Conceptual Autonomy of Agents

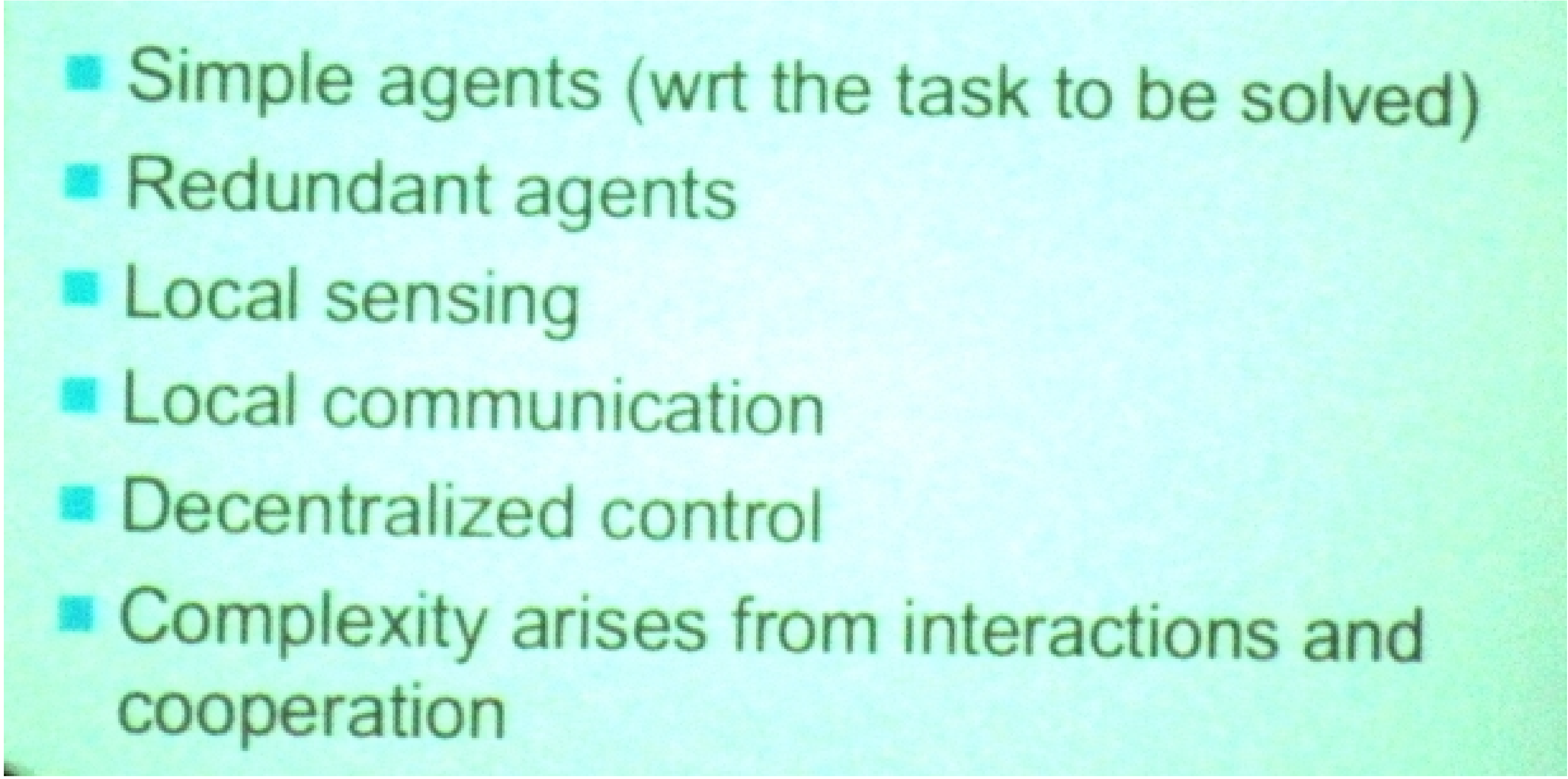
Timo Honkela

Helsinki University of Technology
Adaptive Informatics Research Center
Cognitive Systems Research Group

ICAART'09
21st of January 2009

Part I: Introduction

Swarm intelligence principles
(Marco Dorigo yesterday)



- Simple agents (wrt the task to be solved)
- Redundant agents
- Local sensing
- Local communication
- Decentralized control
- Complexity arises from interactions and cooperation

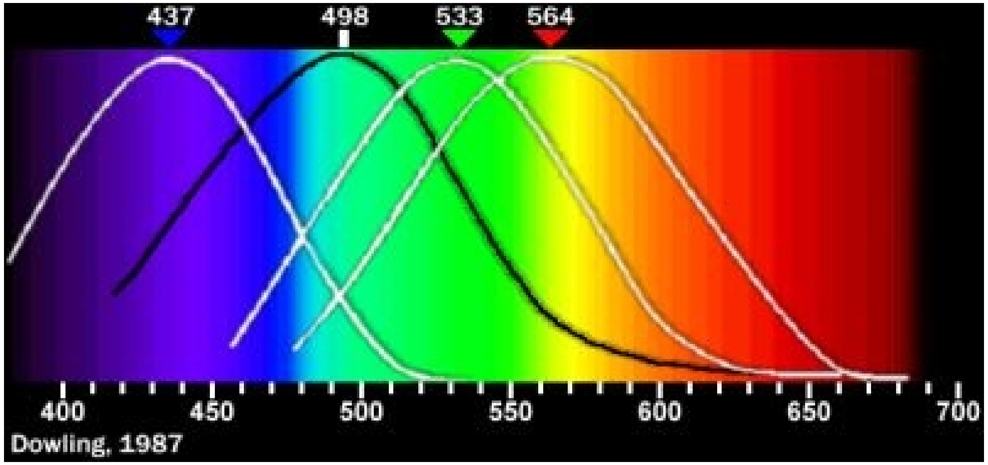
Human intelligence principles

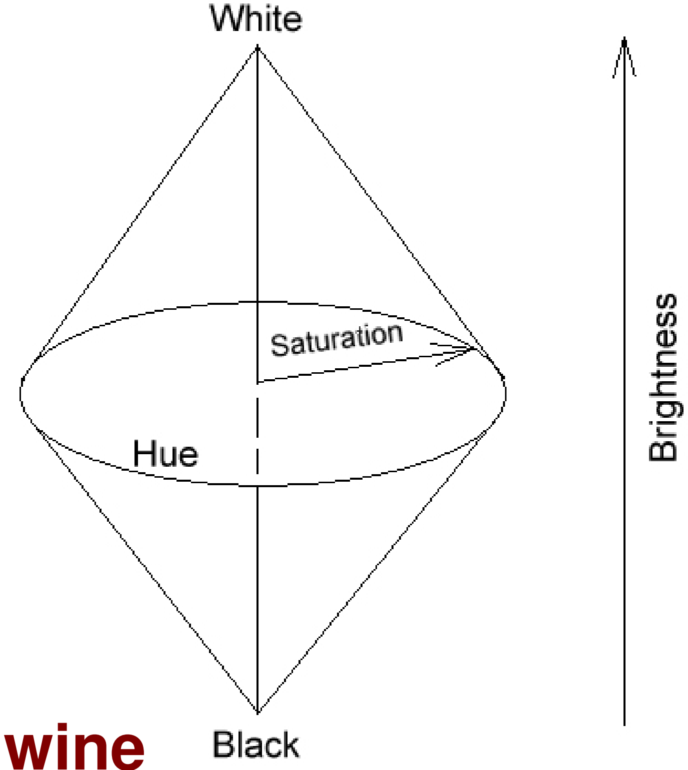
- Rather complex agents but with bounded rationality
- Individual life paths
- Varying skills, knowledge and experiences
- Experiences inherently multimodal
- Only indirect means to share cognitive contents
- Mostly decentralized control

Challenge of communication

- As people learn language through individual life path and thus end up having different (subjective) ways understanding and producing language how is successful communication possible?

