Cognitive
Science

Cognition Meaning and Action

Lodz–Lund Studies in Cognitive Science

Cognition
Meaning
and Action

Lodz–Lund
Studies in Cognitive Science

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Cognition Meaning and Action

Lodz–Lund
Studies in Cognitive Science

eds. Piotr Łukowski, Aleksander Gemel, Bartosz Żukowski

Łódź–Kraków 2015

Piotr Łukowski, Aleksander Gemel, Bartosz Żukowski – University of Łódź
Faculty of Educational Sciences, Institute of Psychology, Department of Cognitive Science
91-433 Łódź, Smugowa 10/12

Published by Łódź University Press & Jagiellonian University Press

First edition, Łódź–Kraków 2015

ISBN 978-83-7969-759-5 – paperback Łódź University Press

ISBN 978-83-233-3920-5 – paperback Jagiellonian University Press

ISBN 978-83-7969-760-1 – electronic version Łódź University Press

ISBN 978-83-233-9201-9 – electronic version Jagiellonian University Press

Łódź University Press
8 Lindleya St., 90-131 Łódź
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
phone +48 (42) 665 58 63

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Distribution outside Poland

Jagiellonian University Press
9/2 Michałowskiego St., 31-126 Kraków

phone +48 (12) 631 01 97, +48 (12) 663 23 81, fax +48 (12) 663 23 83

cell phone: +48 506 006 674, e-mail: sprzedaz@wuj.pl
Bank: PEKAO SA, IBAN PL 80 1240 4722 1111 0000 4856 3325
www.wuj.pl

JAGIELLONIAN
UNIVERSITY PRESS

CONTENTS

THE CROSSROADS OF COGNITIVE SCIENCE

The monograph *Cognition, Meaning and Action. Lodz-Lund Studies in Cognitive Science* collects papers written by the members of two Cognitive Science Departments: of Lund and of Lodz. It presents a range of issues currently examined in both centers. Some texts are written in collaboration as the result of collective research.

The opening article "Cognitive science: From computers to ant hills as models of human thought" (Peter Gärdenfors) offers an introduction to the history of ideas in cognitive science as it has been developing throughout last decades. Much of the contemporary mind theories derive from Descartes' *res cogitans* and *res extensa* distinction, and to some extent they may be seen as a continuation of rationalist-empiricist debate. The dawn of computer science is kept in quite rationalist fashion. The fundamental concept of computer science is the theoretical construct of Turing's machine. Inspired by Turing's concept, John von Neumann proposes a general architecture for modern computer based on logic circuits. The transfer of these findings to a theory of how the mind works was only a matter of time. Soon after von Neumann's proposal, McCulloch and Pitts interpreted neurons as a logic circuits combining information from other neurons according to some logical operations. This leads directly to one conclusion: the entire brain is a huge computer – and so the foundational metaphor for cognitive science was born.

Cognitive science can be said to emerge in 1956, the year in which Noam Chomsky, in response to the behaviourist concept of language, presented his proposition of *transformational grammar*. His central argument is based on the claim that processing the grammar of natural language requires a sort of algorithm as used in Turing machine. Also in 1956 Newell and Simon demonstrated the first computer program constructing logical proofs from a given set of premises and, finally, the concept of Artificial Intelligence was used for the first time. The philosophical assumption of the AI approach to cognitive processes is that the representation of mental content and processing is essentially *symbol manipulation*: only logical relations connect different symbolic expressions in a mental state of a person. The meaning of symbols is not part of the process of thinking, since they are manipulated exclusively on the basis of their form.

This quite rationalist manner of representing the cognitive process gave rise to several forms of criticism. One of them – derived from empiricism – was a new model of cognition called connectionism. Connectionist systems, also called *artificial neuron networks*, consist of large number of simple but highly interconnected units ("neurons"). According to the connectionists' point of view, thinking is not manipulation of meaningless symbols run and controlled by a central processor computer-like program, but it rather occurs in parallel neuronal processes distributed all over the brain, which is seen as a *self-organizing system*.

However, as it is claimed in the first paper, there are aspects of cognitive phenomena for which neither symbolic representation nor connectionism seems to offer appropriate "modelling tools". Those aspects include: mechanisms of concept acquisition, concept learning, and the notion of *similarity*. They turned out to be problematic for the symbolic and associationist approaches. To deal with them, a third form of representing information was proposed based not on symbols or connections between neurons, but rather on *geometrical* or *topological* structures. These structures generate mental *spaces* that represent various domains, and allow for modelling similarity in a very natural way as, for example, with the function of distance in such a space.

The topics of all other papers oscillate around the eponymous subject from the point of view of communication and its efficiency. The philosophical perspective of thinking, typical for the research on cognition, meaning, and action, is here replaced by psychological as well as neurophysiological benchmarks. The concept of the meaning of natural language expressions presented in "Two procedures expanding a linguistic competence" (Piotr Łukowski) is the result of two approaches, of the logical and of the one known in the cognitive psychology as *exemplary theory of meaning*. It employs *model example*, *function of sufficient similarity*, *accidental* and *essential similarities* and *zone of proximal development*. From such a perspective, the meaning inevitably appears to be a social, dynamic, and temporal phenomenon. Furthermore, since cognitive psychology is firmly founded on neuroscientific research, the properties of the presented understanding of *notions* can be partially linked to their neurophysiological correlates, as outlined in the following chapter: “Neurobiological basis for emergence of notions” (Konrad Rudnicki).

Comparative studies of *feature lists*, (dynamic) *frames*, and *conceptual spaces* as models for the representation of scientific conceptual knowledge is the aim of “Similarity as distance: Three models for scientific conceptual knowledge” (Frank Zenker). It is shown that the concepts arising from and giving rise to the exact measurement – mainly scientific ones – are properly represented in conceptual spaces. Also in the paper “The Approximate Numbers System and the treatment of vagueness in conceptual spaces” (Aleksander Gemel, Paula Quinon) the advantages of this model are successfully confirmed for the representation of concepts whose character is far from being scientific, i.e. vague concept of number.

Interpersonal communication defines the context of analyses for the next two papers: “To tell and to show: the interplay of language and visualizations in communication” (Jana Holsanova, Roger Johansson, Kenneth Holmqvist) and “Communication, cognition, and technology” (Peter Gärdenfors, Jana Holsanova). The main topic of both texts concerns various kinds of visualization with particular focus on how they influence communicational effectiveness. *Structuralist semiotics* and naturalistic, *computational concepts of language* are traditionally considered as being in conflict. Yet, closer analysis reveals their complementarity. In the paper “Semiotics, signaling games and meaning” (Aleksander Gemel, Bartosz Żukowski) some reconciliation of these two paradigms is proposed, which results in a coherent model preserving the advantages of the both concepts. The hybrid model requires, however, a formal tool to organize the semantic structure of the cultural system. To this aim *content implication* is introduced.

Starting from the following paper, rational action is the leading problem for all texts. The first of them, “Out of the box thinking” (Dorota Rybarkiewicz) explains in terms of the theory of metaphor how to break natural, standard borders – our typical *canyons* of thought – in order to find a better solution of a given problem. Procedures of decision making are analyzed in two papers closing the volume: “The everyday of decision-making” (Annika Wallin) and “Short- and long-term social interactions from the game theoretical perspective: A cognitive approach” (Magdalena Grothe, Bartosz Żukowski). In the former, the study of human everyday practice becomes the source of truths (information) about what a real and rational decision process looks like and of ideas about how to improve this process. In the latter, the rationality of decision making is steeped in the game theory. The well-known results established for the models of prisoner's dilemma and those with an indefinite time framework are related to the social interactions which are consistent with the cooperative equilibrium over a longer time.

Peter Gärdenfors (*Department of Cognitive Science, Lund*)
Piotr Łukowski (*Department of Cognitive Science, Łódź*)
Łódź, March 2015

PETER GÄRDENFORS

COGNITIVE SCIENCE: FROM COMPUTERS TO ANT HILLS AS MODELS OF HUMAN THOUGHT

1. Before cognitive science

In this introductory chapter some of the main themes of the development of cognitive science will be presented. The roots of cognitive science go as far back as those of philosophy. One way of defining cognitive science is to say that it is just *naturalized philosophy*. Much of contemporary thinking about the mind derives from René Descartes' distinction between the body and the soul. They were constituted of two different substances and it was only humans that had a soul and were capable of thinking. According to him, other animals were mere automata.

Descartes was a *rationalist*: our minds could gain knowledge about the world by rational thinking. This epistemological position was challenged by the *empiricists,* notably John Locke and David Hume. They claimed that the only reliable source of knowledge is sensory experience. Such experiences result in *ideas,* and thinking consists of connecting ideas in various ways.

Immanuel Kant strove to synthesize the rationalist and the empiricist positions. Our minds always deal with our inner experiences and not with the external world. He introduced a distinction between the thing in itself (*das Ding an sich*) and the thing perceived by us (*das Ding an uns*). Kant then formulated a set of *categories of thought,* without which we cannot organize our phenomenal world. For example, we must interpret what happens in the world in terms of cause and effect.

The favourite method among philosophers of gaining insights into the nature of the mind was *introspection*. This method was also used by psychologists at the end of the 19th and the beginning of the 20th century. In particular, this was the methodology used by Wilhelm Wundt and other German psychologists. By looking inward and reporting inner experiences it was hoped that the structure of the conscious mind would be unveiled.

However, the inherent subjectivity of introspection led to severe methodological problems. These problems set the stage for a scientific revolution in psychology. In 1913, John Watson published an article with the title "Psychology as the behaviourist views it" which has been seen as a *behaviourist* manifesto. The central methodological tenet of behaviourism is that only objectively verifiable observations should be allowed as data. As a consequence, scientists should prudently eschew all topics related to mental processes, mental events, and states of mind. Observable behaviour consists of *stimuli* and *responses.* The brain was treated as a black box. According to Watson, the goal of psychology was to formulate lawful connections between such stimuli and responses.

Behaviourism had a dramatic effect on psychology, particularly in the United States. As a consequence, animal psychology became a fashionable topic. Laboratories were filled with rats running in mazes and pigeons pecking at coloured chips. An enormous amount of data concerning *conditioning* of behaviour was collected. There was also a behaviourist influence in linguistics: the connection between a word and the objects it referred to was seen as a special case of conditioning.

Analytical philosophy, as it was developed in the early 20th century, contained ideas that reinforced the behaviourist movement within psychology. In the 1920s, the so-called Vienna circle formulated a philosophical programme which had as its primary aim to eliminate as much as possible of metaphysical speculations. Scientific reasoning should be founded on an *observational* basis. The observational data were obtained from experiments. From these data knowledge could only be expanded by using logically valid inferences. Under the headings of *logical empiricism* or *logical positivism,* this methodological programme has had an enormous influence on most sciences.

The ideal of thinking for the logical empiricists was logic and mathematics, preferably in the form of *axiomatic systems*. In the hands of people like Giuseppe Peano, Gottlob Frege, and Bertrand Russell, arithmetic and logic had been turned into strictly formalized theories at the beginning of the 20th century. The axiomatic ideal was transferred to other sciences with less success. A background assumption was that all scientific knowledge could be formulated in some form of *language*.

2. The dawn of computers

As a part of the axiomatic endeavour, logicians and mathematicians investigated the limits of what can be computed on the basis of axioms. In particular, the focus was put on what is called *recursive functions*. The logician Alonzo Church is famous for his thesis from 1936 that everything that can be computed can be computed with the aid of recursive functions.

At the same time, Alan Turing proposed an abstract machine, later called the *Turing machine*. The machine has two main parts: an infinite tape divided into cells, the contents of which can be read and then overwritten; and a movable head that reads what is in a cell on the tape. The head acts according to a finite set of instructions, which, depending on what is read and the current state of the head, determines what to write on the cell (if anything) and then whether to move one step left or right on the tape. It is Turing's astonishing achievement that he proved that such a simple machine can calculate all recursive functions. If Church's thesis is correct, this means that a Turing machine is able to compute everything that can be computed.

The Turing machine is an abstract machine – there are no infinite tapes in the world. Nevertheless, the very fact that all mathematical computation and logical reasoning had now been shown to be mechanically processable inspired researchers to construct real machines that could perform such tasks. One important technological invention was the so-called logical circuits that were constructed by systems of electric tubes. The Turing machine inspired John von Neumann to propose a general architecture for a real computer based on logic circuits. The machine had a central processor which read information from external memory devices, transformed the input according to the instructions of the program of the machine, and then stored it again in the external memory or presented it on some output device as the result of the calculation. The basic structure was thus similar to that of the Turing machine.

In contrast to earlier mechanical calculators, the computer *stored* its own instructions in the memory coded as binary digits. These instructions could be modified by the programmer, but also by the program itself while it was operating. The first machines developed according to von Neumann's general architecture appeared in the early 1940s.

Suddenly there was a machine that seemed to be able to think. A natural question was then to what extent computers think like humans. In 1943, McCulloch and Pitts published an article that became very influential. They interpreted the firings of the neurons in the brain as sequences of zeros and ones, by analogy with the binary digits of the computers. The neuron was seen as a logic circuit that combined information from other neurons according to some logical operator and then transmitted the results of the calculation to other neurons.

The upshot was that the entire brain was seen as a huge computer. In this way, the metaphor that became the foundation for cognitive science was born. Since the von Neumann architecture for computers was at the time the only one available, it was assumed that the brain too had essentially the same general structure.

The development of the first computers occurred at the same time as the concept of *information* as an abstract quantity was developed. With the advent of various technical devices for the transmission of signals, such as telegraphs and telephones, questions of efficiency and reliability in signal transmission were addressed. A breakthrough came with the mathematical theory of information presented by Claude Shannon. He found a way of measuring the amount of information that was transferred through a channel, independently of which code was used for the transmission. In essence, Shannon's theory says that the more improbable a message is statistically, the greater is its informational content (Shannon, Weaver, 1948). This theory had immediate applications in the world of zeros and ones that constituted the processes within computers. It is from Shannon's theory that we have the notions of bits, bytes, and baud that are standard measures for present-day information technology products.

Turing saw the potentials of computers very early. In a classical paper from 1950, he foresaw a lot of the developments of computer programs that were to come later. In that paper, he also proposes the test that nowadays is called the *Turing test*. To test whether a computer program succeeds in a cognitive task, such as playing chess or conversing in ordinary language, let an external observer communicate with the program via a terminal. If the observer cannot distinguish the performance of the program from that of a human being, the program is said to have passed the Turing test.

3. 1956: Cognitive science is born

There are good reasons for saying that cognitive science was born in 1956. That year a number of events in various disciplines marked the beginning of a new era. A conference where the concept of *Artificial Intelligence* (AI) was used for the first time was held at Dartmouth College. At that conference, Alan Newell and Herbert Simon demonstrated the first computer program that could construct logical proofs from a given set of premises. This event has been interpreted as the first example of a machine that performed a cognitive task.

Then in linguistics, later the same year, Noam Chomsky presented his new view of *transformational grammar* that was to be published in his book *Syntactic Structures* in 1957. This book caused a revolution in linguistics and Chomsky's views on language are still dominant in large parts of the academic world. A central argument is that any natural language would require a Turing machine to process its grammar. Again we see a correspondence between a human cognitive capacity, this time judgements of grammaticality, and the power of Turing machines. No wonder that Turing machines were seen as what was needed to understand thinking.

Also in 1956, the psychologist George Miller published an article with the title "The magical number seven, plus or minus two: Some limits on our capacity for processing information" that has become a classic within cognitive science. Miller argued that there are clear limits to our cognitive capacities: we can actively process only about seven units of information. This article directly applies Shannon's information theory to human thinking. It also explicitly talks about cognitive processes, something which had been considered to be very bad manners in the wards of the behaviourists that were sterile of anything but stimuli and responses. However, with the advent of computers and information theory, Miller now had a *mechanism* that could be put in the black box of the brain: computers have a limited processing memory and so do humans.

Another key event in psychology in 1956 was the publication of the book *A Study of Thinking*, written by Jerome Bruner, Jacqueline Goodnow, and George Austin, who had studied how people group examples into categories. They reported a series of experiments where the subjects' task was to determine which of a set of cards with different geometrical forms belong to a particular category. The category was set by the experimenter, for example the category of cards with two circles on them. The subjects were presented one card at a time and asked whether the card belonged to the category. The subject was then told whether the answer was correct or not. Bruner and his colleagues found that when the concepts were formed as conjunctions of elementary concepts like "cards with red circles", the subjects learned the category quite efficiently; while if the category was generated by a disjunctive concept like "cards with circles *or* a red object" or negated concepts like "cards that do *not* have two circles," the subjects had severe problems in identifying the correct category. Note that Bruner, Goodnow, and Austin focused on *logical* combinations of primitive concepts, again following the underlying tradition that human thinking is based on logical rules.

4. The rise and fall of artificial intelligence

Newell and Simon's program was soon to be followed by a wealth of more sophisticated logical theorem-proving programs. There was great faith in these programs: in line with the methodology of the logical positivists, it was believed that once we have found the fundamental axioms for a particular domain of knowledge we could then use computers instead of human brains to calculate all the consequences of the axioms.

But thinking is not logic alone. Newell and Simon soon started on a more ambitious project called the General Problem Solver that, in principle, should be able to solve any well-formulated problem. The General Problem Solver worked by means-end analysis: a problem is described by specifying an initial state and a desired goal state and the program attempts to reduce the gap between the start and the goal states. However, work on the program was soon abandoned since the methods devised by Newell and Simon turned out not to be as general as they had originally envisaged.

The first robot programs, like for example STRIPS developed at Stanford Research Institute, also followed the symbolic tradition by representing all the knowledge of the robot by formulas in a language that was similar to predicate logic. The axioms and rules of the program described the results of various actions together with the preconditions for the actions. Typical tasks for the robots were to pick up blocks in different rooms and stack them in a chosen room. However, in order to plan for such a task, the program needed to know all the consequences of the actions taken by the robot. For instance, if the robot went through the door of a room, the robot must be able to conclude that the blocks that were in the room did not move or ceased to exist as a result of the robot entering the room. It turned out that giving a complete description of the robot's world and the consequences of its actions resulted in a combinatorial explosion of the number of axioms required. This has been called the *frame problem* in robotics.

The optimism of AI researchers and their high-flying promises concerning the capabilities of computer programs were met with several forms of criticism. Already in 1960, Yehoshua Bar-Hillel wrote a report on the fundamental problems of using computers to perform automatic translations from one language to another. And in 1967, Joseph Weizenbaum constructed a seductive program called ELIZA that could converse in natural language with its user. ELIZA was built to simulate a Rogerian psychotherapist. The program scans the sentences written by the user for words like “I”, “mother”, “love” and when such a word is found, the program has a limited number of preset responses (where the values of certain variables are given by the input of the user). The program does very little calculation and understands absolutely nothing of its input. Nevertheless, it is successful enough to delude an unsuspecting user for some time until its responses become too stereotyped.

Weizenbaum’s main purpose in writing ELIZA was to show how easy it was to fool a user that a program has an understanding of a dialogue. We are just too willing to ascribe intelligence to something that responds appropriately in a few cases – our human-centred thinking extends easily to computers. Weizenbaum was horrified because some professional psychiatrists suggested ELIZA as a potential therapeutic tool that might be used in practice by people with problems.

In spite of the critics, AI lived on in, more or less, its classical shape during the 1970s. Among the more dominant later research themes were the so-called *expert systems* that have been developed in various areas. Such systems consist of a large number of symbolic rules (that have normally been extracted from human experts) together with a computerized inference engine that applies the rules recursively to input data and ends up with some form of solution to a given problem.

The most well-known expert system is perhaps MYCIN, which offers advice on infectious diseases (it even suggests a prescription of appropriate antibiotics). MYCIN was exposed to the Turing test in the sense that human doctors were asked to suggest diagnoses on the basis of the same input data, from laboratory tests, that was given to the program. Independent evaluators then decided whether the doctors or MYCIN had done the best job. Under these conditions, MYCIN passed the Turing test, but it can be objected that if the doctors had been given the opportunity to see and examine the patients, they would (hopefully) have outperformed the expert system.

However, expert systems never reached the adroitness of human experts and they were almost never given the opportunity to have the decisive word in real cases. A fundamental problem is that such systems may incorporate an extensive amount of knowledge, but they hardly have any knowledge about the *validity* of their knowledge. Without such meta-knowledge, a system cannot make valid *judgements* that form the basis of sound decisions.

5. Mind: the gap

A unique aspect of our cognitive processes is that we experience at least part of them as being *conscious*. The problem of what consciousness is has occupied philosophers for centuries, and there is a plethora of theories of the mind.

Cartesian dualism, which treats the body and the mind as separate substances, has lost much of its influence during the 20th century. Most current theories of the mind are *materialistic* in the sense that only physical substances are supposed to exist. But this position epitomizes the question of how conscious experiences can be a result of material processes. There seems to be an unbridgeable gap between our physicalistic theories and our phenomenal experiences.

A theory of the mind that has been popular since the 1950s is the *identity theory*, which claims that conscious processes are identical with material processes in the brain. As a consequence, the phenomenal is in principle *reducible* to the physical. It should be noted that according the identity theory it is only processes in the brain that can become parts of conscious experiences.

However, the new vogue of cognitive theories based on the analogy between the brain and the computer soon attracted the philosophers. In 1960, Hilary Putnam published an article with the title “Minds and machines” where he argued that it is not the matter of a brain or a computer that determines whether it has a mind or not, but only what *function* that brain or computer performs. And since the function of a computer was described by its program, the function of the brain was, by analogy, also identified with a program. This stance within the philosophy of mind has become known as *functionalism*.

The central philosophical tenet of the AI approach to represent cognitive processes is that mental representation and processing is essentially *symbol manipulation*. The symbols can be concatenated to form expressions in a *language of thought* – sometimes called Mentalese. The different symbolic expressions in a mental state of a person are connected only via their logical relations. The symbols are manipulated exclusively on the basis of their form – their *meaning* is not part of the process.

The material basis for these processes is irrelevant to the description of their results – the same mental state can be realized in a brain as well as in a computer. Thus, the paradigm of AI clearly presupposes the functionalist philosophy of mind. In brief, the mind is thought to be a computing device, which generates symbolic expressions as inputs from sensory channels, performs logical operations on these sentences, and then transforms them into linguistic or non-linguistic output behaviours.

However, functionalism leaves unanswered the question of what makes certain cognitive processes conscious or what gives them *content*. As an argument against the strongest form of AI that claims that all human cognition can be replaced by computer programs, John Searle (1980) presents his “Chinese room” scenario. This example assumes that a person who understands English but no Chinese is locked into a room together with a large set of instructions written in English. The person is then given a page of Chinese text that contains a number of questions. By meticulously following the instructions with respect to the symbols that occur in the Chinese questions, he is able to compose a new page in Chinese that comprises answers to the questions.

According to functionalism (and in compliance with the Turing test) the person in the room who is following the instructions would have the same capacity as a Chinese-speaking person. Hence functionalism would hold that the person *together with* the equipment in the room understands Chinese. But this is potently absurd, claims Searle. For analogous reasons, according to Searle, a computer lacks *intentionality* and can therefore not understand the meaning of sentences in a language. Searle’s argument has spawned a heated debate about the limits of functionalism and what it would mean to *understand* something.

6. First heresy against high-church computationalism: thinking is not only by symbols

6.1. Artificial neuron networks

For many years, the symbolic approach to cognition was totally dominant. But as a result of the various forms of criticism which led to a greater awareness of the limitations of the “symbol crunching” of standard AI programs, the ground was prepared for other views of the fundamental mechanisms of thinking. We find the first signs of heresy against what has been called “high-church computationalism”.

For empiricist philosophers like Locke and Hume, thinking consists basically in the forming of *associations* between “perceptions of the mind.” The basic idea is that events that are similar become connected in the mind. Activation of one idea activates others to which it is linked: when thinking, reasoning, or daydreaming, one thought reminds us of others.

During the last few decades, associationism has been revived with the aid of a new model of cognition – *connectionism*. Connectionist systems, also called *artificial neuron networks*, consist of large numbers of simple but highly interconnected units ("neurons"). The units process information in *parallel* in contrast to most symbolic models where the processing is serial. There is no central control unit for the network, but all neurons act as individual processors. Hence connectionist systems are examples of *parallel distributed processes* (Rumelhart, McClelland, 1986).

Each unit in an artificial neuron network receives activity, both excitatory and inhibitory, as input; and transmits activity to other units according to some function of the inputs. The behaviour of the network as a whole is determined by the initial state of activation and the connections between the units. The inputs to the network also gradually change the strengths of the connections between units according to some *learning rule*. The units have no memory in themselves, but earlier inputs are represented indirectly via the changes in strengths they have caused. According to connectionism, cognitive processes should not be represented by symbol manipulation, but by the *dynamics* of the patterns of activities in the networks. Since artificial neuron networks exploit a massive number of neurons working in parallel, the basic functioning of the network need not be interrupted if some of the neurons are malfunctioning. Hence, connectionist models do not suffer from the computational brittleness of the symbolic models and they are also much less sensitive to noise in the input.

Some connectionist systems aim at modelling neuronal processes in human or animal brains. However, most systems are constructed as general models of cognition without any ambition to map directly to what is going on in the brain. Such connectionist systems have become popular among psychologists and cognitive scientists since they seem to be excellent *simulation tools* for testing associationist theories.

Artificial neuron networks have been developed for many different kinds of cognitive tasks, including vision, language processing, concept formation, inference, and motor control. Among the applications, one finds several that traditionally were thought to be typical symbol processing tasks like pattern matching and syntactic parsing. Perhaps the most important applications, however, are models of various forms of *learning*.

Connectionist systems brought a radically new perspective on cognitive processes: cognition is *distributed* in the system. In contrast, a von Neumann computer is controlled by a central processor. In favour of this architecture it has been argued that if the brain is a computer, it must have a central processor – where would you otherwise find the "I" of the brain? But the analogy does not hold water – there is no area of the brain that serves as a pilot for the other parts: there is no one in charge. The neuronal processes are distributed all over the brain, they occur in parallel and they are to a certain extent independent of each other. Nevertheless, the brain functions in a goal-directed manner. From the connectionist perspective, the brain is best seen as a *self-organizing system*. Rather than working with a computer-like program, the organization and learning that occur in the brain should be seen as an *evolutionary* process (Edelman, 1987).

On this view, the brain can be seen as an *ant hill*. The individual neurons are the ants who perform their routine jobs untiringly, but rather unintelligently, and who receive signals from other neurons via their dendrite antennas. From the interactions of a large number of simple neurons a complex well-adapted system like an ant hill *emerges* in the brain. In other words, cognition is seen as a holistic phenomenon in a complex system of distributed parallel processes.

Along with the development of symbolic and connectionist programming techniques, there has been a rapid development in the *neurosciences*. More and more has been uncovered concerning the neural substrates of different kinds of cognitive processes. As the argument by McCulloch and Pitts shows, it was thought at an early stage that the brain would function along the same principles as a standard computer. But one of major sources of influence for connectionism was the more and more conspicuous conclusion that neurons in the brain are not logic circuits, but operate in a distributed and massively parallel fashion and according to totally different principles than those of computers. For example, Hubel and Wiesel's (1968) work on the signal-detecting functioning of the neurons in the visual cortex were among the path-breakers for the new view of the mechanisms of the brain. It is seen as one of the strongest assets of connectionism that the mechanisms of artificial neuron networks are much closer to the functioning of the brain.

Another talented researcher who combined thorough knowledge about the brain with a computational perspective was David Marr. His book *Vision* from 1982 is a milestone in the development of cognitive neuroscience. He worked out connectionist algorithms for various stages of the visual processing from the moment the cells on the retina react, until a holistic 3D model of the visual scene is constructed in the brain. Even though some of his algorithms have been questioned by later developments, his methodology has led to a much deeper understanding of the visual processes during the last two decades.

6.2. Non-symbolic theories of concept formation

Both the symbolic and the connectionist approaches to cognition have their advantages and disadvantages. They are often presented as competing paradigms, but since they attack cognitive problems on different levels, they should rather be seen as complementary methodologies.

However, there are aspects of cognitive phenomena for which neither symbolic representation nor connectionism seems to offer appropriate modelling tools. In particular it appears that mechanisms of *concept acquisition*, which is paramount for the understanding of many cognitive phenomena, cannot be given a satisfactory treatment in any of these representational forms. Concept learning is closely tied to the notion of *similarity*, which has also turned out to be problematic for the symbolic and associationist approaches.

