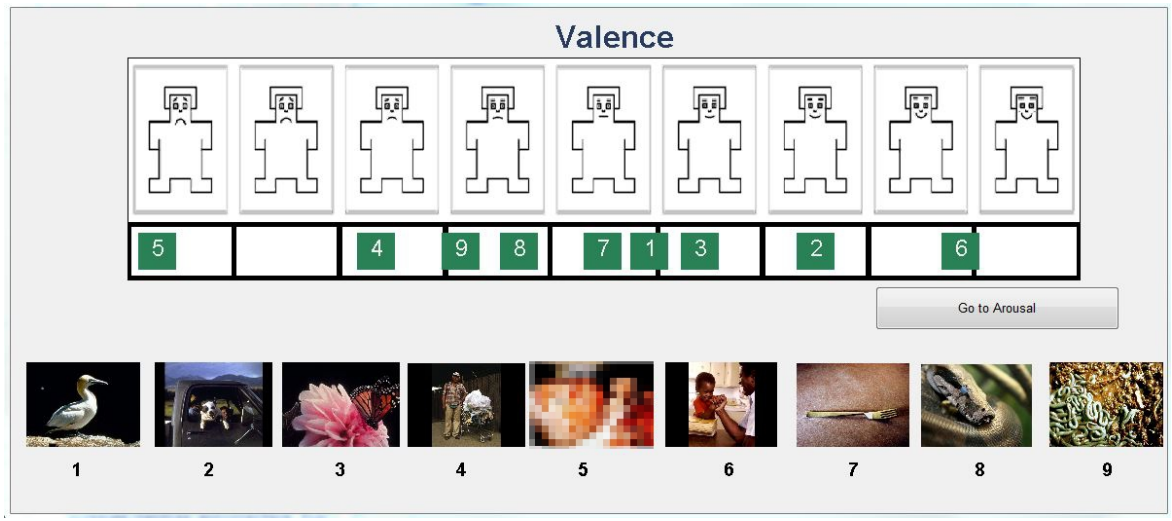


Fig. 3. Example of a ranking of calibration stimuli in terms of valence.

Personality computing

- Mapping individuals to personality traits

ID	Question	Trait
1	I am reserved	Ext
2	I am generally trusting	Agr
3	I tend to be lazy	Con
4	I am relaxed, handle stress well	Neu
5	I have few artistic interests	Ope
6	I am outgoing, sociable	Ext
7	I tend to find fault with others	Agr
8	I do a thorough job	Con
9	I get nervous easily	Neu
10	I have an active imagination	Ope

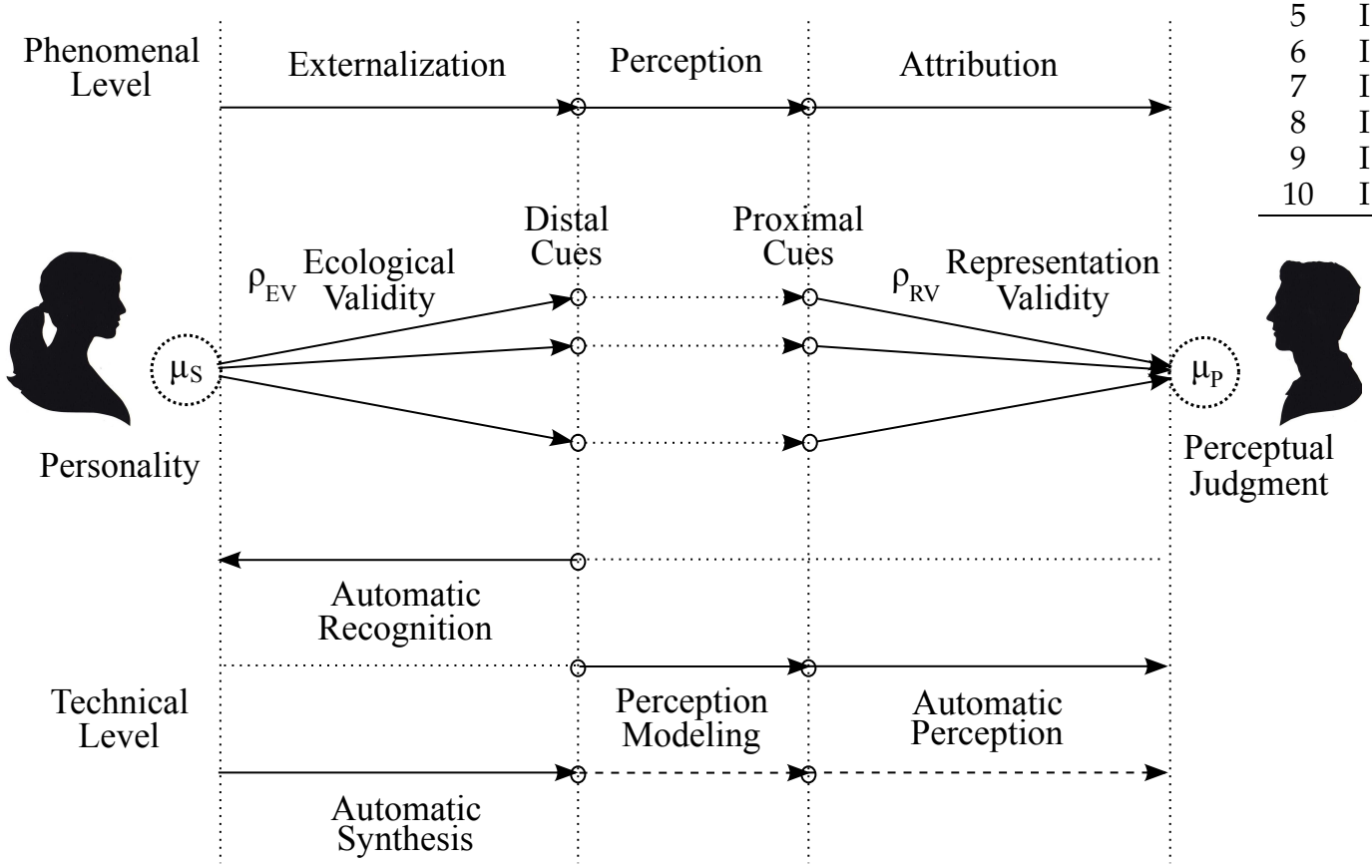
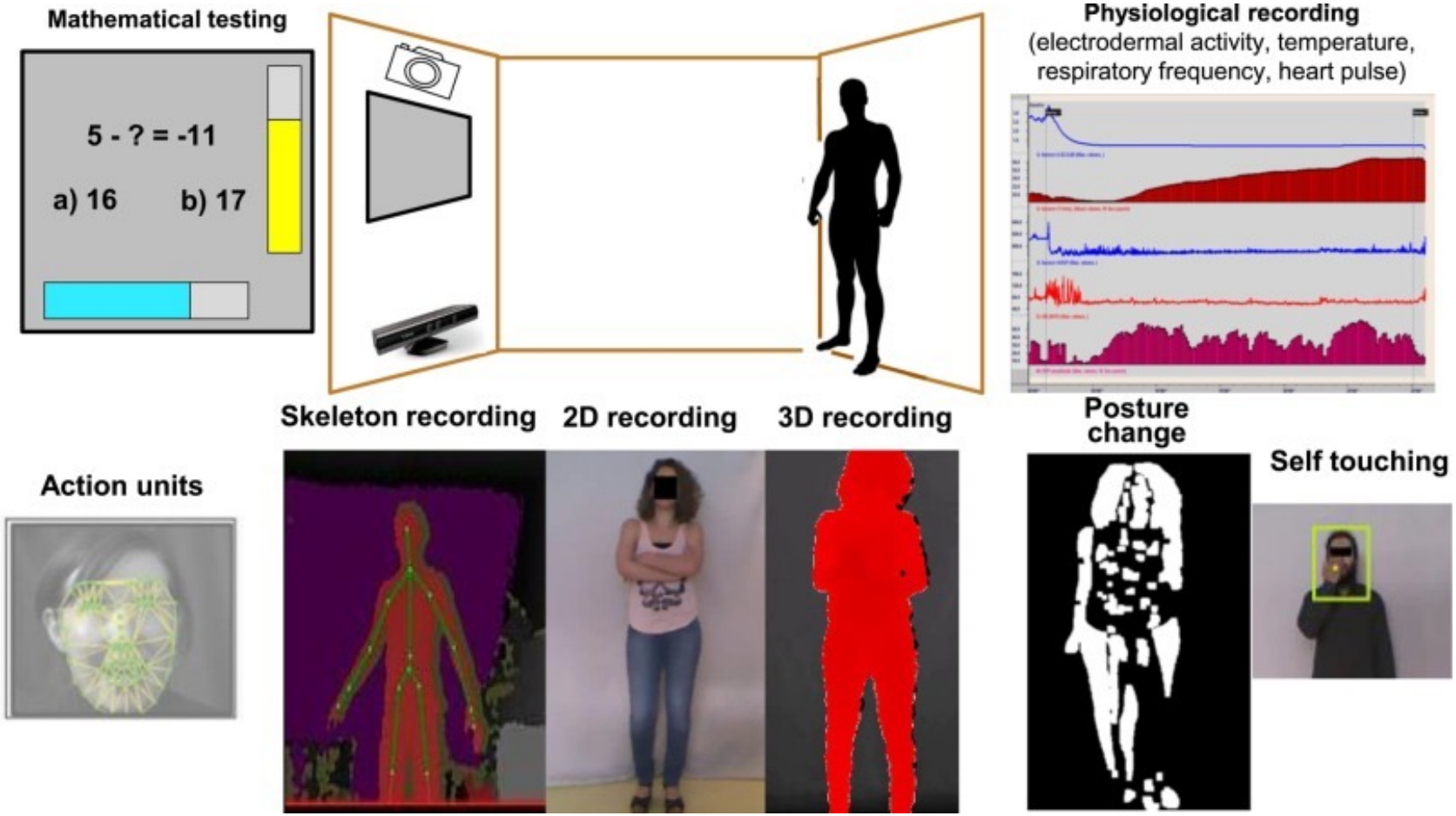