Simple challenge: Color naming





Human vision: rods, cones,...
Physical reasons for color
Contextuality of naming

red wine
red skin
red shirt

→ Hardin
→ Gärdenfors

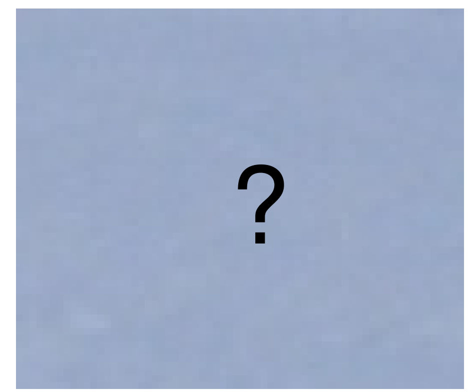
Color naming, cont'd

gray sky

blue water

white snow

blueish cloud



Color naming, cont'd

gray sky

blue water

→ white snow

blueish cloud

Example: concept of computation

Example:
concept of computation

Subjectivity of interpretation

Meaning negotiation

Complex challenge: different cultural contexts

?

?

"Shall I compare thee
to a summer's day?"

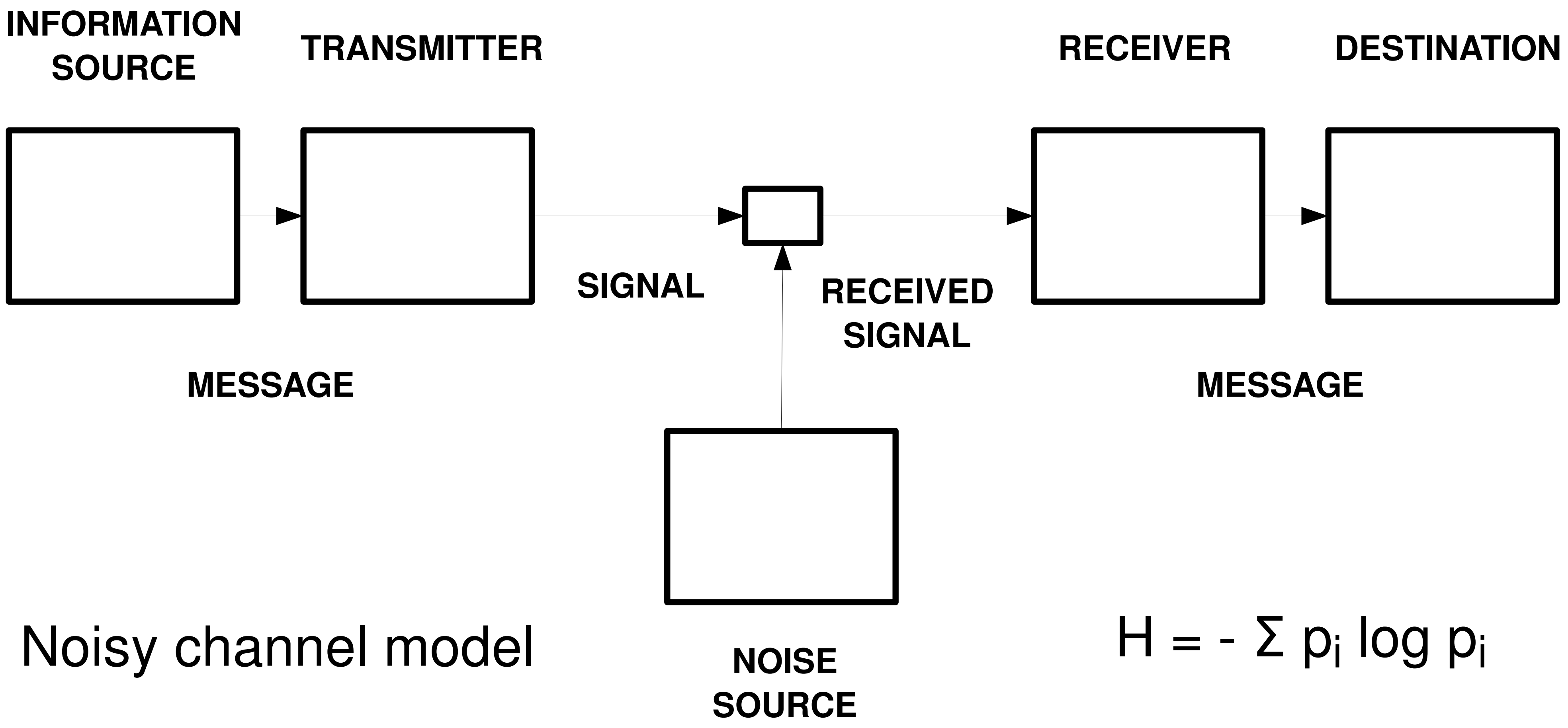
Experiential grounding of human knowledge

Human understanding of the world and of the relationship between language use and perception and action within the world is based on a long active and interactive learning process for which the genotype gives a certain basis but which is mainly determined by the individual interaction with the world including other human beings and the social and cultural context

Meaning, communication and information

- In the following, we can remind ourselves how information theory deals with communication and meaning

General communication system and measuring information (Shannon & Weaver)



Noisy channel model

$$H = -\sum p_i \log p_i$$

Weaver on Shannon

- “Relative to the broad subject of communication, there seem to be problems at three levels. [...]
  - LEVEL A. How accurately can the symbols of communication be transmitted? (The technical problem)
  - LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem)
  - LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem)”
- “The semantic problems are concerned with the identity, or satisfactorily close approximation, in the interpretation of meaning by the receiver, as compared with the intended meaning of the sender.” (1949, p. 4)

Traditional formalization of meaning

- Formalisms like first-order predicate logic have widely been used as a basis for theories of meaning (epistemology); consider also contemporary efforts such as Semantic Web
- These formalisms, however, seem to provide only limited means for creating in-depth theories of how **language** is **understood**

Traditional formalization of meaning, cont'd

- Traditional logic provides means e.g. for modeling quantification, connectives, analytical truths and conceptual hierarchies
- However, many semantic phenomena are matters of degree. Various proposals that apply Bayesian probability theory or fuzzy sets deal with this.