To handle concept formation, among other things, a third form of representing information that is based on using *geometrical* or *topological* structures, rather than symbols or connections between neurons, has been advocated. This way of representing information is called the *conceptual* form. The geometrical and topological structures generate mental *spaces* that represent various domains. By exploiting distances in such spaces, judgements of similarity can be modelled in a natural way.

In the classical Aristotelian theory of concepts that was embraced by AI and early cognitive science (for example, in the work of Bruner, Goodnow, and Austin presented above) a concept is defined via a set of *necessary and sufficient properties*. According to this criterion, all instances of a classical concept have equal status. The conditions characterizing a concept were formulated in linguistic form, preferably in some symbolic form.

However, psychologists like Eleanor Rosch showed that in the majority of cases, concepts show *graded membership*. These results led to dissatisfaction with the classical theory. As an alternative, *prototype theory* was proposed in the mid-1970s. The main idea of this theory is that within a category of objects, such as those instantiating a concept, certain members are judged to be more *representative* of the category than others. For example robins are judged to be more representative of the category “bird” than are ravens, penguins, and emus; and desk chairs are more typical instances of the category “chair” than rocking chairs, deck chairs, and beanbag chairs. The most representative members of a category are called *prototypical* members. The prototype theory of concepts fits much better with the conceptual form of representing information than with symbolic representations.

6.3. Thinking in images

When we think or speak about our own thoughts, we often refer to inner scenes or pictures that we form in our fantasies or in our dreams. However, from the standpoint of behaviourism, these phenomena were unspeakables, beyond the realm of the sober scientific study of stimuli and responses. This scornful attitude towards mental images was continued in the early years of AI. Thinking was seen as symbol crunching and images were not the right kind of building blocks for computer programs.

However, in the early 1970s psychologist began studying various phenomena connected with *mental imagery*. Roger Shepard (Shepard, Metzler, 1971) and his colleagues performed an experiment that has become classical. They showed subjects pictures representing pairs of 3D block figures that were rotated in relation to each other and asked the subjects to respond as quickly as possible whether the two figures were the same or whether they were mirror images of one another. The surprising finding was that the time it took the subject to answer was linearly correlated with the number of degrees the second object had been rotated in relation to the first. A plausible interpretation of these results is that the subjects generate mental images of the block figures and *rotate* them in their minds.

Stephen Kosslyn (1980) and his colleagues have documented similar results concerning people’s abilities to imagine maps. In a typical experiment, subjects are shown maps of a fictional island with some marked locations: a tree, a house, a bay, etc. The maps are removed and the subjects are then asked to focus mentally on one location on the map and then move their attention to a second location. The finding was that the time it takes to mentally scan from one location to the other is again a linear function of the distance between the two positions on the map. The interpretation is that the subjects are *scanning a mental map*, in the same manner as they would scan a physically presented map.

Another strand of mental imagery has been developed within *cognitive semantics*. In the Chomskian theory of linguistics, syntax is what counts and semantic and pragmatic phenomena are treated like Cinderellas. In contrast, within cognitive semantics, as developed by Ron Langacker (1987) and George Lakoff (1987) among others, the cognitive representation of the meaning of linguistic expressions is put into focus. Their key notion for representing linguistic meanings is that of an *image schema*. A common assumption is that such schemas constitute the representational form that is common to perception, memory, and semantic meaning. The theory of image schemas also builds on the prototype theory for concepts. Again, this semantic theory replaces the uninterpreted symbols of high-church computationalism with image-like representations that have an inherent meaning. In particular, our frequent use of more or less conventional *metaphors* in everyday language can be analysed in an illuminating way using image schemas.

7. Second heresy: cognition is not only in the brain

7.1. The embodied brain

The brain is not made for calculating – its primary duty it to control the body. For this reason it does not function in solitude, but is largely dependent on the body it is employed by. In contrast, when the brain was seen as a computer, it was more or less compulsory to view it as an isolated entity. However, there is little hope that such a scenario would ever work. As a consequence, there has recently been a marked increase in studies of the *embodied* brain.

For example, the eye is not merely seen as an input device to the brain and the hand as enacting the will of the brain, but the eye-hand-brain is seen as a *coordinated* system. For many tasks, it turns out that we think faster with our hands than with our brains. A simple example is the computer game Tetris where you are supposed to quickly turn, with the aid of the keys on the keyboard, geometric objects that come falling over a computer screen in order to fit them with the pattern at the bottom of the screen. When a new object appears, one can mentally rotate it to determine how it should be turned before actually touching the keyboard. However, expert players turn the object faster with the aid of the keyboard than they turn an image of the object in their brains. This is an example of what has been called *interactive thinking*. The upshot is that a human who is manipulating representations in the head is not the same cognitive system as a human interacting directly with the represented objects.

Also within linguistics, the role of the body has attracted attention. One central tenet within cognitive semantics is that the meanings of many basic words are embodied, in the sense that they relate directly to bodily experiences. George Lakoff and Mark Johnson show in their book *Metaphors We Live By* (1980) that a surprising variety of words, for instance prepositions, derive their complex meaning from a basic embodied meaning that is then extended by metaphorical mappings to a number of other domains.

7.2. Situated cognition

There is one movement within cognitive science, known as *situated cognition*, which departs even further from the traditional stance. The central idea is that in order to function efficiently the brain does not only need the body but also the surrounding world. In other words, it is *being there* that is our primary function as cognitive agents (Clark, 1997). Cognition is not imprisoned in the brain but emerges in the interaction between the brain, the body and the world. Instead of representing the world in an inner model the agent in most cases uses the world as its own model. For example, in vision, an agent uses rapid movements of the eyes to extract what is needed from a visual scene, rather than building a detailed 3D model of the world in its head.

In many cases it is impossible to draw a line between our senses and the world. The captain of a submarine "sees" with the periscope and a blind person "touches" with her stick, not with the hand. In the same way we "think" with road signs, calendars, and pocket calculators. There is no sharp line between what goes on inside the head and what happens in the world. The mind leaks out into the world.

By arranging the world in a smart way we can afford to be stupid. We have constructed various kinds of artefacts that help us solve cognitive tasks. In this way the world functions as *scaffolding* for the mind (Clark, 1997). For example, we have developed a number of memory aids: we "remember" with the aid of books, DVDs, hard-disks, etc. In this way, memory is placed in the world. For another practical example, the work of an architect or a designer is heavily dependent on making different kinds of sketches: the sketching is an indispensable component of the cognitive process (Gedenryd, 1998).

The emphasis on situated cognition is coupled with a new view of the basic nature of the cognitive structures of humans. Instead of identifying the brain with a computer, the *evolutionary origin* of our thinking is put into focus. The key idea is that we have our cognitive capacities because they have been useful for survival and reproduction in the past. From this perspective, it becomes natural to compare our form of cognition with that of different kinds of animals. During the last decade, *comparative cognition* has grown considerably as a research area. The methodology of this branch is different from that of traditional cognitive psychology. Instead of studying subjects in laboratories under highly constrained conditions, evolutionary psychology focuses on data that are *ecologically valid* in the sense that they tell us something about how humans and animals act in natural problem-solving situations.

7.3. The pragmatic turn of linguistics

The role of culture and society in cognition was marginalized in early cognitive science. These were problem areas that were to be addressed when an understanding of individual cognition had been achieved. However, when the focus of cognitive theories shifted away from symbolic representations, semantic and pragmatic research reappeared on the agenda. Pragmatics consists of the rules for linguistic actions; semantics is conventionalized pragmatics; and finally, syntax adds grammatical markers to help disambiguate when the context does not suffice to do so. This tradition connects with several other research areas such as anthropology, psychology, and situated cognition,

This shift of the linguistic programme can also be seen in the type of data that researchers are considering. In the Chomskian research programme, *single sentences* presented out of context are typically judged for their grammaticality. The judgements are often of an introspective nature when the researcher is a native speaker of the language studied. In contrast, within the pragmatic programme, recordings of actual *conversations* are recorded or video-taped. For the purpose of analysis, they are normally transcribed by various methods. The conversational analyses treat language as part of a more general interactive cognitive setting.

7.4. Robotics

The problem of constructing a robot is a good test of progress in cognitive science. A robot needs perception, memory, knowledge, learning, planning, and communicative abilities, that is, exactly those capacities that cognitive science aims at understanding. Current industrial robots have very little of these abilities – they can perform a narrow range of tasks in a specially prepared environment.

In contrast, nature has, with the stamina of evolution, solved cognitive problems by various methods. Most animals are *independent* individuals, often extremely flexible. The simplest animals are classified as *reactive systems*. This means that they have no foresight, but react to stimuli as they turn up in the environment. So, given nature's solutions, why can we not construct machines with the capacity of a cockroach?

The current trend in robotics is to start from reactive systems and then add higher cognitive modules that amplify or modify the basic reactive systems. This methodology is based on what Rodney Brooks calls the *subsumption architecture*. One common feature of such robots is that that they *learn by doing*: linguistic or other symbolic inputs play a minor role in their acquisition of new knowledge. One factor that was forgotten in classical AI is that animals have a *motivation* for their behaviour. From the perspective of evolution, the utmost goals are survival and reproduction. In robotics, the motivation is set by the constructor.

8. The future of cognitive science

The goal of contemporary cognitive science is not primarily to build a thinking machine, but to increase our understanding of cognitive processes. This can be done by various methods, including traditional psychological experiments, observations of authentic cognitive processes in practical action, or by simulating cognition in robots or programs. Unlike the early days of AI when it was believed that one single methodology, that of symbolic representation, could solve all cognitive problems, the current trend is to work with several forms of representations and data.

Furthermore, the studies tend to be closely connected to findings in neuroscience and in other biological sciences. New techniques of *brain imaging* will continue to increase our understanding of the multifarious processes going on in the brain. Other techniques, such as *eye-tracking*, will yield rich data for analysing our cognitive interaction with the world and with the artefacts in it.

As regards the practical applications of cognitive science, a main area is the construction of *interfaces* to information technological products. The aim is that IT products should be as adapted to the demands of human cognition as possible. In other words, it should be the goal of information technology to build scaffolding tools that enhance human capacities. To give some examples of already existing aids, pocket calculators help us perform rapid and accurate calculations that were previously done laboriously with pen and paper or even just in the head. And word processors relieve us from the strain of retyping a manuscript.

Donald Norman (1988) started a tradition in *cognitive design* with his classical book *The Design of Everyday Things*. He showed by a wealth of provocative examples that technical constructors very often neglect the demands and limitations of human cognition. The user-friendliness of computer programs, mobile phones, remote TV controls, etc., has increased, but there is still an immense potential to apply the findings of cognitive science in order to create products that better support our ways of thinking and remembering.

Another area where cognitive science ought to a have a great impact in the future is *education*. There is a strong trend to equip schools at all levels with more and more computers. Unfortunately, most efforts are spent on the technical and financial aspects and very little on the question of *how* the computers should be used in schools. A number of so-called educational computer programs have been developed. With few exceptions, however, these programs are of a drill-and-exercise character. In particular, various kinds of *simulation* programs may be supportive for the learning process. For example, when teaching physics, a program that simulates the movements of falling bodies and displays the effects on the screen, allowing the student to interactively change the gravitational forces and other variables, will give a better grasp of the meaning of the physical equations than many hours of calculation by hand. Another promising area is the use of *virtual agents* in virtual environments.

For the development of truly educational computer programs, collaboration with cognitive scientists will be mandatory. Those who design the programs must have a profound knowledge of how human learning and memory works, of how we situate our cognition in the world and of how we communicate. Helping educationalists answer these questions will be one of the greatest challenges for cognitive science in the future.

As a last example of the future trends of cognitive science, I believe that research on the processing of sensory information will be useful in the development of *tools for the handicapped*. The deaf and blind each lack a sensory channel. Through studies of multimodal communication, these sensory deficits can hopefully be aided. If we achieve better programs for speech recognition for example, deafness can be partly compensated for.

In conclusion, we can expect that in the future, cognitive science will supply man with new tools, electronic or not, that will be better suited to our cognitive needs and that may increase the quality of our lives. In many areas, it is not technology that sets the limits, but rather our lack of understanding of how human cognition works.

References

Bruner, J.S., Goodnow, J.J. and Austin, G.A. (1956). *A Study of Thinking*. New York: Wiley.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Clark, A. (1997). *Being there*. Cambridge, MA: MIT Press.

Edelman, G.M. (1987). *Neural Darwinism.* New York, NY: Basic Books.

Gedenryd, H. (1998). *How Designers Work*. Lund: Lund University Cognitive Studies 75.

Hubel, D.H., Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology* 195: 215–243.

Kosslyn, S. (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago, IL: University of Chicago Press.

Lakoff, G., Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Langacker, R.W. (1987). *Foundations of Cognitive Grammar*, vol. 1. Stanford, CA: Stanford University Press.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysic*s 5: 115–133.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81–97.

Norman, D.A. (1988). *The Design of Everyday Things*. New York, NY: Basic Books.

Putnam, H. (1960). Minds and machines. In: S. Hook (ed.), *Dimensions of Mind*. New York, NY: New York University Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104: 192–233.

Rumelhart, D.E., McClelland, J.L. (1986). *Parallel Distributed Processing*, vol. 1–2. Cambridge, MA: MIT Press.

Searle, J.R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences* 3: 417–457.

Shannon, C.E., Weaver, W. (1948). *The Mathematical Theory of Communication*. Chicago, IL: The University of Illinois Press.

Shepard, R.N., Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171 (3972): 701–703.

Turing, A. (1950). Computing machinery and intelligence. *Mind* 59: 433–460.

Watson, J.B. (1913). Psychology as the behaviourist views it. *Psychological Review* 20: 158–17*7.*

Weizenbaum, J. (1967). *Computer Power and Human Reason*. New York, NY: Freeman and Company.

Piotr Łukowski

TWO PROCEDURES EXPANDING A LINGUISTIC COMPETENCE

1. Introduction

On the ground of psychology there are many various theories explaining what a notion is. All these approaches can be grouped in three classes:

(A) *classical* theories which present a notion as a compilation of features in common for many particular objects.

(B) *probabilistic* theories in which notions are derived from the higher structure – an individual theory of the world. Notions are elements of the structure.

(C) theories in which a notion is strictly connected with its references and with other notions.

Theories of the first class are a little bit naïve today, and theories of the second class fail because of the error of circularity – in order to explain what a notion is, we need to explain what a structure is, but in order to explain what a structure is, we need to know what a notion is (Maruszewski, 2001: 297–299). Theses of theories of the third class are obviously compatible with simple everyday observations and with the logical knowledge about the properties of natural language expressions.[1] Thus, the approach presented in the paper coincides with the main ideas of the theories of this last class.

---

[1] In his book *Concepts, Kinds and Cognitive Development* Keil considers the process of concepts creation in child's mind. Both the dialogues to be presented below, well known from our life, as well as the conclusions of my paper coincide with basic thesis by Keil. For example, in his theory Keil employs known in psychology notion "zone of proximal development" (created by famous psychologist Lev Vygotsky) as a name for all those people which help us to understand and to correctly use new (for us) words. Also, the notion of similarity, used by us in the paper, is more precisely analyzed by another famous psychologists Inhelder and Piaget. They consider a development of a child's mind

Dialogue 1

During a stroll with his father a boy asks: – *What is it, daddy?*

– *A building, honey* – the father answers.

– *Oh…*

– *And this is another building, isn't it?* – after a while the child adds pointing at next object.

– *No, it isn't. This is a shed.*

– *Why this is not a building?* – the boy asks.

– *Because it is much smaller than a building, it is constructed from ordinary boards, it is not for living, just for garden's tools…*

– *Uh-huh…*

Dialogue 2

During a stroll with his father a child asks: – *What color is this leaf?*

– *It is brown* – the father answers.

– *It's impossible! Your jacket is brown and the leaf's color is different.*

– *My jacket is brown and the leaf is also brown. Both colors are a little bit different but both are brown* – the father states.

– *And what about that car's color?* – the boy asks.

– *I'm afraid the car is not brown* – the father answers – *that car's color has a delicate red shade.*

Every story is well known and typical for strolling with a small child. Both present a similar process of expanding the linguistic competence of a child. In both cases the child is taught by his father how to use words correctly. In the first story the child recognizes a correct use of the word "building" and the correct use of the word "shed". In the second story the child is taught by his father how to correctly use the word "brown". In spite of similarities both processes fundamentally differ. In the first scene, two objects named by the child (with the father's help) differ in some clear, easy to notice way, and that is why the difference can be defined. Obviously, the definition can be partial only. As far as the terms are not mathematical defining cannot be complete – it can be complete only for the terms of a mathematical, strict sense. In the second scene the differences between colors of objects are subtle and difficult to describe. In this case, a correct use of the name depends on the accurate recognition of resemblance to the model object.

---

as a process in which concept creation firstly bases on "accidental similarities" of objects, but later on "essential similarities", see (Keil, 1989: 5–7, 9), also (Piaget, Inhelder, 1974: 37).

Both simple and standard stories illustrate a process of learning of the correct use of new words. Thanks to them, in a schematic way, the most basic truths are shown about the process. The careful examination of both situations should deliver some important hints about the modeling of this process for automata. Thus, first of all let us notice that the entire process depends on the permanent teaching of the child by the parent. In other words, it is necessary to continuously verify the process. The process is controlled from outside and depends on the regular provision of new information about the correct use of words. Sometimes, a child is informed about a name of an object new for them, in other times an example is provided, *i.e.* an object which can be used by the child as a model object, an object which is a case of the correct use of the given word. It means that there is another important element of the process: an ostensive definition. Since the process in question consists in the permanent confrontation of words with non-linguistics objects, an ostensive defining must be a constant element of the process. During a conversation, a new word appears together with a gesture of pointing to an object which is an example either of the correct or of the wrong use of the word. Usually, in the ostensive defining the same person utters the word and shows an object. In our stories this kind of defining splits into two persons: the first one utters the word and the second points to an object. Sometimes the first person is a child, and the second a father, sometimes it is the other way round. Undoubtedly, in all those cases there is still ostensive defining.

Another constant element of the procedure that extends the linguistic competence is a partial defining of the meaning of new words. This fact is evident in the case of the first dialogue. However, the second situation is not free from the partial defining – here, the partial defining is of the ostensive type. In the first dialogue, the meaning (the content) of the word "shed" is distinguished from the meaning (the content) of "building" by using only few arbitrary features. Indeed, after the initial exchange at the beginning of the dialogue, a child believes that he knows how to correctly use the word "building". Then, he uses the word in a wrong way, calling "a building" an object that is not a building. The father corrects the boy calling a new object "shed" and giving some spontaneously selected features typical for sheds but not for buildings. Since this set of features is arbitrary it could be replaced by some other set. Of course, none of such sets can be used for a complete definition of the shed, because such a complete content (*i.e.* set of features) does not exist for words without mathematical sense. Obviously, such a step is repeated many times during a lifetime of each of us. We use some word already known to us in all cases similar to the model (for us) case, as long as somebody would correct us. Similarly, the child from the first dialogue possesses in his memory a model of a building being a picture of some real object, which was correctly named as "building". Then, the child tries to compare every newly encountered object to the picture of the building from his memory. If a similarity of both objects is in some sense sufficient, the child will call the new object, "a building". If not, the child will try to find an appropriate name for the new object. Probably, this will be done with the help of somebody: the father, the mother, some colleague, etc. In the first dialogue, the father gives his child a model example for the correct use of the word "shed". From now on, the boy possesses at least two model pictures: one representing "typical" building and another representing also "typical" shed. Moreover, a difference between a "typical" building and a "typical" shed is defined selectively, *i.e.* partially. During one's life, model examples evaluate (are changed many times). They are replaced by better, "more typical" objects. This permanent and endless improvement of the model cases is the result of the permanent and endless expansion of word's meaning understanding. Moreover, usually one model case for a given word is not enough and so the correct use of this word is recognized by more than one exemplary object. Thus, it should be supposed that after some time the child from the above stories (as well as everybody) possesses more than one example for the "typical" building and also more examples for the "typical" shed – after all, there are plenty of various buildings and many diverse sheds.

A partial defining appears also in the second dialogue. The child tries to know which color can be called "brown". As a result, the child possesses in his memory pictures of few brown objects: the father's jacket, the leave. He also knows that the car is not brown. It is an evident case of partial defining of the color. It is impossible for the child to know all shades of brown color, so all those cases are "partially" represented in the child's memory by those three examples: two positive and one negative.

In order to use the model examples for word's understanding and to use them correctly the child needs to competently recognize which features should be compared and which should be ignored – some features are essential, others are inessential from the point of view of the word's meaning. In the case of the first dialogue, the child's successful use of the words "building" and "shed" depends on their ability to recognize which features are essential, and which are not, to distinguish buildings from sheds. Although the boy goes beyond the ostensive defining at this stage of the process, he employs all the results that have been engraved into his memory by the former ostensive defining.

Let us emphasize that the procedure that expands a linguistic competence by mathematical notions is quite different. Since those notions are precise and sharp, the parent or teacher has to present to the child their complete definitions. An incomplete definition would define some other notion, for example, a partial definition of the square "it is a two-dimensional figure with all four angles right and with opposite sides parallel" would define the rectangle.

2. Generality – one face of the tolerance of expression

A definition of a tolerant expression uses a class of functions of similarity thanks to which it is possible to recognize if a given new still unnamed object can be named by this expression *i.e.* if it is sufficiently similar to some of the images of objects being model examples of the correct use of this expression. Every function establishes the degree of similarity to at least one from all model examples from our memory. If there is, in our memory, some model example for the same name to which a new object is similar (it means that there is a sufficient degree of similarity), then this new object can be called by this name. Sometimes, for some names it is quite difficult to find only one model example. To illustrate, this is the case with the name "temple". Neither a concrete church nor a concrete synagogue nor a concrete mosque, nor a Buddhist temple, nor … can be the model example for "temple". Pointing out to some concrete church as the model example of temple would suggest that neither a synagogue and nor a mosque is a temple. That is why in the case of some names we have more than one model example. It can be even assumed that we have more than one model example for most names we use. In the case of the name "temple" it is much easier to possess several model examples representing "typical" church, "typical" synagogue, "typical" mosque, "typical" Buddhist temple, and so on. The problem discussed here is typical for general names. A name is *general* if it has more than one reference. In other words the set of designates (a *scope*) of the general name has more than one element. Obviously, a set with more than one element can be divided into non-empty subsets. For every general name there are various principles of division of its scope. These principles depend on our interest. We can choose some interesting for us feature and divide the scope on elements possessing the feature or not. Thus, for the name “child” one can use the feature “born in Europe”. Then, the set of all children will be divided into two subsets: all children born in Europe, and all other children. If the feature is gradable, we can divide the scope into subsets associated with each degree of the feature. For the same name “child” one can use gradable feature “born in the year *n*”. In this case, the set of all children will be divided on subsets: children born in 2014, children born in 2013… It is easy to notice that every subset of the scope has its own name subordinate to the initial name. In general, the scope *a* of the general name *A* can be divided into subsets $a_1, \ldots, a_n$, associated with names $A_1, \ldots, A_n$, being subordinate to *A*. Thus, the name *A* is superior to every name A_i, for $i = 1, \ldots, n$. In Aristotelian logic, there is a tradition to call *a* – the scope of the name *A* – a *genus*, and its every subset a_i, for $i = 1, \ldots, n$, *a species*. If such division of the genus is possible it means than the name of the genus is general. It is the second understanding of the generality of the name. Obviously, both understandings are equivalent. Indeed, if the name has more than one designate it means that its scope can be divided into non-empty subsets, and conversely, if the scope of the name can be divided into non-empty subsets it means that it is a set with more than one element.

For every species (genus), there is a set of features or qualities such that every member of this species (genus) and no other object possesses all of these features or qualities. Every such feature/quality is called a feature/quality *characteristic* for the species (genus). The use of a general name is vulnerable to the *error of generality* (Łukowski, 2012: 92–93). This error has two variants. In the first one, a user of a general name *A* attributes some feature or quality characteristic for one species a_i to members of some other species a_j (with $j \neq i$), because both species belong to the same genus *a*. In another version, a user attributes some feature or quality characteristic for one species a_i to the entire genus *a*, because species a_i belongs to *a*. For example, if somebody observed that his/her daughter likes to play with dolls, he/she believes that the boy in the neighborhood also likes to play with dolls (the first version of the error) or that all children like to play with dolls (the second version of the error). It seems that it is easier to avoid the error of generality in the situation when one model example for correct use of the name is replaced by several different model examples. Of course, such an approach does not guarantee avoiding this error – usually, the set of examples is not sufficient. On the other hand, even in the case of "temple" it is possible to partially define a temple as a building in which people gather to pray. Such species-free definition of the temple can be used in parallel with the definition using various species represented by the set of model examples. Obviously, both ways of defining are partial.

One of the most important features of a majority of expressions of the natural language is simple: every such expression must possess more than one reference – object, situation, etc. In other words, it must be useful in more than one case. This multiusage should be limited in some reasonable way – an expression which would fit all objects, situations, properties, and so on, would be useless. If the expression A can be correctly used in the cases $P_1, \ldots, P_n$, then it should be correctly used also in all cases sufficiently similar to at least one of $P_1, \ldots, P_n$ (Łukowski, 2011: 133–134). Of course, the sufficient similarity is due to the meaning of the expression A. This is a definition of the *tolerance* of the expression A. This definition is strictly connected with the problem of generality. All cases $P_1, \ldots, P_n$ can be and should be understood as the model examples. On the other hand, every nonmathematical general name must be tolerant. Cases $P_1, \ldots, P_n$ mentioned in the definition can represent various species but do not have to – all of them can be model examples of one species only. But the most natural situation is one in which every P_i, for $i = 1, \ldots, n$, represents a distinct species. In the case of the name A = "temple", P_1 can be a concrete church, P_2 some concrete synagogue, P_3 a concrete mosque, and so on.

Obviously, every species of the genus can also be divided into some subspecies. Then, the divided species becomes a new genus and its subspecies, new species. Thus, for example, the scope of the general name "church" cab be divided into subsets composed of, respectively catholic churches, evangelical churches, orthodox churches… Obviously, in the next step, one can divide any scope of the new species, *e.g.* the set of catholic churches into subsets: catholic churches built in France, catholic churches built in Italy… The division of the scope of the general name is impossible when the scope is a one-element set, *e.g.* the scope of the name "church located on the Wawel Hill". Usually, such a one-element set is represented by a proper name – in this case by "Wawel Cathedral".

In the definition of tolerant expression, there is another important element: *similarity* besides model objects/cases. This notion has two essential components: *sufficiency* of similarity and similarity *due to the meaning* of the (tolerant) expression. Both issues can be more easily explained on the colored objects. For this purpose let us imagine three paper leaves: P_1 – big white leaf overprinted by black letters on one side; P_2 – red leaf of the same size and shape as P_1 overprinted in the same way as P_1; and P_3 – small crumpled pale red leaf unwritten on both sides. It is obvious that every two leaves from the set $\{P_1, P_2, P_3\}$ are similar from some reasonable point of view. Moreover, every similarity is even sufficient in some sense but it does not mean that every two leaves are similar in general. Let us consider the name A = "red leaf of paper". Surely, the second object can be called A, and so it can be also a model example for correct use of the name A. In order to know if P_1 or P_3 can be called A, we need to recognize if they are sufficiently similar to P_2, of course, **due to the meaning** of the general name A. Leaves P_1 and P_2 are similar – they are identical in size, both are overprinted in the same way by the same text or pattern. But P_1 is white while P_2 is red. Thus, if the similarity would be settled due to the color, they are not similar at all, and it does not matter if this non-similarity is sufficient or not. It means that, if P_2 is correctly called A, then P_3 cannot be called A – since P_1 and P_2 are not sufficiently similar due to the red color, only one of these cases can be called red. However, the second comparison is not as easy as the first one. Obviously P_1 is red, but what about P_3 which is pale red? In this situation everything depends on our opinion based on our experience, knowledge, intuition, etc. If the pale-red color is in our opinion closer to red than to non-red, we will recognize that P_1 and P_3 are sufficiently similar due to the red color, and so we will call P_3 as red. However, it is also possible that our decision will be opposite. Then, since P_1 and P_3 would not be sufficiently similar due to the red color, P_3 would be called non-red. It seems necessary to emphasize that our decision does not need to be final – we can change it. There are many doubtful and difficult cases. In such cases it is possible that at one moment, two objects can be recognized by us as sufficiently similar due to the meaning of the name, and not recognized as such at some other moment. Moreover, objects observed by us just before the decision is taken influence the decision. Usually in such a situation, a pale red object in the presence of intensively red objects is recognized as non-red. Otherwise, white, black or green objects can provoke us to the opposite attitude. Finally, it is worth noticing that the lack of the sufficient similarity of some object to the given model examples of the correct use of the expression A does not mean that the object is sufficiently similar to the model examples of the correct use of the expression *not-A*. Somebody can be for us neither bald nor not-bald. These states make us introduce (create) a new word(s) for adequate naming of all such dubious cases. The third leaf of paper from the example, can be called neither red, nor not-red but just “pale red”, “almost red”, “pink”, “dimly red”, “softly red”, “white-red”, etc.