Fig. 2. The Figure shows the relationship between the Brunswik Lens and the three main problems addressed in Personality Computing. Automatic Personality Recognition is the inference of self-assessments (μ_S in the figure) from distal cues, Automatic Personality Perception is the inference of assessments (μ_P in the figure) from proximal cues, Automatic Personality Synthesis is the generation of artificial cues aimed at eliciting the attribution of predefined traits.

A. Vinciarelli et al. : A Survey of Personality Computing. *IEEE Trans. on Affective Computing (2014)*

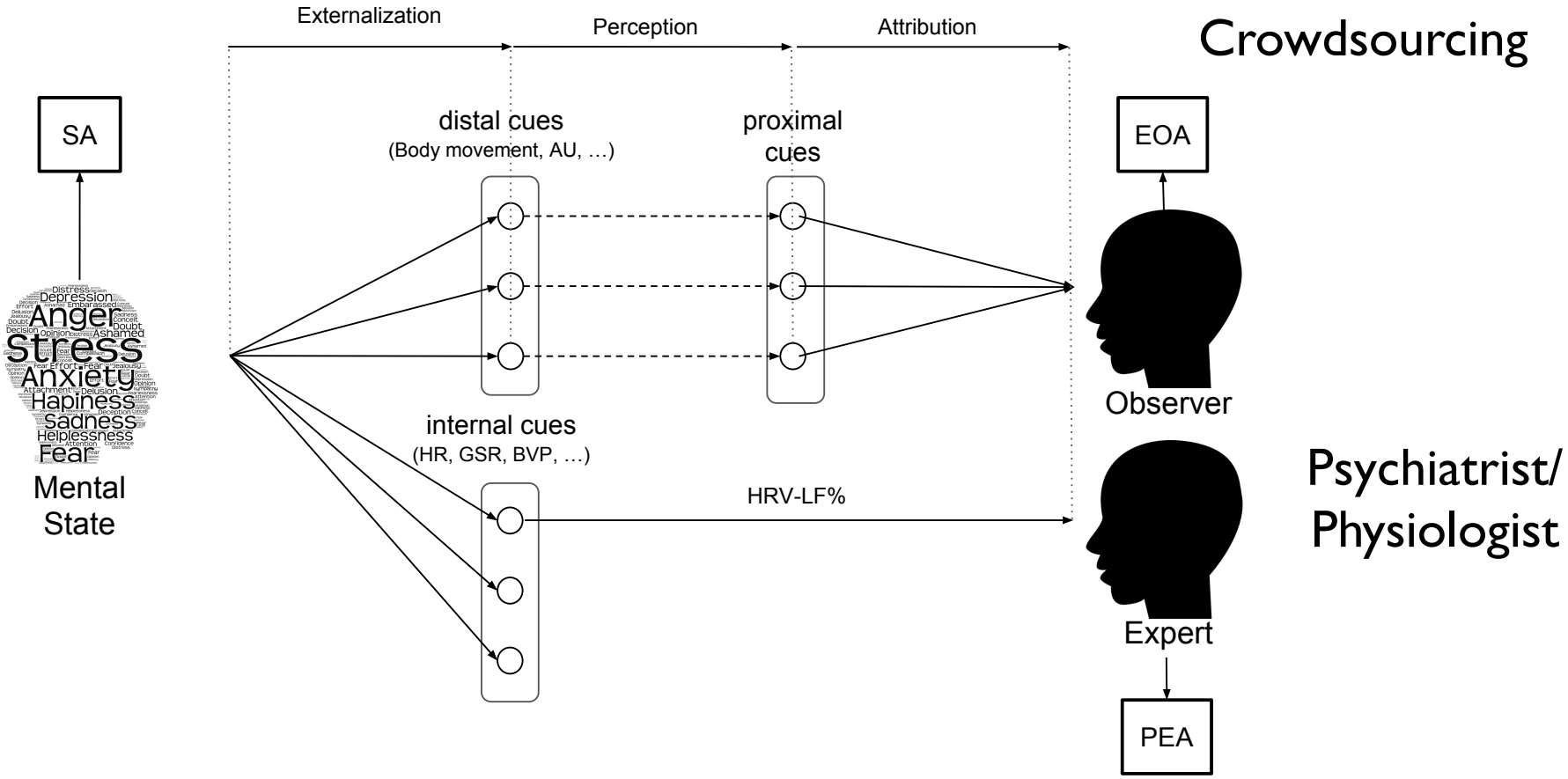
Affective States



Aigrain et al. : Multimodal stress detection from multiple assessments. *IEEE Trans. on Affective Computing (2016)*