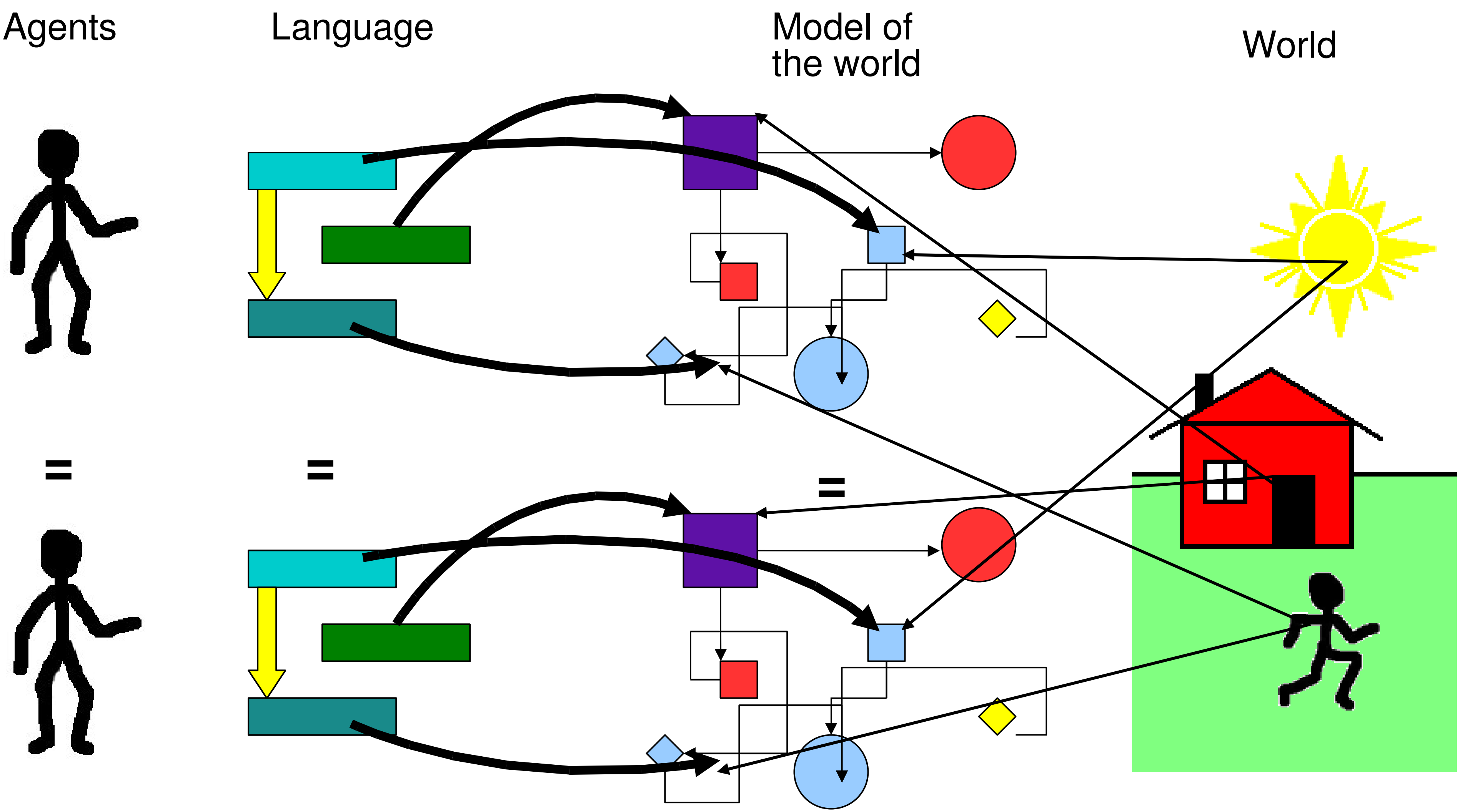
Neglected pattern recognition processes

- Even these methodological extensions do not suffice if the pattern recognition processes are not taken into account
- The world is not straightforwardly experienced as discrete objects and events but there are complex underlying cognitive processes involved

The role of cognition(s)?

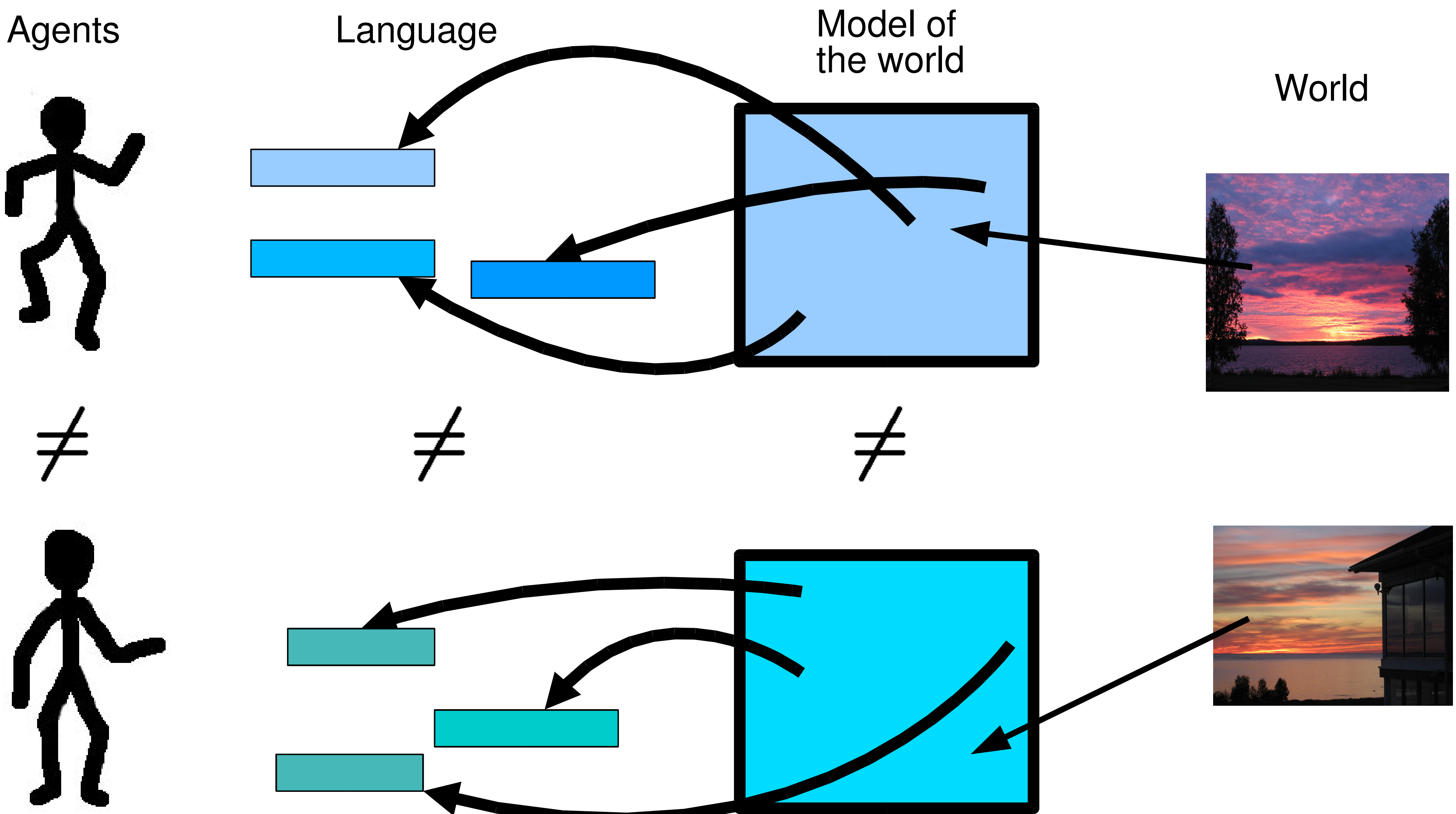
- Moreover, many philosophers and formalizers have been keen on denying the role of **individual cognition** or the **society of cognitions** in the theory of meaning
- Nevertheless, language is primarily a means for communication (between different individuals and over time)

