Multimodality

Multimodal Social Signal Processing - Lecture 19

Prof. Alessandro Vinciarelli
School of Computing Science &
Institute of Neuroscience and Psychology

http://www.dcs.gla.ac.uk/vincia
Alessandro.Vinciarelli@glasgow.ac.uk

This lecture is based on the following text (available on Moodle):

- Vinciarelli & Esposito, "Multimodal Analysis of Social Signals", in "The Handbook of Multimodal-Multisensor Interfaces", Oviatt et al. (eds.), 203-226, ACM, 2018;
- Partan & Marler, "Issues in the Classification of Multimodal Communication Signals", The American Naturalist, 166(2), pp. 231-245, 2005.

Outline

- Multimodality (Psychology & Neuroscience)
- Multimodality (Communication & Life Science)
- Multimodality (Computing Science & AI)
- Conclusions

Outline

- Multimodality (Psychology & Neuroscience)
- Multimodality (Communication & Life Science)
- Multimodality (Computing Science & AI)
- Conclusions

Gestalt Theory (I)

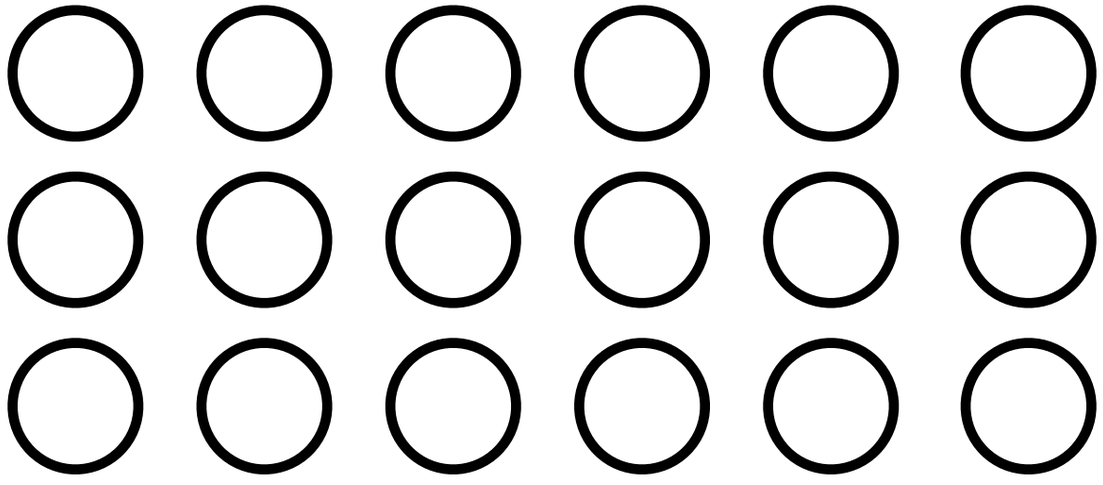
"It has been said: The whole is more than the sum of its parts. It is more correct to say that the whole is something else than the sum of its parts, because summing is a meaningless procedure, whereas the whole-part relationship is meaningful"

Koffka, "Principles of Gestalt Psychology", 1935

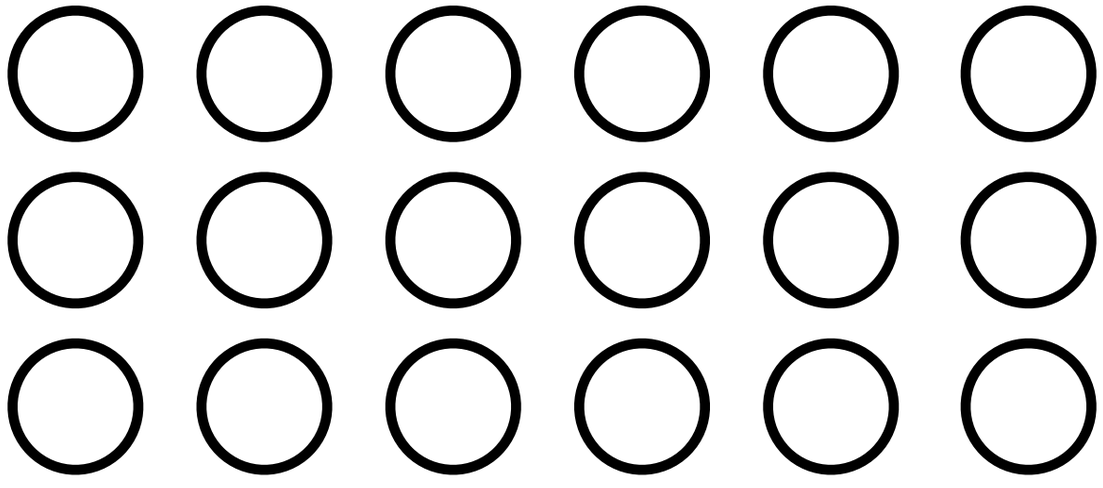
Gestalt Theory (II)