Affective States

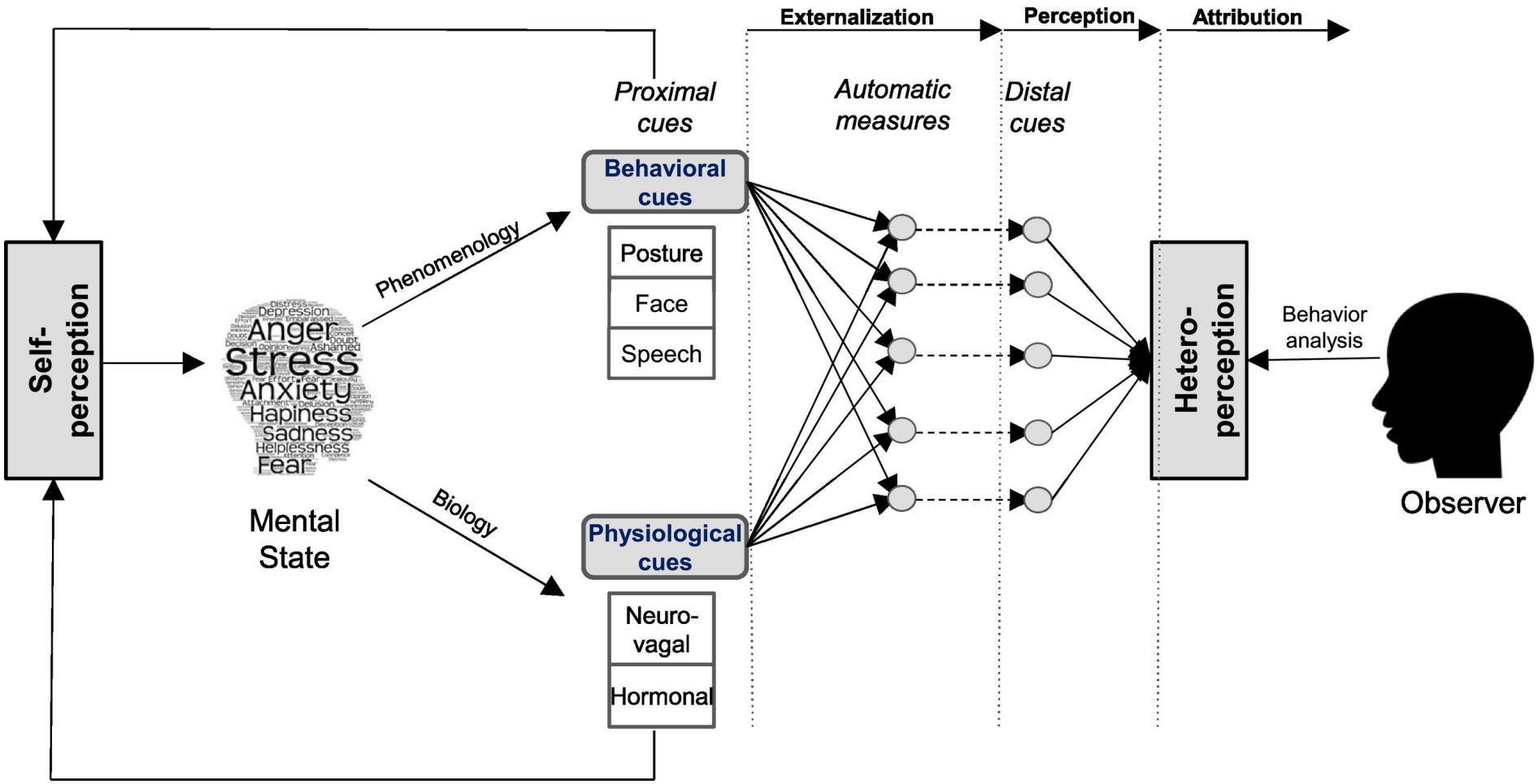
- Combining Multiple assessments



Aigrain et al. : Multimodal stress detection from multiple assessments. *IEEE Trans. on Affective Computing (2016)*