It has been already mentioned that our opinion in such cases depends on our experience, knowledge, intuition, and probably on some other factors. But all the factors listed and not listed here depend on model examples we have in our memory or we had at one time. On the other hand, our opinions how to correctly use the name *A* can be identified with our understanding of the meaning of *A*. Thus, our understanding of the meaning of *A* depends on all our model cases of the correct use of the name *A*. In this sense, our understanding of the meaning of words is defined by model cases we have/had in our memory. Since other users of the language influence our choice of the model cases representing in our memory the meaning of words, meanings have social character. In this sense they are created by the community we live in. In two introductory stories (at the beginning of the paper) the role of community is played by the father.

Returning to the example with the three leaves of paper, due to the meaning of the name “leaf of paper” every two (from all three) objects are sufficiently similar, and due to the name “a leaf of paper overprinted by black letters on one side” objects P_1 and P_2 are sufficiently similar, but P_1 and P_3 as well as P_2 and P_3 are not sufficiently similar. The meaning of the name forces us to ignore all features, which are not related to the meaning in consideration. That is why, from the point of view of the meaning of some name *A*, two objects can be sufficiently similar although at first view they seem to be completely and undoubtedly dissimilar.

All these remarks are related to the process of knowing the meaning of these words, which already exist in the language, as well as of the extension of the language by new words naming objects from penumbra of some vague expressions. However, it is not settled how the natural language appears. That is why, there is assumed some “social” knowledge about the correct use of words, and a process of expansion of that knowledge. This “social” knowledge consists of two fundamental components: the sufficient similarity to the model examples and the always only partially given content of expressions. All above remarks dealt mainly with the first component. However, the second one is a fact, too. Unfortunately, we are not aware that accepted by us defining usually is partial only. We believe in the correctness and accuracy of definitions of many words, and all logical arguments (like *sorites*) which contradict these beliefs become a real shock. Defining meaning is partial because it is nothing more than the “verbal” counterpart of the "visual" concrete model example. For its concreteness, every visual model example can only partially represent all possible referents of the word. Similarly, every description of the meaning of the expression defined by words only partially explains the meaning. Aristotle's definition of a man as a rational animal is precise in a similar way as the father's definition of the shed: it is an object standing on the ground smaller than building, constructed from ordinary boards, not for living, just for garden's tools. Every partial defining refers to some concrete observable case and creates a picture of this case. Aristotle thought about people living around him, probably forgetting about people with heavy mental diseases. Similarly, the father looking at the concrete shed did not think about many other various sheds which do not coincide with his description. Usually, this partial defining of expressions leads to no serious problems. Moreover it is economical – communication is faster and simpler. Preciseness can always be increased depending on needs and context. Partial defining causes problems in one case only, when we try to treat the partial definition as the complete one. However, for no natural language expression *E* there exists a complete definition, *i.e.* sentence correct definition of the form: *for any x, x is E if and only if x satisfies* $b_1, \ldots, b_k$. Every finite characteristics of *E* generates gaps in the scope of possible referents of *E*. This problem is closely related with the vagueness of natural language expressions, and is discussed in the next paragraph. This partial defining has some interesting result. Although we strongly believe in correctness, it means completeness of many definitions of words, in any problematic situation we know "very well" how to use the word totally ignoring the definition. For example, we believe that the man is a rational animal, but thinking about some deeply mentally subnormal man, who surely is unable to think, we ignore the definition and without doubts know very well, that he/she is a man.

This defining is also relative – it appeals to the meaning of already known words. Of course, already known words have meaning as precise as those partially defined. All non-mathematical expressions – the already known as well as the new once – have only partial understanding and are mutually relative. Unlike the shed, the building is enough big, it is possible to live in it, etc. The shed unlike the building is rather small and it is impossible to live in it. Such a mutually relative defining is more operative than the hierarchy of definitions with undefined primitive terms at the base. Always, understanding of new words appeals to the understanding of old ones. In this sense it is relative. But unlike the mathematical language, the natural one does not contain the so called primitive terms – terms which are introduced without definition. Although defining is relative in mathematics, not all terms are defined. In the natural language all terms can be defined, if we only wish. This defining is always relative and partial. That is why, unlike in mathematics, natural language expressions have approximate and rough understanding. It is not a problem, because it is temporal – understanding of words is still developed and improved. During their life the child will change old model examples onto better new model pictures of objects, and old partial definitions onto more precise but still partial new verbal characteristics. This process never ends and for two people neither proceeds in the same way nor results exactly in the same understanding of words.

3. Vagueness – another face of the tolerance of expression

In some moments in the above considerations there has appeared the problem of vagueness as a reason for inability of complete understanding/defining of all those natural language expressions which have no mathematical sense. Let us recall that the name *N* is *vague* if and only if there are objects about which nobody can know if they are designates of *N* or not. It means that for some object *O* nobody can reasonably decide that it is a designate of *N* or not a designate of *N*. More generally, an expression *E* is *vague*, if and only if there are some cases that nobody can know (or reasonably decide) if *E* or *not-E* can be correctly used as a name for these cases or not. All those doubtful cases create the so called *penumbra* of the expression also called the *set of border-line cases* (Russell, 1923, also Łukowski, 2011: 141–151). In fact, "penumbra" is a better name because this area of "vague" objects or cases is not a set in the mathematical sense as nobody can even know if some object or case belongs to penumbra or not. Penumbra has no sharp boundaries separating it from both extensions: positive and negative. *Positive extension* of *E* consists of all those objects which are designates of *E* or all those cases which can be correctly named by *E*. The *negative* extension of *E* contains all those objects which are not designates of *E* or all those cases which cannot be correctly named by *E*. Of course, penumbra of *E* has no sharp borders if and only if at least one of the extensions of *E* has no sharp borders. Moreover, an expression *E* is vague if and only if an expression *not-E* is vague.

There is a procedure, called *sorites*, which enables to verify whether the expression is vague. *Sorites* is an argumentation that uses two premises typical for mathematical induction. The first premise states that the expression *E* applies to some object or case. Let us call of the property ascribed by *E* to the object/case a "feature". The second premise allows to transfer the feature from any object/case to the next one – all objects or cases are previously linearly ordered in such a way that the first object should be correctly named by *E* (because it is *E*) and the last one should be correctly named be *not-E* (because it is *not-E*). It is clear that, if both premises are satisfied the entire argumentation results in contradiction: the last object in the sequel is named by *E*, but it is *not-E*. *Sorites* consists of so small steps that passing from *E* to *not-E* cannot be stopped at any other step because of the obvious correctness of every step. Thus, everything in the argumentation is obviously true with the exception of the conclusion which, as it leads to a contradiction, is obviously false. If the sorites applied to the expression *E* leads to contradiction, it means that *E* is vague. Russell, in his famous lecture given in 1923, pointed out the vagueness of many various natural language expressions (Russell, 1923).[2] As it was showed by Black, all names of real, inanimate objects are vague (Black, 1949: 433).[3] Chwistek showed that all names for living beings are vague (Chwistek, 1934: 10–11).[4] Many years later, Sorensen showed that some abstract names also are vague (Sorensen, 1990).[5] Shortly speaking almost all natural language names and predicates are vague, with the exception of: "motion", "rest" (*i.e.* not-motion), Parmenides's "being" and "not-being". It is impossible to pass from the case of movement to the case of rest through cases which are neither motion nor rest. Similarly, there is nothing which would be neither being nor not-being. The only surely sharp (*i.e.* not-vague) expressions are mathematical.

A tolerance is an essence of the vagueness. An expression cannot be vague not being tolerant. Tolerance of the expressions makes the second inductive premise of *sorites* true. In the sequel, every object or case and its successor as well as predecessor are sufficiently similar due the understanding of *E*. That is why, the reasoning cannot be stopped at any step. Such objects are sufficiently similar because a difference – due to the meaning of *E* – between them is *insignificant*. If such difference between two objects or cases is insignificant, then only

---

[2] Russell tried to prove that even logical terms are vague.

[3] The argument showing, that the not-chair (a lump of the chair's leg) must be called a chair.

[4] Chwistek asked his famous question: "Is human's mother a human?". In the light of the theory of evolution the answer on this question is not obvious.

[5] Sorensen showed that "murder" is a vague name. Other abstract terms are also vague, *e.g.* "longing", (Łukowski, 2008: 32–33).

two situations are reasonable: either both of them will be called *E* by us or both of them will be called *not-E*. For insignificant difference it is excluded that one object/case will be called *E* and the second *not-E*. Otherwise, if the difference is significant both objects/cases have to be called differently: if one is *E*, then, the second is *not-E*. It is clear that the inductive premise and the principle of the tolerance are strictly connected. Two names “sufficient similarity” and “insignificant difference” refer to the same. That is why the vagueness is a kind of display of the tolerance.

The second dialogue shows how the boy is taught by his father how to correctly use the vague name “red”. Logical knowledge about vagueness warns us that such a use cannot be consequent and precise. Real cases are named by us by tolerant words whose understanding is established by model examples. However, all such examples can compose a sequent from the reasoning *sorites*. It is a real, unsolvable problem. If anybody from us will be called “human being”, then his/her mother also has to be called “human being” – it is obvious consequence of our meaning of the word “human being”. However, all such selected ancestors compose the sequence imminently leading us to non-human beings. It is an effect of the confrontation of the real world with our limited natural language.

4. Coexistence of generality and vagueness

As it was shown, tolerance of natural language expressions occurs in two forms: generality and vagueness. In the case of vagueness, two objects are sufficiently similar due to some feature, if the feature is gradable in barely recognizable way – degrees of the feature should be small enough. Somebody with n hairs on the head is sufficiently similar due to the understanding of the name “bald” to anybody with *e.g.* $n + 36$ hairs. Then, if one of these persons will be called bald (not bald), then the second one should be called by the same name. This is a “vague” case of the tolerance. However, tolerance does not need to be connected with only gradable features. Then, the tolerance has a face of generality of expressions. Let us assume that *A* is a father of a girl, while *B* is a father of a boy. Due to the understanding of the name “father” *A* and *B* are sufficiently similar. The scope of this name can be divided into sub-scopes: “father of one girl”, “father of two girls”…, “father of a boy”, “father of two boys”…, “father of a girl and a boy”, “father of two girls and one boy”… It is easy to notice, that the possible order of the set of all these sub-scopes can be partial, but not linear. That is why, it is difficult to consider the feature "to be a father" as gradable. However, regardless of the possible order, all sub-scopes consist of cases which are sufficiently similar: due to the meaning of the word "father" the father of two girls and one boy is sufficiently similar to the father of three boys – a difference due to the meaning (*i.e.* correct understanding) of the name "father" is insignificant. Even, if all these sub-scopes would have sharp boundaries there would be still a tolerance – a "general" case of the tolerance. However, from the practice of medicine we know that all these sub-scopes have not sharp but vague borders – a difference between two fathers does not need to be gradable. It means that the name "to be a father" of the feature is a case of tolerance of the mix type simultaneously coming from vagueness and from generality of the word "father".

Intermingling both kinds of tolerance can be also easily exemplified by the word "red". Two objects a red t-shirt and a red car are sufficiently similar due to the understanding of the word "red" – it does not matter that one is a t-shirt and another a car. Both objects belong to the scope of the general name "red (object)". This scope can be divided into many subsets (sub-scopes) represented by names: "red car", "red t-shirt", "red dress", "red umbrella"… However, it does not matter from which subsets two objects come, both are sufficiently similar due to the word "red". Apart from this "general" case of the tolerance there is another, "general" case. Indeed, the scope of the word "red" can be divided in another, "vague" way into subsets represented by names: "bright red", "dark red", "deep dark red", "red with yellow tone", "red with slight yellow tone"… Surely, these sub-scopes are not distinguished by any sharp borders. It is not a difficult exercise to imagine the situation joining both kinds of tolerance: the light-red car and the washy-red t-shirt.

Fortunately, usually we meet such objects and cases which can be clearly distinguished from the point of view the understanding of words. In the first dialogue, the child is looking at the shed which differs gradually from other objects standing around. It is not difficult to imagine a long sequence of objects changing in barely recognizable way from the concrete shed to the concrete building. Then, it would be a problem to name all such objects. Fortunately, in our life we have simpler cases that explicitly differ from each other, like the child, for whom the building clearly differs from the shed. Because of this fact, the father can partially (and relatively) define both names "building" and "shed" – such defining is sufficient.

While observing a new object we need to compare it due to the meaning of the given name with some model example in our memory. Depending on the kind of similarity we have to do with the "vague" or "general" tolerance. Although the name can be simultaneously vague and general (of course, each from another point of view) one of the tolerances should be ignored. We treat the name either as tolerant in the sense of vagueness or as tolerant in the sense of generality. Our choice depends on the character of the feature due to which we recognize the similarity.

5. Conclusion

Procedures expanding linguistic competence can be divided into two kinds. The first teaches us how to correctly use general expressions, the second how to correctly use vague expression. Both procedures differ in some essential way, because there is an essential difference between generality and vagueness. However, both procedures have some common features. **Firstly**, both use linguistic competence of the community. In this sense, our understanding and use of words has a social character. No already existing word can be successfully used in violation of the existing use of this word in the community. **Secondly**, both procedures have an ostensive character – it does not matter, if it is a red color or a building and a shed. Every procedure appeals to the model examples we have in our memory. All these examples play a role of useful, concrete types of these cases in which we can undoubtedly and correctly use the word. **Thirdly**, both procedures only lead to a partial characterization of the correct use of words. One procedure prefers visual characterization, while the other verbal one. Both characterizations are not adequate to the full range of the correct use of the words and so they are not correct. Because of this fact, they are still improved and frequently changed – still with the help of members of the community. So, partiality implies temporality.

Procedure extending linguistic competence by vague expressions will be here called the "**procedure for vagueness teaching**". Its schema has been already discussed above. Recapitulating, if someone would like to take a correct decision to use or not to use the word "red" in a given case, he/she would compare the color of the new object to the model examples existing in the memory. If the similarity seems to be sufficient, the name "red" can be used. In other situations, it is necessary to use some extended method. Since the color of the new object can be neither red nor not-red, it means that it is a tone of some other color, *e.g.* brown. Then, it is useful to recognize to which from these two colors (red and brown) represented by some model examples the new color is more similar. Then, we can say that the new color is brown-red but it is more red that brown – it is more similar to the model of the red color than to the model of the brown color.

Fig. 1

This procedure has some concrete exemplification in the form of the so called *conceptual spaces* invented by Gärdenfors and based on *Voronoi tessellation* (also named *Voronoi partition*) (Gärdenfors, 2000, 2014, see also Voronoi, 1908). Thanks to Gärdenfors's modification it is an interesting scheme to consider the penumbra of vague names. Initially, the distinction between two colors is sharp, exactly due to the Voronoi partition, being an axial symmetric line between two points representing the model examples.

Fig. 2.1

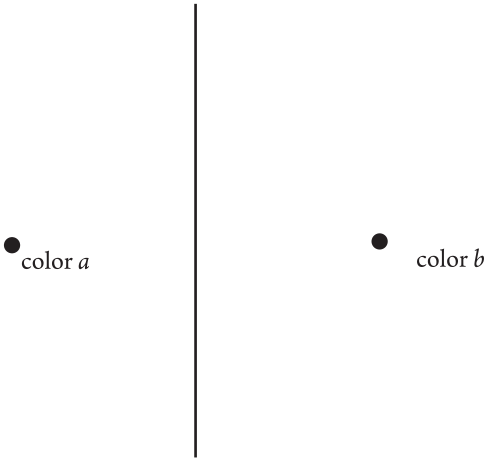

Fig. 2.2[6]

However, it suffices to use more than one model example for at least one color (for example, only red, and not brown) and sharp border is replaced by the bundle of several sharp borders, which taken together give a vague border, *i.e.* penumbra.[7]

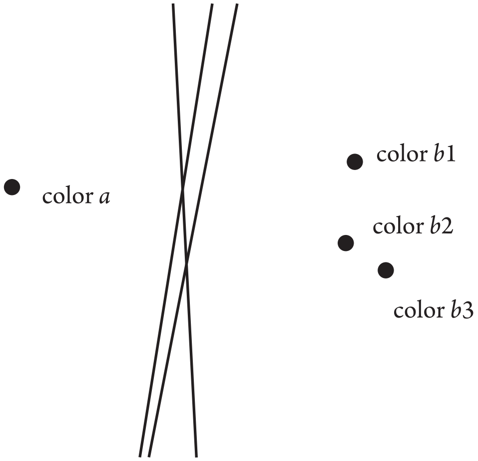

Fig. 3

---

[6] Fig. 2.2 is a schema of the previous fig. 2.1.

[7] This graphic model precisely coincides with the class of precisifications – every sharp line from the graph is just a one precisification function (see: van Frassen, 1966, 1968; Fine, 1975).

Unfortunately, it is impossible to "return" to one colored picture related to the fig. 3. Instead of one figure it should be here, three pictures, respectively to pairs: $(a, b1)$, $(a, b2)$, $(a, b3)$. All of them taken together should create a three-layer picture, where every layer presents its own sharp border between a and bi, for $i = 1, 2, 3$.

Summarizing, let us reconstruct the scheme of the procedure on the example of the second story. The agent (the child) gives information from the community (from the father) that the color of the father's jacket is brown. In the agent's memory there appears a record of the model example of the brown color. In a similar way, it appears in his/her memory another model example of the brown color (leaf). Both examples are different, so in the memory color brown is represented by more than one model example. Thus, the brown color has not one rigid, precise understanding. Since all these examples are given by the community, the agent's understanding of the correct use of the word "brown" is social. After some time, the agent needs to recognize the color of that car. The color of the new object is not represented in the agent's memory, but it is similar to the brown. With the help of the father, the child tries to decide, if the color is more similar to brown or to red – at this stage of the procedure the help of the father is not necessary. If the new color will be more similar to red, the agent will call it red, if it will be more similar to brown, it will be named red. But, there is another third possibility. The agent can decide that the new color is brown-red. However, the decision is not easy, because neither red nor brown are represented by single model examples. It means that there is a vague area of all these cases which are neither red nor brown (see fig. 3). After some time, the agent is able to introduce to his memory new model examples of the brown color – all these colors which are (in the picture) between model examples of the brown.

As it has been already mentioned, colors are not the only vague terms. There are plenty of vague expressions in the natural language. However, the procedure of recognizing of the correct use of the vague name/predicate seems to be standard and similar to the just presented. The problem of the vagueness should be in the procedure appropriately separated from other problems, *e.g.* generality. If we need to recognize which of two objects, a house and a chair, is wooden, we ignore all differences coming from the fact that one is a house and the other a chair. In both cases it can be difficult to decide, if some of them is wooden or not.

Procedure extending a linguistic competence by general expressions will be here called the "**procedure for generality teaching**". Similarly to the previous procedure, this one also employs model examples, mainly at the first moment, but not only. Thus, firstly, the agent calls the object by the name "building" because of some features common to the new object and to the picture of the building. It is a mistake, which is immediately revised by the father – a member of the community. In the next step, the father calls the new object by the name "shed". In the case of the vague term it could be the end of teaching. Here, however, the father formulates a list of features which make an average building distinct from an average shed.[8] This partial, relative and verbal defining of the shed will constitute the correct use of the name "shed" as well as that of "building" in the future. Of course, probably this partial characterization will be improved by a better list, which also will be a partial only defining. Such a partial defining is a logical necessity although it is the reason of mistakes occasionally made by people using it. Let us assume that O is partially defined as a name of an object being a_1, a_2, a_3, a_4, and not being b_1, b_2, b_3 (*i.e.* being $\neg b_1$, $\neg b_2$, $\neg b_3$). Since it is partial defining of O the set $\{a_1, a_2, a_3, a_4, \neg b_1, \neg b_2, \neg b_3\}$ does not suffice as a complete characterization. It means that in the future we might meet an object which should be named O, and its characterization is a little bit different: a_1, a_2, a_5, a_4, $\neg b_1$, $\neg b_2$, $\neg b_3$, $\neg b_4$, *i.e.* instead of a_3 there are added a_5 and $\neg b_4$. Now, we have two alternative definitions, and both are partial – improvement of the partial definition leads to the next partial definition, because the complete definition of O does not exist. Obviously, all definitions are temporal – in the future every definition can be made more precise or even rejected.

Two, considered here, procedures expanding a linguistic competence have some important similarities although the first one deals with vagueness, while the second with generality. Inhelder and Piaget noticed that during our childhood original *accidental similarities* are consequently replaced by more conceptually advanced *essential similarities* (see: Keil, 1989: 9, also Piaget, Inhelder, 1974: 37). Every procedure:

1) is a social phenomenon thanks to Vygotsky's *zone of proximal development* (Keil, 1989: 5–7);
2) is based on model examples – "verbal" or "visual" – permanently replaced by better once;
3) leads to a merely partial defining, which can be "verbal" or "visual";
4) is dynamic and temporal.

---

[8] To speaking the truth, also in the previous case the father tries to explain the difference formulating the feature which is an essence of the difference: that color is not brown because it has a red savor.

From this perspective, a meaning of the natural language expression is identified with the correct, *i.e.* accepted by the community, use of this expression. It means that the meaning of words is **social**, **partial**, and **temporal**, where partial means imprecise and misleading, while temporal means dynamic and changing.

References

Black, M. (1949). *Language and Philosophy. Studies in Method*, Ithaca–New York: Cornell University Press.

Chwistek, L. (1934). Granice zdrowego rozsądku. In: L. Chwistek, *Pisma filozoficzne i logiczne*, vol. 2, Warszawa: PWN, 1963.

Fine, K. (1975). Vagueness, truth and logic. *Synthese* 30: 265–300.

Gärdenfors, P. (2000). *Conceptual Spaces*. Cambridge MA: MIT Press.

Gärdenfors, P. (2014). *Geometry of Meaning. Semantics Based on Conceptual Spaces*. Cambridge MA: MIT Press.

Keil, F.C. (1989). *Concepts, Kinds and Cognitive Development*. Cambridge: MIT Press.

Łukowski, P. (2008). *Paradoksy nieostrości i zmiany*. Toruń: Wydawnictwo Rolewski.

Łukowski, P. (2011). *Paradoxes*. London–New York: Springer.

Łukowski, P. (2012). *Logika praktyczna z elementami wiedzy o manipulacji*. Warszawa: Wolters Kluwer.

Maruszewski, T. (2001). *Psychologia poznania*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Piaget, J., Inhelder B. (1974). *The child's construction of quantities*. London: Routledge and Kegan Paul.

Russell, B. (1923). Vagueness. *The Australasian Journal of Psychology and Philosophy* 1 (4): 84–92.

Sorensen, R. (1990). Vagueness implies cognitivism. *The American Philosophical Quarterly* 27 (1): 1–14.

van Frassen, B.C. (1966). Singular terms , truth value gaps and free logic. *Journal of Philosophy* 63: 481–495.

van Frassen, B.C. (1968). Presupposition, implication and self-reference. *Journal of Philosophy* 65: 136–152.

Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. *Journal für die Reine und Angewandte Mathematik* 133: 97–178.

Konrad Rudnicki

NEUROBIOLOGICAL BASIS FOR EMERGENCE OF NOTIONS

1. Introduction

Even though psychology and logic share several topics of common interest, they operate within different paradigms and therefore encounter difficulties in mutual communication of their advances. One of the most dire problems shared by both is the question of acquisition and structure of notions. Both of these issues have been minutely addressed from perspective of logic in the previous article entitled “Two procedures expanding a linguistic competence” (Łukowski, 2015). The theory proposed there for development of linguistic competence and ontogenesis of notions, has significant impact on foundations of semantics. Its claims could not be made if not for several neuroscientific insights into functioning of the central nervous system that resulted in refining the models of cognitive processes. It is apparent that emergence of notions in an individual is a direct product of brain activity. Even though philosophers still contest the character of relationship between neural function and psychological phenomena, it is no longer controversial in science that the structure of the first determines the structure of the latter. Following that, any theory expressed in language of psychology or philosophy should comply with inferences concerning the architecture of thinking that is derived from neurophysiological data. Although there is a vast amount of information concerning neurobiological correlates of psychological phenomena some of them are of special interest in the discourse on notions and the *meaning* of words. This chapter will focus especially on biological substructure underlying vagueness, generality, dynamicity and temporality of notions.

2. Basic concepts

Every neuronal network, be it artificial or natural, requires constant flow of new inputs (information) to properly function and produce useful output. In case of the human brain inputs are provided by receptors of sensory systems. Contrary to the popular belief, humans have more than five senses. Their actual number is estimated to be around twenty, depending on adopted definition of *sense*. When matter of learning is raised, it is customary in philosophy to concentrate the argument around visual perception. That practice is not groundless as majority of sensory information received by human brain is visual. 10^{10} bits of information is coded by retina each second. Considering that human cortical, sensory network is approximately convergent (Foxe, Schroeder, 2005), only a fraction of the mentioned amount of information reaches cerebral cortex. Around 10^4 bits/sec ultimately reaches the fourth layer of primary visual cortex (V1) (Raichle, 2010). V1 processes visual stimuli at the most rudimentary level. Its cells are sensitive to the most basic kinds of stimuli, for instance straight lines or dots, with respect to their spatial orientation. After preprocessing, signal is propagated to cortical areas where neurons respond to more complex sets of properties. In terms of notion creation it is crucial to note that information, coded by higher-level cortical networks, is derived from activity associated with the simplest stimuli. Pivotal role in the transition between initial processing and emergence of meaning is played by the association areas that integrate information from different modalities. It is still a matter of debate whether the most important role in the generation of conscious thought and concepts is played by some specialized areas or interconnectivity along with the interplay of a whole network. However, the activity of these association areas is definitely essential to this problem (Freeman, 1998). Apart from cortices on the borders between lobes these areas include thalamus and cerebellum, while a central position in that network is occupied by prefrontal cortex. At all stages of the information processing, back-propagation of activity from higher to lower areas is observed. This *reafference* process is partially responsible for a currently established paradigm of constructivism in regard to perception. In simple words: people do not perceive things purely by means of an incoming stimuli, but as they expect/understand/reckon them to be. The most famous process associated with this phenomenon is called *priming*. It would appear that when primarily prompted with stimulus from a given category, people are able to recognize further stimuli from that category faster than from other categories. This process have been demonstrated in many different kinds of material, ranging from verbal (McNamara, 2005) to emotional (Hart *et al.*, 2010). This concept leads to the second important principle of brain functioning.

While activation triggered by stimuli is the first component to be considered in generation of a single *state of the network*, there are others to be accounted for. Equal importance is held by pre-existing architecture of the neural network, determined by biological development and all information processed in the past (Freeman, 2004). All neural networks are shaped through learning processes governed by synaptic plasticity. Exemplary mechanisms of synaptic plasticity are *long term potentiation* and *long term depression*. The first constitutes strengthening of the synaptic connection between two neurons in response to repeated occurrence of action potentials, while outcome of the latter is opposite. In conjunction, these two are among the most prominent processes shaping architecture of the neuronal network (Tsumoto, 1992). There are multiple more mechanisms that co-determine neuronal web; all of them utilize exogenous or endogenous stimuli as triggers, which prepares the system to recept future ones. Following from mentioned principles, a meaningful state is understood as a dynamical activity pattern evoked by stimulus, constructed with respect to pre-existing state of the network, partially incorporating stimulus properties (Freeman, 2004). Activity that constitutes these *states* can be further convergently processed to produce their incomplete representation in the form of an expression (verbal if necessary) that can be communicated in order to elicit similar state in another human being. Incompleteness of linguistic representation stems from convergent architecture of processing which determines that verbalization codes less information than neural state. Even though these *states* are highly dynamical and difficult to grasp, it turns out that they occur in discrete stages, where each stage begins with a transition of the whole network activity to a new spatiotemporal pattern (Freeman, 1998). What is also worth mentioning is the fact that each past state co-determines future states through management of new stimuli influx by means of behavior, further expanding the importance of constructivism in perception. It might appear that some special properties are required from each meaningful state to discern them from non-meaningful states. It would however be controversial on the grounds of biology to contest the presence of meaning even in states unrelated to language (Atlan, Cohen, 1998), since what constitutes a meaning is a response of an organism to any representation with potential informational load. Essentially, for the immune system an antigen is information-bearing representation, which through interaction with immune cells conveys that information and elicits immune response specifically dependent on the type of antigen in question. Be that as it may, when the problem of notions is concerned, language is necessary and crucial.