"Gestalt Theory describes different laws or principles for perceptual grouping of information into a coherent whole, including the laws of proximity, symmetry, similarity, closure, continuity […]"

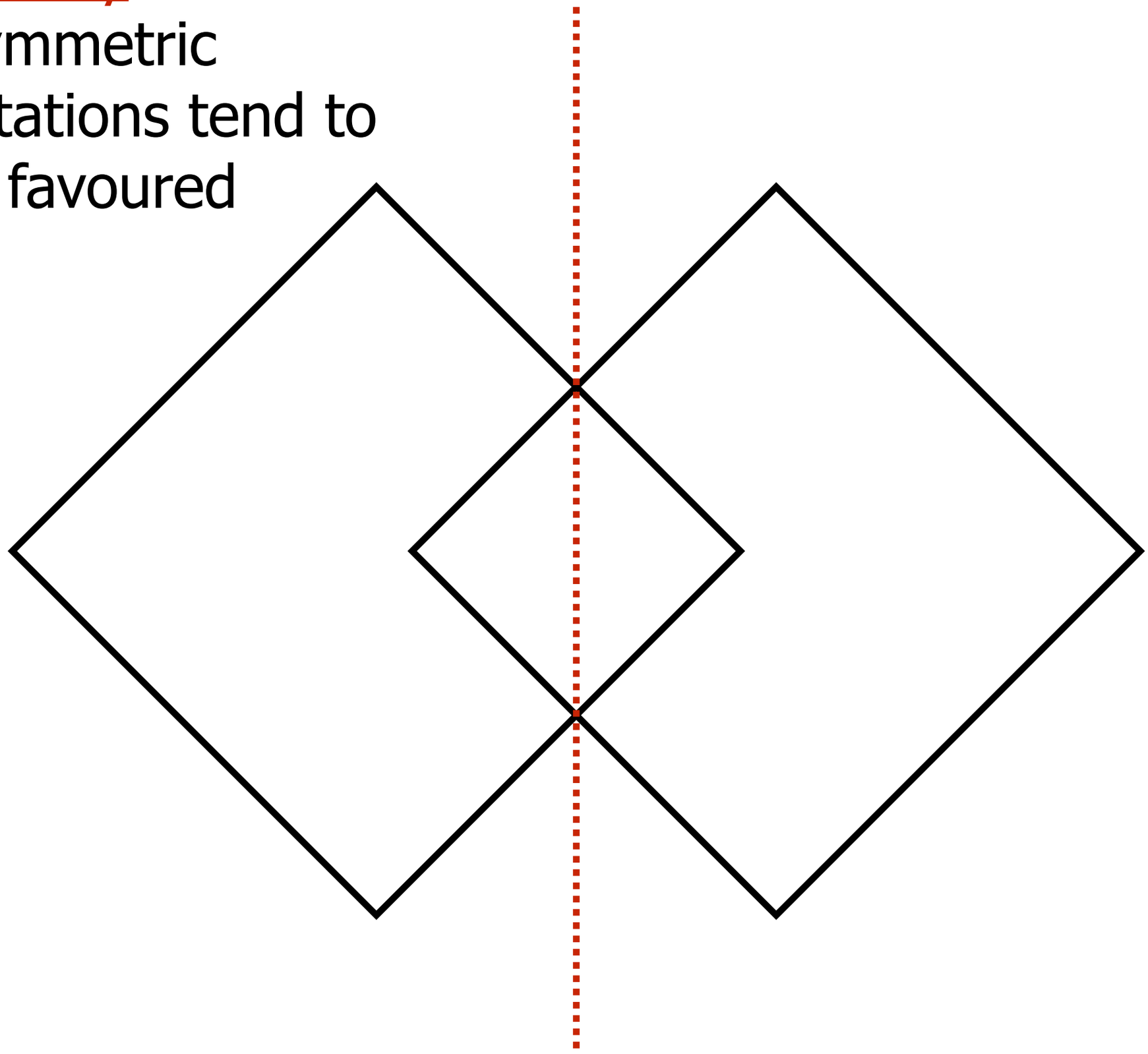
Oviatt, "Theoretical Foundations of Multimodal Interfaces and Systems", in "The Handbook of Multimodal-Multisensor Interfaces", pp. 20-50, 2018.