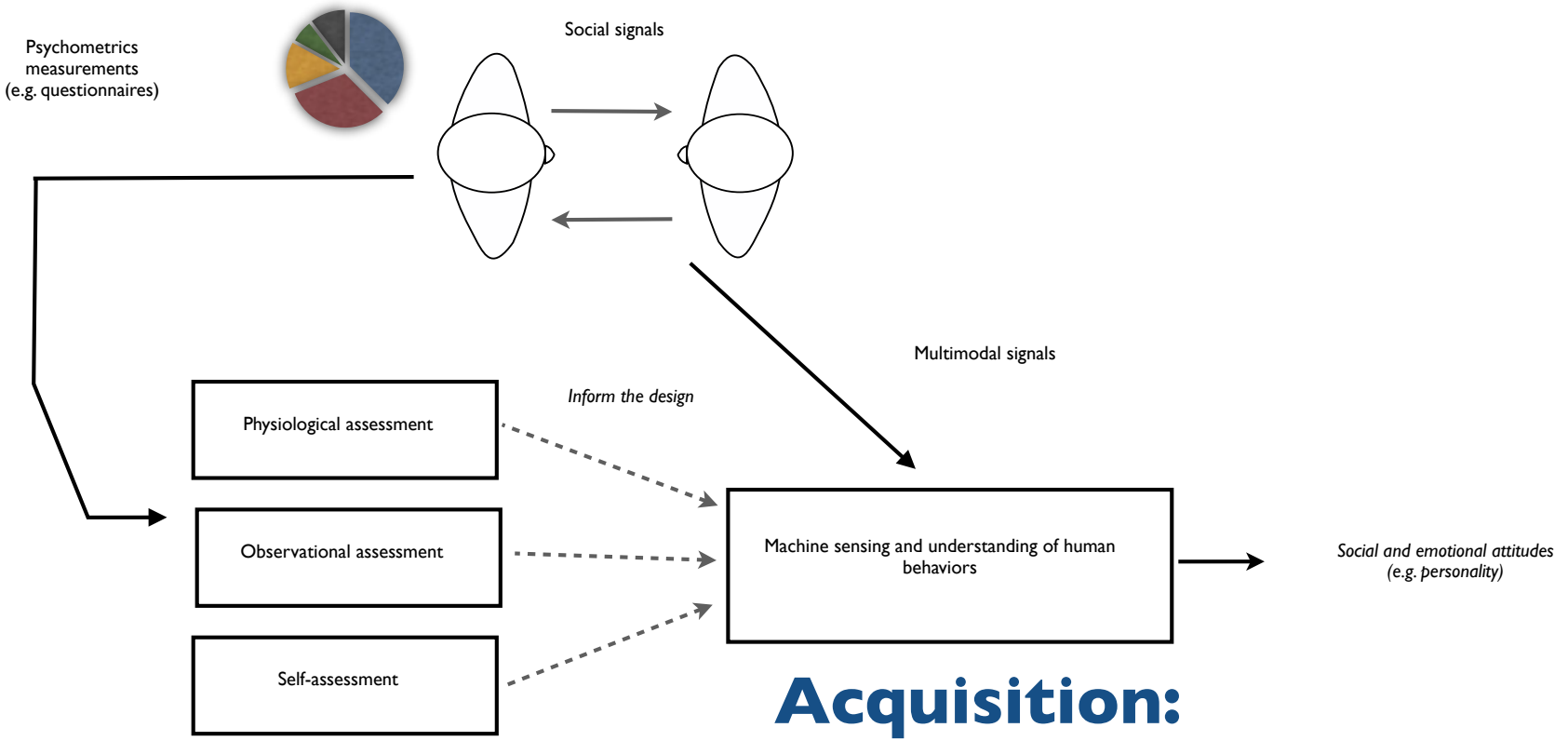
Stress

- Mapping individual to personality traits



Aigrain et al. : Multimodal stress detection from multiple assessments. *IEEE Trans. on Affective Computing (2016)*

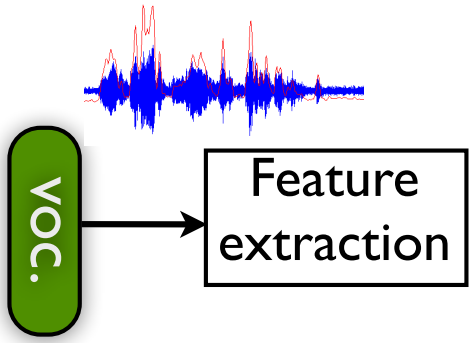
Affective Computing



Acquisition:

- Objective: acquisition of emotional / social signals
- Constraints: noise, devices, naturalness, …
- Techniques: multiple sources, filtering

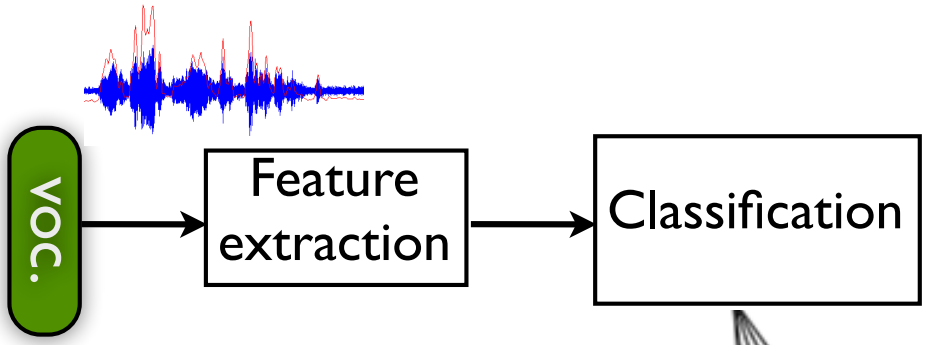
Affective Computing



Feature extraction

- Objective: compute compact and discriminant representation of emotions
- Constraints: intra / inter individual variabilities, context…
- Techniques: homogeneous extraction, speaker normalization, feature selection

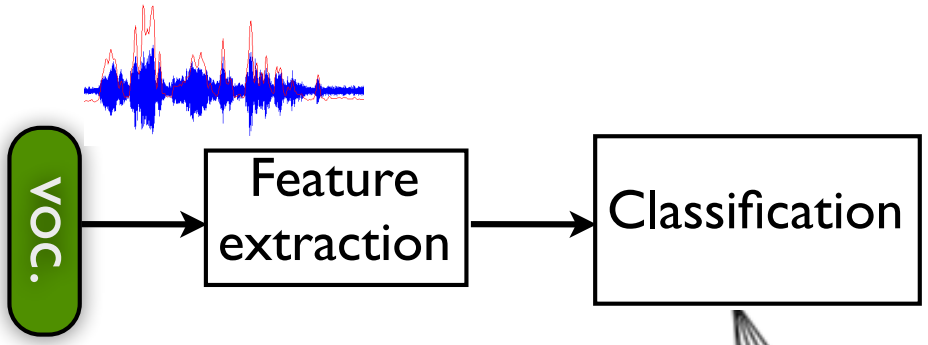
Affective Computing



Classification

- Objective: classify feature
- Constraints: few samples per emotion, classifiers,
- Techniques: machine learning, cross-validation, transfer learning.