Neurolinguistic studies already partially described networks engaged in language processing. Most commonly localized in the left hemisphere, collectively called *perisylvian cortex*, these areas include inferior frontal cortex, superior temporal cortex, inferior parietal cortex, operculum and insula. Any damage to these areas produces deficits in language production or understanding (*aphasias*). These areas are densely connected to each other, receive information and partially overlap with higher-order multimodal areas (Catani, Jones, 2005), forming strong links between language and action (Bedny, Caramazza, 2011) what is vital in regard to generation of meaningful states. *Embodied cognition* view even suggested that experience of meaning was embedded in the activation of sensory-motor control systems (Hauk *et al.*, 2004). However, it was experimentally proven that the activation of the modal-specific circuits is not absolutely necessary for comprehension, but is instead involved in learning and intentional action, along with networks responding preferentially to the semantic aspect of the language (Bedny, Caramazza, 2011; Chatterjee, 2010). Nonetheless, it is established that repeated perception of objects from the same category creates a typical network activity pattern which emerges from the generalization of properties of multiple exemplars, and as such can be approximately extracted through statistical analysis. That pattern should be understood in terms of *attractor basin* for dynamical neural system (Duch, Dobosz, 2011), not as set of features of the prototypical object around which a category is layered (Foo, Low, 2008). Note that different networks present in the brain code different levels of abstraction and not in all cases the activation of higher-level areas is necessary (Binder, Desai, 2011). Aforementioned perisylvian areas seem to categorize information in an abstract form, partially corresponding to linguistic categories such as: nouns, verbs, events, closed-class words, etc. (Martin, 2007; Price, 1998). Figure 1 schematically depicts axonal connections between all three mentioned networks. It is yet important to note that the process of sentence comprehension is by all means not a simple addition of single word meanings corrected for the context. It involves numerous streams that process information in regard to: abstract semantic representations, syntactic structures and extra-linguistic sources of information. These streams are specifically sensitive to hundreds of different sentence properties. Their functioning is studied with evoked potentials, functional magnetic resonance and magnetoencephalography (Panizza, 2012).

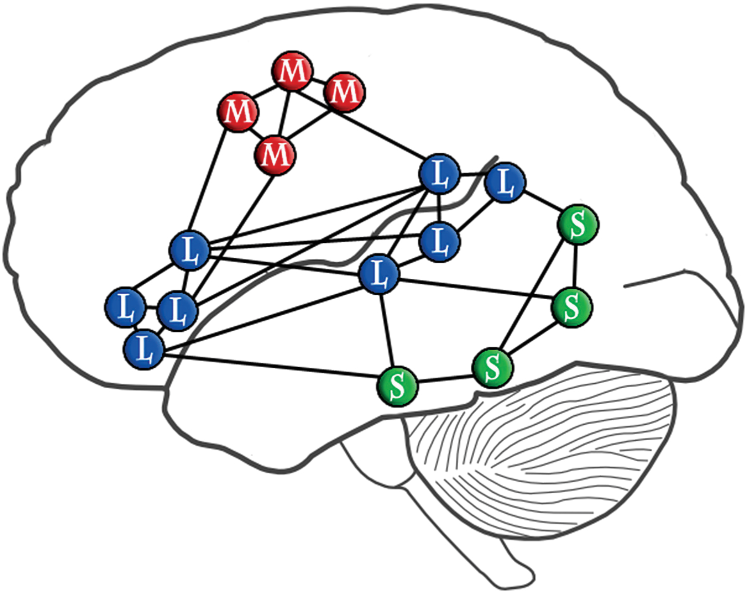

Fig. 1. Schematical visualization of connections between: L – perisylvian cell assemblies, R – semantic circuits in the inferior-temporal object perception stream and M – fronto-central motor systems (Pulvermüller, 2012)

Even though spatial and temporal resolution of methods currently used for the visualization of brain activity is still lacking, neuroscientists resiliently push forward our understanding of cognitive networks by experimentally extracting crucial components of thought. Only recently, a way to communicate with people in a vegetative state, by means of functional magnetic resonance, was found. Experimenters instructed patients to imagine playing tennis if they wanted to answer *yes*, or imagine their old apartment if they wanted to answer *no*, even though there was no guarantee that patients even heard the question. It turned out that approximately 20% of patients in vegetative state were conscious and able to answer questions by imagining concepts while experimenters were able to decode their answers from patterns of brain activity (Owen *et al.*, 2006). Even extremely complex notions are currently described with their substantial neural correlates, as it would appear that the same parts of cortical network are active in experience of beauty when people view classical art and when mathematicians see allegedly beautiful equations (Zeki *et al.*, 2014). This exemplifies contemporary, spectacular advancements in fields investigating neural activation underlying the emergence of complex notions.

3. Dynamicity and zone of proximal development

The acquisition of linguistic competence, a process responsible for emergence of notions was said to be "controlled from outside and depend on the regular provision of new information about the correct use of words" (Łukowski, 2015). Neuronal networks require frequent stimulation of all their compartments, otherwise connections between cells will deteriorate and stored information will vanish. It is apparent though, that networks filter all incoming information, based on its relevance. Otherwise chaotic influx of stimuli would be comparable to a neuronal *noise* (resting-state) and network would take random shape. Aside from the attentional processes responsible for constant governance of explorative behavior, social factors influence the magnitude of each new information acquired by the brain. In dialogues presented in the previous chapter, role of the father was to ascribe the weights to information in the process of ostensive defining, in which neuronal network was engaged. Through reafference, new information, about "buildings" provided by visual system, with addition of verbal information provided by auditory system, was graded as crucial due to the presence of the father – who in this case represented the *zone of proximal development*. As a result, cells and connections that might have been associated with inessential properties of the notion will deteriorate, while those essential will be provided with means of strengthening (for instance, neurotrophins). Moreover, management of the attributed importance of information includes control of the *function of similarity*. Through feedback in social situations people learn what volume of change in perception calls for change in linguistic description.

Zone of proximal development illustrates the environment and its conditions in which a child feels safe and has the ability to explore. In the course of development, shape of the zone changes and impact size of different elements varies (for example in adolescence emphasis shifts from parents to peers). For that reason even though meaning of a word is an incidental neural phenomenon it is co-determined *by the community we live in*. Closely related to that concept is the dynamicity of notions. Following from the fact that each state of the network changes the network itself, it is evident that there can never be two situations in which a notion is understood in exactly the same way. Only after reduction to its linguistic representation which is an incomplete derivative of the incidental mental state, it might seem that notion has its complete and definite description.

4. Model examples and generality

In analysis of neural dynamics, the *model example* for a given word would be detectable if no further stimulus is provided besides the word itself. It would then be possible to describe it as "a prototype vector or a specific distribution of semantic layer activations" (Duch, Dobosz, 2011). Linguistic similarity between words can be assessed in many different ways, and only some of them bear relationship to architecture of the brain's semantic system (Carlson *et al.*, 2014). Nevertheless, it would appear that linguistically similar words have similar *attractor basins*, which are a way to describe neural activity dynamics. For instance, concrete words require a lot more activity from the network than abstract words, since more properties are required to be represented. Consequently, concrete words reach their *attractor basins* faster (Duch, Dobosz, 2011). "Decision" of the network (the outcome of processing expressed *i.e.* as a verbalization) on the membership of the stimuli to one of the categories stems from the *degree of similarity* of the activation elicited by that stimuli to a *prototype vector*. The *similarity* is however not judged, but rather is the cause for a certain outcome (further neural activity). Even though these processes are highly dynamic and temporal, it would seem that access to their shape is possible, since it is possible to distinguish instantaneous, recurring, stable states of synchronization (Fingelkurts, Fingelkurts, 2001). Given repeated exposition to various uses of the word "temple", it comes to no surprise that information coded by a *model example* in a form of neural activation yields no resemblance to a description of any particular type of temple. Similarly, in the research by Zeki *et al.* (2014) there is no attempt to define "beauty", since what is characteristic for all beautiful things might be impossible to articulate, even though we know that it is processed specifically by medial orbitofrontal cortex. That is why the generality of notions arises from inherent properties of information processing and language acquisition (fig. 2).

It follows that the *sufficient similarity* is determined through production of the same communicational outcome (*i.e.* a word) in every particular instance separately. After being perceived by another person (with his/her own architecture of the neural network), it gives raise to the *partiality* of a given content of the expression. That phenomenon falls in line with several other manifestations of the economics in cognitive processes. It would require vast amounts of time and resources from the network or might even be impossible, to produce a language without convergent processing and partial sacrifice of information.

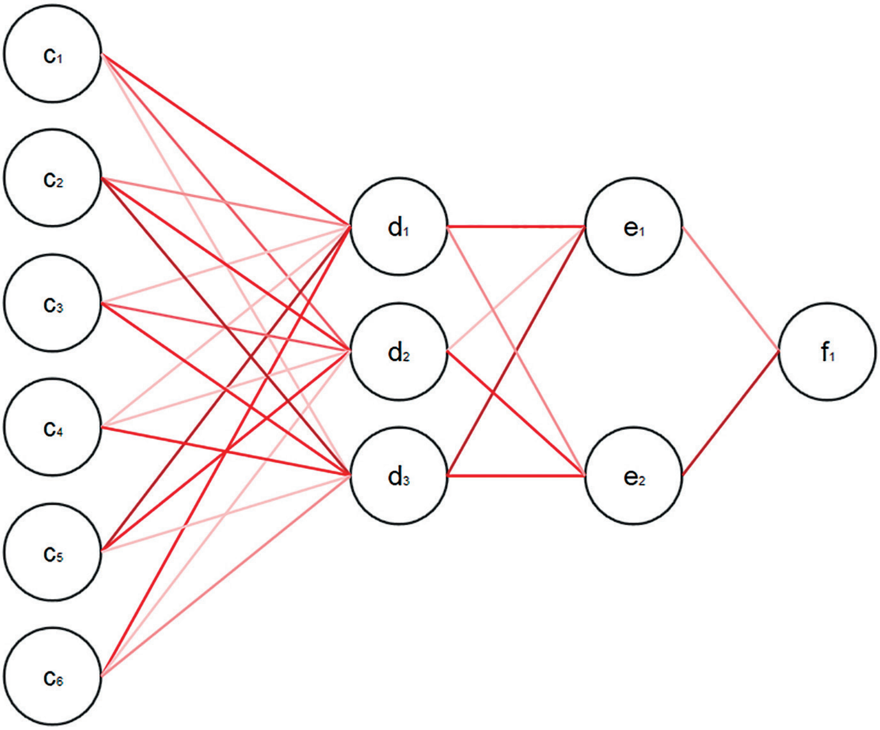

Fig. 2. Schematical visualization of a hypothetical neural network

In the presented in fig. 2 neural network *N*, cells C_n respond preferentially to the simplest properties of the stimuli, and the complexity of detected properties rises with successive layers. Color intensity of a connection symbolizes its strength. Properties of the model example for the notion that is associated with activation of f_1 might be then derived by following a path of strongest connections. In possible, the subsequent process of language production, the language network would receive stimulation convergently, from multiple layers of N. Note that this figure is an oversimplification which does not include reafference, processing in multiple parallel overlapping networks (*i.e.* emotional) and several other mechanisms. Separation of any singular network from the brain requires artificial creation of crisp boundaries, whereas in reality they are fuzzy.

5. Vagueness

There are two neuroscientific concepts critical to the understanding of vagueness and *sorites* argumentation, namely *all-or-none* law and *just noticeable difference* (JND) principle. However, these concepts originate from different levels of description of cognitive processes – one from cellular neurobiology while the other from psychophysics.

All-or-none law states that all action potentials generated by neurons are approximately of the same size (expressed as a value of cellular membrane depolarization) and do not convey information concerning the magnitude of the stimulus that evoked them. Instead, strength of any stimulus might be coded as a frequency of action potentials. It follows, since there is no gradation in singular responses of a cell, that there is a sensory threshold of a stimulus strength at which a cell will generate an action potential 50% of the time. The value of that threshold varies depending on a type of receptor and its history of excitation, since repeated exposition to stimuli changes sensory threshold through sensitization. Therefore, any subliminal stimulus does not provide any information for the network to process. This constitutes the most elementary limitation of human cognition in relation to the *sorites* argument.

One of the earliest laws of psychophysics concerns the smallest difference between two stimuli that people are able to detect. Weber-Fechner law states that JND value is a constant ratio of value of the preceding stimulus. In other words, the amount that is required to be added to the stimulus, in order for the change to be detectable, is a set fraction of that stimulus ($\frac{\Delta I}{I} = k$; where I is the strength of the stimulus, ΔI is the smallest detectable change in the stimulus strength and k is constant). That law does not hold true for all types of stimuli and is in fact only an approximation. However, the general idea of JND relativity encompasses all types of sensory perception. It is easiest to grasp it with the help of an example. Imagine two envelopes and that one of them contains a coin. It is trivial to identify the envelope with a coin, through manual examination of its weight. Yet, if we put a coin inside a shoe it would be impossible to discern its weight from an empty one.

When confronted with these laws, it is clear that *sorites* argumentation puts our mind not only in an unnatural situation, but also impossible to process on grounds of perception. It forces us to refer to purely formal, mathematical procedure and to infer basing on logical, not perceptual understanding of the *insignificant difference*. In terms of cognitive psychology, the *significant difference* starts not on the level of sensory threshold and not even on the level of just noticeable difference, but where neural network architecture put its boundary conceived in the process of linguistic competence acquisition. Furthermore, continuous character of the *sorites* procedure forces us to acknowledge its conclusion as an absolute by virtue of rules governing stability and dynamicity of cognitive processes. Human's perception is constantly dependent on preceding stimuli and characterized by the attention-driven bias in judgment towards what was previously perceived (serial dependence) (Fischer, Whitney, 2014). Imagine a slightly different setting of *sorites* procedure, where all considered ambiguous cases would be randomly mixed and always paired with a unambigous one. After collection of responses from a person, experimenters could sort them back again in the original way. Emerging distribution of responses would reveal the penumbra, while decisions concerning membership to both cases *A* and ~*A* could be described as fuzzy sets. Then, the examination of neurobiological correlates of decisions taken deep in the penumbra would partially reveal variables that influence judgment in ambiguous situations. They would definitely cover individual differences in properties of neuronal noise, excitability and most importantly, an interpretation of situational context which would *i.e.* influence people's tendency to equalize the number of *A* and ~*A* answers. Note that heavy influence of a context on any categorization is definitely part of an adaptive behavior, considering that in an environment different mistakes carry different costs with them. Thus, in the end a decision is always made, and can always be changed, making vagueness a problem of linguistics and mathematical interpretation of crisp sets containing notion designates, not of the perception or language production.

6. Conclusion

Theories of logic in their reasoning typically do not pertain to empirical studies. However, if they encompass mental processes, this lack of grounding in principles of cognitive psychology and neurobiology prevents any theory from being truly complete. The contrary can be seen on the example of linguistic competence where processes described on an epiphenomenal level share their structure with underlying biological counterparts. Concepts of *sufficient similarity*, *model examples* and *partiality* of an expression content, among others, are possible to be better explained through the addition of principles that describe the functioning of neural networks.

References

Atlan, H., Cohen, I.R. (1998). Immune information, self-organization and meaning. *International Immunology* 10 (6): 711–717.

Bedny, M., Caramazza, A. (2011). Perception, action, and word meanings in the human brain: The case from action verbs. *Annals of the New York Academy of Sciences* 1224 (1): 81–95.

Binder, J.R., Desai, R.H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences* 15 (11): 527–536.

Carlson, T.A., Simmons, R.A., Kriegeskorte, N., Slevc, L.R. (2014). The emergence of semantic meaning in the ventral temporal pathway. *Journal of cognitive neuroscienc*e 26 (1): 120–131.

Catani, M., Jones, D.K. (2005). Perisylvian language networks of the human brain. *Annals of neurology* 57 (1): 8–16.

Chatterjee, A. (2010). Disembodying cognition. *Language and cognition* 2 (1): 79–116.

Duch, W., Dobosz, K. (2011). Visualization for understanding of neurodynamical systems. *Cognitive Neurodynamics* 5 (2): 145–160.

Fingelkurts, A.A., Fingelkurts, A.A. (2001). Operational architectonics of the human brain biopotential field: Towards solving the mind-brain problem. *Brain and Mind* 2 (3): 261–296.

Fischer, J., Whitney, D. (2014). Serial dependence in visual perception. *Nature neuroscience* 17: 738–743.

Foo, N., Low, B.T. (2008). A Note on Prototypes, Convexity and Fuzzy Sets. *Studia Logica* 90 (1): 125–137.

Foxe, J.J., Schroeder, C.E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport* 16 (5): 419–423.

Freeman, W.J., (1998). The neurobiology of multimodal sensory integration. *Integrative physiological and behavioral science* 33 (2): 124–129.

Freeman, W.J. (2004). How and why brains create meaning from sensory information. *International journal of bifurcation and chaos* 14 (2): 515–530.

Hart, S.J., Green, S.R., Casp, M., Belger, A. (2010). Emotional priming effects during Stroop task performance. *Neuroimage* 49 (3): 2662–2670.

Hauk, O., Johnsrude, I., Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41 (2): 301–307.

Łukowski, P. (2015). Two procedures expanding a linguistic competence. In: P. Łukowski, A. Gemel, B. Żukowski (eds.), *Cognition, Meaning and Action. Lodz-Lund Studies in Cognitive Science*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Reviews Psychology* 58: 25–45.

McNamara, T.P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S., Pickard, J.D. (2006). Detecting awareness in the vegetative state. *Science* 313 (5792): 1402.

Panizza, D. (2012). Formal neurosemantics. Logic, meaning and composition in the brain. *Journal of Neurolinguistics* 25 (5): 460–488.

Price, C.J. (1998). The functional anatomy of word comprehension and production. *Trends in cognitive sciences* 2 (8): 281–288.

Pulvermüller, F. (2012). Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. *Journal of Neurolinguistics* 25 (5): 423–459.

Raichle, M.E. (2010). Two views of brain function. *Trends in cognitive sciences* 14 (4): 180–190.

Tsumoto, T. (1992). Long-term potentiation and long-term depression in the neocortex. *Progress in neurobiology* 39 (2): 209–228.

Zeki, S., Romaya, J.P., Benincasa, D.M., Atiyah, M.F. (2014). The experience of mathematical beauty and its neural correlates. *Frontiers in human neuroscience* 8: 1–12.

Frank Zenker

SIMILARITY AS DISTANCE: THREE MODELS FOR SCIENTIFIC CONCEPTUAL KNOWLEDGE

1. Introduction

Starting with Minsky (1975) and more widely recognized since Barsalou (1992), (dynamic) *frames* have become a well-accepted tool to model conceptual knowledge. We begin our comparison with its predecessor model, the *feature list,* and trace its development into the (dynamic) frame model. Subsequently, we summarize extant frame-style reconstructions of taxonomic change as a paradigmatic application within history and philosophy of science, thus addressing the incommensurability of frameworks, or world-views, from a cognitive historical perspective.

We hold that a frame is a sophisticated feature list whose application centrally supports the claim that historically successive taxonomies are comparable. But we criticize that the frame model yields little insight beyond taxonomic change. Introducing *conceptual spaces* as an alternative model, we discuss dimensions, their combinations, how to recover frames by analogue expressions, similarity as geometric distance across diachronically varying spaces, and the status of scientific laws.

2. Feature lists, frames, spaces

Originating in Aristotelian philosophy (Taylor, 2003), the feature list constitutes both an entrenched and a somewhat outdated tool for reconstructing conceptual knowledge. Paradigmatically instantiated by taxonomic knowledge, *e.g.*, in biology, it can be summarized by three assumptions: (i) the representation of a concept provides a summary of the class of its instances (*i.e.*, the things falling under the concept); (ii) the binary features representing a concept are individually necessary and jointly sufficient to define the concept; (iii) features are nested in subset relations such that, if concept X is a subset of concept Y, then the defining features of Y are nested in those of X. Such features are regularly referred to as *defining* or *essential* (see Kuukkanen, 2006: 88).

On the classical view, combinations of binary features (aka attributes) provide a definition of that which falls under, or instantiates, a concept if and only if the features are present in, or true of, the thing. Features are normally rendered in natural language, typically by nouns or adjectives. The classical example sees MAN being analyzed as [+biped, +rational, +animal]. Standardly, problems arise upon observing, for instance, that a Para-Olympic athlete may thus fail to instantiate MAN – which is somewhat absurd. Short of allowing *ad hoc* modifications, or throwing individually necessary and jointly sufficient features over board, however, the absurdity is not easily remedied. The model's merits are worlds neatly cut along the patterns that such features generate, where the choice of features may always be a matter of convention, and particular conventions may be contingent upon contexts. To categorize champagne, vodka, fruit juice, and soda water, for instance, why not borrow from chemistry and use [+/–C_6] alcohol, [+/–CO_2] carbon-dioxide.

Barsalou and Hale (1993) demonstrate that feature lists contain rich relational information, primarily with respect to truth, as attributes count as true or false of a thing. Secondly, whatever a feature names, if true of the thing, will name one of its aspects. Thirdly, as set-members, a concept's defining features obey the logical relation of conjunction, just as several concepts obey exclusive disjunction. Fourthly, contingent relational information may be read off a feature list, thus allowing for strict or probabilistic predictions such as: consumers of items in the +C_6 category (likely) need a designated driver. Finally, nesting of concepts accounts for the analytic character of "A bachelor is a man", as BACHELOR, when analyzed as [+man], [–married], is subordinate to MAN.

Exemplar and (weighed) prototype models are mathematically refined extensions of the feature list model that seek to remedy the absurdity of the Para-Olympic example, above. They are in part motivated also by empirical investigations into human categorization (Labov, 1973; Rosch *et al.*, 1976; see Jaeger, 2010) which strongly suggest that humans do not, invariably across contexts, categorize via necessary and sufficient features. Whether all models operate at the *symbolic* level, *i.e.*, presuppose an explicit language, may be debated. Such refined models, at any rate, remain grounded in feature lists, but abandon the strictness with which the (possibly weighed) presence of features projects into category-membership. So in principle, but see below, considerations of similarity rather than identity may govern concept boundaries (see Barsalou, Hale, 1993: 103–124).

The above relations give rise to the frame model of concepts, introduced in the next subsection. Generally, the frame model qualifies as an extension of the traditional feature list, reached by allowing non-binary features (such as large, medium, small) and the relations of constraint and invariance.

2.1. A frame is a sophisticated feature list

When one suspends the additional elements introduced by frames, and moreover restricts attributes to binary values, then the frame model *collapses* into the feature list model, rather than approximating some model analogous to it. This should become clear when appreciating that frames may be stepwise generated from feature lists.

The first step beyond feature lists requires understanding a feature as the value of some attribute. For example, [+blue], [+green] are binary values of the attribute "color" and [+long], [+round] are binary values of the attribute "shape". The additional structure (above that of feature lists) consists in using more than two values to define an attribute, *i.e.*, allowing such values to be *n*-ary. Therefore, an additional relation (which a feature list model does not allow to represent) is that between an attribute and its value(s), called the type-relation (informally: the is-a relation) such as "square" is a type of shape, "blue" is a color, etc. In a second step, one takes attributes to display structural invariants which "specify relations between attributes that do not vary often across instances of a concept" (Barsalou, Hale, 1993: 125). Moreover, and in contrast, constraints represent relations between attribute values "which instead vary widely across the instances of a concept" (*ibid.*: 125). One thus reaches the notion of a simple frame, defined as "a co-occurring set of multi-valued attributes that are integrated by structural invariants" (*ibid.*: 126), such that constraints hold across values and "produce systematic variability in attribute values" (Barsalou, 1992: 37). For instance, a comparatively slim person will normally not be heavy (relative to her height). In combination with structural invariants, then, constraints *generate structure* for the purpose of representing a concept(-instance), giving rise to the notion frame-pattern. Both constraints and structural invariants take on a pivotal role in reconstructing scientific conceptual change (see below).

The main advantage that the frame model can so far claim over feature lists is that "the addition of 'attribute-value relations' and 'structural invariants' increases their expressiveness substantially" (*ibid.*: 127), once given the means to model both stable and variable relations across attributes and values. Consequently, one may view the representation of a concept to proceed *primarily* via structural invariants and constraints. Structural invariants, as it were, tell us which attributes (are likely to) collect, or bind, into a concept, and constrained values identify concept instances.

In a final step, by recursion, the components used in conceptual representation (*i.e.*, attributes, values, structural invariants and constraints) are taken to be represented not by linguistic means ("words"), but by frames: "[T]his recursive process can continue indefinitely, with the components of these more specific frames being represented in turn by frames themselves" (*ibid.*: 133). Whenever conceptual knowledge shall include not just things, but also relations (such as is-part-of, or requires), again, frames are employed recursively: "At any level of analysis, for any frame component, there is always the potential to note new variability across exemplars of the component and capture it in a still more specific frame" (*ibid.*: 134). So there is no principled limit to finding new attributes, "simply by noting variance across the component's exemplars and representing this variance with a new attribute-value structure" (*ibid.*: 133f.). Such recursion results in the more specific structure representing the less specific one, while retaining a symbolic representation nevertheless. After all, both attributes and values are primarily identified through natural language terms. We return to this aspect when comparing frames to conceptual spaces, which is a non-symbolic model, below.

A frame's attributes and values normally arise empirically from querying experimental participants. The choice of attributes, supposedly, is always influenced by "goals, experience and intuitive theories" (Barsalou, 1992: 34). The examples discussed in the literature therefore count as partial representations, including event frames (aka scripts), which are sequential adaptations of the object-frames discussed here. In the scientific case, the identification of attributes, values, etc. is based on the material under study, and so will draw on science historical work, but itself constitutes a more a systematic contribution to the philosophy of science.

2.2. Applying frames to taxonomic change

To further appreciate the frame model, a simple example from ornithology may be helpful. Based on Chen (2002), it does without iteration, employs binary features, and concerns the late 18th century discovery of a South American species of bird (commonly called "screamer") that "has webbed feet like ducks but a pointed beak like chickens" (*ibid.*: 7). This particular combination of features was "not allowed", and so constituted an anomaly to the then-standard taxonomy of Ray (1678). The misfit violated the constraint that the attributes foot and beak always go together. "This anomaly forced [ornithologists] to alter the frame of *bird* and the associated taxonomy, because it made a very important constraint relation between *foot* and *beak* invalid" (Chen, 2002: 7). Figures 1 and 2 illustrate partial frames of the predecessor taxonomy by Ray (1678), and the revised taxonomy by Sundevall (1889), respectively. Notice that, to distinguish WATER from LAND-BIRD, Ray's taxonomy employs the attributes beak (values: round or pointed) and foot (webbed or clawed), connected by a structural invariant (double-headed arrow); see Chen (2002: 5).

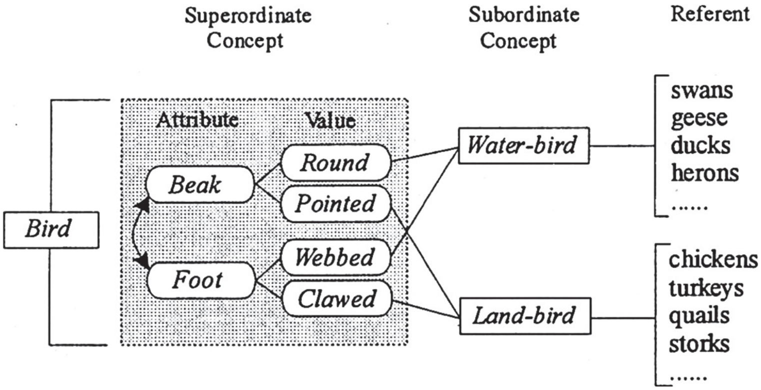

Fig. 1. Partial frame for Ray's (1678) bird concept (Chen, 2002)

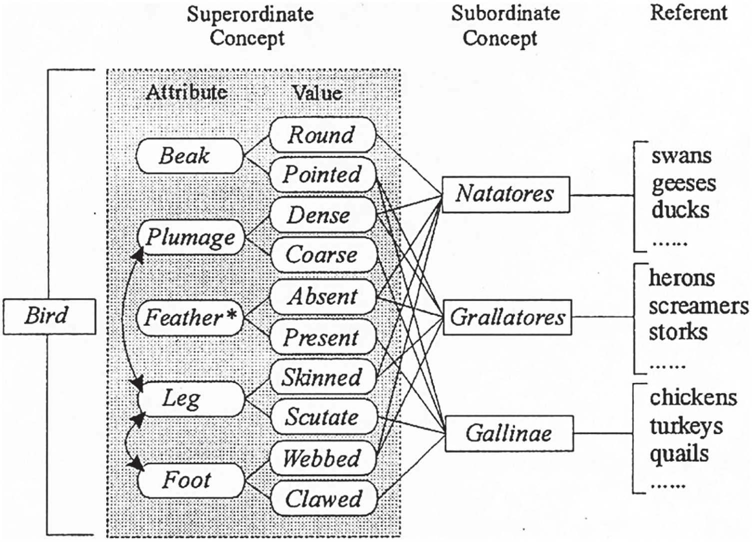

Fig. 2. Partial frame for Sundevall's (1889) bird concept (Chen, 2002)

The historical transition between the representations in figures 1 and 2 is standardly interpreted as a redefinition of the concept BIRD. To frame-theorists, moreover, the example demonstrates scientific taxonomic change to be reconstructable as a *motivated* revision.