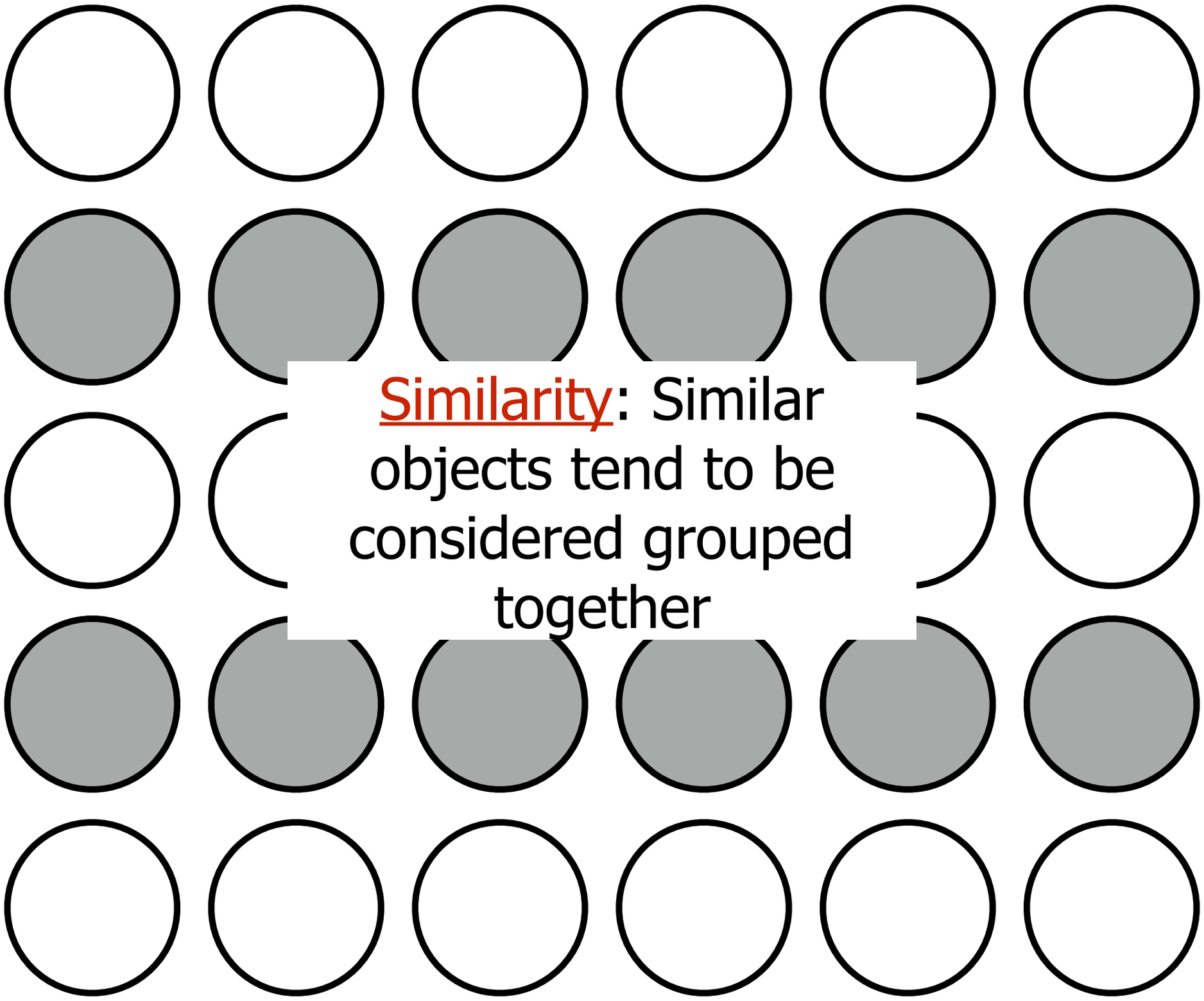


Proximity: Objects close to each other tend to be considered grouped together

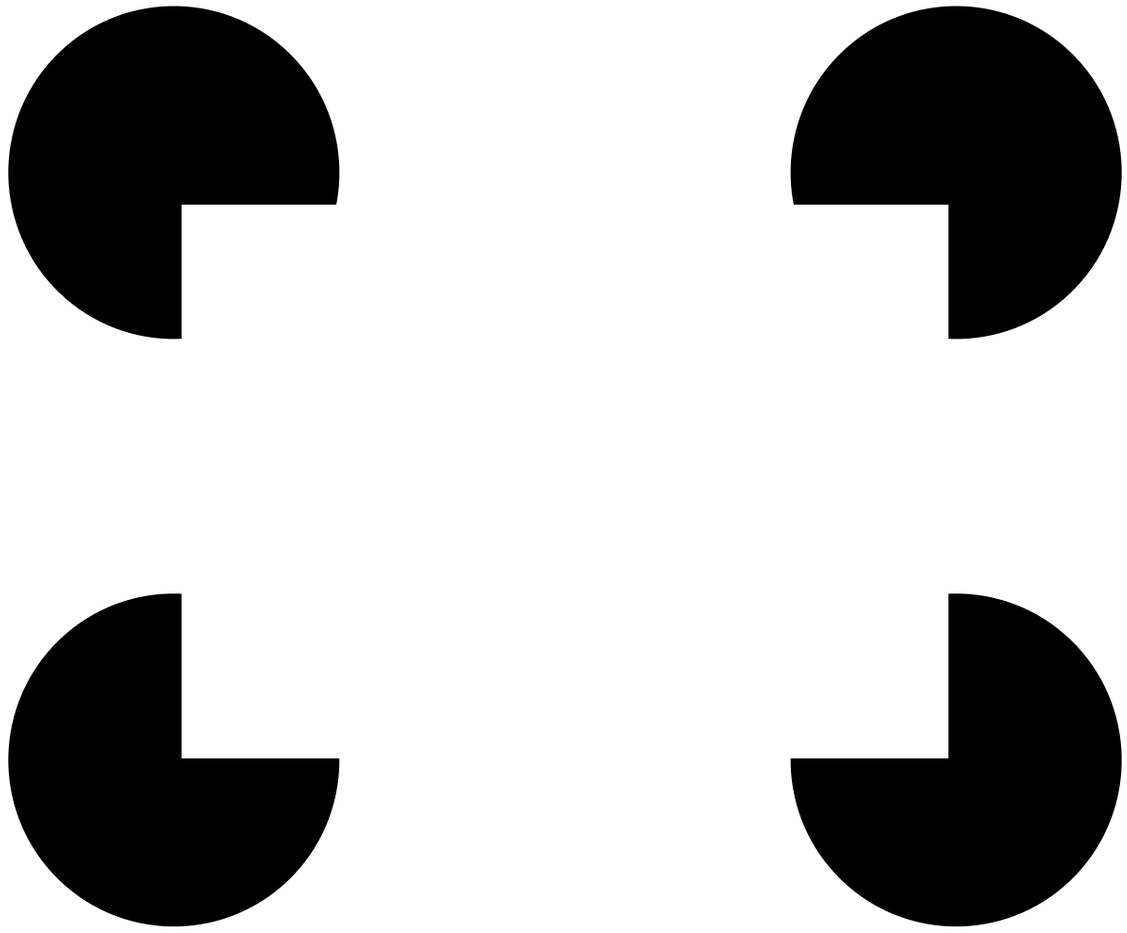


Symmetry: More symmetric interpretations tend to be favoured

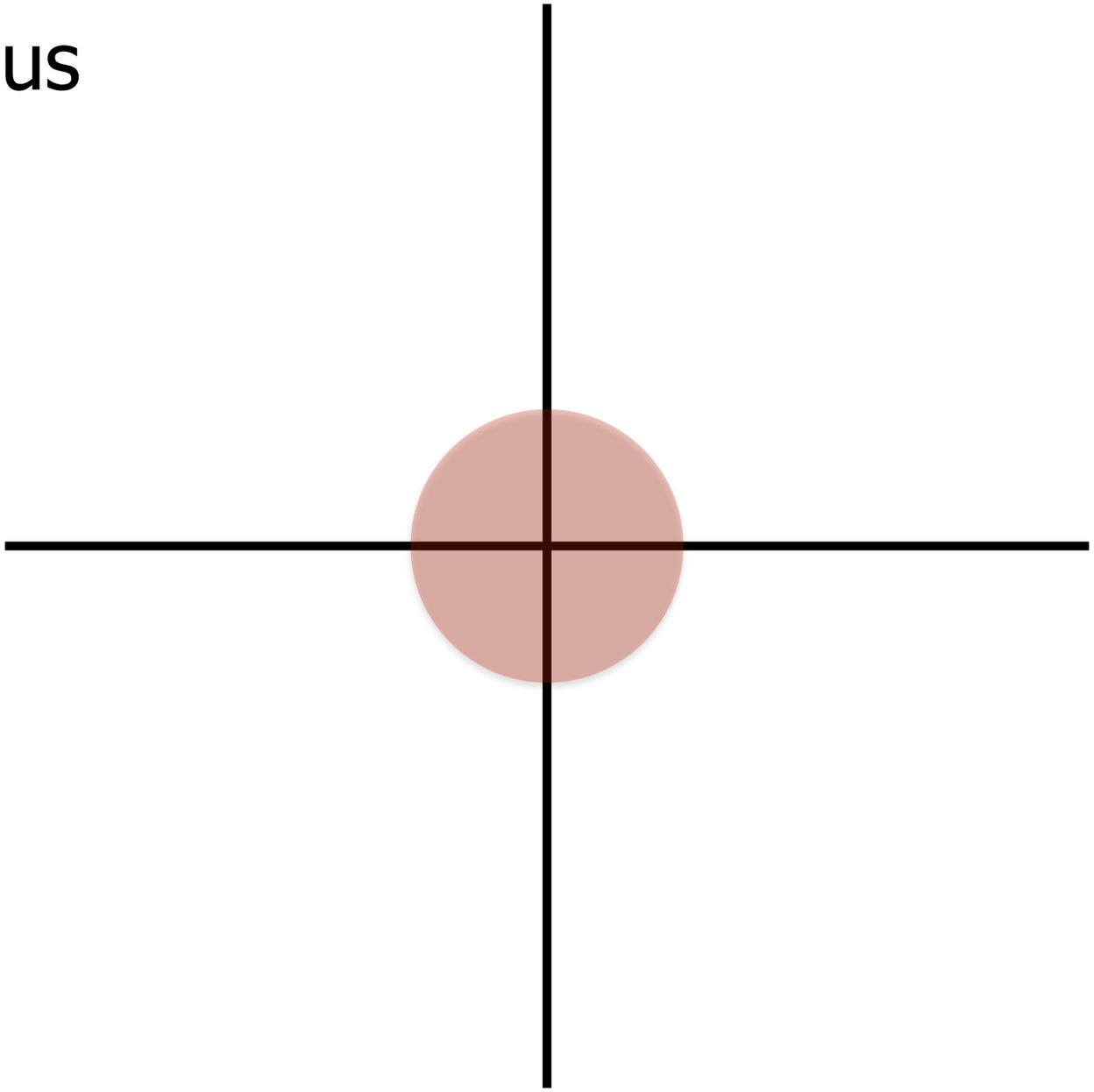



Similarity: Similar objects tend to be considered grouped together

Closure: close and complete figures tend to be perceived in absence of continuous lines



Continuity: lines that follow one another tend to be considered continuous



Multy-Sensory Perception

"[…] brain processing fundamentally involves multi sensory perception and integration […] which cannot be accounted for by studying the senses in isolation."

Oviatt, "Theoretical Foundations of Multimodal Interfaces and Systems", in "The Handbook of Multimodal-Multisensor Interfaces", pp. 20-50, 2018.