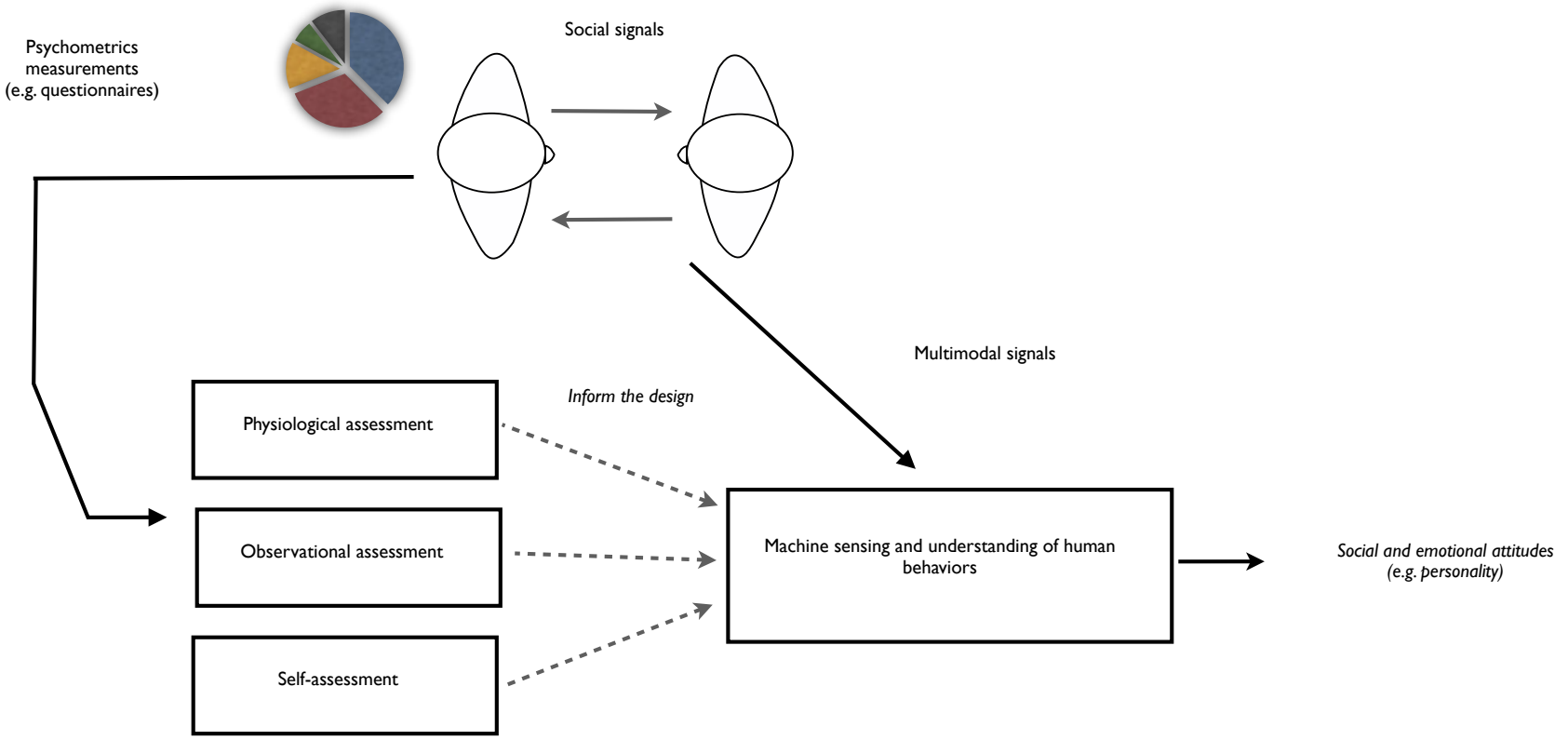
Affective Computing



Classification

- Objective: classify feature
- Constraints: few samples per emotion, classifiers,
- Techniques: machine learning, cross-validation, transfer learning.

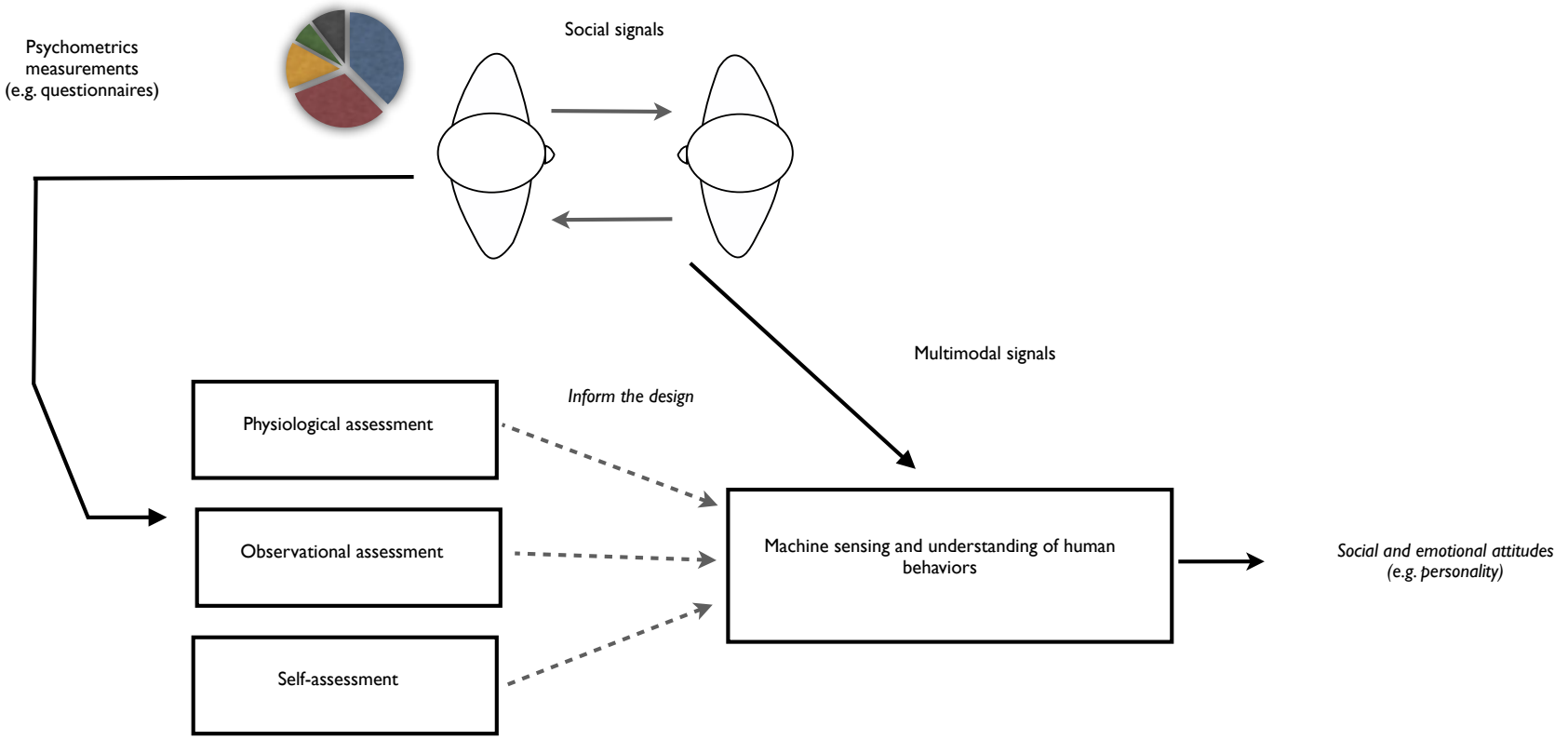
Affective Computing



Challenges

- Speaker traits: gender, age, style, pathology,
- Annotation, inter-annotator agreement, perception test
- Affective states, attitudes, personality traits

Affective Computing



Challenges

- Feature representation
- Machine learning

Personality computing

- Mapping individuals to personality traits

ID	Question	Trait
1	I am reserved	Ext
2	I am generally trusting	Agr
3	I tend to be lazy	Con
4	I am relaxed, handle stress well	Neu
5	I have few artistic interests	Ope
6	I am outgoing, sociable	Ext
7	I tend to find fault with others	Agr
8	I do a thorough job	Con
9	I get nervous easily	Neu
10	I have an active imagination	Ope