Traditional logico-analytic viewpoint



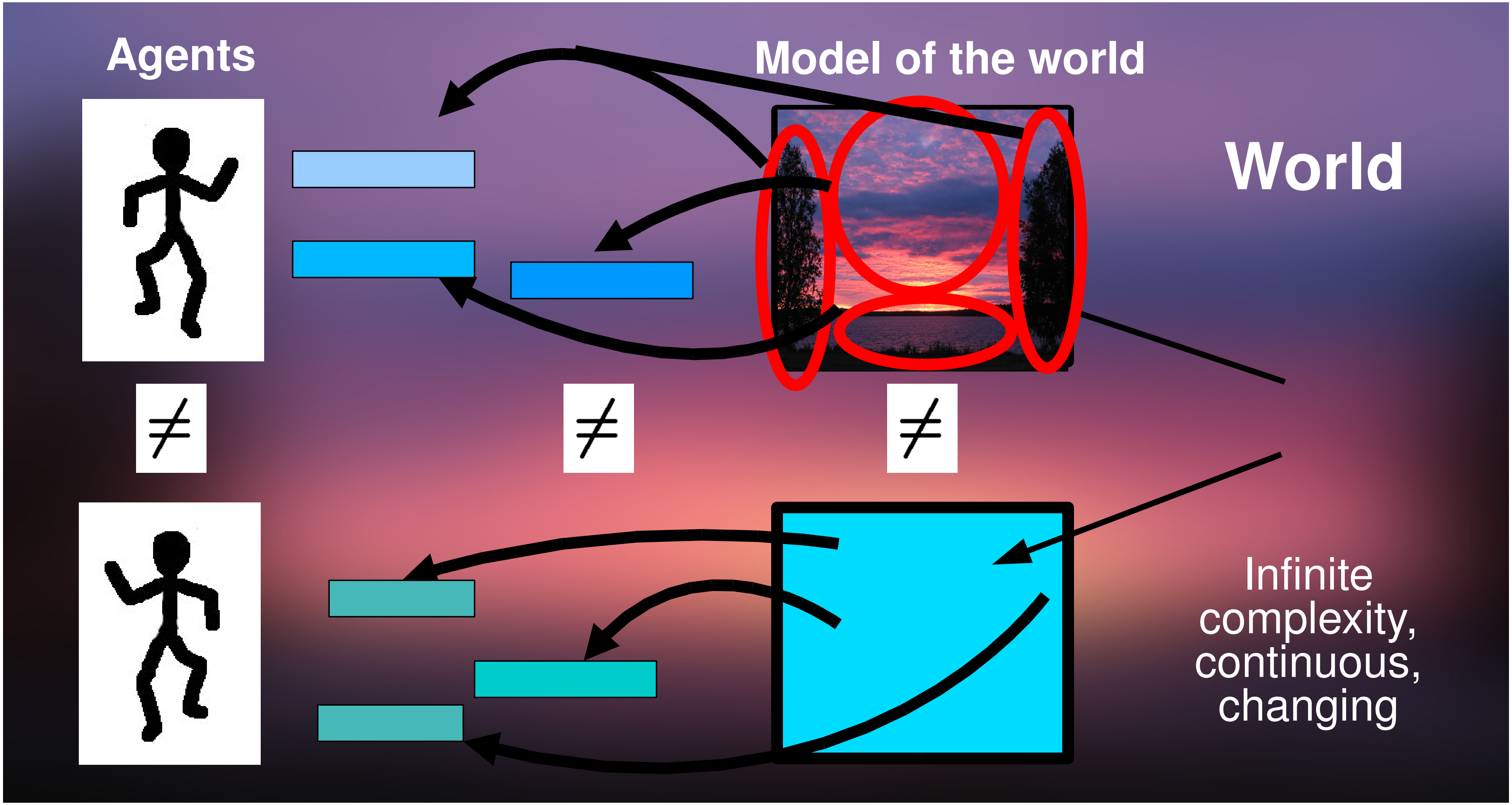
Emergentist viewpoint

(Emphasis on pattern recognition and learning)



Emergentist viewpoint

(Emphasis on constructivism)



Relevance?

- A large proportion of modern human activity in its different forms (science, industry, society, culture, etc.) is based on the use of language
- There are at least 6000 languages in the world and many more dialects
- Each language has the order of 10^5 to 10^{10} different word forms
- Each word is understood differently by each speaker of that language at least to some degree

Relevance, cont'd

- The formal basis of (in practice) all information systems does not take this basic phenomenon into account
- The assumption of shared meanings is simply not adequate
- A level of socio-cognitive modeling is needed
- In the following, a tentative version of Intersubjective Meaning Spaces (IMS) is outlined to serve as a basis for a novel kind of formalization of meaning

Part II: Drafting Methodology

Word space models

- Word space models represent meaning as points or areas in a high dimensional vector space
  - Self-Organizing Semantic Maps
    (Ritter & Kohonen 1989, Honkela, Pulkki & Kohonen, 1995)
  - LSA (Landauer & Dumais 1997)
  - HAL (Lund & Burgess 1996)
  - Conceptual Spaces (Gärdenfors 2000)
  - Word ICA (Honkela, Hyvärinen & Väyrynen 2004)
  - etc. etc.