> Sundevall's *bird* no longer entails a constraint relation between *beak* and *foot*; instead, new constraint relations are formed between *foot* and *plumage*, as well as between *foot* and *leg covering*. [T]hese are physical constraints imposed by nature, resulting from the adaptation to the environment. The new superordinate concept inevitably alters the taxonomy by expanding the conceptual field at the subordinate level. (Chen, 2002: 8)

Further contrasting fig. 2 with the successor taxonomy developed by Gadow (1893), in fig. 3, the subsequent transition to it may be interpreted as a more radical shift than that from Ray's to Sundevall's. After all, "Darwin discovered that species are not constant, and therefore affinity among species must be founded on their common origin" (Chen, 2002: 12). Gadow's taxonomy, moreover, had been developed in response to Sundevall's which "emphasized the dissimilarities between screamers and waterfowl" (*ibid.*: 12) rather than their similarities.

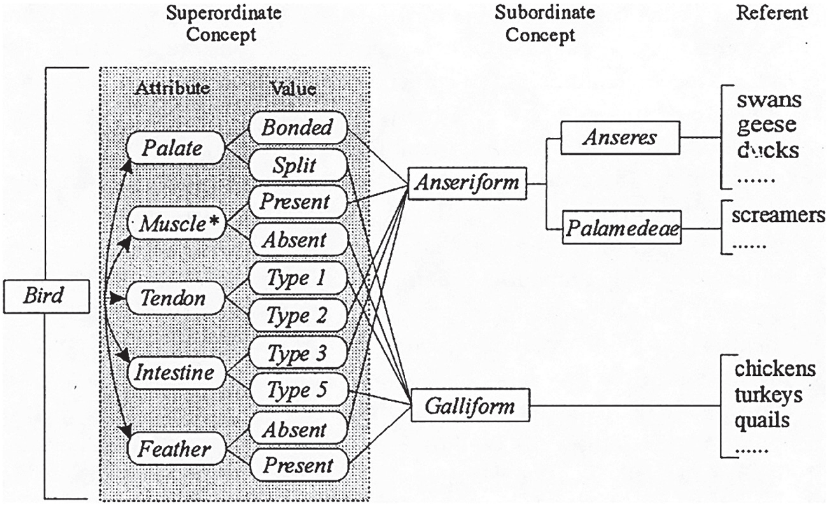

Fig. 3. Partial frame for Gadow's (1893) bird concept (Chen, 2002)

Except for the attribute feathering arrangement, Gadow's attributes pick out (radically) different morphological features than those in figures 1 and 2. Each attribute, moreover, is connected by a constraint. Constant across the three taxonomies remains the use of body parts, motivated by their cognitive salience (Tversky, Hemenway, 1984; see Chen, 2002: 16f.).

2.3. Frames and historical scientific change

The reconstruction and comparison of taxonomies as partial frames has been carried out with particular regard to Kuhn's (1970) incommensurability of taxonomies.[1] *Contra* Kuhn, a frame-reconstruction is said to provide some form of comparability. Firstly, applying the frame model facilitates a representation that helps explain why Sundevall's and Gadow's taxonomies are mutually incommensurable: the pair violates Kuhn's no overlap-principle for kind terms (Kuhn, 1993; Chen, 1997). The no-overlap principle can be rendered as "concepts

---

[1] Further applications of the frame model to scientific change can be found, among others, in Andersen, Barker and Chen (1996); Chen *et al.* (1998); Chen (2003, 2005); Chen and Barker (2000), and the book-length Andersen *et al.* (2006). Cases range from the wave vs. the particle theory of light, over astronomy and nuclear physics, to the transition from Maxwell's to Einstein's conception of electro-dynamic action. The latter is briefly discussed below.

belonging to the same subordinate group cannot overlap in their referents" (Barker *et al.*, 2003: 226). This, in turn, shall also help explain why "communication obstacles were bound to occur between the followers of the two systems [...] who could not find an equivalent native term with referents that do not overlap those of the foreign one" (Chen, 2002: 9). Secondly, in developing a consensus on the superiority of Gadow's taxonomy over Sundevall's, the scientific community of ornithologist could chose in an instrumentally rational way – or so the reconstructive method supports – because both Gadov and Sundevall used spatial features (body parts). Hence, contrary to the incomparability-interpretation of "incommensurable" – being one the mature Kuhn had come to reject (Chen, 1997; Kuhn, 1983; Hoyningen-Huene, 1993; Zenker, Gärdenfors, 2015) – criteria could have been rationally compared. Therefore, "the preference for spatial features in attribute selection could have functioned as a cognitive platform for the rational comparison of the Sundevall and the Gadow systems [...]" (Chen, 2002: 18).

Allegedly incommensurable taxonomies, thus, may have cut nature along different features, but it were spatial features nevertheless. Such cuts, therefore, need not result in incomparable taxonomies despite violating the no-overlap principle. Consequently, historical transitions between taxonomies that *prima facie* support the incommensurability thesis (because a given pair of taxonomies is thought to undermine choice-rationality whenever it consists of mutually incomparable alternatives) may – namely upon comparing their frame reconstructions – turn out to be either reconcilable with, or in violation of, one or more standard maxims of choice-rational action, *e.g.*, a mini-max principle (see Zenker, Gärdenfors (2015) on choice-rationality vs. communicative rationality in scientific change). This result, Chen suggests, draws on distinctly *cognitive* mechanisms.

> [T]axonomic change is rooted deeply in the cognitive mechanisms behind the processes of classification and concept representation. These cognitive mechanisms determine the process of mutual understanding and rational comparison during taxonomic change. In fact, the cognitive platforms for rational comparison identified in our historical cases, that is, compatible contrast sets and attribute lists, were the products of such cognitive mechanisms as the relational assumptions adopted in classification and the preference for body parts developed in concept representation. (Chen, 2002: 19)

We now turn to some aspects of the critical reception of the frame model.

2.4. Incomparability and non-translatability

As Stanford (2008) observes, although they are presumably fine for descriptive purposes, frames do not appear to improve our understanding of incommensurable world-views as cognitive phenomena. But a slightly more drastic consequence might obtain. As far as taxonomic knowledge is concerned, when we employ frames then the incommensurability as incomparability of world-views is seemingly reconstructed away, while incommensurability as non-translatability of world-views is reconstructively confirmed. After all, as we saw, the frame-reconstruction confirms that historically succeeding taxonomies did overlap in their referents; so its items are not open to a one-to-one translation.

The fact seems to be this: using a frame-model, one reaches a representation where the comparison of the conceptual structures of two or more world-views consists in nothing but a comparison of their (partial) frames. This allows tracing the requisite constraint violation and observing if, and how, anomalies are resolved in a different frame, or not. If this is a fact, then one may use it in at least two ways. One option is to object, as Stanford did, that the frame model leaves the genesis and the effects of incommensurable world views in the dark; frames merely facilitate a different view on the non-translatability side of the issue – which thus speaks against what Nersessian (1995) dubbed the "cognitive historical approach". But one may also seek to mount the following more far-reaching claim: since the cognitive historical reconstruction renders allegedly incommensurable taxonomies comparable, it follows that incommensurability as incomparability is *false* as a claim regarding the cognitive representation of concepts. Moreover, if insights into causes and effects of the incommensurability of world-views are needed, then – as far as a cognitive account of conceptual representation is concerned – might not such insights just as well lie outside of it?

One thus suggests that causes and effects of this phenomenon need not be treated as genuine issues of conceptual representation. Instead, incommensurability as non-translatability of world-views, and the communication breakdown that goes along with it, may – perhaps more straightforwardly, too – be explained by citing human imperfections.[2] Presumably, this option won't sit well with

---

[2] For instance, one might cite psychological deficits, in the sense of having remained, or become, unable to adopt or switch between different views, or strong biases in the sense of no longer considering, *e.g.*, that claims to one ultimate ontology, or a final description of the world,

everybody. Vis-à-vis the comparability claim, which can be supported both by the frame model and other models of conceptual representation, however, this option may be harder to resist. So, comparability being provided, translatability appears to be a less pressing issue. I take frame models to support that, as a thesis on the incomparability of conceptual structures, incommensurability is a false claim. And as a claim on the non-translatability of world-views, the plausibility of incommensurability can largely, though perhaps not entirely, be accounted for by drawing on factors that do not pertain to human conceptual representation.[3]

2.5. Conceptual spaces

The expressive power that frames gain over feature lists is notable, but overall perhaps rather meager. In support of this claim, frames will now be compared to *conceptual spaces* (Gärdenfors, 2000). The latter is argued to be a more useful model in application to scientific concepts, because measurement-theoretic considerations that, for instance, underlie nominal, ordinal, interval, and ratio scales are native to the model, as it is generally "closer" to mathematics than the other two. Therefore alone, a spatial representation should sit much better with the intuitions of working scientists. Furthermore, reference to the empirical world is of lesser importance; ontological finality is neither implied, nor precluded, by the model. Particularly whether a measurement structure picks out a real structure, or not, isn't a pressing question (see below).

Conceptual spaces provide a geometric or topological model of conceptual knowledge. An assumption that seems basic to the frame model – that concepts are represented primarily symbolically, and only then through the frame-structures themselves – is here discarded. Instead, information is modeled at a level *between* the symbolic and the subconceptual one. The symbolic forms of mathematical physics, for instance, do therefore not represent concepts, but specify mathematical relational structures. Scientific axioms and laws can therefore be viewed as the symbolic expression of constraints on the distribution of points in a space, a view that will become clearer below.

---

may be dogmatic, or group-sociological/institutional, in the sense that actors are rationally uncompelled to consider alternatives to some research program that they have been investing in, and have perhaps also profited from.

[3] For a more upbeat review of Andersen *et al.* (2006), see Botteril (2007). For further criticism, see Bird (2012).

Past Stevens' (1946) influential work (to whom the above classification of four differentially informative measurement scales goes back), in "mature" measurement theory, mathematical relational structures are embeddable into empirical relational structures, *i.e.*, principally projectable into an ultimate ontology (structures may therefore be called "real"). Stevens did not, in any detail, treat the conditions that such empirical structures should satisfy (see Diez, 1997a: 180). From the point of view of the conceptual spaces model, however, this is fine. After all, the postulated dimensions aren't in any good sense "out there" either.[4]

We now provide a non-technical summary of conceptual spaces. Rigorous treatments include Aisbett and Gibbon (2001) and also Adams and Raubal (2009), an overview of applications of conceptual spaces being provided in Zenker and Gärdenfors (2015).

2.5.1. Dimensions

A conceptual space is built up from a number of quality dimensions. Examples include temperature, weight, brightness, pitch, as well as the three ordinary spatial dimensions (height, width, depth). In science, moreover, one regularly finds quality dimensions of an abstract non-sensory character such as mass, force, or energy. The notion of a dimension may be taken literally. Each quality dimension is assumed to be endowed with a certain *geometrical* structure. Figure 4 illustrates the weight-dimension (one-dimensional with a zero point); it is isomorphic to the half-line of the non-negative numbers. That there are no negative weights is a basic assumption commonly made, *e.g.*, in physics, that is far from trivial; the non-negativity of the weight dimension is a historical contingency. As an *ad hoc* assumption, for instance, the fire-substance *phlogiston* (a theoretical entity) had for some time been assumed to have negative weight, and in the late 18th century gave way to the oxygen-account (McCann, 1978; Gärdenfors, Zenker, 2013).

Fig. 4. The weight dimension

---

[4] Some mathematical relational structures have been claimed as being constitutive of empirical relational structures, and so count as methodologically a priori (see, *e.g.*, Friedman 2001). This Neo-Kantian aspect is addressed in Zenker and Gärdenfors (2015).

As a second example, following Munsell (1915), the cognitive representation of color can be modeled by three dimensions (fig. 5). The first is *hue*, represented by the familiar color circle (red via yellow to green, blue and back to red). The topological structure of this dimension differs from the dimensions that represents *time* or *weight* (which are both isomorphic to the real line). The second dimension is *saturation* (or chromaticity), ranging from grey (zero color intensity) to increasingly greater intensities; it is isomorphic to an interval of the real line. The third dimension is *brightness*, varying from white to black; it is a linear dimension with end points. Together, these three dimensions, one with circular and two with linear structure, constitute the color domain which is a subspace of our perceptual conceptual space. This domain is regularly illustrated by the color spindle (two cones attached at their bases). Brightness is shown on the vertical axis; saturation is represented as the distance from the center of the spindle; and hue is represented by the positions along the perimeter of the central circle. The circle at the center of the spindle, moreover, is tilted so that the distance between yellow and white is shorter than that between blue and white.

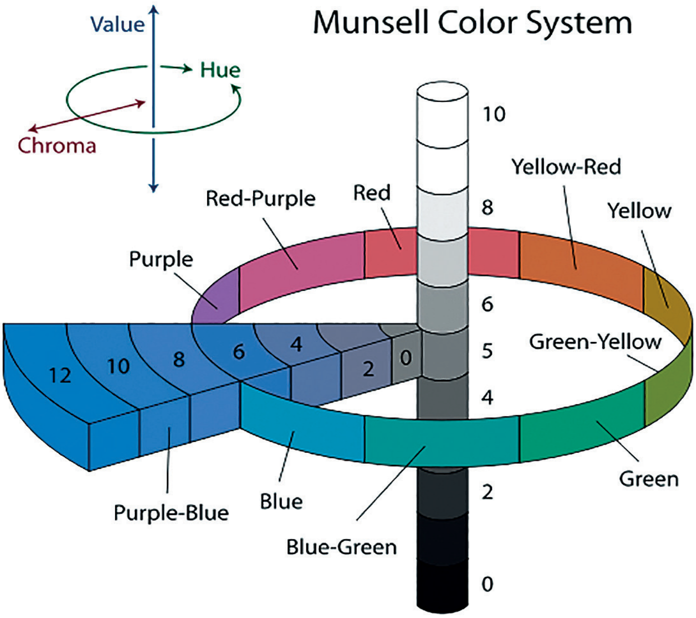

Fig. 5. The Munsell color system

2.5.2. Combinations

A conceptual space can be defined as a collection of quality dimensions. However, the dimensions of a space should not be viewed as being totally independent. Rather, they are *correlated* in various ways since the properties of the objects modeled in the space co-vary. For example, in the domain of fruits the ripeness and the color dimensions co-vary. It is presumably not possible to give a complete list of the quality dimensions that make up the conceptual spaces of humans. Some of these dimensions seem to be innate and to some extent "hardwired" (*e.g.* color, pitch, force, and probably ordinary three dimensional space). Others are presumably learned, yet others are introduced by science (see Gärdenfors, Zenker, 2013).[5]

In modeling a scientific concept, the requisite dimensions have to be identified and values that report distances on these dimensions, *i.e.*, a *metric* (see Berka, 1983: 93), must be assigned. For instance, day and nighttime are standardly modeled by one circular dimension with 24 equally-sized intervals called *hours*. Before the invention of mechanical clocks, however, the two points on this dimension that separate twelve hours of nighttime from twelve hours of daytime were commonly coordinated locally to sunrise and sunset. As both points shifted, again locally, over the course of one year, their distance changed, thus giving rise to a variable temporal metric which, meanwhile, has given way to one of constant clock intervals. The same occurred in the case of ordinary space, but in the inverse direction – from constant to variable – when the Euclidian metric assumed by Newton was first replace by the Minkowskian metric (used in special relativity) and then the Schwarzschild metric (general relativity), both of which generalize the Euclidian one. It is easy to see that a change of metric leads to a change in the symbolic expressions used to calculate with these distances.

If it is not possible to describe an object fully by assigning a value on one dimension without also assigning a value on another, then such dimensions are said to be *integral*; otherwise they are called *separable*. For instance, an object cannot be given a brightness value without also giving it a hue; the pitch of a sound always goes along with its loudness. Or take Newtonian mechanics, where an object is fully described only when it is assigned values on eight dimensions: 3-D space, 1-D time, 3-D force, 1-D mass.[6]

---

[5] For criticism, see Gauker (2007) and, primarily as to the necessity of positing such spaces, Decock (2006).

[6] Since $F = ma$ holds, some values can be inferred, *e.g.*, for the three force dimensions.

On these distinctions, the notion of a *domain* can be defined as a set of integral dimensions separable from all other dimensions. More precisely, domain C is separable from D in a theory, if and only if the invariance transformations of the dimensions in C do not involve any dimension from D; and the dimensions of a domain C are integral, if and only if their invariance class does not involve any other dimension. For instance, until the advent of relativity theories in physics, the three spatial dimensions were separable from the time dimension. So, the spatial coordinates x, y, z were separable from the time coordinate t in Galilean, but not in Lorentz transformations. Moreover, *mass* is separable from everything else in Newton's theory, but is no longer separable from *energy* in special relativity (see Gärdenfors, Zenker, 2013).

As the criterion for identifying a domain, we propose the independence of the respective measurement procedures (see Diez, 1997a: 183f.). For instance, in classical mechanics, the measurement of distance and duration (trigonometry and chronometry, respectively) are independent, as light signals are tacitly assumed to propagate instantaneously rather than at finite speed.

2.6. The conceptual spaces model recovers the frame model

A comparison between frames and conceptual spaces in case of taxonomic knowledge is straightforward. The main result is that the notions attribute, value, structural invariance and constraint (see above) can be provided with analogues. In particular, the structural invariants and constraints of the frame model arise naturally from the geometry of the conceptual space (*e.g.*, category membership is a matter of occupying regions of a space). It is therefore reasonable to view conceptual spaces as a refinement, or improvement, of the frame model. Using one or the other tool may, at times, be a matter of convenience, related to the complexity of what is to be modeled. For taxonomic knowledge, for instance, conceptual spaces appear over-powerful. Put differently, representing conceptual knowledge at nominal level (through n-ary features) is under-complex.

2.6.1. Analogues

What on the frame-model is an attribute will correspond to a single dimension or, as the case may be, to combinations thereof. For instance, each color can be represented as a sub-region of the space spanned by *hue*, *saturation* and *brightness*, rather than by natural language terms. Further, the value of an attribute corresponds to a point or, as the case may be, to an interval on one or several dimensions. Generally, the metric of a dimension corresponds closest to the attributes' values. On assumption of being an equal distance apart, for instance, the values "large", "medium" and "small" of the attribute "size" will be modeled by an interval-scaled dimension, otherwise by an ordinal scaled one (when this assumption is relaxed), or a nominal-scaled one (when no such ordering relation is assumed), or, in the case of the most informative metric, an ratio scaled one (where the value zero is meaningful, which it is only if what could be measured can in fact be absent). Yet, unlike the conceptual spaces model – where *betweenness* is meaningful by virtue of the dimensions' geometric properties – nothing in the frame model represents in a *motivated* way that "medium" is between "big" and "small". Model users know as much, the model doesn't!

The purpose of a constraint, as we had seen, was to rule out, or make unlikely, some among the logically many attribute-value combinations. Constraints thus result from the particular selection of values that define a subordinate category. To mimic this in conceptual spaces, where instances of a concept are represented as points or vectors in an *n*-dimensional space, one may speak of a space's sub-regions being comparatively unpopulated. The notion structural invariant, finally, corresponds to a correlation of dimensions, which means that a number of dimensions represent jointly.[7]

2.6.2. Similarity as distance

To appreciate how distances may be exploited in accounting for similarity, reconsider the attributes and values of Sundevall's taxonomy (fig. 2), with abbreviations in brackets:

---

[7] Structural invariants have been interpreted to represent synthetic a priori knowledge, *i.e.* knowledge about the empirical world that originates in a (taxonomic) structure constitutive of experience, but itself not based in it. Structural invariants, for instance, are claimed to account for such synthetic *a priori* knowledge claims as: "There are no [normal] birds with legs that attach to their necks" (Barker *et al.*, 2003: 225f.). Denial of this claim may lead a hearer to assume that a speaker does not understand the concept BIRD. The synthetic *a priori* status of such knowledge can, in principle, be saved in conceptual spaces, assuming one has identified it. At the same time, it remains unclear if singling out these rather than those elements as synthetic *a priori* is helpful, or necessary (see Zenker, Gärdenfors, 2015).

beak (BE)	round (ro), pointed (po)
plumage (PL)	course (co), dense (de)
feather (FE)	absent (ab), present (pr)
leg (LE)	skinned (sk), scutate (sc)
foot (FO)	webbed (we), clawed (cl)

One may here treat each attribute as a nominal-scaled dimension. All values are binary, *e.g.*, the beak is either round or pointed; a third value does not apply. So each dimension gives rise to a scale with two parts (nominal order), yielding a total of five integral dimensions. The natural kinds *Natarotes, Grallatores,* and *Gallinea* are thus reconstituted as three distinct regions of a mostly unpopulated 5-D space (tab. 1).

Table 1. Comparison of dimensions in Sundevall's taxonomy

Kind terms	Nominal dimensions	Instances
Natarotes	BE-*ro*, PL-de, FE-ab, LE-sk, FO-we	swans, geese, ducks
Grallatores	BE-po, PL-de, FE-ab, LE-sk, FO-we	herons, screamers, storks
Gallinae	BE-po, PL-*co*, FE-pr, LE-*sc*, FO-*cl*	chickens, turkeys, quails

Table 1 readily conveys that *Natarotes* and *Grallatores* are similar up to the beak-dimension (BE-*ro* vs. BE-*po*). Their similarity remains rather hidden in a frame model, but would be immediate in a feature list, or in a conceptual space, here in virtue of *Natarotes* and *Grallatores* occupying neighboring regions, or hyperplanes, in a 5-D space. In a frame and a feature list model, moreover, it is unclear how to measure by virtue of the tool employed the comparative distance between *Natarotes, Grallatores* and *Gallinae*. In the idiom of conceptual spaces, in contrast, the *Gallinae* region is maximally distant from the *Natarotes* region, as it differs on four dimensions from *Grallatores*. That this distance cannot be expressed more informatively is a result of employing binary features. Generally, moreover, when seeking to express taxonomic difference as distance, then the conceptual space model has been invoked implicitly.

Because the attribute (read: dimension) feather is retained with identical values, one may describe the change from Sundevall's to Gadow's taxonomy as a replacement, or a revision, of four dimensions (*cum* invariants and constraints). This yields a somewhat trivial, but a no less correct reconstruction of conceptual change. Such is perhaps easier to accept when the incommensurability of world-views is not treated as a genuine issue of conceptual representation (see above). The partial frame of Gadow's new taxonomy, moreover, features five dimensions, not all of which take binary values. One may therefore say that complexity, as measured by the number of dimensions, is not constant. Gadow uses four *new* dimensions; featuring also one region less, in this respect, his taxonomy is simpler than Sundevall's. The types of intestines (type 3 and 5), by contrast, suggest that complexity has increased. The same seems to hold for the tendon dimension. *Prima facie*, these dimensions still constitute nominal scales.

By defining change-operations on the dimensions and their mode of combination, the conceptual spaces model may also be applied dynamically. In increasing order of the severity of a revision, these operations are: (i) addition/deletion of laws,[8] (ii) change in scale/metric, (iii) change in integrality/separation of dimensions, (vi) change in importance (or salience) of dimensions; (iv) addition/deletion of dimensions (for examples, see Gärdenfors, Zenker, 2010; 2013). A more informative pre- vs. post-change reconstruction would seek to employ the comparative distance between taxonomic items. Relative distance between reconstitutions of dimensional points within (regions of) spaces would thus measure, for instance, whether screamers had become more similar to ducks, or not. *Severity of scientific change* then comes out as *distance between spaces* (rather than distance within one and the same space), *i.e.*, as a function of the above types of changes, and some second-order distance measure.

2.7. Laws as constraints

In the following, we view scientific laws as symbolic expression of constraints on conceptual spaces. Consequently, historical transitions to a posterior space are in principle continuously reconstructable through a modification of the prior space's dimensions, leaving little room for incommensurability to seriously trouble a cognitive account of scientific conceptual knowledge representation.

The foundations for a theory of measurement in the modern sense arise with Helmholtz (1887), were provided with – some say, insufficient – systematization by Stevens (1946), and have been further developed by Krantz, Luce, Suppes and

---

[8] Note that the first type of change is not strictly a change of the space; a change of law merely effects how the points are distributed in it.

Tversky (1971, 1989, 1990). For an overview and the caveat in Steven's work, see Diez (1997a, b), Hand (2004). When dimensions are fine grained, one approaches scientifically exact measurement. Here, shortcomings in the information conveyed by the frame model's attribute-value structure may be observed that suggest a revision of this model. When attributes are not bi-, but *n*-ary, then any attempt at using frames to model ordering relations between values presumably incurs a revision towards conceptual spaces. For instance, when modelling scientific concepts, *e.g.*, in physics, dimensions tend to (though they need not) be ratio-scales; one will want to make sense of the fact that empirical theories and their mathematical laws commonly depend on, and give rise to, measurement results at this level of scale.

As indicated above, one may also attempt to motivate the *symbolic* character of scientific laws by virtue of the representational tool. Andersen and Nersessian (2000), for instance, state that they "believe that [frame] analysis can be extended to represent the similarity class of problem situations for *nomic* concepts" (*ibid.*, 230), *i.e.*, those obeying law-like generalizations. In their electromagnetism example (fig. 6), the Lorentz force-treatment is distinguished from the electromotive force-treatment; frame-style, the attributes conductor, ether, and magnet (values: moving or at rest) are coordinated to the respective force laws, whose symbolic forms however differ strikingly, and implausibly so, as the application situation is identical. (In modern terms, applications pertain to the *relative* motion of a magnet vis-à-vis that of a conductor.) "[I]n Maxwellian electrodynamics, although the resultant electromagnetic induction is the same whether it is the magnet or the conductor that is moving and the other at rest, these are interpreted as two *different kinds* of problem situations" (*ibid.*: 237, italics added). By now, the point of their example will be familiar: by suspending the attribute ether, Einstein's revision of Maxwell's electrodynamics removed a "total overlap" (*ibid.*) between the two treatments.

As can be observed in fig. 6, the laws are appended, rather than motivated by the frame model. Frames therefore seem to apply to scientific laws without providing insight into their status as symbolic generalizations. Such strikingly different formulae, being standardly treated as evidence of "symbolic rupture" in radical scientific change, however, can also be viewed as the symbolic expressions of constraints over different conceptual spaces. Indeed, scientific laws may be viewed as nothing but the symbolic forms of such constraints. This demotes the importance of laws in scientific change vis-à-vis the dominant view (*e.g.*, Dorato, 2005).

Consequently, the continuity of mathematical structure that is reconstructively achieved in limiting case reduction (see Batterman, 2003) – an achievement that, following Worrall (1989), structural realists tend to cite as strong evidence in disfavor of incommensurability claims – need not exclusively be treated as a matter of scientific laws or axioms, either. Instead, if empirical theories are primarily characterized through the scale-type of the dimensions – or, more contemporaneously, the admissible transformation of a scale (see Diez, 1997b) – and their mode of combinations (integral vs. separable), then "continuity in scientific change" denotes the continuous generation of a predecessor into a successor space (see Gärdenfors, Zenker, 2013; Zenker, Gärdenfors, 2015).