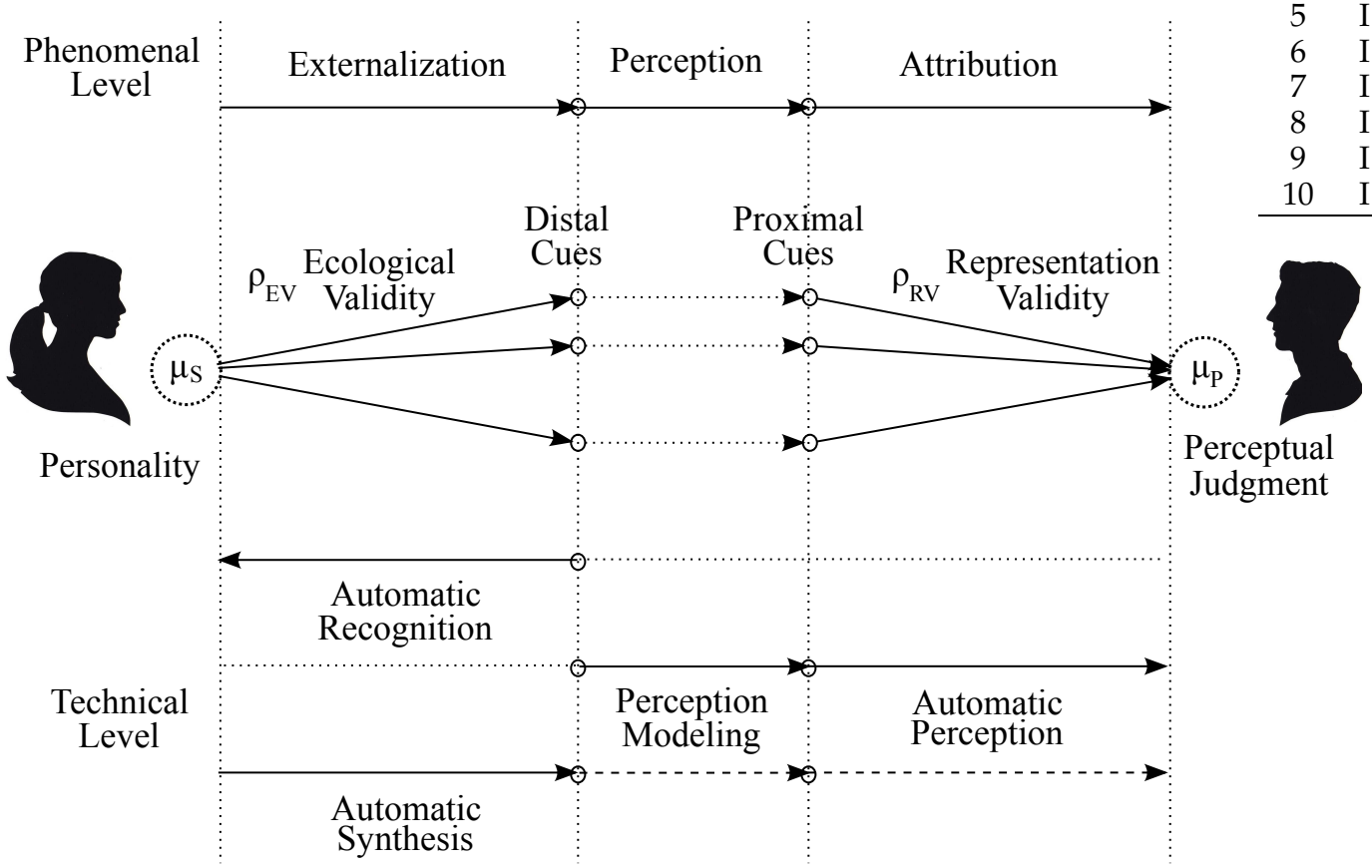
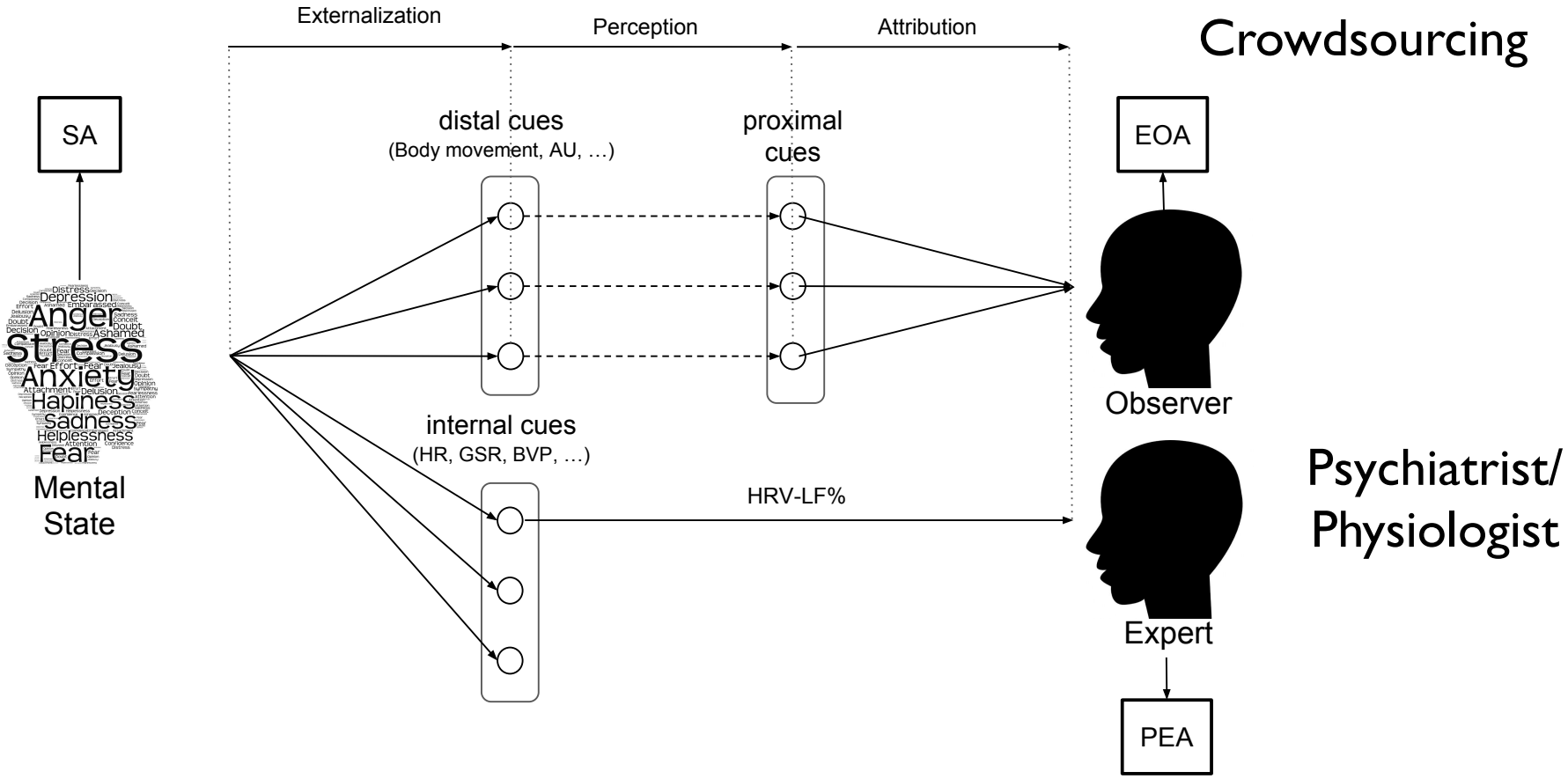


Fig. 2. The Figure shows the relationship between the Brunswik Lens and the three main problems addressed in Personality Computing. Automatic Personality Recognition is the inference of self-assessments (μ_S in the figure) from distal cues, Automatic Personality Perception is the inference of assessments (μ_P in the figure) from proximal cues, Automatic Personality Synthesis is the generation of artificial cues aimed at eliciting the attribution of predefined traits.

