UNIVERSITY OF SKÖVDE

Human cognition & HRI

Serge Thill
serge.thill@his.se

Contents

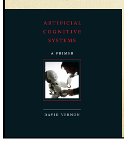
Contents

(from stampauctionnetwork.com)

"Es ist schon alles gesagt worden; nur noch nicht von jedem."
– Karl Valentin (1882 – 1948)

(Everything has already been said; just not by everyone.)

Social cognition



Borrowed to some degree from Ch 9 of this

What agents do

- Agents have
  - Goals and intentions

What agents do

- Agents have
  - Goals and intentions
  - An agent can have a certain state of affairs as a goal and a specific set of intended actions in mind to achieve this goal

What agents do

- Agents can
  - Act
  - Perceive and attend

What agents do

- Agents can
  - Act
  - Perceive and attend
  - Attention is intentionally-directed perception
    - "the temporally-extended processes whereby an agent concentrates on some features of the environment to the (relative) exclusion of others." (Kaplan and Hafner 2006)

Interaction

A single agent has

Interaction

A single agent has

- Actions
- Goals
- Intentions
- Attention

Interaction

A single agent has

- Actions
- Goals
- Intentions
- Attention

Social interaction between 2 or more has

Interaction

A single agent has

- Actions
- Goals
- Intentions
- Attention

Social interaction between 2 or more has

- **Joint** actions
- **Shared** goals
- **Shared** intentions
- **Joint** attention

What abilities does a social agent need?

What abilities does a social agent need?

- Reading faces
- Detecting eye gaze
- Recognizing emotional expressions
- Perceiving biological motion
- Paying joint attention
- Detecting goal-directed actions
- Discerning agency, imitation, deception, empathy
- ...

What abilities does a social agent need?

- **It needs to be aware of the cognitive state of the other agent**

How does one infer the cognitive state of another?

How does one infer the cognitive state of another?

- Cognitive and bodily states are linked
  - When an agent perceives a social stimulus, this perception produces bodily changes in the perceiving agent
  - The perception of bodily states in other agents frequently evokes a tendency to mimic those states
  - The agent's own body states trigger affective states in the agent
- Consequently, the posture, movements, and actions of an agent
  - Convey a great deal about its cognitive and affective disposition
  - Influence how another agent will behave towards it

Theory of mind

- A cognitive agent's act must anticipate the actions of an agent that itself is already anticipating what it is going to do
- That is, an agent must anticipate the intentions of other agents
  - Predict what they will do
  - Possibly, why they want to do it
- ToM is *the ability to infer what someone else is thinking and wants to do*

What to infer

- Humans infer different types of intention
  - Interpreting movements (lower level intentions)
    - **What** is the desired state?
  - Interpreting actions (higher level)
    - **Why** is it desired? (underlying motive)
- Grasp a cup vs. he is thirsty

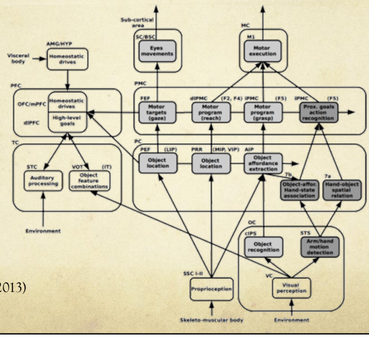
How to infer

- Internal simulation
  - A mechanism that predicts the consequences of the agent's own actions based on its own intentions by internal simulation using forward models
  - Input: overt or covert motor commands
  - Output: the likely sensory consequences of carrying out those commands
- When observing another agent's actions, the same mechanism can be used
  - associate observed movements with likely, i.e. intended, sensory consequences
- (Cf. the ideo-motor principle and the mirror-neuron system)

Mirror neurons and intention understanding

- Exist in PMC areas (F5) and IPL
- May help in understanding actions of others
- May encode action goals
- Do *not* encode detailed motor commands (see Umiltà et al., 2008)
- May be relevant for language (Arbib, 2005; Chersi et al., 2010)

Brain areas of interest



Thill et al. (2013)

Empathy

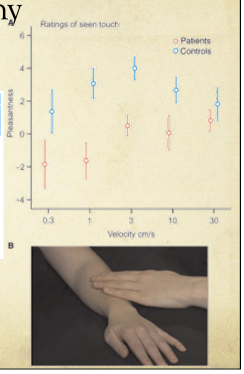
doi:10.1093/brain/awr011 Brain 2011: 134; 1116–1126 | 1116

BRAIN
A JOURNAL OF NEUROLOGY

Reduced C-afferent fibre density affects perceived pleasantness and empathy for touch

India Morrison,[1,2] Line S. Löken,[3] Jan Minde,[4] Johan Wessberg,[2] Irene Perini,[2] Inger Nennesmo[5] and Håkan Olausson[1,2]

1 Department of Clinical Neurophysiology, Blå stråket 7, Sahlgrenska University Hospital, S-413 45, Gothenburg, Sweden
2 Institute of Neuroscience and Physiology, University of Gothenburg, S-413 90, Gothenburg, Sweden
3 Oxford Centre for functional MRI of the Brain (FMRIB), University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK
4 Department of Surgery, Unit of Orthopedics, Perioperative Sciences, Umeå University Hospital, S-901 85, Umeå, Sweden
5 Department of Pathology, Karolinska University Hospital, Huddinge, S-141 86, Stockholm, Sweden