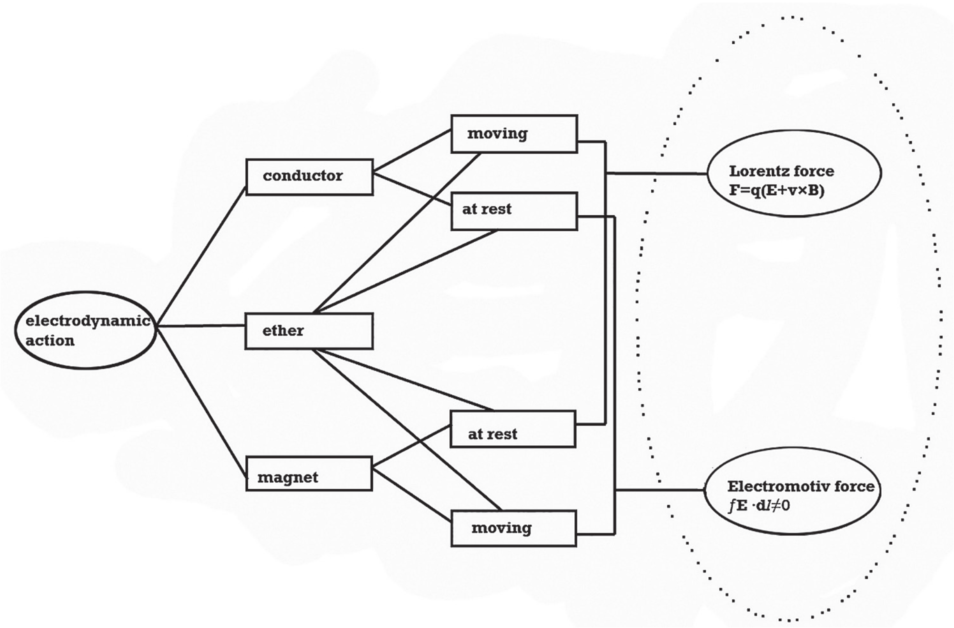

Fig. 6. Partial frame for Maxwellian "electrodynamic action" (Andersen, Nersessian, 2000: S238)

Zenker and Petersen (2014), for instance, have applied conceptual spaces to the conceptual development of 19th-century fluid dynamics, tracing the changes to the underlying spaces from the Euler equation to the Navier-Stokes equation. On their analysis, the developmental process can be viewed as an instance of normal science, in the sense of Kuhn (1970), because the most radical change, the deletion of dimensions, did not occur. Consequently, diachronically prior successful applications of fluid dynamics were always retained as limiting cases of posterior models. (This paper also provides a spatial view on the limiting-case relation.)

To achieve a representation of scientific concepts that is sensitive to the empirical theories in which the concepts feature, but also informed by their historical and potential future dynamics, similarity measures over *diachronically* related spaces remain wanting. This would hold especially for cases that witness the deletion of dimensions; definitions of such measures remain open to discussion. Extant treatments of conceptual dynamics that project, or transform, conceptual spaces by and large do so through *synchronic* (see, *e.g.*, Raubal, 2004; Kaipainen, Hautamäki, 2015) rather than diachronic variation. The above types of changes, the definition of a domain, as well as the metric of a dimension should presumably feature as building blocks of such measures.

3. Conclusion

The frame model is presumably applicable whenever the conceptual spaces model is. But the latter model gains in applicability to concepts that arise from, and give rise to, exact measurement. Having reviewed the development of feature lists into the frame model, and how to recover frames by conceptual spaces, the latter model's advantages were seen to be gained by making key notions of modern measurement theory native to it. Attempts at achieving as much in the frame model, presumably, would lead to something looking very much like the conceptual space model.

Correspondences between frames and conceptual spaces were pointed out, and it was suggested that using one or the other model may be a mere matter of convenience. For the representation of taxonomic knowledge, in particular, conceptual spaces may well appear over-complex. When addressing the question whether taxonomic items have become more (or less) *similar* through taxonomical change, one should admit to implicitly presupposing the conceptual spaces model. After all, as we saw, neither frames nor feature lists yield a readily meaningful notion of similarity as geometric distance.

By questioning the assumption that the rationality of a scientific change is inherently a symbolic matter, *i.e.*, has to be demonstrated in symbols, moreover, we have sought to support the claim that conceptual spaces can provide a model for scientific conceptual change across disciplines. Hence, the assumption that a conceptual space is *not* an intrinsically symbolic model remains indispensable.

Acknowledgements

This is an extended version of Zenker (2014), originally presented at CTF 2009, University of Düsseldorf, Germany. I would like to thank the audience at this event and the volume's editor, Aleksander Gemel. Research was funded by the Swedish Research Council (VR).

References

Adams, B., Raubal, M. (2009). A metric conceptual space algebra. In: S. Hornsby, K.C. Claramunt, M. Denis, G. Ligozat (eds). *Spatial Information Theory* (Proceedings of the 9th International Conference, COSIT 2009, in Aber Wrac'h, France, September 2009), Lecture Notes in Computer Science 5756 (pp. 51–68). Berlin: Springer.

Aisbett, J., Gibbon, G. (2001). A general formulation of conceptual spaces as a meso level representation. *Artificial Intelligence* 133: 189–232.

Andersen, H., Nersessian N. (2000). Nomic concepts, frames, and conceptual change. *Philosophy of Science* 67: S224–S241.

Andersen, H., Barker, P., Chen, X. (1996). Kuhn's mature philosophy of science and cognitive psychology. *Philosophical Psychology* 9: 347–364.

Andersen, H., Barker, P., Chen, X. (2006). *The cognitive structure of scientific revolutions*. Cambridge: Cambridge University Press.

Barker, P., Chen, X., Andersen, H. (2003). Kuhn on concepts and categorization. In: T. Nickles (ed.). *Thomas Kuhn* (pp. 212–245). Cambridge: Cambridge University Press.

Barsalou, L. (1992). Frames, concepts, and conceptual fields. In: A.Lehrer, E.F. Kittay (eds). *Frames, Fields, and Contrast* (pp. 21–74). Hillsdale, NJ: Erlbaum.

Barsalou, L., Hale, C.R. (1993). Components of conceptual representation: From feature list to recursive frames. In: I. Van Mechelen, J. Hampton, R. Michalski, P. Theuns (eds). *Categories and Concepts: Theoretical Views and Inductive Data Analysis* (pp. 97–144). San Diego, CA: Academic Press.

Batterman, R.W. (2003). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. New York, NY: Oxford University Press.

Berka, K. (1983). *Measurement: Its Concepts, Theories, Problems*. Boston Studies in the Philosophy of Science, vol. 72. Dordrecht: Reidel.

Bird, A. (2012). What can cognitive science tell us about scientific revolutions? *Theoria* 75: 293–321.

Botteril, G. (2007). Review of H. Andersen, P. Barker, and X. Chen (2006). The cognitive structure of scientific revolutions. *Notre Dame Philosophical Reviews*, http://ndpr.nd.edu/news/25236-the-cognitive-structure-of-scientific-revolutions (accessed: 05.2014).

Chen, X. (1997). Thomas Kuhn's Latest Notion of Incommensurability. *Journal for General Philosophy of Science* 28: 257−273.

Chen, X. (2002). The 'platforms' for comparing incommensurable taxonomies: A cognitive-historical analysis. *Journal for General Philosophy of Science* 33 (1): 1–22.

Chen, X. (2003). Object and event concepts: A cognitive mechanism of incommensurability. *Philosophy of Science* 70: 962–974.

Chen, X. (2005). Transforming temporal knowledge: Conceptual change between event concepts. *Perspectives on Science* 13: 49–73.

Chen, X., Barker, P. (2000). Continuity through revolutions: A frame-based account of conceptual change during scientific revolutions. *Philosophy of Science* 67: S208–S223.

Chen, X., Andersen, H., Barker, P. (1998). Kuhn's Theory of Scientific Revolutions and Cognitive Psychology. *Philosophical Psychology* 11: 5–28.

Decock, L. (2006). A physicalist reinterpretation of 'phenomenal' spaces. *Phenomenology and the Cognitive Sciences* 5: 197–225.

Diez, J.A. (1997a). A hundred years of numbers. A historical introduction to measurement theory 1887–1990. *Studies in History and Philosophy of Science* 28: 167–185.

Diez, J.A. (1997b). A hundred years of numbers. A historical introduction to measurement theory 1887–1990. *Studies in History and Philosophy of Science* 28: 237–265.

Dorato, M. (2005). Why are most laws of nature mathematical? In: J. Faye, P. Needham, U. Scheffler, M. Urchs (eds). *Nature's Principles* (pp. 55−75). Amsterdam: Springer.

Friedman, M. (2001). *Dynamics of Reason.* Stanford, CA: CSLI Publications.

Gärdenfors, P. (2000). *Conceptual Spaces*. *The Geometry of Thought*. Cambridge, MA: MIT Press.

Gärdenfors, P., Zenker, F. (2010). Using conceptual spaces to model the dynamics of empirical theories. In: E.J. Olsson, S. Enqvist (eds.). *Belief Revision meets Philosophy of Science* (pp. 137–153). Berlin: Springer.

Gärdenfors, P., Zenker, F. (2013). Theory change as dimensional change: Conceptual spaces applied to the dynamics of empirical theories. *Synthese* 190: 1039–1058.

Gauker, C. (2007). A critique of the similarity space theory of concepts. *Mind and Language* 22 (4): 317–345.

Hand, D.J. (2004). *Measurement Theory and Practice*. London: Arnold.

Helmholtz, H.V. (1887). Zahlen und Messen erkenntnistheoretisch betrachtet. In: H.V. Helmholtz (1920). *Schriften zur Erkenntnistheorie* (pp. 70–109). Berlin: Springer (transl. in: H.V. Helmholtz (1930). *Counting and Measuring* (pp. 72–114). Princeton, NJ: Van Noestrand).

Hoyningen-Huene, P. (1993). *Reconstructing Scientific Revolutions. Thomas S. Kuhn's Philosophy of Science*. Chicago: University of Chicago Press.

Howard, D. (2009). "Let me briefly indicate why I do not find this standpoint natural". Einstein, General Relativity and the Contingent Apriori. In: M. Domski, M. Dickson (eds).

Discourse on a New Method. Reinvigorating the Marriage of History and Philosophy of Science (pp. 333–355). Chicago and La Salle, ILL: Open Court.

Jaeger, G. (2010). Natural color categorization are convex sets. *Lecture Notes in Computer Science* 6042: 11−20.

Kaipainen, M., Hautamaki, A. (2015). A perspectivist approach to conceptual spaces. In: F. Zenker, P. Gärdenfors (eds.). *Applications of Geometric Knowledge Representation*, Synthese Library, vol. X. Dordrecht: Springer.

Krantz, D.H., Luce, R.D., Suppes, P., Tversky, A. (1971, 1989, 1990). *Foundations of Measurement*, vol. I–III. New York: Academic Press.

Kuhn, T.S. (1970). *The Structure of Scientific Revolutions*. Chicago, ILL: University of Chicago Press.

Kuhn, T.S. (1983). Commensurability, comparability, and communicability. In: P. Asquith, T. Nickles (eds.). *PSA 1982*, vol. II (pp. 669–688). East Lansing, MI: Philosophy of Science Association.

Kuhn, T.S. (1993). Afterwords. In: P. Horwich (ed.). *World Changes: Thomas Kuhn and the Nature of Science* (pp. 311–341). Cambridge, MA: MIT Press.

Kuukkanen, J.-M. (2006). *Meaning Change in the Context of Thomas S. Kuhn's Philosophy*. Dissertation, University of Edinburgh, http://hdl.handle.net/1842/1259 (accessed: 07.2011).

Labov, W. (1973). The boundaries of words and their meanings. In: C.J. Bailey, W.W. Shuy (eds.). *New Ways of Analyzing Variation in English* (pp. 340–373). Washington, DC: Georgetown University Press.

McCann, H.G. (1978). *Chemistry Transformed: The Paradigm Shift from Phlogiston to Oxygen*. Norwood, NJ: Ablex Publishing Corporation.

Minsky, M. (1975). A framework for representing knowledge. In: P. Winston (ed.). *The Psychology of Computer Vision* (pp. 211–277). New York, NY: McGraw-Hill.

Munsell, A.H. (1915). *Atlas of the Munsell Color System*. Boston, MA: Wadsworth-Howland.

Nersessian, N.J. (1995). Opening the black box: Cognitive science and history of science. *Osiris* 10: 195–211.

Raubal, M. (2004). Formalizing conceptual spaces. In: A. Varzi, L. Vieu (eds). *Formal Ontology in Information Systems*. *Proceedings of the Third International Conference* (FOIS 2004), vol. 114 (pp. 153–164). Amsterdam: IOS Press.

Rosch, E., Merris, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology* 8, 382–439.

Stanford, P.K. (2008). Review of H. Andersen, P. Barker, and X. Chen (2006). The cognitive structure of scientific revolutions. *British Journal for the History of Science* 41: 116–117.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science* 103: 677–680.

Thagard, P. (1999). *How Scientists Explain Disease*. Princeton, NJ: Princeton University Press.

Thagard, P. (2009). Review of H. Andersen, P. Barker, and X. Chen (2006). The cognitive structure of scientific revolutions. *British Journal for the Philosophy of Science* 60: 843–847.

Taylor, J.R. (2003). *Linguistic Categorization*. Oxford: Oxford University Press.

Tversky, B., Hemenway K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General* 113: 169−193.

Worrall, J. (1989). Structural realism. *Dialectica* 43: 99−124.

Zenker, F. (2014). From feature lists via frames to spaces. In: T. Gamerschlag, D. Gerland, R. Osswald, W. Petersen (eds.). *Frames and Concept Types: Applications in Language and Philosophy* (pp. 69−89). Dordrecht: Springer.

Zenker, F., Gärdenfors, P. (2014). Modeling diachronic changes in structuralism and in conceptual spaces. *Erkenntnis* 79 (8): 1547–1561.

Zenker, F, Gärdenfors, P. (2015). Communication, rationality, and conceptual changes in scientific theories. In: F. Zenker, P. Gärdenfors, (eds.). *Applications of Geometric Knowledge Representation*, Synthese Library, vol. X. Dordrecht: Springer.

Zenker, F., Petersen, G. (2014). From Euler to Navier-Stokes: A spatial analysis of conceptual changes in 19th-century fluid dynamics. *International Studies in the Philosophy of Science* 28 (3): 1–19.

ALEKSANDER GEMEL
PAULA QUINON

THE APPROXIMATE NUMBERS SYSTEM AND THE TREATMENT OF VAGUENESS IN CONCEPTUAL SPACES[1]

1. Introduction

The theory of conceptual spaces is intended to provide a framework for models of both symbolic and non-symbolic representations of knowledge and information (Gärdenfors 2000). As such, it seems to us to be very clearly suited to modeling pre-verbal representations belonging to a so-called core cognition, whose existence is postulated by cognitive developmental psychologists (Feigenson *et al.*, 2004; Carey, 2009). In this paper we propose a treatment of representations of quantity that embraces both symbol-based and pre-verbal numerical concepts.

The representations we aim to study are related to the Approximate Number System (ANS). Cognitive psychologists claim that humans share with animals an abstract sense of quantity: they have a so-called "number sense". To "number sense" amounts two core systems of representations, which get activated by different core mechanisms: the ANS is one of these mechanisms (Dehaene 1997, 2008; Gallistel 1993; Feigenson *et al.*, 2004; Carey, Sarnecka, 2006).

The ANS is a core system in the sense that it is present in human apprehension of quantities before verbal conceptual these of quantities appears. But, it is

---

[1] The core idea of this paper was first formulated in a private conversation of the second author (the authors' names are listed in alphabetical order) with Jakub Szymanik in the summer 2010. The research of the second author has been supported by the IEF FP7 Marie Curie Fellowship "Numbers" (PIEF-GA-2011-301470). We wish to express our gratitude to Peter Gärdenfors for comments on an earlier version of this paper.

claimed, it also underlies symbolic number-concept creation and it is active in the processing of quantitative information for the entire life of an individual. Firstly it reacts to nonverbal input and this remains the case through the whole lifespan (Halberda, Feigenson, 2008), it also gets activated through symbolic input. To exemplify the functioning of the ANS and the ANS-related representations, let us consider the following case: the reaction to a set of 17 elements corresponds to the seventeenth element of the ANS-sequence, *i.e.* a set containing 17 with some specific neighboring area.

It is not our objective to postulate anything about the relation between *exact* number names, *i.e.* numerals, and representations related to the ANS (Halberda, Feigneson, 2008; Mussolin *et al.*, 2012; Wagner, Johnson, 2011; see also Nagen, Sarnecka, 2015). Instead, we observe that numerals are in certain contexts interpreted as vague terms (Krifka, 2009; Solt, 2011). It might seem surprising, as at first glance there is nothing less vague than a number. “Seven” refers to seven; “twelve dwarfs” are twelve entities of the type “dwarf”, etc. Naturally, there are some numerical expressions that embrace vagueness, like expressions consisting of a numeral and a modifier (*e.g.*, “roughly ten”, “more or less a hundred”, or “about twenty”) or generalized numerical quantifiers (*e.g.*, “more” or “little”, when applied to discrete sets of entities, etc.); but vagueness is usually not attributed to exact number names (*i.e.* numerals). Nevertheless, when the use of language is studied more carefully, one can see that in some contexts numerals behave like vague terms. This is the case, for example, in expressions like “there are 100 people in the room” or “two thousand people participated in the demonstration”, in which case the intention of the speaker is to approximate the number of people (Dehaene, 1997, 2008).

Our aim in this paper is to conduct a preliminary study of a possible correlation between three concepts: exact numerals, in particular when used as vague quantifiers, ANS-related representations as a (naturalized) semantics for such expressions, and a model of the ANS-related representations in the framework of conceptual spaces. Our study amounts to an overall understanding of the structure of the ANS.

It should be clear from the beginning that we work under several arbitrary assumptions. First, we work in a paradigm in which “number sense” exists. Second, we are primarily interested in theoretical investigation and no experimental study has been proposed to verify whether there really exists a correlation between ANS-related representations and symbolic representations of exact numerals used as vague quantifiers. Conducting such a study would be very valuable not only for the current endeavour; it would also shed additional light on the nature of the ANS.

2. Two core systems of numerical representation and exact numerals

According to a paradigm defended by many cognitive scientists (Spelke, *et al.*, 2007; Carey, 2009), at the basis of human numerical abilities lay two innate core cognitive systems responsible for pre-verbal representations of quantitative content, which in some way support interpretation of numerical expressions when the symbolic concept of a number is created: *the system of parallel individuation* (which allows identification in an exact way of sets up to four elements); and *the system of approximate numbers* (which allows approximation of cardinality of sets bigger than five elements) (Feigenson *et al.*, 2004).

2.1. Approximate Number System

The system of approximate numbers (ANS) is a cognitive system that provides a mental representation of the approximate quantity of elements in the set. The size of the ANS-related representation is proportional to the number of perceived units and for this reason ANS-related representations are often referred to as analogue magnitude representations (ANRs). The analogue characteristic of the ANS means that all input provided by the system are represented by continuous data. In other words, a mental representation of a set consisting of 20 elements will be about twice as large as the mental representation of a set consisting of 10 elements.

The ANS-related representations work as an interconnected whole: an approximation of the cardinality of a given set is possible because it is put into correspondence with other representations. All estimations of size happen through comparison to the rest of elements. Comparisons of sizes of multiple sets are based on the function of their ratio. In other words, the distinction between a 10-element set and a 20-element set is as difficult to make as the distinction between 20- and 40-element sets, since in both cases the ratio of the amount of elements in the sets is 1:2. The straightforward consequence of this observation is that the difficulty of discriminating between two sets is not dependent on absolute difference between their cardinalities. This seems quite intuitive when we analyze a case in which the absolute difference is the same. Let us take two pairs of sets: the first pair consisting of a 10-element set and a 20-element set, and the other consisting of a 200-element set and a 210-element set. Although the absolute difference between the amounts of the elements of two sets is 10, it is intuitively correct, and also experimentally proven, that it is much easier to distinguish between the first pair of sets than the second.

Two effects describe behaviour of the ANS:

1) the *size effect*, when the absolute difference between the two pairs of sets is constant, then sets with a smaller number of elements are easier to discriminate. This relation is illustrated by the example described above;

2) the *distance effect*, according to which it is easier to discriminate sets that differ by a significantly large number of elements. For instance, it is easier to notice the difference between 20 and 35 elements in the set then between 20 and 23 elements in the set.

2.2. Parallel Individuation System

The Parallel Individuation System (PIS) is not a quantity-specific, since it is not directly intended to represent numbers. Instead it is intended to create and sustain, in working memory, mental models consisting of a small number of objects – such as things, sounds, events, *etc*. The small number of individuals constituting the mental representation is a result of human cognitive limitations: in particular, the capacity of the working memory is usually limited to 4 objects at a time. Unlike the ANS, which is responsible for analogue representations of approximate cardinalities of sets as continuous wholes, the PIS is responsible for the representation of separate individuals. That is, the parallel individuation system provides no common symbol for “two” or “three” elements, but offers two or three separate symbols for each element respectively (each of the separate symbols may also carry information concerning its properties, or specifying its type).

2.3. Linguistic Number System

None of the presented above innate cognitive systems can provide representations of the exact-number concept. The PIS provides representations only for up to four elements, and is not quantity-specific. The ANS accounts for approximate values. Cognitive scientists claim that exact numbers representations appear in consequence of functioning of the symbolic cognitive system: the Linguistic Number System (LNS). This system consists of numerals and also of various types of quantitative quantifiers. Symbols get their interpretation through the core cognitive systems, but also through interactions with other concepts in the process called conceptual “bootstrapping”.

3. The ANS and vagueness

An inherent feature of the ANS is the approximate or vague character of both the inputs it processes and the representations these inputs activate. Numeral expressions that first come to mind in relation to the ANS are generalized quantifiers such as “many”, “little”, “more”, “less”, or “equally”; or numerical quantifiers used with a modifiers such as “roughly 20” or “about 1000”. In some contexts exact numerals are also used to approximate quantities. This paper is principally devoted to the third case. A prototypical example of use of an exact numeral to approximate a quantity is, for instance, when one says “this morning there were 100 people in the lecture hall” in order to estimate the number of people participating in some gathering.

In this paper we assume that the correct semantics for exact natural numbers, especially when used as vague quantifiers, is the ANS-related system of representations. It seems plausible since, as it is claimed by cognitive scientists, it is exactly this system of representations that gets activated by any quantitative input asserted without counting elements. This means that exact numbers used in a specific context and for a specific purpose (*i.e.* to estimate cardinalities bigger than four) may be considered vague terms. For instance, when one says “there are 100 people in the room” or “there are 1000 leaves on the tree”, or “4000 people participated in the event”, the objective of the speaker clearly does not consist in providing the exact cardinality of the set, but just in approximating it.

3.1. Vagueness

Traditionally, the concept of vagueness relates to terms. A vague term is a term that has borderline cases of interpretation: when we say that the term "red" is vague it means that there are cases of redness that are difficult for a human to classify as red. In other words the semantics for vague terms is characterized by the existence of so-called borderline cases. According to Shapiro, who adopts the definition given by McGee and McLaughlin, an "object *a* is the borderline case of predicate *F*, if *Fa* is 'unsettled' *i.e.*, if *a* is not determinately an *F*, nor is *a* determinately non-*F*" (Shapiro, 2006: 7). The technical term "determinately", used in the above definition, was introduce by McGee and McLaughlin as follows: "to say that an object *a* is determinately an *F* means that the thoughts and practices of speakers of the language determine conditions of application for *F*, and the facts about *a* determine that these conditions are met" (McGee, McLaughlin, 1994).

Besides vague terms related to a quality (ex. bald, old), there are numerical quantifiers that are vague ("many", "little", "more", "less", "about 100"). Moreover, and this is what we study in this paper, even exact numerals, when used in specific contexts, can also be considered vague terms. Strictly speaking we are interested in understanding semantics underlying exact numerals when these are used to estimate cardinalities bigger than four, and behave as vague expressions.

As we mentioned above, a term is vague if it has borderline cases of interpretation, and some object can be called a borderline case of some predicate if the thoughts and practices of speakers of the language don't determine the perfect conditions of application for this predicate, and if facts about this object do not determine whether these conditions are met. Thus, in the context of the relationship between the ANS and vagueness, the question is whether the phrase "one hundred", used to estimate the approximate number of leaves on the tree, has borderline cases of interpretation. It seems obvious that the answer to this question is yes, since there are senses of "100" that are difficult for a human to classify as 100 when she is using the ANS. For instance, if someone asks a person to estimate the number of leaves on two trees, consisting of 100 leaves on the first tree and 103 leaves on the second, the answer would be probably "100" for both. The numeral "100" is used as a vague quantifier and it is so independently of speaker's intentions (whether she aimed at providing precise information or whether she just wanted to approximate).

Observe that since the representations provided by the ANS are vague, the size of the representation of the number 3 can't be *exactly* half as long as the representation of number 6. Moreover, the vague nature of the ANS means that the boundaries of representations are blurred, which entails the possibility of confusion between two adjacent numerical representations. This phenomenon is in line with the characteristics of ANS size and distance effects. As demonstrated by numerous studies of the ANS, these effects are reflected in the difficulty of distinguishing, for example, 20- and 21-element sets from each other. Therefore, the first difficulty in modeling an ANS-related representation system in conceptual space consists in the necessity of considering the blurred nature of the boundaries of these representations. As we discussed above, the effect of blurred boundaries on the conceptual spaces model is specific to models of vague terms, where one uses clusters of prototypes.

3.2. Spatial arrangement of ANS numerical representations

Due to the analogue nature of the ANS, its numerical representations do reflect quantitative intuitions in such a way that they are best described with spatial categories. The type of representations that ANS generates is called analogue magnitude representations (AMRs).[2] They are analogue because the neural entities that they activate are direct analogues of the magnitudes they represent. The particular type of AMRs that we are modeling activates in reaction to discrete quantitative magnitudes.

In the construction of the model, we use the idea that ANS-related representations are best characterized by spatial dimensions (Dehaene 1997, 2008; Carey, Sarnecka 2006). Thus the natural tool for modeling the representational structure of ANS-magnitudes appears to be the geometric representation model proposed by Peter Gärdenfors, known as the theory of conceptual spaces (Gärdenfors 2000).

4. Conceptual spaces

The main thesis of the theory of conceptual spaces developed by Gärdenfors is that the meanings of words can be represented as an organized spatial geometrical structure. Such a structure supports representation of many words

---

[2] Beck (2014) provides a philosophical introduction to AMRs.

simultaneously and provides a geometrical model for language representation. Concepts are represented as regions containing points in a one-dimensional or multidimensional space. Objects belonging to the extension of reference of the term being modeled are mapped onto these points. There is a metric (or a distance function) within the spaces. The distance between points expressed by the metric corresponds to some kind of similarity between the qualities of the points.

In the most general terms a metric is defined in the following way:

Definition:[3] Consider a space S. A function f defined on a space S is a metric for S *iff* for all points $a, b, c \in S$:

(i) $f(a, b) \geq 0$, with $f(a, b) = 0$ iff $a = b$;
(ii) $f(a, b) = f(b, a)$;
(iii) $f(a, b) \leq f(a, c) + f(c, b)$.

The metric is used to measure a degree of similarity between two points in the conceptual space. Quality dimensions, which can be taken into account, are of different types. They can be one of the three ordinary spatial dimensions – height, width, and depth – or some other physically measurable quality such as temperature, weight, brightness, or pitch. Finally they can have an abstract non-sensory character. The closer the two points are to one another, the more similar they are: if point x is closer to point y than to point z, then x is more similar to y than to z. In other words, the closer that representations of two points are placed in the space, the more similar are the objects represented by these points. Similarity can hence be defined as a monotonically decreasing function of the distance expressed by a metric, defined on the space.

Finally, concepts are represented as regions of conceptual space (sets of points). For example, *red* is a certain region in color space. But there is an important condition: not just any set of points in a space is a region of conceptual space in natural language, but only those regions that tend to have a specific topological feature. Namely, the concepts and properties have to correspond to convex regions in the given space. Convexity is defined as follows:

Definition: A region R is convex *iff* for any two points x and y in R, all points between x and y are also in R.

---

[3] This very general definition of metric is satisfied by many different metrics used in mathematics, for instance by the Euclidean metric. For our current endeavour particular differences between metrics are unimportant and hence will be ignored.

This means that, as long as we consider a domain containing objects *x* and *y* as having some property, any object that is located between *x* and *y* will also have this property. For a convex region, one can describe positions within it as being more or less central.

Decisions about which points belong to one region are made on the basis of similarity to a distinguished point representing the prototypical example of the reference. One of the hypotheses of the theory of conceptual spaces is that (for many predicates) there exist prototypical examples of their reference (the prototype effect) (Gärdenfors 2000).