Mc Gurk Effect

https://www.youtube.com/watch?v=G-lN8vWm3m0

Recap

- Our brain does not process multiple (possibly multi-sensory) signals individually;
- The initial intuitions proposed in the early 20th century (Gestalt Psychology) have been later confirmed by neuroscience;
- When it comes to perception, "the whole is other than the sum of the parts" (Koffka, 1935).

Outline

- Multimodality (Psychology & Neuroscience)
- Multimodality (Communication & Life Science)
- Multimodality (Computing Science & AI)
- Conclusions

Multimodal Social Signals (I)

"Multimodality [...] not a recent discovery. The ancient rhetors and theorists Cicero and Quintilian stressed the importance of voice, gesture and face in the delivery of discourse very early on."

Poggi, "Mind, Hands, Face and Body",
Weidler Buchverlag, 2007.

Multimodal Social Signals (II)

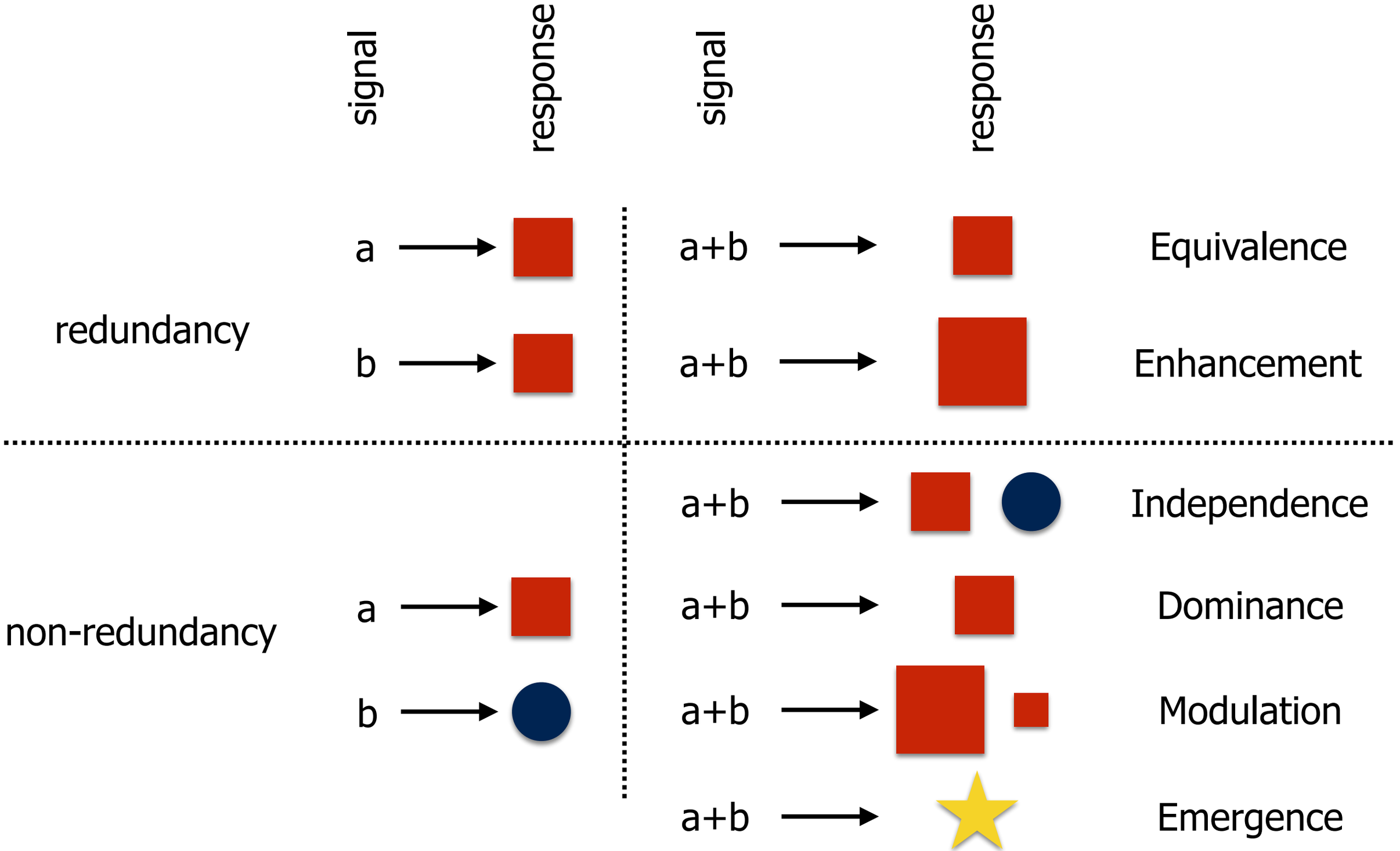
"[monkeys] utter a reiterated sound […] accompanied by vibratory movements of their jaws or lips, with the corners of the mouth drawn backwards and upwards, by the wrinkling of the cheeks, and even by the brightening of the eyes."