Collaboration between two agents

- Share the same goal, intend to act together, and coordinate their actions to achieve their shared goal through joint attention

Joint action

- Characteristics:
  - Mutual responsiveness
  - Commitment to joint activity
  - Commitment to mutual support
- Each agent must be mutually responsive to the intentions and actions of the other and must must be committed to supporting the efforts of the other agent to play their role
  - Consequently, each agent behaves in a way that is guided partially by the behaviour of the other agent
- Both agents have the same intention but they need not have the same reason for engaging in the activity
  - The outcome of the collaboration is the same for both agents
  - But the reason for adopting the goal of achieving that outcome need not be the same

Shared intentions

- Not simply a collection of individual intentions
  - Agents with a shared intention represent the overall shared goal between them
  - Each agent only has its own partial sub-plans
  - An individual agent with a shared intention does not need to know the other agent's partial plan
    - However, they do need to share the overall goal

Realisation of a shared intention and the execution of a joint action

- Each agent must represent its own actions and their predicted consequences **and**
  - the goals, intentions, actions and predicted consequences of the other agent
  - the effect that their actions have on the other agent
- It must also have at least a partial representation of how actions combine to achieve the overall goal **and**
  - an ability to predict the effects of their joint actions (so that it can monitor progress towards the overall goal and adjust its actions to help the other agent)
  - shared perceptual representation (joint attention)

Joint attention

- Each agent must monitor, understand, and direct the attentional behaviour of the other agent
- Both agents must be aware that this is going on
- An agent must:
  1. Detect and track the attentional behaviour of the other agent
  2. Influence the attentional behaviour of the other agent
  3. Engage in social coordination to manage the interaction
  4. Understand that the other agent has intentions

Take home message

- Social interaction requires many abilities that are not at all trivial.
- When building interactive machines, all of this matters from at least two angles:
  - The machine's own abilities must be sufficient for its part in the interaction
  - The machine must be designed so that a human can collaborate with it (which requires an understanding of the human's cognition)

But wait…

But wait…

- … there is more!

Social cognition and social interaction

Interactive explanation:
"One that relies on social interaction playing an enabling or constitutive role. Example: an infant behaves differently in a 'live' interaction because the coupling is more dynamically stable and disposes the infant to keep interacting as opposed to a playback of a previous interaction, which is dynamically less stable and easier to disengage from."

Opinion

Cell PRESS

Can social interaction constitute social cognition?

Hanne De Jaegher[1], Ezequiel Di Paolo[2] and Shaun Gallagher[3,4]

[1] Marie Curie Project DISCOS, Department of Psychiatry, University of Heidelberg, VossStrasse 4, D-69115 Heidelberg, Germany
[2] Ikerbasque: Basque Science Foundation, Department of Logic and Philosophy of Science, University of the Basque Country (UPV/EHU), Av de Tolosa 70, 20080 Donostia – San Sebastián, Spain
[3] Department of Philosophy, Institute of Simulation and Training, University of Central Florida, Orlando, FL 32816-1352, USA
[4] Humanities Division, University of Hertfordshire, Hatfield, AL10 9AB, UK

Social interaction

- Social interaction is "*two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed and their relational dynamics acquire an autonomy of their own*"

  (deJaegher et al. 2010)

Social interaction

- Social interaction is "*two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed and their relational dynamics acquire an autonomy of their own*"

  (deJaegher et al. 2010)

Social interaction

- Social interaction is "*two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed and their relational dynamics acquire an autonomy of their own*"

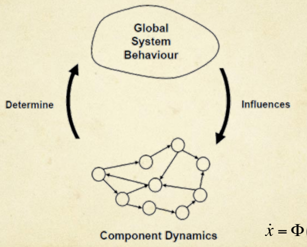
  (deJaegher et al. 2010)

Social interaction

- Social interaction is "*two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed and their relational dynamics acquire an autonomy of their own*"

  (deJaegher et al. 2010)

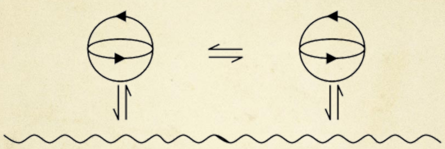
Circular causality in autonomous systems



$$\dot{x} = \Phi(x, \rho, \eta)$$

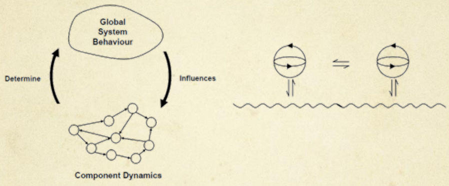
E.g. Kelso (1995), Clark's CRC (1997, 1998), Bickhard's recursive self-maintenance (2000)

Interaction between two agents



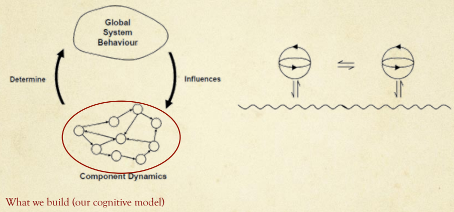
Maturana and Varela (1987)