Distributional hypothesis

- Two words are semantically similar to the extent that their contextual representations are similar (Miller & Charles 1991)
- The meaning of words is in their use (Wittgenstein)

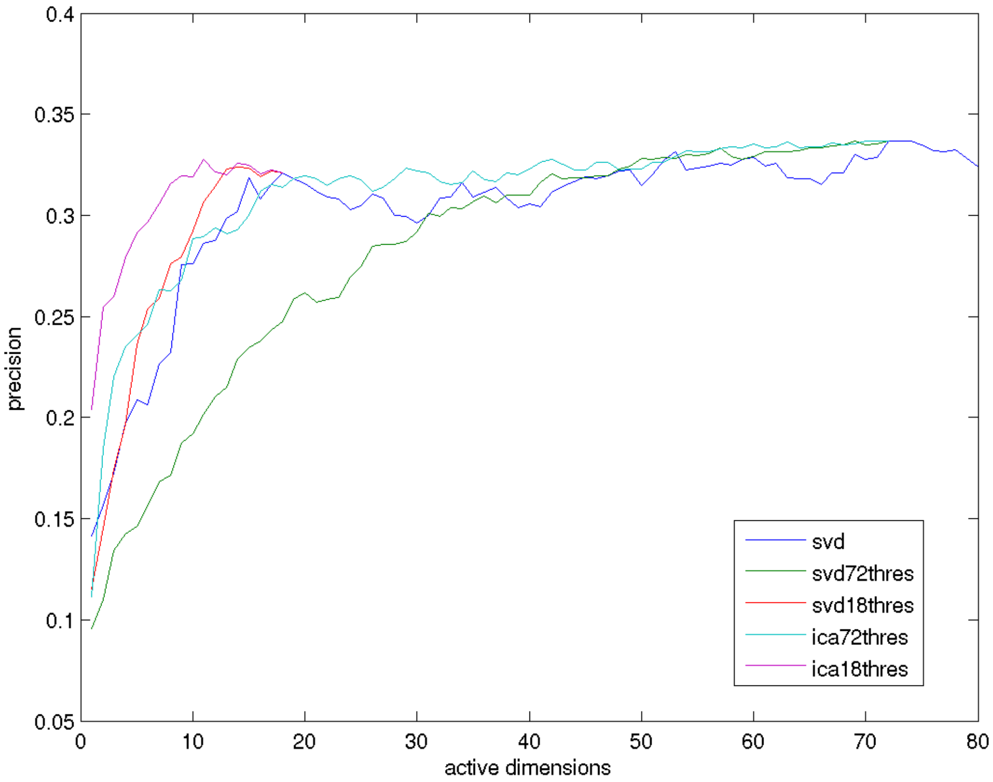
Language as dimensionality reduction?