Darwin, "The Expression of Emotions in Animals and Men",
John Murray, 1872.

Multimodal Social Signals (III)

"Multimodal [...] communication is defined as communication via composite signals received through more than one sensory channel. We use the word "signal" [...] to refer to the entire set of communicative features [...] of an animal's behavior that occur simultaneously."

Partan and Marler, "Issues in the Classification of Multimodal Communication Signals", The American Naturalist, 166(2), pp. 231-245, 2005.



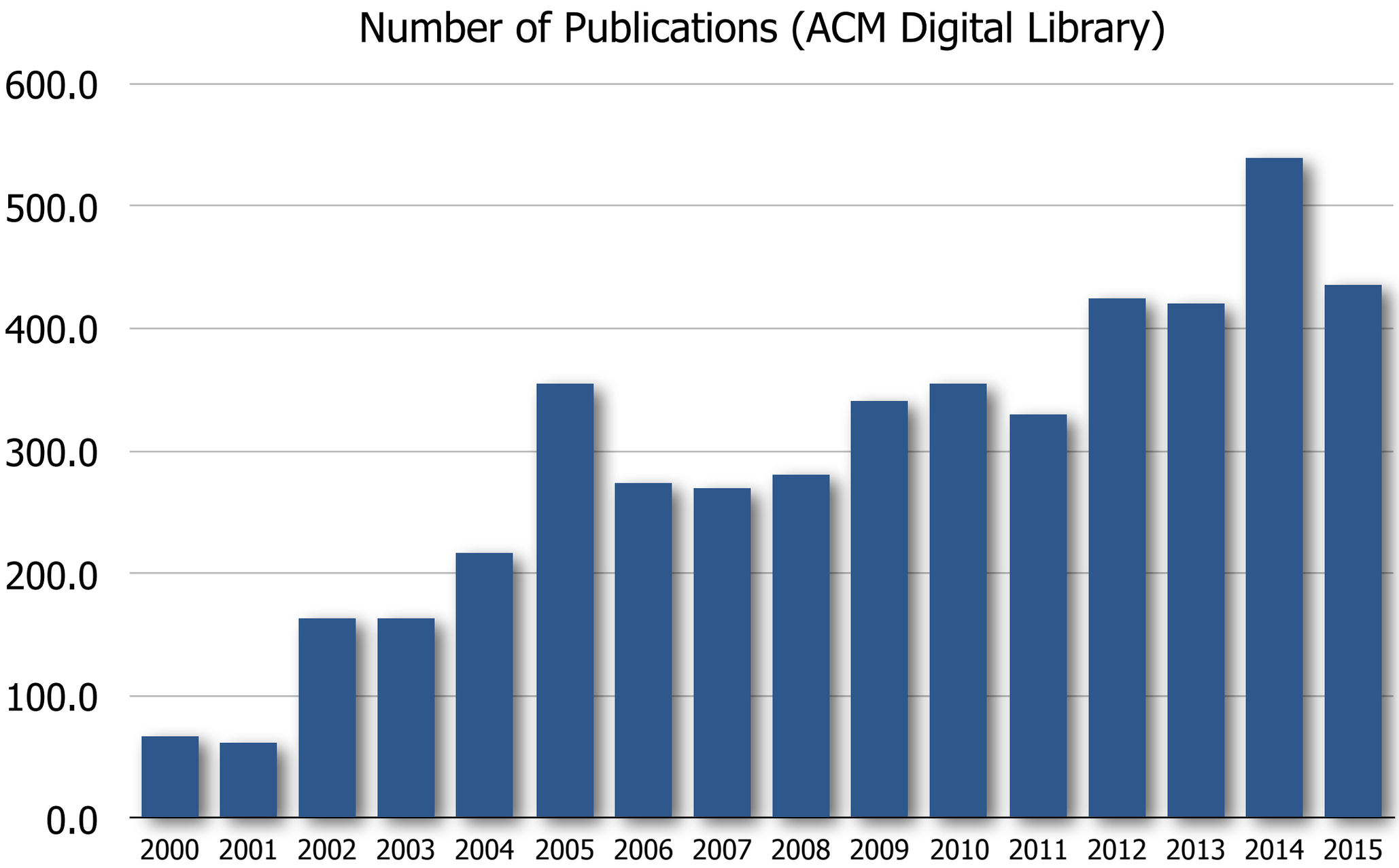
Partan and Marler, "Issues in the Classification of Multimodal Communication Signals", The American Naturalist, 166(2):231-245, 2005.