From interaction to cognition

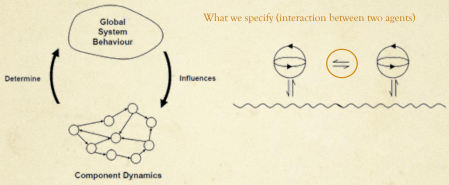


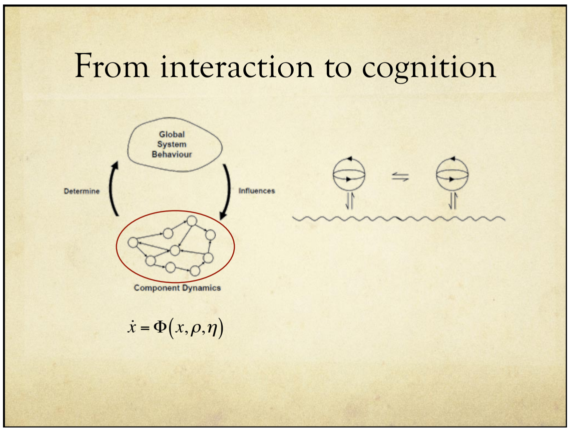
(Thill & Vernon, 2015)

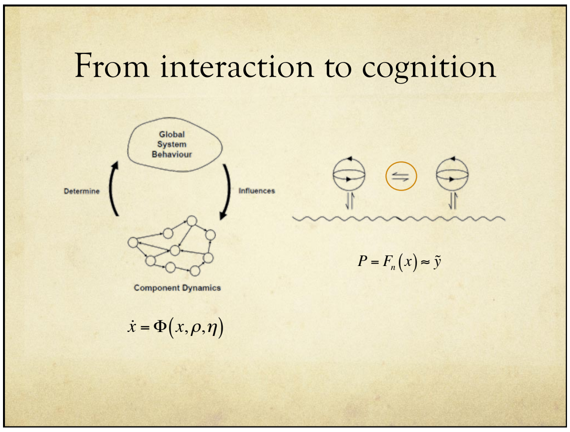
From interaction to cognition

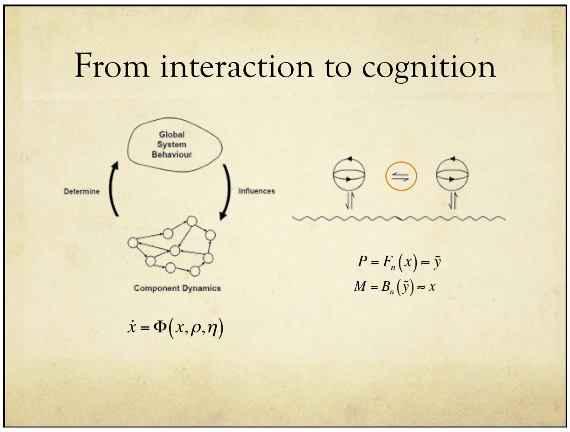


From interaction to cognition

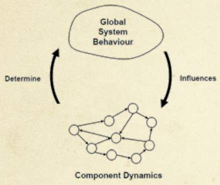


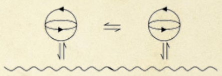

From interaction to cognition
Global System Behaviour
Determine
Influences
Component Dynamics
$\dot{x} = \Phi(x, \rho, \eta)$


From interaction to cognition
Global System Behaviour
Determine
Influences
Component Dynamics
$P = F_n(x) \approx \tilde{y}$
$\dot{x} = \Phi(x, \rho, \eta)$


From interaction to cognition
Global System Behaviour
Determine
Influences
Component Dynamics
$P = F_n(x) \approx \tilde{y}$
$M = B_n(\tilde{y}) \approx x$
$\dot{x} = \Phi(x, \rho, \eta)$

From interaction to cognition



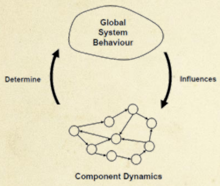


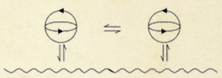
$$P = F_n(x) \approx \tilde{y}$$
$$M = B_n(\tilde{y}) \approx x$$

$$\dot{x} = \Phi(x,\rho,\eta)$$

Can be learned:
Theory of mind
Internal simulations

From interaction to cognition



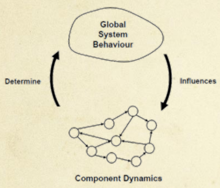


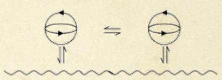
$$P = F_n(x) \approx \tilde{y}$$
$$M = B_n(\tilde{y}) \approx x$$

$$\dot{x} = \Phi(x,\rho,\eta)$$

Specified by the task / system designer;
refinements can be learned

From interaction to cognition





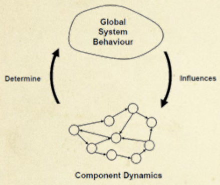
$$P = F_n(x) \approx \tilde{y}$$
$$M = B_n(\tilde{y}) \approx x$$

$$\dot{x} = \Phi(x,\rho,\eta)$$

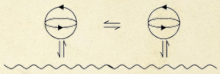
Objective function cast in terms of interaction:

$$O = \Psi\left(F_n(x), B_n(\tilde{y})\right)$$

From interaction to cognition



$$\dot{x} = \Phi(x, \Psi, \eta)$$

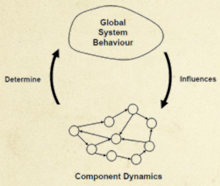


$$P = F_n(x) \approx \tilde{y}$$
$$M = B_n(\tilde{y}) \approx x$$

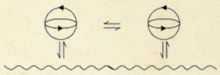
Objective function cast in terms of interaction:

$$O = \Psi\big(F_n(x), B_n(\tilde{y})\big)$$

From interaction to cognition



$$\dot{x} = \Phi(x, \Psi, \eta)$$



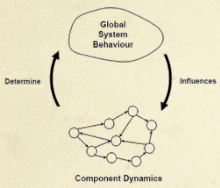
$$P = F_n(x) \approx \tilde{y}$$
$$M = B_n(\tilde{y}) \approx x$$