ICA of word contexts; nonlinearity through thresholding

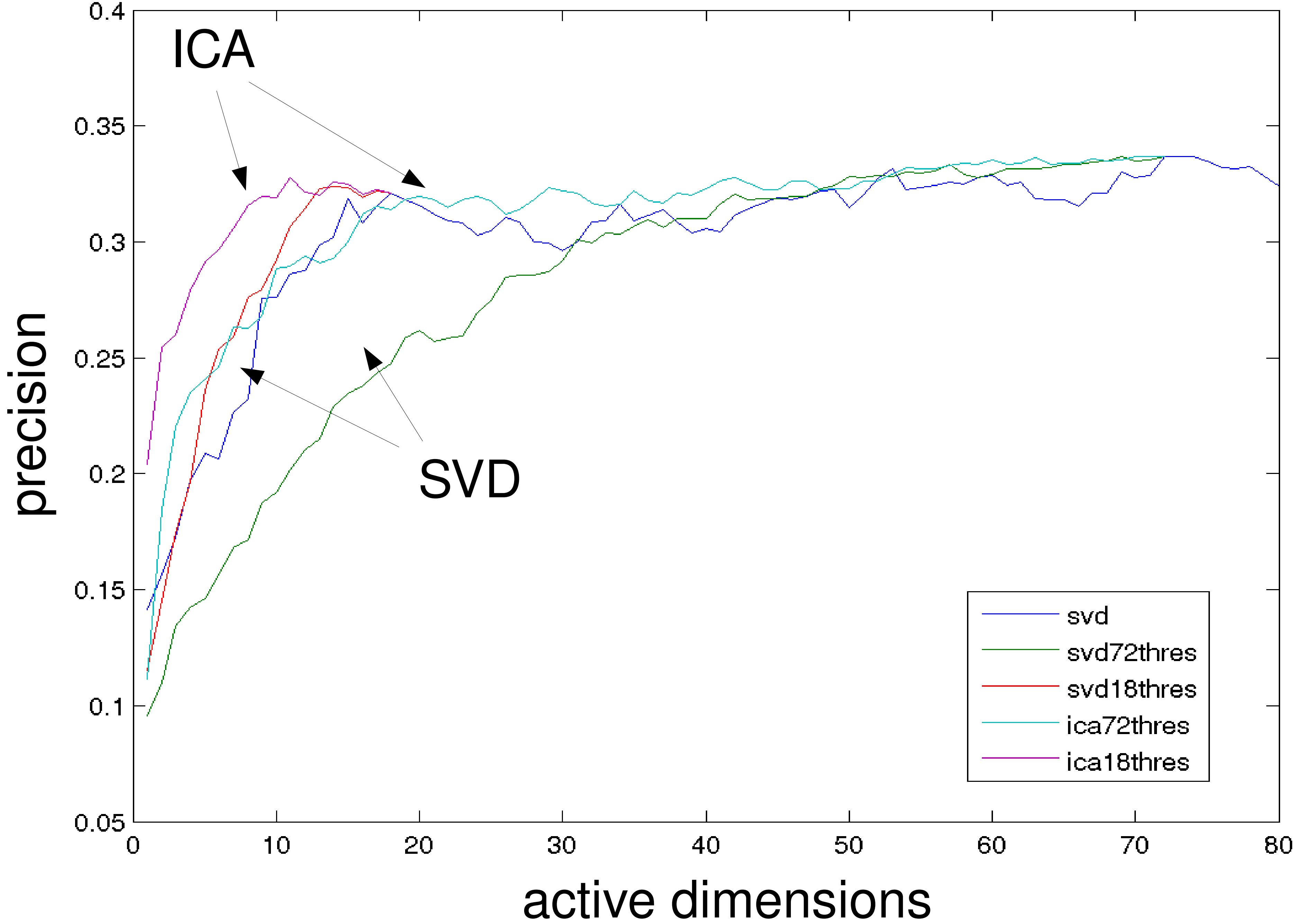
Comparison with SVD/LSA

Effect of sparseness and meaningful emergent components

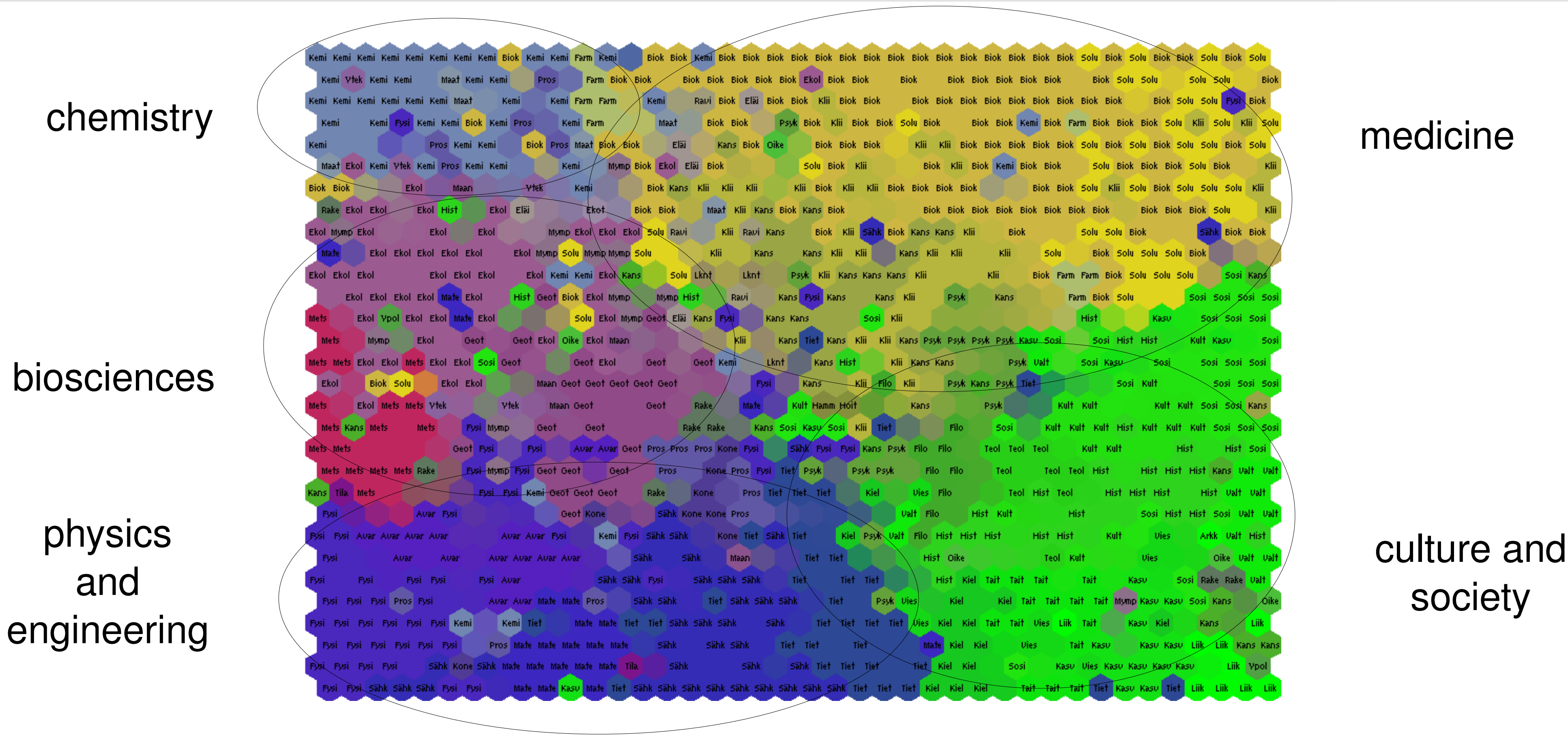
Data: TOEFL tests



(Väyrynen, Lindqvist, Honkela 2007)


ICA
SVD
precision
active dimensions
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
10
20
30
40
50
60
70
80
svd
svd72thres
svd18thres
ica72thres
ica18thres

Application example: Document map of research



A map of 3224 funding applications to Academy of Finland

(Honkela & Klami, 2007)

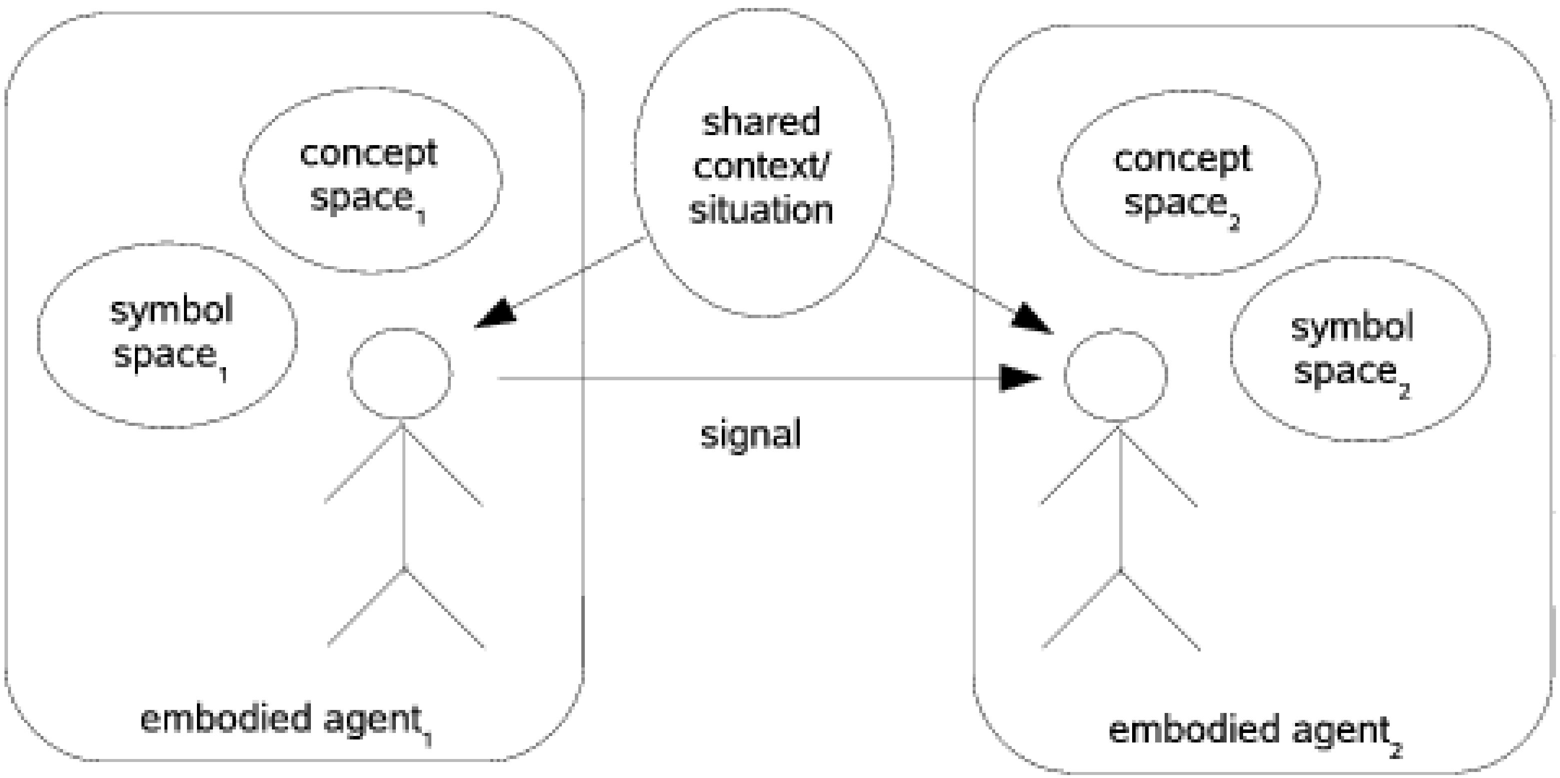
Complication: Subjectivity of Meaning

- As indicated already earlier, almost all formal or computational theories of meaning are based on the assumption that meanings are shared
- Weaver was aware of this limitation in information theory (cf. A, B, C)