Mathematical counterparts of conceptual spaces are Voronoi diagrams (or tessellations), which exemplify a technique of dividing metrical space into cells. Every cell has a center, called a seed or generator, and contains all and only those points that lie closer to the given seed than to any other. Seeds model prototypes. In other words, a set of prototypes $P = \{p_1, \ldots, p_n\}$ generates a set of convex regions $C = \{c_1, \ldots, c_n\}$. Figure 1 is an example of this:

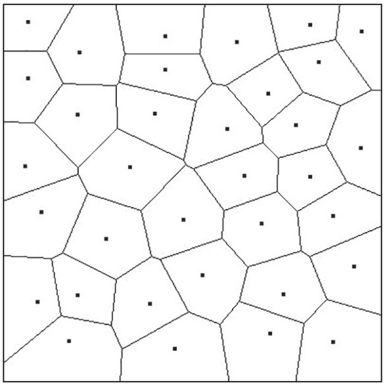

Fig. 1. A Voronoi tessellation

In the case of certain terms it can happen that it is more adequate to consider a set P* of clusters of points $\{\{p1_1, \ldots, p1_n\},\{p2_1, \ldots, p2_n\}, \ldots ,\{pn_1, \ldots, pn_n\}\}$ instead of a set P of points. As claimed by Decock, Douven (2014), see also Douven *et al.* (2013), this would be the case for vague concepts. Decock, Douven (2014) use the standard example of colors as requiring this cluster modeling: color names are vague, and modeling conceptual space with clusters of prototypes is specific to vague concepts. In other words, the meaning of some terms is more adequately modeled when one considers the division of the corresponding conceptual space as generated not by single points but by groups of points corresponding to prototypical representations of meaning.[4]

It is traditionally assumed that when the partition of a space is generated by a set P* of clusters of points, these clusters form circles. That is, prototypical areas are circular, as shown in fig. 2.

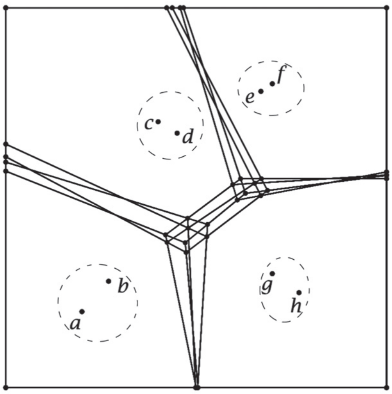

Fig. 2. Voronoi Diagrams generated by P*

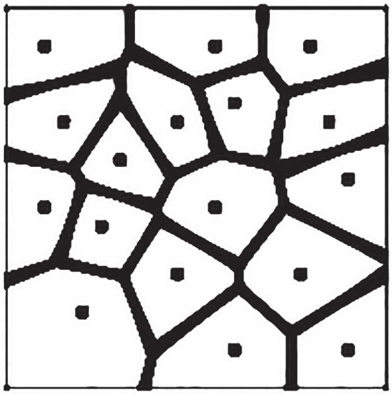

Fig. 3. Use of P* to model vague concepts (Decock, Douven, 2014)

The tessellation made by the cluster of points formed in a circle (so-called vague prototypes), allows blurring of boundaries between concepts in the space.

---

[4] How to think about diagrams in this setting is presented in Gärdenfors and Williams (2001). See in addition Gärdenfors (2000).

Every point lying on the boundary of two or more terms is considered a borderline case of these predicates. The tessellation made by vague prototypes is presented in fig. 3.

4.1. The ANS in the conceptual spaces framework

Our central endeavor in this paper is to define a conceptual space that models a structure of ANRs related to the ANS. This system of ANRs is activated by discrete quantitative inputs. It is magnitude sensitive; that is, the bigger the quantity of the input, the bigger the magnitude represented in the system.[5] The boundaries between elements of the ANS are not sharp: it is not clear whether 39 belongs to the scope of 45, or whether this scope stops at 40. Moreover, the analogue magnitude representations obey the so-called Weber's Law: "the ability to discriminate two magnitudes is determined by their ratio. As the ratio of two magnitudes approaches 1:1 they become harder to discriminate, and beyond a certain threshold determined by the subject's 'Weber constant' cannot be discriminated at all" (Beck, 2014). On the neural level, a ratio 1:1 of two magnitudes in a noisy system makes it almost impossible to discriminate the neural entities. This effect of ratio-sensitivity occurs every time human adults have to discriminate magnitude when explicit counting is not possible (Barth *et. al.*, 2003).

In this paper we target a particular type of AMRs' system. Therefore, there are multiple general aspects of such systems that we shall ignore in the proposed model. For example, there are various AMRs that are activated through different input – which underlies important differences, and gives rise to essentially different models, even though these AMRs share multiple common features and traits of functioning. Moreover, even the system of AMRs related to quantities is activated by input of various types (visual, aural, kinetic, or tactile). In this paper we want to avoid discussing the common nature of AMRs and the extent to which various inputs that activate ANS-related representations differ across types of AMR systems. We focus on AMRs that are activated in response to constant visual input consisting of various sets composed of different quantities of identical blue dots evenly distributed on a whiteboard. Even if the targeted objective of our research is to understand semantics for exact numerals, we will model the

---

[5] It is shown that analogue magnitudes related to the ANS get bigger in logarithmic progression (Dehaene, 2008; Berteletti *et al.*, 2010).

non-conceptual content of ANS-related representations, which occurs in this situation automatically before any other type of representation supporting quantitative information. Exact numerals and sensory quantitative, discrete input are claimed to activate the same system of representations (the ANS-related system of AMRs), so starting by proposing a sketch of a model for a simple sensory input seems to be a natural step to take.

The discussion of how exactly analogue magnitude representations are instantiated in the brain (are larger quantities represented by more neurons' firings or by faster firing of a fixed population of neurons?) does not fall within the scope of this paper. What is important for us is that "their psychophysical signatures strongly suggest analogue type of representational scheme" (Carey, 2009: 458). One simple way of thinking about this kind of representation is proposed by Carey and Sarnecka (2006: 477) and Carey (2009: 118): "[there exists] a helpful analogy to the following external system of analogue number representations. [...] Line length is a direct analogue of number. [...] Suppose our brains deploy magnitude representations that are likewise analogue". Note that we do not take a position in the debate about how representations really look – that is irrelevant to our endeavor.

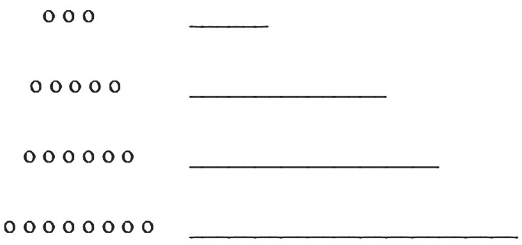

Fig. 4. The representation of quantities in the ANS

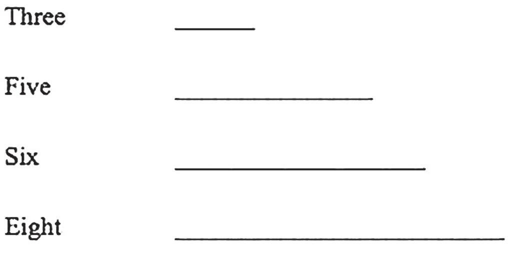

Fig. 5. Representation of numbers in the ANS

The authors' original example focuses on symbolic input (fig. 5).

Indeed, it is claimed that when an agent acquires names for numbers, her neural reaction and activation of ANS-related representations will also take place when input is symbolic. Observe that the point we are making inscribes this line of reasoning in an even more straightforward way: our aim is to provide a model for vague-exact quantifiers.

We do not take up this challenge here, but it seems to us that there is a natural way of generalizing our model to other analogue magnitude representation systems. As put by Carey and Sarnecka (2006: 477),

> [n]umber is not the only dimension of experience represented by analogue magnitudes – other examples include brightness, loudness, and temporal duration. In each case, as the physical magnitudes get bigger, it becomes increasingly harder to discriminate values that are the same absolute distance apart. [...] [T]he discriminability of any two values is a function of their ratio [...].

4.2. The ANS and prototypicality

According to prototype theory, the cognitive process of categorization has a graded nature, *i.e.* some members of a category are better, more straightforward, or more central representatives of this category than other members (Rosch, 1973; Lakoff, 1987; Taylor, 2003). There exists serious empirical evidence confirming the prototype effect in human cognitive processes. One of the most famous series of experiments is Brent Berlin and Paul Kay's study of color categorization (Berlin, Kay, 1969). In the first series, respondents (competent users of different languages) were asked to divide a color continuum according to system of color categories functioning in their language. The results showed that ranges of color categories differ considerably from one language to the next. In a second series of experiments, subjects were asked to point out the best copy of a color among over three hundred color plates. Regardless of the different ranges of categories in their languages and their linguistic area, all participants pointed to the same color, *i.e.* same shade of, for instance, red. The results of Berlin and Kay's study show that even if languages categorize colors in different ways, for every color there exists a prototypical example, in common and similar across all languages.

As we said above, the ANS-related AMRs system represents the numerosity of a set by an analogue extension. Remember that in this paper we concentrate on a specific type of input, that is, we focus on AMRs that are activated in response to constant visual input consisting of various sets, which are composed of different quantities of identical blue dots distributed on a whiteboard in such a way that every dot is not further from some other dot than its diameter. For the moment we assume that dots can be organized in different shapes (spatially ordered or randomly distributed). Later, for the sake of model simplification, we will consider only these shapes, which are representable in the Cartesian coordinate system.

We argue that certain types of distribution of visual input are cognitively privileged. They activate ANS-related AMRs in the same way as the exact symbolic numerals do. This provides an opportunity to incorporate the phenomenon of prototypicality for model of ANS-related representations in the conceptual spaces framework. The distinctiveness of distribution can be studied, as we do it in this paper, from theoretical perspective, but we agree that the most significant results might be achieved through empirical investigation. Simply speaking, it is a plausible hypothesis that some shapes may be psychologically prototypical cases. In other words, in analogy to Berlin and Kay's color categorization experiment, some arrangements of blue dots on a board may be considered prototypical or more salient instances of a given cardinality than others.

Hypotheses presented in this paper could be tested empirically in experiments inspired by Berlin and Kay's work. One of our ambitions is to prepare theoretical and conceptual frameworks for such studies. To begin with, let us consider a two-dimensional conceptual space – with height and length as quality dimensions – in which we consider spatial arrangements of identical blue dots evenly distributed on a whiteboard. According to Berlin and Kay's result for color-predicates, there exists a language-independent set of prototypes. Subjects from the study converged on some unified set of examples when asked "what is a prototypical red?" Following Berlin and Kay, we offer the hypothesis that some of these arrangements will be more easily recognized by the subjects (competent users of the given language) as corresponding to a given numeral. The situation we want to describe is the following: subjects look at different quantities of identical dots. These dots can be arranged in different ways. The following figures contain just 4 dots, but this can be generalized to greater quantities. Moreover, as we said above, in the model construction we consider only spatially ordered shapes.[6]

---

[6] We are aware of limitations that this constraint entails and of the necessity of adjusting the proposed model to empirical results on prototypically.

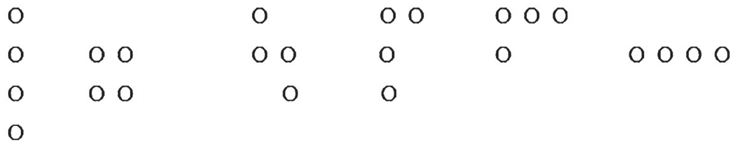

Fig. 6. Examples of arrangements of the four-element sets of dots in two-dimensional space

Similarly to the fig. 6, the extended, continuous, and approximate representations of the ANS also can be distinguished in terms of the space occupied by their shape:

Fig. 7. The arrangements of shapes of 4-element sets of dots

Each of these possible arrangements varies according to height and length parameters and as such, they can be straightforwardly instantiated in conceptual spaces as ANS-related AMRs for "four". The hypothesis is the following: as in the case of colors, it seems pretty plausible that humans converge to some unified set of prototypes. Before an empirical study is conducted and a set of distribution recognized as prototypes psychologically grounded, we propose that in our model the role of the prototype is played by dots which are the most evenly distributed.[7] The rest of the configurations are, in the proposed model, ordered following the degree of similarity to the prototype or prototypes.

To support our hypothesis that evenly distributed dots are more prototypical, one can refer to Tversky's studies provided in „Features of similarity" (Tversky, 1977). In two-dimensional conceptual space (height/length), the element of equal dimension ratio 1:1 is certainly a natural candidate for a prototype. The choice of an equilateral shape representation for the prototype is supported by Tversky's studies devoted to determining the degree of prominence of geometric

---

[7] Again, it is just a simplification necessary for initiation discussion on conceptual spaces models for ANS-related representations. We use "evenly" with an undefined meaning, but we are aware that more exact definition is necessary.

figures (Tversky, 1977).[8] As Tversky puts it: „A major determinant of the salience of the geometric figures is goodness of form. Thus, a ‘good figure’ is likely to be more salient than a ‘bad figure’, although the latter is generally more complex” (*ibid.*: 334). Generally the more prominent or salient figure, the more regular and symmetric shape it has. The example of good and bad in form figures, given by Tversky, is presented on following figure:

Fig. 8. More salient figure (left), and less salient one (right) (Tversky, 1977: 334)

For the sake of model construction, we propose to order the rest of the, configurations of 4-element sets of dots (mentioned above), following the degree of similarity to the prototype or prototypes. In other worlds the location of the other elements of the set is then determined by the degree of similarity to the prototype with regard to one of the dimensions, *i.e.* length and height. This is shown in the following figure:

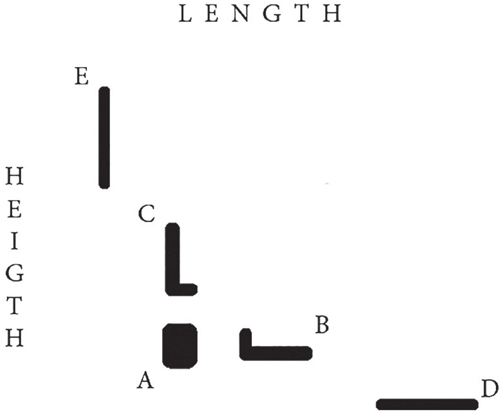

Fig. 9. Example of an ordered set of the arrangements of 4-element sets

---

[8] It can be argued that dots distributed in straight lines are in reality prototypical as well. In this paper we will not investigate this possibility.

Element marked as A represents the prototype, the element marked as B differs from the prototype with regard of length dimension by 1 point on similarity scale. The element C also differs from the prototype by 1 on the similarity scale, but with regard of height dimension. D and E differ from prototype by 2 points with regard of length and height dimension respectively.

In the similar way we can model the sets of ANS-related representations of other magnitudes. Models for bigger numbers will differ from the above, because these representations take into their scope neighboring cardinalities. So, for example, some inputs that activate the representation of 17 also activate some (non prototypical) representations of 16 and 18. The number of representations taken into account by the AMR for each number grows with magnitudes of elements from the number line.

The application of Voronoi tessellation to the conceptual spaces model allows us to map the entire conceptual representation structure of size provided by the ANS. The set of generators of the tessellation will be constituted by the collection of prototypes of each subsequent quantity represented by ANS magnitudes (*i.e.* prototypes of 4, 5, 6, and so on). Therefore, if we treat points as representations of successive prototypes of numeric quantities, we will get the following picture of the conceptual structure of the ANS.

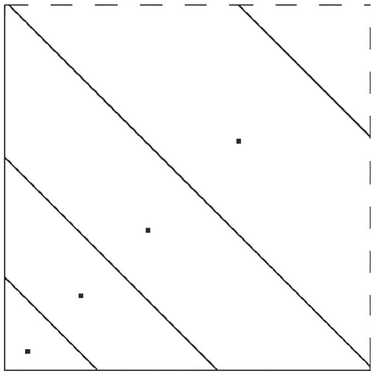

Fig. 10. The idealized conceptual structure of the ANS

However, the above-presented conceptual structure of ANS representations is certainly not vague. It should be noted that modeling the approximate nature of ANS-related representations involves taking into account the blurred borders in the above scheme. This objective can be achieved by using the fuzzy prototype model (the so-called prototype groups) proposed by (Douven *et al.*, 2013). In short, this model requires treating cases of objects adjacent to the prototype also as prototypical cases, and using them as generators of subsequent tessellations. As a result, we get the following diagram:

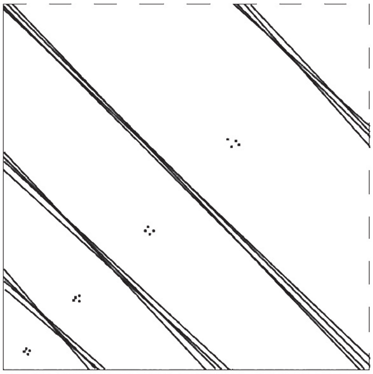

Fig. 11. The conceptual spaces structure of the ANS

Thicker boundaries of tessellation model boundary cases where the number of elements in the set might or might not activate the representation which are activated by prototypes. Figures 9 and 10 contain a preliminary visualization of the proposed model. We suspect that for larger numerical values regions of tessellation will differ from the picture above. Regions representing larger numerical values are not only contiguous upon the regions representing two closest values in numeral progression, but also are contiguous upon other proximate regions. We say that region A is contiguous upon the region B, *iff* a ratio of numerical values represented by both regions is close as possible to 1:1. The exact value of the ratio should be ascertained experimentally, but for the purposes of the model, we can venture to say that the deviation should not exceed 0.1. This rule explains why the region representing 5 in conceptual space is contiguous only upon regions denoting 4 and 6, since the ratio between the 5 and 4 and between 5 and 6 is respectively 1.25 (deviation = 0.25) and 0.83 (deviation = 0.23). While the region representing for instance 33 is contiguous upon 6 regions (*i.e.* representing 30, 31, 32, 34, 35, 36), since deviation of a ratio between 30 and other numbers is smaller than 0.1.

Understanding the contiguousness of regions in terms of the ratio of values representing by them is the consequence of the characteristic for ANS size and distance effects. According to the latter it is easier to discriminate – with respect to a given set – a set that contains more elements than a set with a similar number of elements to the comparison set. In other words, if one compares two numbers to the same target, the number that is further away from the target should be easier to discriminate. For instance, it is easier to tell the difference between 10 and 15 than between 10 and 11. Concepts arranged close to each other in conceptual space, since they have a common vague area (*i.e.* a fuzzy boundary), are much more difficult to distinguish than concepts arranged further away from each other in space.

According to the size effect, when the absolute difference between the two pairs of sets is constant, it is easier to compare sets with a lower number of elements. This effect can be quite intuitively achieved in the presented model. As a matter of fact, the size effect occurs simply due to the way that the magnitude of representation is modeled in conceptual space. This is because in the proposed model, the magnitude of the representation binds directly with the amount of variants of the objects' spatial orientations. Both values (*i.e.* magnitude of the representation and amount of variants of the objects' spatial orientations) have a direct impact on the size of representation and the number of prototypes, and as a consequence also on the size of the prototype group. The size of the prototype group has a direct bearing on the size of the penumbra – *i.e.* the range of the blurred boundary between adjacent terms in the series. In other words, the larger the sets that are compared, the easier it is to make a mistake, since the absolute difference between sets is small enough to be located within the penumbra area.

It is important to note that the proposed model of the conceptual framework of ANS-related representations is easily and intuitively reconciled with the ANS's distance and size effects. Both effects show that similarity between ANS-related representations in conceptual space is consistent with Weber's law. This means that the degree of similarity is designated by the inverse ratio of two magnitudes being compared with each other. In other words, the smaller the ratio between the magnitudes, the more similar objects are.

5. Conclusions and openings

In this paper we discussed the ANS-related system of representations. These representations firstly get activated by preverbal, empirical input, and then continue to be activated when symbols from number line acquire meaning. The model based on conceptual spaces that we proposed, aims at accounting for the structure of these representations. In consequence, additional light is shed on the relation between exact numerals, especially when they are used as vague quantifiers, ANS-related representations as a (naturalized) semantics for such expressions and a model of the ANS-related representations in the framework of conceptual spaces.

The model we propose is based on the idea that tessellations corresponding to ANS-related representations are generated not by a single prototype, but by clusters of prototypes. In consequence boarders between concepts are fuzzy. We argue that this model accounts for size and distance effects, characteristic for ANS.

We believe that further investigation into the exact structure of this model is necessary, in particular necessary is to take into account experimental data disclosing sets of empirical inputs, which correspond to prototypical representations. Another interesting opening, which we do not explore in this paper, is the relation between activation of the system of analog magnitude representations by non-symbolic input and activation of this system by symbolic input (language, number names, numerals). Such a study would enhance understanding of the semantic of numerals.

References

Beck, J. (2014). Analogue magnitude representations: A philosophical introduction, *British Society for the Philosophy of Science* 66 (4): 829–855.

Barth, H., Kanwisher, N., Spelke, E. (2003). The construction of large number representations in adults. *Cognition* 86: 201–221.

Berlin, B. Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*, Berkeley: University of California Press.

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., Zorzi, M., (2010). Numerical estimation in preschoolers, *Developmental Psychology* 46 (2): 545–551.

Carey, S., (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind and Language* 16: 37–55.

Carey, S., (2009). *The Origin of Concepts*. New York: Oxford University Press.

Carey, S., Sarnecka, B.W. (2006). The development of human conceptual representations. In: M. Johnson, Y. Munakata (eds.), *Processes of Change in Brain and Cognitive Development: Attention and Performance* XXI (pp. 473–496). New York: Oxford University Press.

Decock, L., Douven, I. (2014). What is graded membership? *Noûs* 48: 653–682.

Dehaene, S. (1997). *The Number Sense.* New York: Oxford University Press.

Dehaene, S. (2008). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In: P. Haggard, Y. Rosetti,

M. Kawato (eds.), *Sensorimotor Foundations of Higher Cognition* (pp. 527–574). New York: Oxford University Press.

Douven, I., Decock, L., Dietz, R., Égré, P. (2013). Vagueness: A conceptual spaces approach, *Journal of Philosophical Logic* 42 (1): 137–160.

Feigenson, L., Dehaene, S., Spelke, E.S. (2004). Core systems of number. *Trends in Cognitive Sciences* 8 (7): 307–314.

Gallistel, Ch. R. (1993). *The Organization of Learning*, Cambridge MA: MIT Press.

Gärdenfors, P. (2000). *Conceptual Spaces*, Cambridge MA: MIT Press.

Gärdenfors, P., Williams M.-A. (2001). Reasoning about Categories in Conceptual Spaces. In: *Proceedings of the Fourteenth International Conference of Artificial Intelligence* (pp. 385–392), San Francisco CA: Morgan Kaufmann.

Halberda, J., Feigenson, L. (2008). Developmental change in the acuity of the 'Number Sense': The approximate number system in 3-, 4-, 5-, 6-year olds and adults. *Developmental Psychology* 44: 1457–1465.

Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In: E. Hinrichs, J. Nerbonne (eds.), *Theory and Evidence in Semantics* (pp. 109–132). Stanford: CSLI Publications.

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*, Chicago: University of Chicago Press.

McGee, V., McLaughlin, B. (1994). Distinctions without a difference, *Southern Journal of Philosophy* 33 (suppl.): 203–251.

Mussolin, C., Nys, J., Leybaert, J., Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends in Neuroscience and Education* 1: 21–31.

Nagen, J., Sarnecka, B. (2015), Is there really a link between exact-number knowledge and approximate number system acuity in young children? *British Journal of Developmental Psychology* 33 (1): 92–105.

Rosch, E.H. (1973) Natural categories, *Cognitive Psychology* 4: 328–350.

Shapiro, S. (2006). *Vagueness in Context.* New York: Oxford University Press.

Solt, S. (2011). Vagueness in quantity: Two case studies from a linguistic perspective. In: P. Cintula, Ch. G. Fermüller, L. Godo, P. Hajek (eds.), *Understanding Vagueness. Logical, Philosophical and Linguistic Perspectives* (pp. 157–174). London: College Publications.

Spelke, E.S., Kinzler, K.D. (2007). Core knowledge. *Developmental Science*, 10: 89–96.

Taylor, J.R. (2003). *Linguistic Categorization*, New York: Oxford University Press.

Tversky, A. (1977). Features of similarity. *Psychological Reviews* 84 (4): 327–352.

Wagner, J.B., Johnson, S.C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition* 119: 10–22.

Peter Gärdenfors
Jana Holsanova

COMMUNICATION, COGNITION, AND TECHNOLOGY

1. Introduction

The most likely form taken by the first human communication was gestures and miming. The body is still an important component in a direct conversation with other people. However, humans are the only animals who have developed a spoken language as our primary tool for communication. *Dialogue* is our most genuine form of interaction and it will be our point of departure when we compare different forms of communication supported by technology.

Since early in the history of *Homo sapiens,* we have used different *media* for communication. The oldest are cave paintings that are about 40 000 years old. But it is above all written language that has influenced our way of mediating thoughts. The oldest forms of writing are about 5000 years old.

The last few centuries have seen rapid development of different technologies for communication. Printing was invented a little more than 500 years ago. Telegraphy and photography are about 150 years old. Bell invented the telephone in 1876, Marconi made the first radio transmissions in 1895, Edison taught us how to record sound 100 years ago, and moving pictures are equally old. In the last 50 years we have seen how the fax, the television, the computer and the mobile phone have radically influenced our ways of communicating with other people.

Imagine, for example that your boyfriend or girlfriend travels to New York to study for a semester. Unfortunately, you are unable to go along. What types of communication technologies will you use to keep your love alive? A handwritten letter is of course very personal, but it will take time before it reaches the addressee. E-mail is an excellent form of communication over long distances for keeping in touch with people; it is fast, cheap and you can write long messages.

Texting an SMS by mobile phone is not a bad solution but is more suitable for short messages. However, if you want to express your feelings and get an immediate response, the telephone (or the videophone) is the superior medium to "feel close to" the other person. A disadvantage is of course the time difference: you cannot call at any time that suits you.

However, the different technologies will influence how the messages are shaped. In most respects, mediated communication is more limited than a normal dialogue, but as we shall see, there is an increasing number of methods to augment human communication with the aid of technology.

We will use face-to-face dialogue as a starting point when we analyse and compare different ways of using technology for communication. In section 2, we will formulate a number of criteria that are useful for the analysis. In section 3, we will then present some of the most common types of technology-supported communication.

In a dialogue, the participants construct a *common ground* that will form the basis for how the conversation develops (Clark, Brennan, 1991). The common ground consists of background knowledge as well as the physical environment and the information generated in the course of the dialogue. Even the expectations that the speaker and the listener have of each other's thoughts, what is called theory of mind, belongs to the common ground. In technology-supported communication, where the participants are not located at the same place or are not communicating at the same time, there is no surrounding environment to build on and therefore the common ground will be more limited.

The speaker takes the common ground as given when she chooses what to say and how to say it. The listener assumes that the message is *relevant* in the sense that it conveys something new in relation to her previous knowledge and that what is communicated is important for the receiver (Grice, 1975; Sperber, Wilson, 1995).

2. Criteria for analysing communication

2.1. Codes

When comparing different forms of communication it is useful to distinguish between the *contents* of the information (what it means), which *code* it is expressed in, and which *medium* is used in the transmission (see *e.g.* Glass, Holyoak, 1986: 8–10). The first factor that we shall consider in our analysis of communication concerns the choice of code.

The dominant code in human communication is of course *language*. There are a (rapidly decreasing) number of natural languages in the world that are used for spoken communication and somewhat fewer that have a corresponding written language. In addition, there are some artificial languages such as Esperanto, Ido, and Klingon, although they have not had any real impact. The grammars and vocabularies of the different languages vary considerably and are therefore different codes. However, there are no substantial differences in what can be expressed by the languages.

But a dialogue does not consist only of words. Gestures and miming are other "codes" that complement spoken language. It is more difficult to identify the different "signs" that are used in gesturing than to identify the words used in language. But for facial expressions there are fairly clear signals such as raising your eyebrows to signal surprise, and a wrinkle between your eyebrows means disliking or not understanding.

An important part of a dialogue is the *eye contact* that is established between the participants. This is essential for turn-taking and feedback in discourse. By eye contact the participants regulate whose turn it is to speak, confirm that they have mutual attention and interest.

Another important part of a conversation is the *prosody* – intonation, rhythm, and emphasis. We can express various emotional states, but also mark what kind of *speech act* is performed (statement, question, imperative, plea, announcement, etc.). In a conversation, a song might be an odd form of communication, but music is obviously also a code.

A conversation where the participants are physically present is *multimodal*: the participants point spontaneously to the objects in the environment, they nod, draw, gesticulate and touch each other. This functions because they have a joint *perceptual space* as an essential part of the common ground (Allwood, 2002; Holmqvist, Holsanova, 2007). The participants may also use other kinds of visual codes than those that belong to body language. For example, they can draw a road description as a clarifying complement to their oral instructions, present a photo of their children when talking about them or point to the timetable when discussing travel plans. Within modern forms of technology-supported communication, pictures and other visual tools play a more important role than they do in ordinary conversations. We will return to this below.