Objective function cast in terms of interaction:

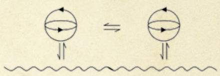
$$O = \Psi\big(F_n(x), B_n(\tilde{y})\big)$$

Defining Φ is left as an exercise to the reader ☺

From interaction to cognition



$$\dot{x} = \Phi(x, \Psi, \eta)$$



$$P = F_n(x) \approx \tilde{y}$$
$$M = B_n(\tilde{y}) \approx x$$

Objective function cast in terms of interaction:

$$O = \Psi\big(F_n(x), B_n(\tilde{y})\big)$$

Defining Φ is left as an exercise to the reader ☺
(but begins to look like a RL problem)

Attributing intentionality to robots

Attributing intentionality to robots

Attributing intentionality

- *"[T]he attributions of intentionality that we make to the robot in this example have nothing to do with formal programs. They are simply based on the assumption that if the robot looks and behaves sufficiently like us, then we would suppose, until proven otherwise, that it must have mental states like ours that cause and are expressed by its behavior and it must have an inner mechanism capable of producing such mental states."*

Attributing intentionality

- *"[T]he attributions of intentionality that we make to the robot in this example have nothing to do with formal programs. They are simply based on the assumption that if the robot looks and behaves sufficiently like us, then we would suppose, until proven otherwise, that it must have mental states like ours that cause and are expressed by its behavior and it must have an inner mechanism capable of producing such mental states."* (Searle, 1980, Rebuttal of the 'combination' reply)

Affect for machines

(Mostly based on Thill & Lowe; 2012)

Functions of emotions

- Three positions (Keltner & Gross, 1999):

1. Emotions have no functions
   - "disorganising forces in human behaviour"
2. Once served functions that are no longer appropriate
   - Like the appendix
3. Serve important functions now
   - What a lot of evidence seems to suggest (Pessoa, 2008; Damasio, 2010; Lowe and Ziemke, 2011; Stapleton, 2011)

An important distinction

- Intrapersonal functions, e.g.
  - Homeostatic regulation (internally and behaviourally; Sterling, 2004)
  - 'cognitive override' in goal-directed behaviour (Oately and Johnson-Laird, 1987; Rolls, 1999; Boureau and Dayan, 2010);
  - behavioural adaptation (e.g. Rolls, 1999) fundamentally concerned with the effects of emotions on learning
- Interpersonal functions, e.g.
  - Communication: emotional expression is high on informational content (Ekman, 2003)
  - social transaction: emotional expression may generally provide a sort of social glue during agent interactions (Griffiths and Scarantino, 2009).

Emotion expression as communication

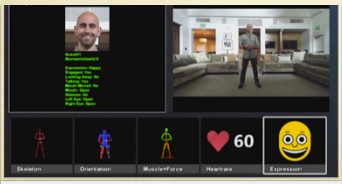
- "in a majority of species, affective states are responsible for the production of communicative signals" (Hauser, 1996)
- "principle of antithesis" (Darwin, 1872): emotional expressions have become disambiguated for the purpose of communication (also Keltner et al., 2003)
- Cross-cultural studies show unambiguously perceived basic emotions (Ekman, 2003)
- But communication implies information exchange. There has to be an advantage for sender and receiver.

Emotion expression as social exchange

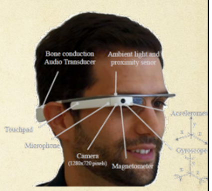
- Different in the sense that it emphasises a "social glue" aspect: disambiguates the roles/needs of the different conspecifics
  - But not necessarily to the equal benefit of all involved
    - "guilt" as a social transaction for reconciliation (Griffiths and Scarantino, 2009)
- Emotions as a way to manipulate the perceiver to the benefit of the expressor's needs and desires (Camras, 2011)

Expressing and recognising emotions for machines

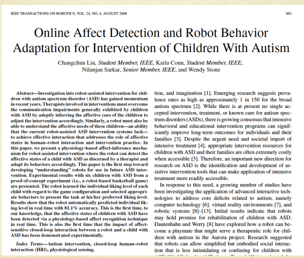
- (Intelligent) machines will have to interact with humans
  - The quality of this interaction might well be a success criterion for the machine's design
- Need the ability to express emotions (to fulfill social needs and exploit social functions of emotions) and recognise them
- Very much still an open research area
  - E.g. no fully mobile robot that can recognise its own emotions or that of others from non-facial cues exists but is needed for seamless integration of an artificial agent into society