Tiered logic for representation of different contexts of agents

Prof. John Crossley earlier today

Theory of Intersubjective Meaning Spaces (IMS)



(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- Agent's internal view of its context, *the concept space*
- The concept space is spanned by a number of features.
- Dimensionalities of the concept spaces can be different for each agent
- Therefore, we denote the features used by agent 1 f_i^1, $i = 1 \ldots N$
- Similarly, agent 2 uses features f_i^2, $i = 1 \ldots M$
- Thus, the concept space of agent 1 is an N-dimensional metric space C^1, and for agent 2, C^2

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- In the framework, there exist two distance measures, namely ω and λ. ω gives a distance between two points inside the concept space of the agent, i.e. $\omega : C^i \times C^i \rightarrow \mathbb{R}, i = 1, 2,$
- λ gives a distance between two points in the concept spaces of the different agents, i.e. $\lambda : C^i \times C^j \rightarrow \mathbb{R}, i \neq j$
- The symbol space S^1 of agent 1 is its vocabulary that consists of discrete symbols

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- An agent i has an individual mapping function ξ^i that maps the symbol $s^i \in S^i$ to C^i
- An agent i expresses each symbol $s^i \in S^i$ as a signal d in the signal space D
- We assume that the signal space D is multidimensional, continuous and shared between the agents
- However, each agent i has an individual mapping function ϕ^i from its vocabulary to the signal space, i.e. $\phi^i : S^i \rightarrow D$ and an inverse mapping ϕ^{-i} from the signal space to the symbol space

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- The sender, agent 1, selects a symbol that corresponds best to the current context by the means of its own conceptual model
- Therefore this model can be seen as an optimal decision problem.
- Let us denote the symbol agent 1 selects and communicates to agent 2 as s^* and the features agent 1 observes corresponding to the current context as f^1. Then agent 1 selects the symbol that corresponds the current observations best by the means of some distance measure ω

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- One suitable choice for ω in that case is for example Euclidean distance
- Formally this can be represented as in Eq. (1).

$$s^* = \arg\min_{s \in S^1} \omega(f^1, \xi^1(s)) \qquad (1)$$

- After symbol selection process, agent 1 communicates the symbol s^* to agent 2, i.e.:

$$d = \phi^1(s^*) \qquad (2)$$

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- When agent 2 observes d, it maps it to some $s^2 \in S^2$ by using the function ϕ^{-2}
- Then it maps the symbol to some point in its conceptual space by using ξ^2

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- If this point is very near of its own observation f^2, we can say that the communication process has succeeded

- Mathematically this is as follows:

$$||\xi^2(\phi^{-2}(d)) - f^2|| \leq \epsilon, \qquad (3)$$

where ϵ is a small constant and $||\cdot||$ is some suitable norm in C^2.

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- The two-agent model relates to the single-agent model very closely but now the sender has some estimate of the receiver's conceptual space available

- This model can be learned from communication samples or it can be known a priori

- The symbol selection process is formally given in Eq. (4)

$$s^* = \arg\min_{s \in \tilde{S}^2} \lambda(f^1, \tilde{\xi}^2(s)) \tag{4}$$

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- In this equation, $\tilde{\xi}^2$ is the model of the receiver
- Note that also the vocabulary of the receiver can be unknown and should be estimated by $\tilde{S}^2$

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- As the dimensionalities of the conceptual spaces C^1 and C^2 can be different, we should use some special method to calculate the distance between points in these two spaces
- Dynamic programming and neural networks may provide solutions to this task
- A neural network solution for the general case when agents have different representations for the same phenomenon has been considered for a supervised learning task by Laakso and Cottrell (2000) and for unsupervised learning by Raitio, Vigário, Särelä and Honkela (2004)

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- Other parts of the communication process remain the same as in the single-agent case
- As agent 1 utilizes the explicit model of agent 2, the symbol selection problem can be seen as a game theoretical problem

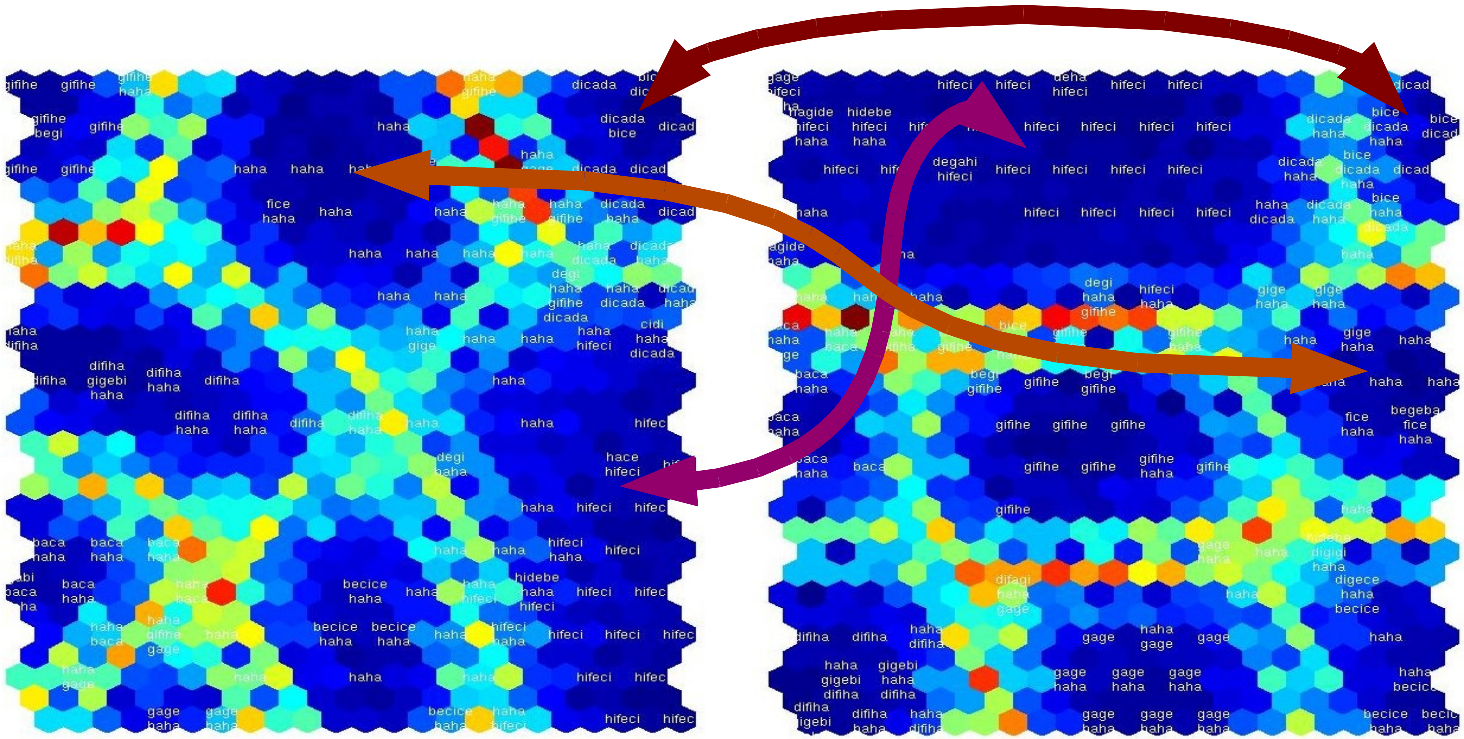
(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