Recap

- Human and animal behaviour are inherently multimodal, with multiple signals used simultaneously;
- The scientific analysis of multimodality starts in the 19th century with Darwin and Duchenne;
- Biology and communication studies shows that multimodality aims at ensuring that a given message is effectively conveyed.

Outline

- Multimodality (Psychology & Neuroscience)
- Multimodality (Communication & Life Science)
- Multimodality (Computing Science & AI)
- Conclusions



Vinciarelli & Esposito, “Multimodal Analysis of Social Signals”, in “The Handbook of Multimodal-Multisensor Interfaces”, Oviatt, Schuller, Cohen, Sonntag, Potamianos and Kruger (eds.), 203-226, ACM Press, 2018.

The Bayes Theorem

A-posteriori probability (or posterior) of class i

Likelihood of class i with respect to the feature vector

$$p(C_i|\vec{x}) = \frac{p(\vec{x}|C_i)p(C_i)}{p(\vec{x})}$$

The evidence

The a-priori probability of class i

Posterior Rule

The expression of the priors according to the Bayes Theorem

$$C^* = \arg\max_{C_k \in C} \frac{p(\vec{x}|C_k)p(C_k)}{p(\vec{x})} =$$

$$= \arg\max_{C_k \in C} p(\vec{x}|C_k)p(C_k)$$

The evidence is the same for all classes and it can be eliminated

Recap

- Maximising the posterior corresponds to minimising the error probability;
- In the case of a zero-one loss function, maximising the posterior corresponds to minimising the Bayes Risk;
- The question is how Bayesian Decision Theory changes in the case of multimodal approaches.

The feature vectors extracted from the data captured with R sensors are concatenated

1

$\ldots$

$$X = \{\vec{x}_1, \ldots, \vec{x}_R\}$$

R

The concatenation of the feature vectors can be treated like any other feature vector

There are no changes for the priors (they do not depend on the input vectors)

$$C^* = \arg\max_{C_k \in \mathcal{C}} p(X|C_k)p(C_k)$$

Early Fusion

Recap

- The early fusion is the concatenation of the feature vectors extracted from the data captured through multiple sensors;
- The concatenation can be treated like any other vector;
- In the early fusion case, there are no changes from a decision theoretic point of view.

There are no changes for the priors (they do not depend on the input vectors)

$$C^* = \arg\max_{C_k \in C} p(\vec{x}_1, \ldots, \vec{x}_R | C_k) p(C_k)$$

The likelihood must be changed to reflect the presence of multiple feature vectors

The assumption is that the input vectors are statistically independent given the class

Product over all sensors

$$p(\vec{x}_1, \ldots, \vec{x}_R | \mathcal{C}_k) = \prod_{j=1}^{R} p(\vec{x}_j | \mathcal{C}_k)$$

If one term is close to zero, the entire product is close to zero

Recap

- The late fusion is the combination of decisions made at the level of individual modalities;
- The individual modalities are assumed to be statistically independent given the class;
- In the late fusion case, there are changes from a decision theoretic point of view.

Outline

- Multimodality (Psychology & Neuroscience)
- Multimodality (Communication & Life Science)
- Multimodality (Computing Science & AI)
- Conclusions

Conclusions

- Multimodality is an inherent characteristic of human perception, from both a psychology and neural point of view;
- In AI, the analysis of multimodal data is performed through early or late fusion;
- In the late fusion case, there are significant changes in the application of Bayes Decision Theory (see next lecture).

Thank You!