How Kinect sees you: a pulsing sack of meat and emotions. (Image from Wired's

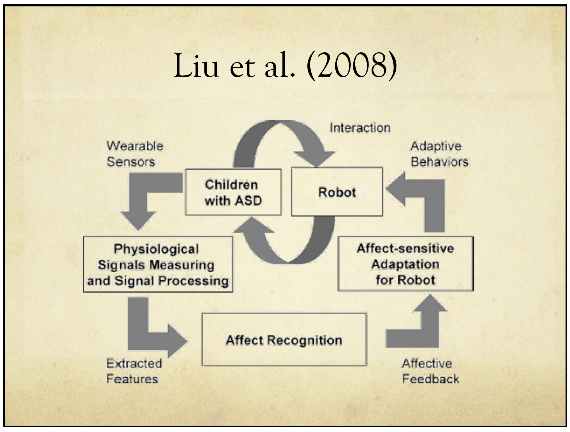


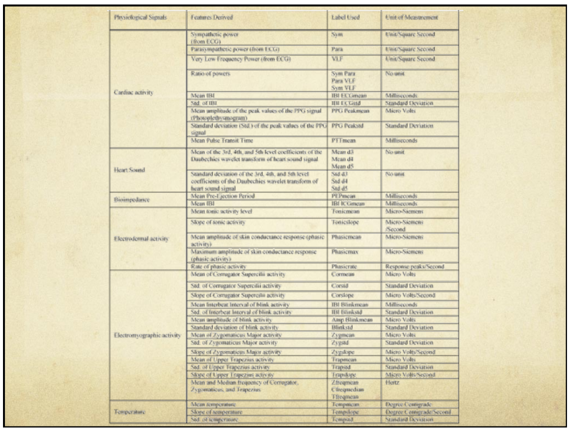
Example



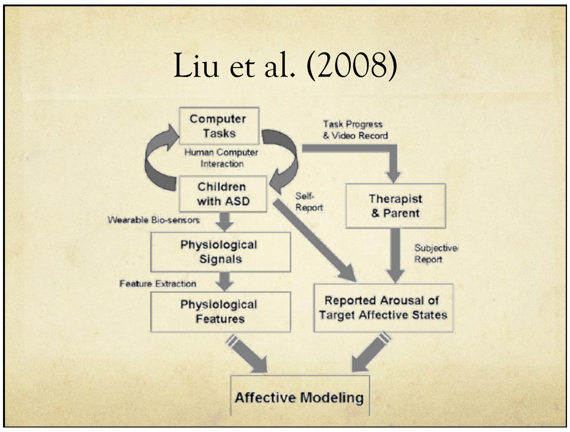
Online Affect Detection and Robot Behavior Adaptation for Intervention of Children With Autism

Changchun Liu, Student Member, IEEE, Karla Conn, Student Member, IEEE, Nilanjan Sarkar, Senior Member, IEEE, and Wendy Stone

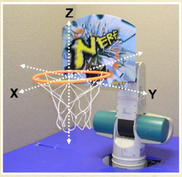


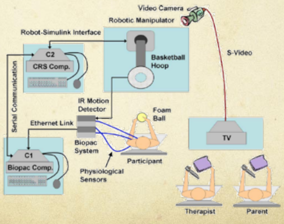


Physiological Signals	Features Derived	Label Used	Unit of Measurement
Cardiac activity	Sympathetic power (from ECG)	Sym	Unit/Square Second
	Parasympathetic power (from ECG)	Para	Unit/Square Second
	Very Low Frequency Power (from ECG)	VLF	Unit/Square Second
	Ratio of powers	Sym Para Para VLF Sym VLF	No unit
	Mean IBI	IBI ECGmean	Milliseconds
	Std. of IBI	IBI ECGstd	Standard Deviation
	Mean amplitude of the peak values of the PPG signal (Photoplethysmogram)	PPG Peakmean	Micro Volts
	Standard deviation (Std.) of the peak values of the PPG signal	PPG Peakstd	Standard Deviation
	Mean Pulse Transit Time	PTTmean	Milliseconds
Heart Sound	Mean of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Mean d3 Mean d4 Mean d5	No unit
	Standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Std d3 Std d4 Std d5	No unit
Bioimpedance	Mean Pre-Ejection Period	PEPmean	Milliseconds
	Mean IBI	IBI ICGmean	Milliseconds
Electrodermal activity	Mean tonic activity level	Tonicmean	Micro-Siemens
	Slope of tonic activity	Tonicslope	Micro-Siemens /Second
	Mean amplitude of skin conductance response (phasic activity)	Phasicmean	Micro-Siemens
	Maximum amplitude of skin conductance response (phasic activity)	Phasicmax	Micro-Siemens
	Rate of phasic activity	Phasicrate	Response peaks/Second
Electromyographic activity	Mean of Corrugator Supercilii activity	Cormean	Micro Volts
	Std. of Corrugator Supercilii activity	Corstd	Standard Deviation
	Slope of Corrugator Supercilii activity	Corslope	Micro Volts/Second
	Mean Interbeat Interval of blink activity	IBI Blinkmean	Milliseconds
	Std. of Interbeat Interval of blink activity	IBI Blinkstd	Standard Deviation
	Mean amplitude of blink activity	Amp Blinkmean	Micro Volts
	Standard deviation of blink activity	Blinkstd	Standard Deviation
	Mean of Zygomaticus Major activity	Zygmean	Micro Volts
	Std. of Zygomaticus Major activity	Zygstd	Standard Deviation
	Slope of Zygomaticus Major activity	Zygslope	Micro Volts/Second
	Mean of Upper Trapezius activity	Trapmean	Micro Volts
	Std. of Upper Trapezius activity	Trapstd	Standard Deviation
	Slope of Upper Trapezius activity	Trapslope	Micro Volts/Second
	Mean and Median frequency of Corrugator, Zygomaticus, and Trapezius	Zfreqmean Cfreqmedian Tfreqmean	Hertz
Temperature	Mean temperature	Tempmean	Degree Centigrade
	Slope of temperature	Tempslope	Degree Centigrade/Second
	Std. of temperature	Tempstd	Standard Deviation