A. Vinciarelli et al. : A Survey of Personality Computing. *IEEE Trans. on Affective Computing (2014)*

Affective States

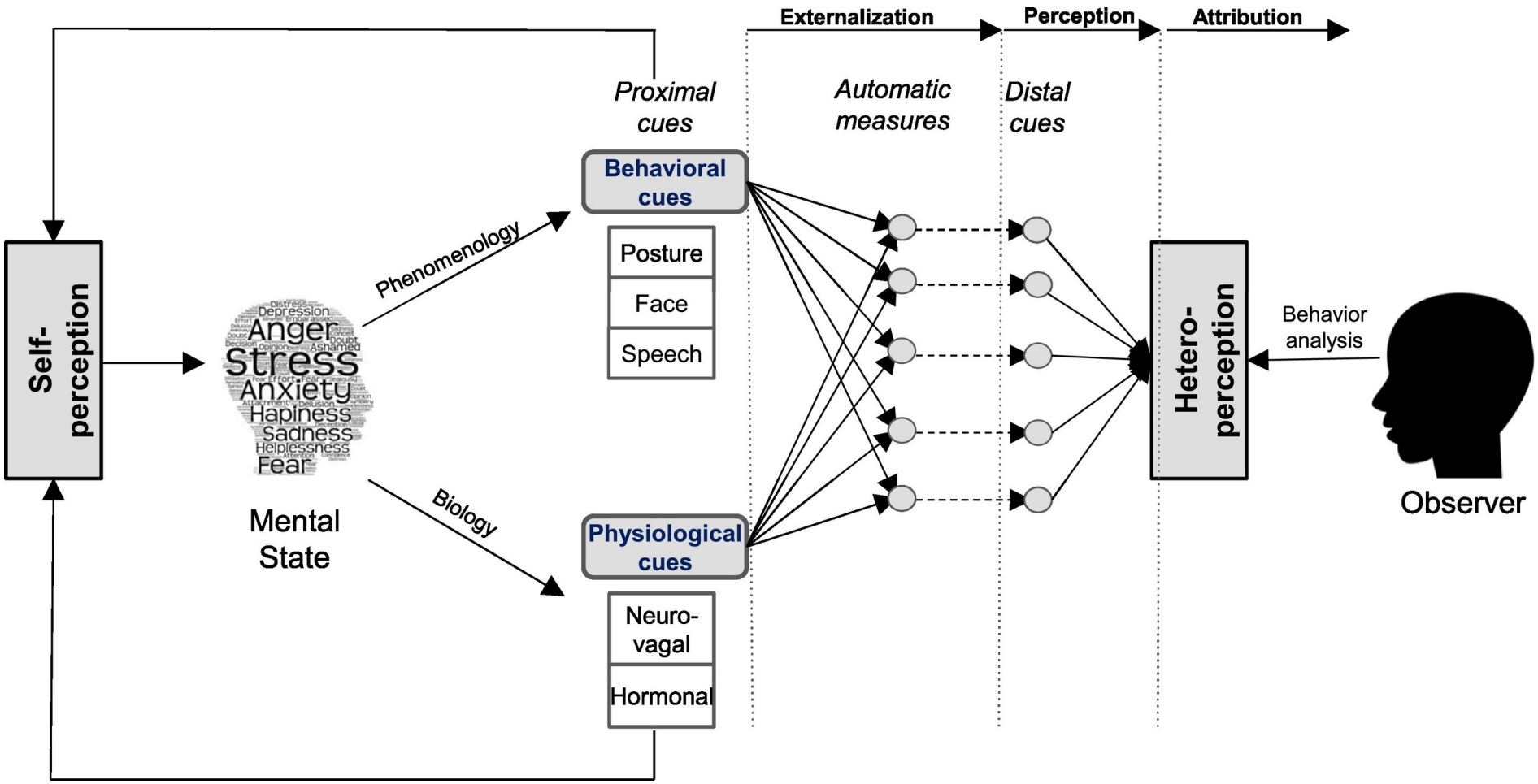
- Combining Multiple assessments



Aigrain et al. : Multimodal stress detection from multiple assessments. *IEEE Trans. on Affective Computing (In press)*

Stress

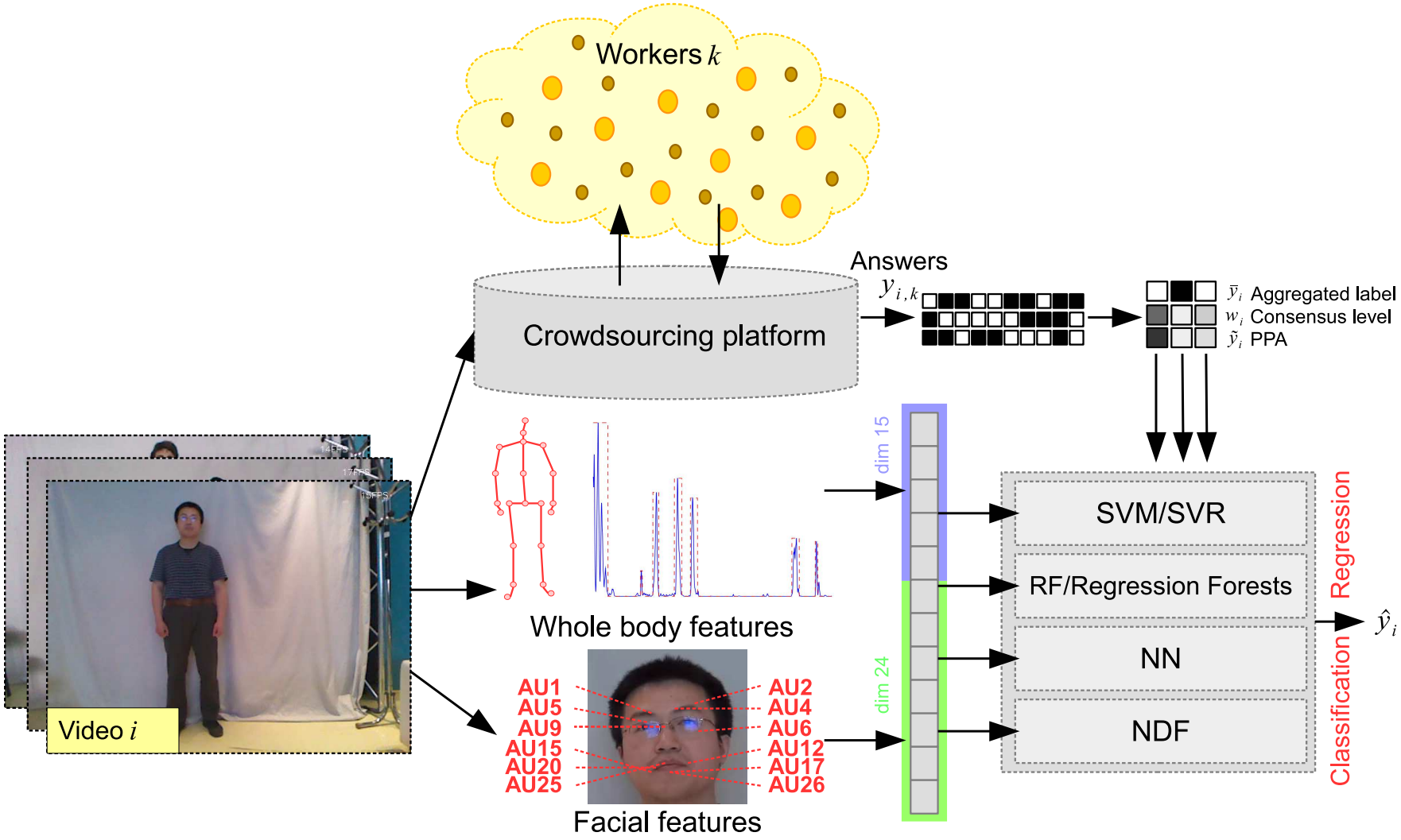
- Mapping individual to personality traits