IMS: Theoretical framework

- As an example related to the vocabulary problem, two persons may have different conceptual or terminological "density" of the topic under consideration
- A layperson, for instance, is likely to describe a phenomenon in general terms whereas an expert uses more specific terms
- In our framework, symbols generate clusters in the conceptual spaces of the agents

(Honkela, Könönen, Lindh-Knuutila & Paukkeri 2008)

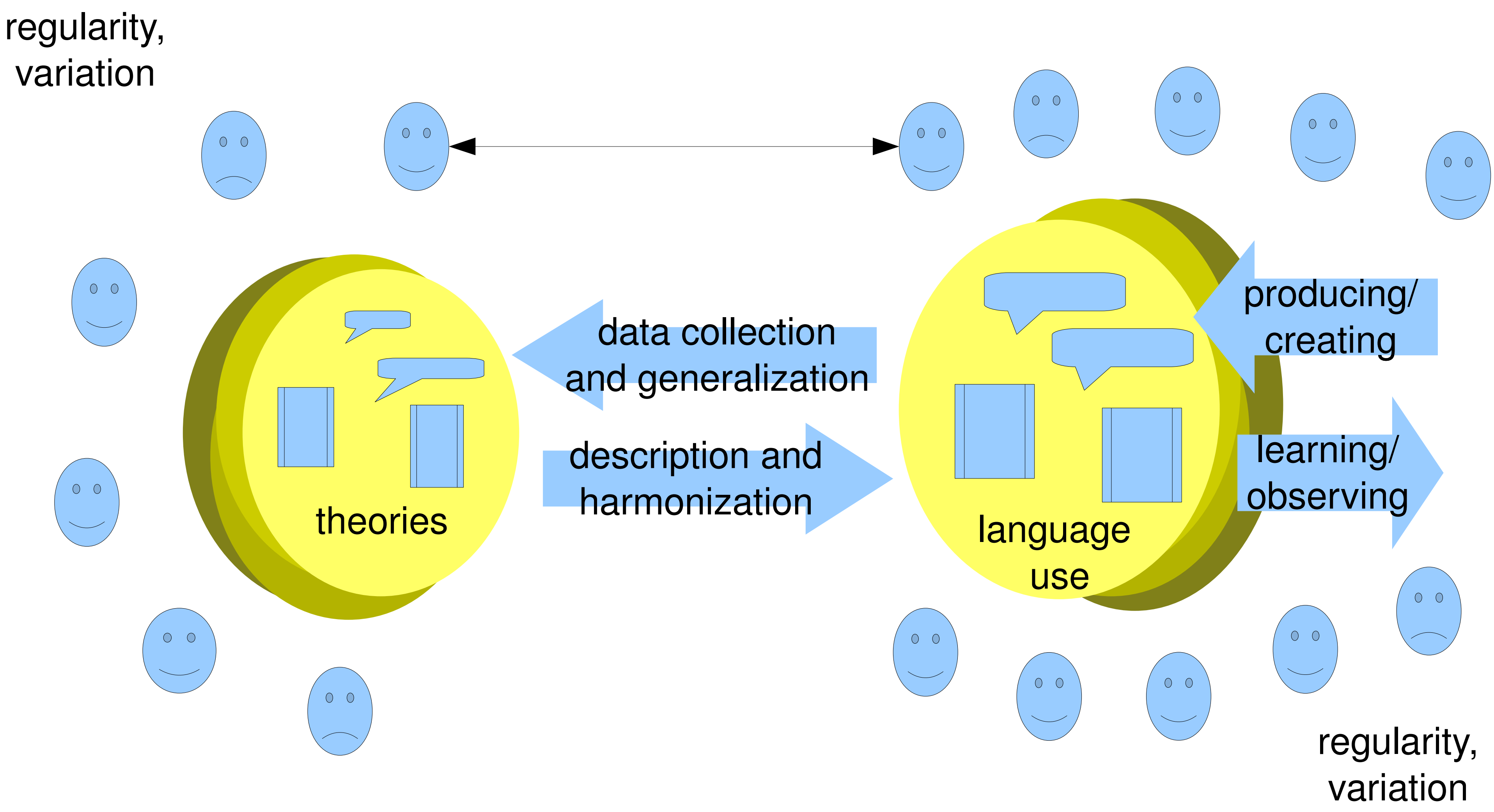
Emergence of a shared lexicon in a community of interacting SOM-based agents



(Lindh-Knuutila, Lagus & Honkela, SAB'06)

Related to e.g. Steels and Vogt on language games

Language use and linguistic theory formation as social phenomena



Practical consequences

- The traditional notion of uncertainty in decision making does not cover the uncertainties caused by differences in conceptual systems of individual agents within a community
- In many transactions, including symbolic/linguistic communication, the differences in the underlying conceptual systems play an important role

Practical consequences, 2

- Serious efforts have been made to harmonize or to standardize the classification systems or ontologies used by agents
- Even if standardization is conducted, there can not be any true guarantee that all participating agents would share the meaning of all the expressions used in the transactions in various context. Thus, the systems need to able to conduct **meaning negotiations**

Some references

Timo Honkela, Ville Könönen, Tiina Lindh-Knuutila, and Mari-Sanna Paukkeri. Simulating processes of concept formation and communication. Journal of Economic Methodology, 15(3):245–259, 2008.

Mikaela Klami and Timo Honkela. Self-organized ordering of terms and documents in NSF awards data. In Proceedings of WSOM'07. University Library of Bielefeld, 2007.

Tiina Lindh-Knuutila, Timo Honkela, and Krista Lagus. Simulating meaning negotiation using observational language games. In P. Vogt et al., editor, Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication, pages 168–179, Rome, Italy, 2006. Springer, Berlin/Heidelberg.

Timo Honkela and Aapo Hyvärinen. Linguistic feature extraction using independent component analysis. In Proceedings of IJCNN'07, pages 279–284, Budabest, Hungary, July 2004.

Thank you!