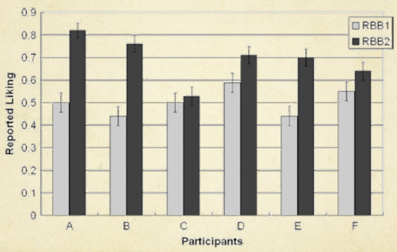


Liu et al. (2008)

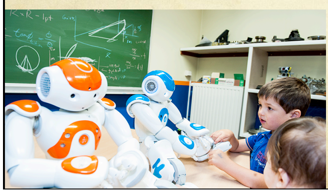




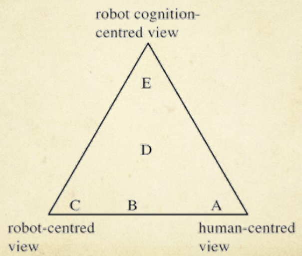
Main result



Evaluating HRI

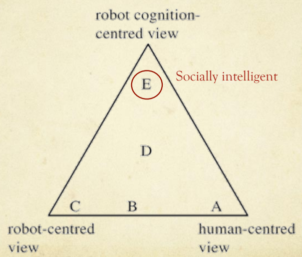


Evaluating HRI



Conceptual spaces of HRI approaches (Dautenhahn, 2010)

Evaluating HRI



Conceptual spaces of HRI approaches (Dautenhahn, 2010)

Evaluating HRI

- Requires (at a minimum) evaluation of
  - interaction characteristics
  - persuasiveness
  - trust
  - engagement
  - compliance

Steinfeld et al. (2006)

Evaluating HRI

- Requires (at a minimum) evaluation of
  - interaction characteristics
  - persuasiveness
  - trust
  - engagement
  - compliance
- The cognitive sciences have methods to investigate these (even if mainly for HHI)

Steinfeld et al. (2006)

Evaluating HRI

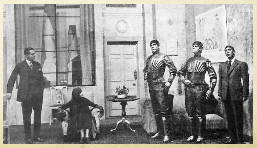
- The question is not whether or not one reserves "social interaction" and methods to measure it for human-human interaction alone…

Evaluating HRI

- The question is not whether or not one reserves "social interaction" and methods to measure it for human-human interaction alone…
- … it is firstly how the artificial agent is actually *perceived* (by the user).

A robot

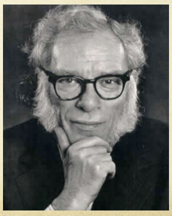
- Named by Josef Capek for his brother Karel's 1920 play "R.U.R. – Rossum's Universal Robots"
- From the Czech *robota*: serf work (i.e. in medieval times, peasant work for the lord who owned the land)



(Robotics)

(Robotics)

- Named by Isaac Asimov in his 1941 short story "Liar!"



Artificial agents are tools

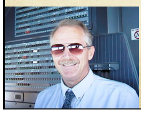
Artificial agents are can be social partners

Cognitive science, tools, and social partners

- Extended mind (à la Clark)
- Distributed cognition (à la Hutchins)
- Social cognition (e.g. à la DeJaegher)

Distributed Cognition

- "The mind is in the world"
  - Need to analyse the interactions between individuals, artefacts, types of representations used, etc.

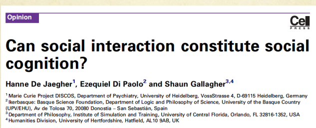


Extended mind

- *"A month ago, I bought an iPhone. The iPhone has already taken over some of the central functions of my brain. It has replaced part of my memory, storing phone [...] I make plans with it, using its calendar to help determine what I can and can't do in the coming months.*
  Friends joke that I should get the iPhone implanted into my brain. But if Andy Clark is right, all this would do is speed up the processing, and free up my hands. The iPhone is part of my mind already."
  ~ Dave Chalmers, foreword to Andy Clark's "Supersizing the mind"

Social cognition

Interactive explanation:
"One that relies on social interaction playing an enabling or constitutive role. Example: an infant behaves differently in a 'live' interaction because the coupling is more dynamically stable and disposes the infant to keep interacting as opposed to a playback of a previous interaction, which is dynamically less stable and easier to disengage from."


Opinion

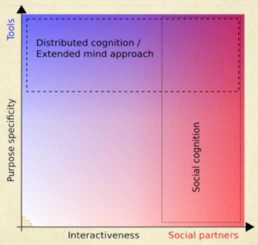
Cell PRESS

Can social interaction constitute social cognition?

Hanne De Jaegher[1], Ezequiel Di Paolo[2] and Shaun Gallagher[3,4]

[1] Marie Curie Project DISCOS, Department of Psychiatry, University of Heidelberg, VossStrasse 4, D-69115 Heidelberg, Germany
[2] Ikerbasque: Basque Science Foundation, Department of Logic and Philosophy of Science, University of the Basque Country (UPV/EHU), Av de Tolosa 70, 20080 Donostia – San Sebastián, Spain
[3] Department of Philosophy, Institute of Simulation and Training, University of Central Florida, Orlando, FL 32816-1352, USA
[4] Humanities Division, University of Hertfordshire, Hatfield, AL10 9AB, UK

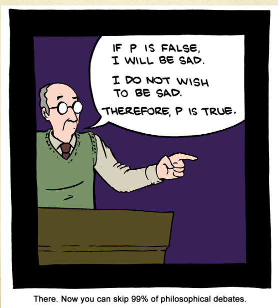
(Just) a schematic



(Thill & Ziemke, 2015)