Aigrain et al. : Multimodal stress detection from multiple assessments. *IEEE Trans. on Affective Computing (In press)*

Affective States

- Combining Multiple assessments



« The picture we like are our image »

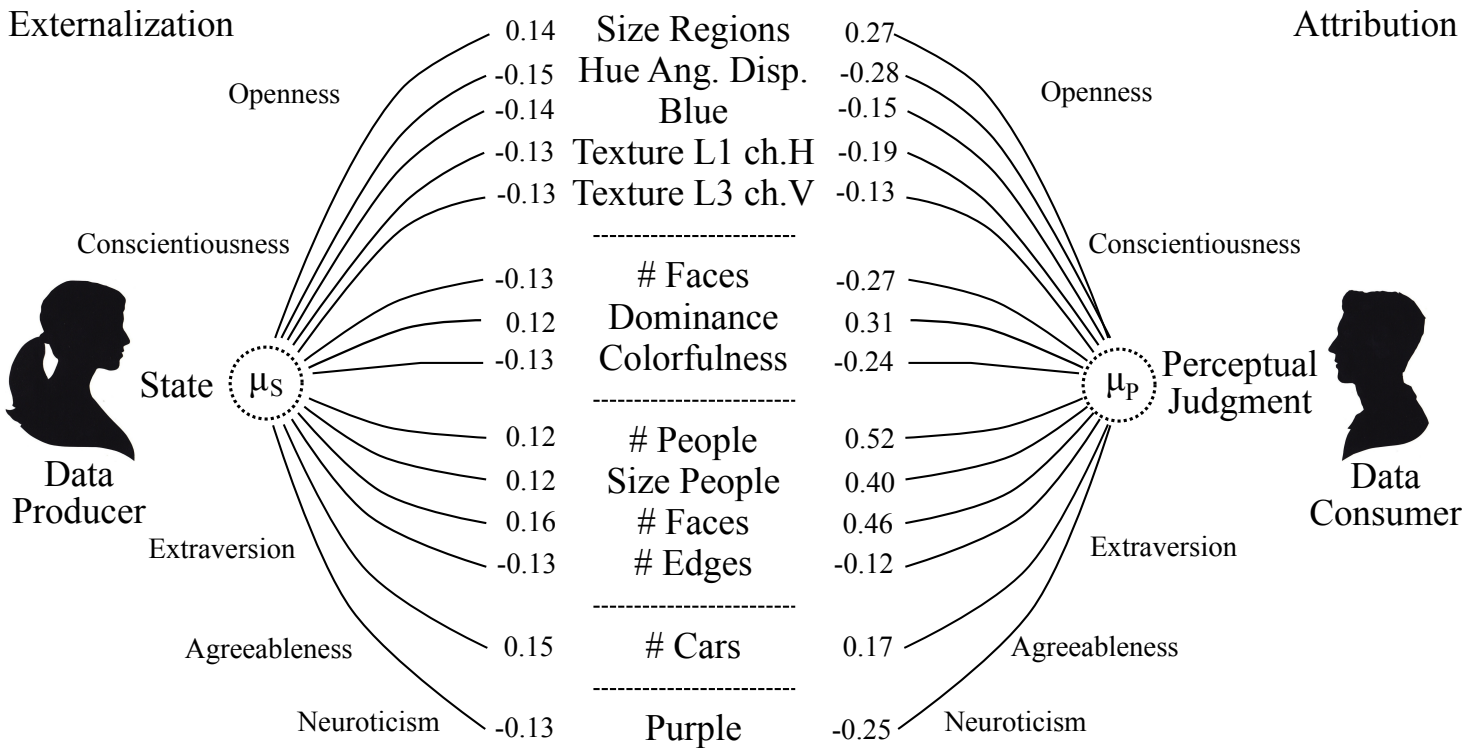


Figure 2: The picture shows the Brunswik Lens model for the PsychoFlickr dataset, where the state corresponds to the Big Five traits (as per assessed with the BFI-10). Ecological and Representation validities are measured with the Spearman Coefficient and the picture shows (for each trait) features for which both values are statistically significant ($p < 5\%$).

Segalin, C., Perina, A., Cristani, M., and Vinciarelli, A. (2016) The pictures we like are our image: continuous mapping of favorite pictures into self-assessed and attributed personality traits. IEEE Transactions on Affective Computing.

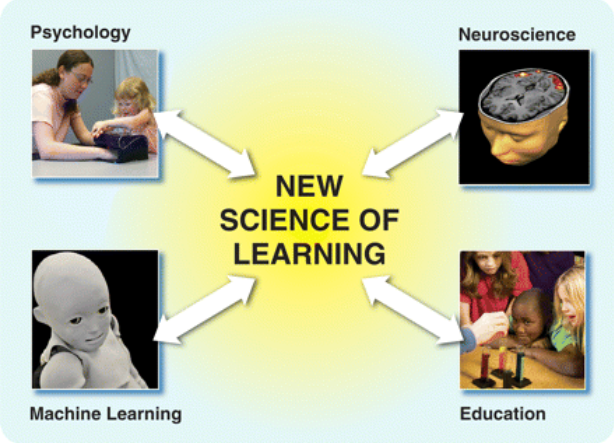
Cristani, M., Vinciarelli, A., Segalin, C., Perina, A.: Unveiling the multimedia unconscious: implicit cognitive processes and multimedia content analysis. In: Proceedings of the 21st ACM International Conference on Multimedia, pp. 213– 222. ACM (2013

Summary

- Social signal processing aims at developing tools to quantify non-verbal behaviors, often subjective, in the context of social interactions

- State-of-the-art research shows that it is possible to automatically map non-verbal cues such as human's body motion into psychometric measurements gathered with appropriate questionnaires (e.g., language-based descriptions of behavior)

- Interplay between computational models, neuroscience and psychology



Social learning
(Meltzoff et al., 2009)

Thank you for your attention

Questions?