A note on embodiment

Warning: May get philosophical for a second



(From smbc-comics.com)

Some (approximately parallel) developments

- Embodied cognition:
  - We cannot explain human cognition while ignoring the body
  - Our models need bodies
  - Look: robots!
  - (Look: simulations!)
  - Embodied models of cognition: models of cognition instantiated in an agent

Some (approximately parallel) developments

- Embodied cognition:
  - We cannot explain human cognition while ignoring the body
  - Our models need bodies
  - Look: robots!
  - (Look: simulations!)
  - Embodied models of cognition: models of cognition instantiated in an agent
- Robotics:
  - Robots become increasingly cheap
  - Engineering issues become increasingly solved
  - Time to find some cool applications!
  - Cool applications require cool behaviour
  - Where do we get that from?
  - Look: human cognition!

Obvious symbiosis going on…

- Embodied cognition:
  - We cannot explain human cognition while ignoring the body
  - Our models need bodies
  - Look: robots!
  - (Look: simulations!)
  - Embodied models of cognition: models of cognition instantiated in an agent
- Robotics:
  - Robots become increasingly cheap
  - Engineering issues become increasingly solved
  - Time to find some cool applications!
  - Cool applications require cool behaviour
  - Where do we get that from?
  - Look: human cognition!

Obvious symbiosis going on…

- Embodied cognition:
  - We cannot explain human cognition while ignoring the body
  - Our models need bodies
  - Look: robots!
  - (Look: simulations!)
  - Embodied models of cognition: models of cognition instantiated in an agent
- Robotics:
  - Robots become increasingly cheap
  - Engineering issues become increasingly solved
  - Time to find some cool applications!
  - Cool applications require cool behaviour
  - Where do we get that from?
  - Look: human cognition!

Obvious symbiosis going on…

- Embodied cognition:
  - We cannot explain human cognition while ignoring the body
  - Our models need bodies
  - Look: robots!
  - (Look: simulations!)
  - Embodied models of cognition: models of cognition instantiated in an agent
- Robotics:
  - Robots become increasingly cheap
  - Engineering issues become increasingly solved
  - Time to find some cool applications!
  - Cool applications require cool behaviour
  - Where do we get that from?
  - Look: human cognition!

But one important difference

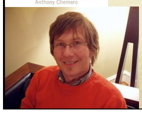
But one important difference

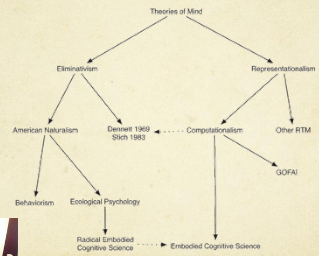
- Interested in **explanatory** models of *biological* cognition

But one important difference

- Interested in **explanatory** models of *biological* cognition
- The only mechanistic requirement is that the robot behaves as specified.
- No need to be biological plausible (but no worries if it is).
- The explanatory requirement goes away

(Note: this is really about "mainstream" embodied cognition)

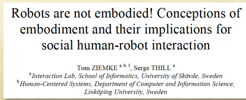




Chemero (2009)

And then something curious happens

- Some researchers in "embodied" AI will
  - Agree with Searle's criticisms of strong AI
  - Think embodied AI solves the issue


Robots are not embodied! Conceptions of embodiment and their implications for social human-robot interaction

Tom ZIEMKE [a, b, 1], Serge THILL [a]
[a] *Interaction Lab, School of Informatics, University of Skövde, Sweden*
[b] *Human-Centered Systems, Department of Computer and Information Science, Linköping University, Sweden*

Like so

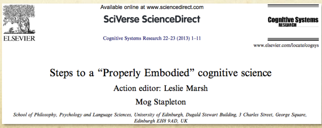
- *"Suppose we put a computer inside a robot, and this computer would not just take in formal symbols as input and give out formal symbols as output, but rather would actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating drinking – anything you like. The robot would, for example have a camera attached to it that enabled it to 'see,' it would have arms and legs that enabled it to 'act,' and all of this would be controlled by its computer 'brain.' Such a robot would [...] have genuine understanding and other mental states"*

Like so

- *"Suppose we put a computer inside a robot, and this computer would not just take in formal symbols as input and give out formal symbols as output, but rather would actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating drinking – anything you like. The robot would, for example have a camera attached to it that enabled it to 'see,' it would have arms and legs that enabled it to 'act,' and all of this would be controlled by its computer 'brain.' Such a robot would [...] have genuine understanding and other mental states"* (Searle, 1980, "The robot reply")

Proper embodiment

- *"[...] recent work in neuroscience and robotics suggests that cognitive systems are not merely superficially embodied in the sense that the sensorimotor interactions with the environment are the only interactions relevant to cognitive behaviour, but that cognitive systems are "properly embodied"; the internal body matters to cognition"*. (Stapleton, 2013)



ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Cognitive Systems Research 22–23 (2013) 1–11

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Steps to a "Properly Embodied" cognitive science

Action editor: Leslie Marsh

Mog Stapleton

School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, George Square, Edinburgh EH8 9AD, UK

CARS

Another type of robot (?)

Why cars?

- Past cars were passive tools for navigating from A to B
- Future cars may be entirely automated vehicles
- Present cars, however, are increasingly intelligent and interactive systems
- This makes them very interesting if you are interested in **interaction and collaboration** between humans and artificial agents

Autonomous cars

- Obvious: they drive themselves
  - Google car; CARS (in Stanford)
- Less obvious: automatically take decisions and actions while you drive (your car may already be doing this)
  - Lane assistance; active safety systems; etc.

Interaction matters even for autonomous cars

"Apparent" intelligence

- **ap·par·ent** *adjective* \ə-ˈper-ənt, -ˈpa-rənt\ :
  - easy to see or understand
  - **seeming to be true but possibly not true**

(Merriam Webster dictionary)

- A system may well have very complicated and clever algorithms that can reasonably be described as intelligent
  - But a user is not necessarily aware of these
- Conversely, humans project cognitive abilities even on simple shapes
  - This doesn't mean that there are any cognitive abilities
- **What people believe about a system may be more important than what is actually true about a system**

Basic motivation

- Humans adapt their interaction in accordance with beliefs about, e.g., the other's cognitive abilities
- This extents to artificial agents (e.g. robots)
  - Vollmer et al. (2011)
  - Kopp (2010)
- Would this remain true for vehicles?
  - Cars in particular

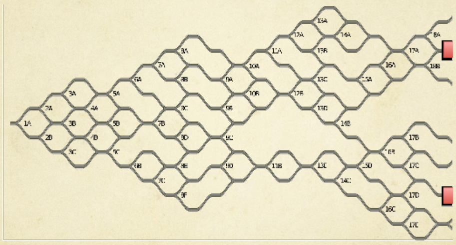
What we did (overview)

- Simulator study
  - AutoSim simulator environment
  - SmartEye 5.90 eyetracker
- Primary task
  - Navigate towards goal locations (with the help of a navigation aid)
- Secondary task
  - Count sounds (errors indicative of cognitive load)
- Manipulation of interest
  - Perceived intelligence of the car (in particular navigation aid)
    - Traffic density (= quality of the roads chosen by the navigation aid)
    - Information provided by the navigation aid
- Question
  - Does driver behaviour change in function of how intelligent they **believe** the system is? (Hypothesis: it does)

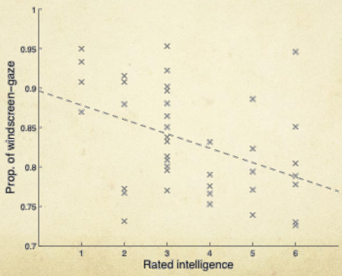
Simulator environment

Navigation aid design

Environment map



Rated intelligence vs. looking ahead



What did people think about the navigation aid?

- **10** preferred arrow + text **specifically** because they liked the additional information
- **4** preferred arrow + text for no clear reason
- **3** preferred arrow **specifically** because they did not feel the need for additional information
- **1** preferred the arrow in the experiment, would prefer the additional information in real life
- **5** preferred arrow, mostly because they did not like the textual modality/having to read but would have liked the information in some other way

Insight 1

- All results support the hypothesis that drivers' behaviour is correlated with beliefs about the systems intelligence
  - Consistent with research on interaction with other artificial agents
- This is pretty cool
  - Cognitive scientists may look at human/car systems if there is an interest in interaction between humans and artificial intelligent agents "in the wild"
    - (as opposed to artificial experiments with prototype robots in the lab)
  - Implications for car design
    - How intelligent it appears may affect driver behaviour
    - Large potential for future research here

Insight 2

- Drivers prefer the informative navigation aid
  - Again in line with previous research
    - Informative trumps minimalist (Vashitz et al, 2008)
    - Informativeness influences trust (Verberne et al, 2012)
- Prefer this information to be concise
  - Text not appreciated (as also noted by e.g. Evans and Stevens, 2007)
  - Icons may be preferable (suggested by some participants)
  - Still, most participants were willing to compromise on conciseness in favour of information received

Vehicle automation and trust

- Automated vehicle's apparent intelligence can be influenced by many distinct systems
  - How people perceive automation is important (e.g. van Driehl et al, 2007; Mahr and Müller, 2011)
- Mismatch between expected and actual abilities is a bad thing
  - Can lead to an inability to detect the limits of the system's abilities (Lee and See, 2004)
  - Expectations can influence trust
    - Note that higher trust is not always more desirable (Helldin et al., 2013)

Example

But does the world want autonomous cars?

Summary

- Social cognition
  - Specifying robot behaviour
- Emotions and affect
  - Improved human robot-interactions
- Evaluations of HRI
  - The relevance of the cognitive sciences
- Embodiment
  - No such thing as an "embodied" AI; most things are as computational as ever
- Interaction with artificial agents
  - Expectations, attribution of abilities

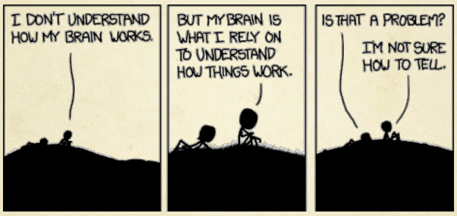
Final take home messages

- Opinion:
  - Human-level cognition is for social interaction

Final take home messages

- Opinion:
  - Human-level cognition is for social interaction
- Social interaction needs a thorough understanding of human cognition
- May lead to better insights on how machines can be sociable
- But it may also lead to a better understanding of human cognition
  - Not because of the necessary cognitive models
  - Rather because they force us to understand how humans interact with machines

Thanks for listening!



(from xkcd.com)