2.2. Medium

The medium of a form of communication is the physical substrate that carries the message. However, “media” is used in a broader sense to refer to the technical systems that “mediate” the communication: radio, television, film, video and newspaper are all called media. As regards spoken dialogue, it is literally the sound waves in the air that are the medium. Later, various technical achievements have conveyed the sound vicariously via telephone or computer wires or via radio waves. The sound of a conversation fades away quickly, but it can be stored on different media: tape, CD, hard disk, electronic memory in answering machines, etc.

Gestures and facial expressions are mediated by light to the eyes and they disappear immediately if they are not stored with the aid of some form of camera. Even eye contact is visual (in the blind it is replaced by tactile contact). Written language is also visual and can be stored and transmitted via paper, photos, or computers as media. The same also applies, of course, to various kinds of pictures, graphs, and diagrams. The tactile sense modality is sometimes used in dialogues, since touching may be part of the body language. The deaf-blind use a sign language that is shaped in the hand of the receiver so that he or she can feel the signs.

2.3. Space

The most important difference between dialogue and mediated dialogue is that the participants need not see each other or be within hearing distance. For modern technology, the distance between the communicators is of marginal significance: The communicators may as well be in two neighbouring rooms as on opposite sides of the planet.

In a virtual world, for example Second Life, a special form of spatial community can be achieved. In this world, the participants communicate via avatars that are present at the same place in the virtual world. In this form of dialogue, the participants can refer to what is present in the virtual environment (which makes the dialogue closer to natural) but they cannot refer to the real spaces they are located at.

2.4. Time

An ordinary dialogue is immediate in the sense that the message reaches the listener directly and the responses of the listener, in the form of eye contact, facial expressions and humming, can have an immediate effect on what the speaker says. Normally, we are not aware of how synchronized we are during a conversation, but a telephone connection where the sound is delayed half a second is sufficient to cause serious disturbances of the dialogue.

In contrast to natural dialogue, the messages are delayed in many technology-supported systems. A letter may be delivered to the addressee weeks after it has been sent and the reply may take an equally long time. Modern communication systems are faster, but the delay still considerably influences the form of the messages. It makes a big difference to talk to somebody on the phone and to talk to each other via answering machines.

An important difference between immediate and delayed communication is that in a direct dialogue, the speaker cannot spend much time in selecting the right wording, while there is time to think through how you want to express yourself in a delayed exchange. It is even possible to change one's mind and rephrase the message before it is sent off. As we shall see in section 3.2, there are many differences between written language and speech, partly as a consequence of the delay in space and time. In other written forms of communication, where the delay is not equally long, such as chat, SMS or e-mail, the linguistic forms are closer to those of spoken language.

2.5. Audience

Even if direct dialogue is the most genuine form of communication, spoken language has, of course, always been used to address large audiences. Also when you speak to many people, you receive feedback from the audience, but there are differences in comparison to what happens in a dialogue: For example, the speaker can not have eye contact with everybody in the audience and it is not possible to build on such a rich common ground as in a dialogue. It is difficult for single listeners to signal that they do not understand or that they want to know more. The larger the audience, the less the overlap between the speaker and the listeners' inner worlds and consequently the less the common ground. Therefore, the expressions used by the speaker must be clearer and less ambiguous so that he can make sure that the listeners follow. If the audience consists of unknown persons, it becomes even more difficult for the speaker to be relevant and to construct a common ground.

The same factors govern mediated communication with one or many. It you write a letter, you can build on the common ground you have with the addressee and hence much can be left implicit. If you write a newspaper article, you can only presume what is commonly known at the time of writing. If you write a book that you expect to be read over a period of many years, you cannot build on dated information. Consequently, more must be stated explicitly in the text. Similar differences apply to telephone conversations in comparison to radio broadcasting.

2.6. Interactivity

An ordinary dialogue is interactive in the sense that one partner can immediately react to the other. The interactivity of a mode of communication can be defined as the possibilities for the receiver to influence the contents (or the form) of the continued communication (Jensen, 1998: 232). A traditional radio broadcast is an example of a mode of communication with low interactivity, but if it is possible for the listeners to ask questions or request, say, a piece of music, interactivity is already improved.

The time delay in turn-taking is also a factor that influences the degree of interactivity. Three types of written exchanges can be compared: An exchange of letters is less interactive than an exchange of e-mails, which in turn is less interactive than chat. Yet another factor that influences the degree of interactivity is the *mobility* of the communication system. A mobile phone provides better interactivity than an ordinary telephone since the people who communicate become less dependent on where they are physically located. In the same way, a laptop provides better interactivity than a stationary computer in relation to e-mail and chat. A communicative disadvantage of portable platforms is that you do not know where your dialogue partner is located. The lack of spatial information in mobile phones results in a more restricted common ground of the communicators. This leads to frequent questions of the type "Where are you?" among users of mobile phone, which are not relevant at all when ordinary phones are used. In general, however, the increased interactivity of a particular mode of communication compensates for the shortcomings that are created by the displacement in space and time.

3. Properties of different kinds of communication

We will now use the criteria presented in the previous section to analyse various types of communication. The rich availability of technical possibilities that exist today makes it possible to choose different kinds of communication media for different types of messages. For example, we choose a medium depending on how quickly we must communicate. E-mail and in particular letters are not used in urgent situations, but then mobile phone conversation or SMS is used since the receiver is supposed to be reached immediately. If not even this works, you yourself or a messenger must physically move to the receiver. Another example is that there is a tendency to send an SMS rather than calling late at night. This is primarily because you will not disturb the receiver (who may be asleep) to the same extent as if you had made a regular call.

3.1. Direct conversation

Dialogue, face to face, is fundamental for human communication. It is important to note, however, that a dialogue is most of the time not a purpose in itself, but it is used in connection with some other activity that the participants are involved in: when they are solving a problem, arguing to reach a joint decision, giving each other instructions, etc. A dialogue is a part of a common "project" (Clark, 1996) – planning a party, shopping together or deliberating on which movie to watch.

A dialogue is an *interactive* process of shared control where the participants must *coordinate* their linguistic and non-linguistic actions while they are performing various common activities. On the one hand, coordination is about how the dialogue should be organised by turn-taking. On the other hand, coordination deals with how various topics are introduced in the discussion and how they are concluded. When speakers make a mistake or perceive that something is not clear, they interrupt themselves and correct what has been said. Such *repairs* may concern pronunciation, the choice of words or grammar. Dialogue is a tool for achieving something common: The participants negotiate what steps to take, what aspects to discuss in order to solve a problem, what criteria to choose to judge the alternatives and to evaluate the solution (Wästerfors, Holsanova, 2005).

A dialogue is a *dynamic* process where the participants catch the thoughts of the other and meet and expand them in their own contributions. The focus of consciousness is continuously moved, primarily by language steering the *attention* of the participants, either towards factors in the environment or towards something in the common "inner world" (Chafe, 1994). The partners successively construct their understanding: statements are confirmed, legitimated, challenged, developed, corrected, declared invalid, etc. (Linell, 2005). Feedback plays a central role for signalling understanding, agreement, encouragement, and continued interest (Allwood *et al.*, 1992).

A dialogue can be seen as a form of *distributed cognition* (Hutchins, 1995; Linell, 2005). Each individual does not have to consider everything by himself or herself since the conversation partners help each other to find words or phrases. In the words of Linell (1998: 224): "In a dialogic situation part of the thought has already been thought by the other and the speaker may exploit this". Because several persons are working on the same problem, the participants can automatically access parallel cognitive processes in the form of increased attention, improved memory and deeper processing of information. Not only is the responsibility for remembering the different steps, solutions and topics distributed between the participants, but they also take turns in taking initiative and in giving feedback. Another advantage is that misunderstandings can be repaired interactively and unclear statements can be corrected immediately, which leads to a quicker development of the dialogue.

3.2. Letters and other uses of written language

In early human history, messengers were used to communicate over long distances. The speaker sent another person who conveyed the message to the receiver. Written language made the personal messenger superfluous. A *letter* is a visual form of communication that is permanent and enables comparatively cheap communication over long distances. A limitation is that feedback is slow. The messages may be long and the text is often well planned. A letter can be read by others and it thereby has a more public character than an ordinary conversation. It is difficult to whisper in a letter.

Written language has limitations, however. A dialogue is bound to a certain context and is expressed with the aid of prosody and facial expressions that may be more intense than the "literal" message. All this is lost when language is fixed on a sheet of paper. It took some time before written language developed into an independent form of communication – it was long seen merely as a support for what was to be *said*. During the Middle Ages, reading was equal to reading aloud – the text was transformed into the oral (McLuhan, 1964: 83). Only after the invention of printing did silent reading develop.

Written language must *compensate*, by various means, for the parts of the communication that are transmitted by prosody, rhythm, and emphasis. Olson (1994) shows that linguistic expression of speech acts, such as "submit", "explain", and "suggest", also arose during the Middle Ages. These markers are not needed in an oral tradition where the speech act is expressed directly using prosody and other tools. Another example is that it is more difficult to express irony in written language than in speech.

A consequence of the permanence of writing and the liberty of taking one's time when formulating a text is that the sentence structures in written communication become more advanced, with more difficult words and heavy syntactic constructions (Linell, 1978). The reader has plenty of time to interpret the text and can in the worst case use a dictionary.

When the *telegraph* was invented, news could reach the audience very fast, which made the whole world more present. It is interesting to note that the form of telegraphic messages in turn influenced the language in newspapers which, having been more like letters, became briefer and more proclaiming. *Fax* has most of the properties in common with letters, except that it is transmitted faster than ordinary mail. This form of communication also tends to be less private – you would hardly send a love letter via fax.

3.3. SMS, e-mail, and chat

SMS is a visual medium that supports *urgent* communication. It is characterized by short planning time and is suitable for quick, short, contact-creating or co-ordinating messages. SMS does not require the simultaneous presence of the receiver and is relatively permanent, depending on how much is saved. Among the disadvantages are that the messages contain a limited number of symbols and that the method of input is cumbersome and time-consuming.

E-mail is a visual medium that allows long messages and is therefore used for *informative* purposes. New e-mail systems and MMS amplify written communication by the possibility to send pictures, graphic information, and sound files. The writer uses a comparatively long time to formulate the message before it is sent, and the partners can reply with a delay. The message is often directed to one or more persons you know and who constitute a homogeneous group with a rich common ground. The language of e-mails is similar to that of ordinary letters. However, the method of writing and the survey of the text is better than in SMS.

Instant messaging (chat) is characterized by *speed* and *spontaneity* and supports written communication in real time with immediate feedback. This makes it come close to dialogue, but in contrast to the fleeting dialogue, instant messaging has more permanence. The messages remain on the screen during the conversation, but are in general deleted when the communication window is closed. It is therefore possible to directly refer to earlier messages – at least as long as the chat continues.

In SMS, e-mail and chat, written language is adapted in order to communicate emotional qualities that are found in a dialogue. Computer-supported language is not just something in between text and speech, but is in many respects very similar to speech (Ko, 1996). The writers feel the pressure to write fast (preferably as fast as they speak) and do not have time to plan or reflect. They compensate for voice quality, facial expressions, and gestures by using smileys (*e.g.* ;o), asterisks for actions (*blinks*, *smiles*), unconventional punctuation (…!?), abbreviations (4U) and capitals for emphasis (SHE gave ME a gift). As a consequence of the rapid turn-taking, the sentences have less linguistic variation and a simpler construction. On the other hand, the users are tolerant and do not care whether the language has the same degree of perfection as in (classical) written language. Politeness is also less of a constraint. A user can, for example, stop a chat very abruptly without being seen as impolite.

3.4. Telephone, mobile phone, and voice messages

The telephone offers an acoustic and fleeting medium that affords simultaneous personal communication over a distance. The medium is suitable for contact-seeking, informative, emotional, and persuasive communication, with direct feedback. The receiver can hear *what* the speaker says and *how* it is said, but does not see the facial expressions, the gestures, and the body language. In long-distance calls there is sometime a time lag that immediately disturbs the feedback and the turn-taking in the conversation.

The mobile phone is a portable tool for communication that drastically increases the possibility of reaching a communication partner at the time you desire – as long as the partner has the phone turned on. As never before, it is now possible for us to have direct dialogues with almost everybody at almost any time. Mobile phones also offer time-independent messages via the voice mail and written communication via SMS or e-mail.

In contrast to the telephone, the voice mailbox is a permanent medium without interactivity or immediate feedback. These properties explain why it is not suitable to break up from your boyfriend via a voice mail: the permanence means that the receiver can replay the message several times; the lack of common context often leads to problems in finding the correct interpretation of the message; and the lack of interactivity does not give the dumped person any answers to his questions.

3.5. Videophone and video-conference

Videophone (*e.g.* Skype) and video-conferences are examples of advanced technology-supported communication with a high degree of interactivity. They allow the use of language, gestures, and body language and thereby provide dialogue-like conversations over long distances. The technology creates the impression that the communication partners share the same room and they may use a complete register of verbal and non-verbal signals in their normal functions. Hand in hand with this come also our expectations of video conversations. Eye contact is an essential part of the non-verbal communication. The communicator who uses a system with video and sound link may believe that the same rules apply as in communication face to face, for example that communication can be initiated and attention drawn via eye contact. However, eye contact does not function in the same way in a video conversation because of the camera placement. Current technology does not allow that you attend to the screen and at the same time look into the camera. If you try to establish eye contact via the screen, your communication partner will see your eyes staring to the floor or out in the air.

If the implicit rules are not followed, irritation will soon follow. Therefore, the users instead try to "stare out" the other one, wave, exaggerate movements, or grimace. The consequence is that technology does not support what it purports to and the users are disappointed (Hutchby, 2001; Heath, Luff, 1991). If, in addition, the bandwidth and computer power is too low to display real-time video in high resolution, turn-taking will not work well, faces will be distorted, body language cannot be perceived and certain gestures can be missed.

In spite of the rapid development of communication technology, we still travel far for various kinds of meetings. For example, when it concerns important business negotiations or marriage proposals, we still want to have direct contact with our dialogue partner. The direct conversation offers certain properties that technology cannot yet replace. Some researchers argue that the smell of the other is an underestimated factor in a dialogue.

3.6. Augmented communication

Already when we use a paper for drawing on as a complement to an oral road description, we use "augmented" communication (Diderichsen, 2006), in the sense that we add a medium that is not necessary for spoken language. Someone who uses a dictionary to understand a letter also augments the communication.

The steadily increasing access to the Internet has provided us with a powerful tool for further forms of augmentation. If, for example, both participants in a video conversation simultaneously are connected to the Internet, they can during the conversation gather facts, check what the partner claims, and supplement the discussion with pictures and other non-verbal information. It becomes more difficult to be relevant in an augmented communication situation.

Within computer-supported collaborative work various programs are used so that persons located at different places can work simultaneously with the same material. For example, two architects can have the same drawing on their screens and communicate via words or point or draw. Nowadays, one finds new techniques based on communicators sharing various interactive displays. There are systems where a projector in the ceiling displays pictures of documents on a table. The system can also detect how hands move over the table and in this way the persons around the table can "pull", "open" and "close", and in other ways interact with the virtual objects that are projected.

In the future we will meet more kinds of augmented communication. There may be virtual food and drink on the restaurant table where you can order by "pulling" the food or drink to your plate and find out about the ingredients by double clicking on it. The interactive environments of *Blade Runner* and other science fiction movies are getting closer.

3.7. Communication with handicap

Technology-supported communication has led to remarkable changes for people with different handicaps. A spellchecking program suffices to make it possible for a dyslectic to become more secure in his or her written communication. Advanced technology aids are available for people with aphasia (Kitzing *et al.*, 2005) and computer-supported communication programs with pictures and symbols facilitate communication for disabled or speech-handicapped persons in general (Rydeman, Zachrisson, 2001). Braille, and its corresponding technology, has made it possible for blind people to replace the visual paper-based communication by a tactile medium. The telephone must have been a revolution for the blind and the mobile phone makes communication over distance even easier for them. During the last few decades speech synthesis programs that can read e-mail, web pages and newspapers have to some extent replaced the function of Braille. For sure, the synthetic voice is still a bit robotic, but it is becoming more and more human.

Deaf persons can of course use letters and text-based media to communicate over a distance. Using the telephone for a long time involved using a messenger. During a period, the text telephone has been a tool for the deaf – it can be seen as an early chat function. Computers with e-mail and chat are of course an improvement but dependent on having a computer available. A paradoxical consequence of the development of mobile phones is that is has radically facilitated the communication of the deaf. SMS allows a rather quick and interactive communication, but above all the videophone has made it possible for the deaf to communicate in real time over long distances via sign language.

References

Allwood, J. (2002). Bodily communication dimensions of expression and content. In: B. Granström *et al*. (eds.), *Multimodality in Language and Speech Systems* (pp. 7–26). Dordrecht: Kluwer Academic Publishers.

Allwood, J., Nivre, J., Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics* 9(1): 1–26.

Chafe, W. (1994). *Discourse, Consciousness, and Time: The Flow of Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H.H., Brennan, S.E. (1991). Grounding in communication. In: L.B. Resnick, J.M. Levine, S.D. Teasley (eds.), *Perspectives on Socially Shared Cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Diderichsen, P. (2006). Augmented communication: The communicative potential of the internet. Lund: *Lund University Cognitive Studies* 132.

Glass, A.L., Holyoak, K.J. (1986). *Cognition*. 2nd edition. New York: Random House.

Grice, P. (1975). Logic and conversation. In: P. Cole, J.L. Morgan (eds.), *Syntax and Semantics*, vol. 3: Speech Acts (pp. 41–58). New York, NY: Academic Press.

Heath, C., Luff, P. (1991). Disembodied conduct: Communication through video in a multi-media office environment. *Proceedings of the SIGCHI conference on human factors in computing systems: Reaching through technology* (pp. 99–103). New Orleans, LA.

Holmqvist, K., Holsanova, J. (2007). Embodied communication and gestural contrast. In: E. Ahlsén *et al*. (eds.), *Communication – Action – Meaning*. Göteborg: Department of Linguistics, Göteborg University.

Hutchby, I. (2001). *Conversation and Technology: From the Telephone to the Internet*. Cambridge: Polity.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Jensen, J.F. (1998). Interaktivitet og interaktive medier. In: J.F. Jensen (ed.), *Multimedier, hypermedier, interaktive medier*, vol. 3 of FISK serien (pp. 199–238). Aalborg: Aalborg Universitetsforlag.

Kitzing, P., Ahlsén, E., Jönsson, B. (2005). Communication aids for people with aphasia. *Logopedics Phoniatrics Vocology* 30: 41–46.

Ko, K.-K. (1996). Structural characteristics of computer-mediated language: A comparative analysis of inter change discourse. In: S. Herring (ed.) *Computer-Mediated Discourse Analysis: The Electronic Journal of Communication* 6 (3).

Linell, P. (1978). Talat och skrivet språk. In: P. Linell, *Människans språk. En introduktion till språk, tänkande och kommunikation* (pp. 230–235). Lund: Gleerups.

Linell, P. (1998). *Approaching Dialogue: Talk, Interaction and Contexts in Dialogical Perspectives.* Amsterdam: John Benjamins.

Linell, P. (2005). En dialogisk grammatik? In: J. Anward, B. Nordberg (eds.), *Samtalets Grammatik* (pp. 231–315). Lund: Studentlitteratur.

McLuhan, M. (1964). *Understanding Media*. New York, NY: McGraw-Hill.

Olson, D.R. (1994). *The World on Paper*. Cambridge: Cambridge University Press.

Rydeman, B., Zachrisson, G. (2001). *Dynamiska kommunikationsprogram och styrsätt för personer med rörelsehinder.* Projektrapport nr 2 – IT i Praktiken. Vällingby: Hjälpmedelsinstitutet.

Sperber, D., Wilson, D. (1995). *Relevance: Communication and Cognition*. 2nd edition. Oxford: Blackwell.

Wästerfors, D., Holsanova, J. (2005). Examples as crucial rhetorical tools in discourse on 'others'. *Text* 25 (4): 519–554.

Jana Holsanova
Roger Johansson
Kenneth Holmqvist

TO TELL AND TO SHOW: THE INTERPLAY OF LANGUAGE AND VISUALIZATIONS IN COMMUNICATION

The use of imitation, gestures and pictures has played a very important role in the evolution of human cognition (Zlatev *et al.*, 2005; Persson, 2008). Another important strand of research has studied the human ability to communicate multimodally, by using a combination of language, mimics, gestures, pictures, and body movements (Allwood, 2002).

Let us have a closer look at the use of various types of visualization that accompany language in communication. Imagine a group of friends involved in a lively conversation about their holiday adventures. Whether they talk about a dangerous bungy-jumping experience, hiking in the mountains, or enjoying the beach and pub life, they behave in a similar way. They do not only use language to describe events and things; they also use gestures, draw sketches, imitate voices, and engage their whole body to re-enact events. In short, when narrating, they not only *tell* but also *show* their experiences, tales, or stories.

Since Plato and Aristotle, these two activities are called *diegesis* (narrators describe things) and *mimesis* (narrators show things). In the following, we will be concerned with mimesis in more detail, following the distinctions made by Clark (2004). The term mimesis has been taken by Plato from music theory and applied to designate scenic performance where actions of persons are imitated and re-enacted. In other words, mimesis is a re-production based on imitation of an action. In communication, mimesis is important for both speakers and listeners. It is a means of showing actions to make them visible (and audible) to the listeners. By using voice, gestures, and drawings, the speakers create scenes, embody and dramatize events, and thereby involve their partners. "Mimesis gives us the sense of reality in fiction, the illusion of access to the reality of personal experience" (Lodge, 1990: 144). The more effective the imagination, the better the possibility for the audience to visualize the speaker's experience.

There is, however, an important distinction to be made between two ways of "showing": on the one hand we can indicate something by pointing at an object, on the other hand we can demonstrate things with the help of our voice, hands, or body. These two senses go back to Charles Peirce (1960), who identified three forms of relations between a representation and an object: an *index* refers to its object by means of a physical connection to it (*e.g.*, a footprint as a representation of a bear), an *icon* resembles the object it depicts (a drawing of a bear), whereas a *symbol* bears no resemblance to what it stands for and is thus arbitrary (the word "bear"). The first sense of showing is thus connected to what Peirce calls indexes, the other sense what he calls icons. According to Clark (2004), it is only the iconic sense of showing that is actually mimetic. In this context, Clark talks about *demonstrations* as selective depictions that "enable others to experience (in part) what it is like to perceive the things depicted" (Clark, Gerrig, 1990: 765). They make it easier for the listeners to imagine objects, scenes and events, what it is like to see, hear, and feel them, etc.

Producing and understanding mimetic devices relies on imagination and pretence (Clark, 2004: 8). When showing (and imitating) a person or demonstrating (displaying) events, the speaker runs a sort of simulation: s/he pretends to be another person and animates this virtual person's actions. But imagination and pretend play would obviously not be possible without the ability to separate real states of affairs from the pretended ones (Frith, 1996). Also, the listeners must be able to understand demonstrations as pretend play to finally reach "joint pretence" together with their partners.

In communication, we can observe a great variety of mimetic devices that stimulate the imagination and involvement of the listeners. They can take form of pictures and illustrations, sound symbolism, quotations, iconic gestures, etc. In the following, we will take a closer look at these imagistic elements of communication and at the interplay between language and visualizations. In particular, we will be concerned with visualizations in different formats of mimesis: both auditive demonstrations, *i.e.* depictions by sound symbolism and quotations, and visual demonstrations, *i.e.* depictions by drawing, gestures, and imagery. The outline of the chapter is as follows: we will proceed from quotations (section 1), gestures (2), drawings (3), to mental imagery (4) and implications for learning (5).

1. To tell and to show with your voice (quotations)

A frequent type of mimetic devices used in conversation is *quotations*. Quotations are audible demonstrations that mediate direct experience and stimulate listeners' imagination. In face-to-face conversation, we often imitate others, play scenes, and integrate others' voices in our presentation in order to achieve certain effects in the current situation. By voice quality and a particular prosody, we animate non-present (virtual) characters and stage their activity. Not only singular speakers are animated but also plural voices, hypothetical speech and thoughts of others (Tannen, 1989; Holsanova, 1998; Adelsvärd *et al.*, 2002).

Look at the following example where speaker B tells his friend M about a telling-off that his girl friend Mary got at her job. He uses quotations and enacts a dialogue between a father of a child and Mary (M) who is working at a kindergarten. Since speaker B imitates two audibly different protagonist voices, which are marked in the transcript (*cf*. Holsanova, 2006b: 253).

Example 1

B so this child has difficulties eating you know, cause she was the only child and they have spoilt her terribly with sweets and things, so that when they serve normal decent food'. she simply doesn't want it . you know, so . they have finally succeeded in making her eat a little and at this moment . her daddy enters, just in the middle of the meal, his kid sitting and eating,

M mm'

B and he says something (VOICE 1) **we'll go home now** right, and Mary said that (VOICE 2) **we are eating now**

M mm'

B (VOICE 1) **yes but we'll go home now** (VOICE 2) **we are eating now and you know what problems we have had with this,** right, (VOICE 2) **and she is eating now and she has to sit here,** . (VOICE 1) **we'll go home now,** and Mary (VOICE 2) **you'll go out till she has finished,**

B (laughing) (VOICE 1) **I want to talk to you,** hehe, (VOICE 2) **sure!** and in front of the door, she's got a damn telling-off, (VOICE 1) *it's me who decides about my children*

Direct quotation can serve the purpose of both dramatization and documentation. In narratives, we rather show what happened in order to increase dramatic intensity and to mediate direct experience. This is the dramatic (*mimetic*) function of direct quotation. By contrast, in reports such as news texts, quotation is used to increase the perceived objectivity and accuracy of the account to the reader. Here the documenting (diegetic) function of quotation is prominent. The writer claims authenticity by implying that he has direct access to the original speech situation (Redeker, 1991). In argumentative discourse, virtual participants in the form of authoritative references are often used to reinforce chosen opinions or to present a mental opponent whose opinions are to be questioned and undermined (Adelsvärd *et al.*, 2002). For listeners, quotations mark an attentional shift towards other real or imaginary scenes, characters, and events (Sanders, Redeker, 1996).

Quotations position those who are animated as speakers in an unflattering light and hide behind in case of sensitive topics (Holsanova, 1998; 2006b), but can also increase the distance between the speaker and the described characters: "with quotations speakers can partly or wholly detach themselves from what they depict" (Clark, Gerrig, 1990: 792).

Another way of depicting persons and events is to use onomatopoeia or sound symbolism (*his heart went ticktock ticktock*). For instance, when characterizing and criticizing politicians the speaker can let them whine like helpless children (*the whining politicians say uhuhu*), or express anger with the help of growling rather than with an extensive verbal description (*she thought mmhrr-mnhrr*). Onomatopoeia is a very effective way to ascribe attitudes without having to describe them verbally.

2. To tell and to show with your hands (gestures)

Probably the most usual way to embody an action or manner of action is through body language, in particular gestures. According to McNeill (1992), planning of utterances involves the interplay of *imaginistic* thinking and *linguistic* thinking (manifested as gestures and speech). There are, however, a number of different opinions regarding the relation between the content of gestures and of concurrent speech (Kita, Özyürek, 2002). The free imagery hypothesis states that gestures are generated from imagery, independently of the language, whereas the lexical semantic hypothesis states that gestures are generated from the semantics of lexical items in the accompanying speech and therefore cannot encode what is not encoded in the concurrent speech. Finally, the interface hypothesis states that gestures originate from an interface representation between language and imagery. According to this last hypothesis, production of speech and gestures is interrelated, and gestures may encode information that is not expressed in speech.

As mentioned earlier, there are two senses of "showing": the first is *indicating* (by pointing) and the other is *demonstrating* (by animation). Prototypically, pointing is understood in its deictic sense: through juxtaposition, pointing gestures instruct the hearer to attend to something beyond the talk and to locate what is being indicated (Goodwin, 2003). On the other hand, pointing gestures can take a step towards iconicity (*e.g.* when a person points at an object and traces the shape of the object that is being pointed at (Streeck, 1996). In the latter sense, when tracing an iconic shape, pointing gestures can be understood as demonstrations.

The information provided by gesture is analogue and depictive (Özyürek, Kita, 1999). The demonstrative function of bodily mimesis can be compared to quotations as demonstrations. Sensitive topics can be made in the gestural channel, keeping the spoken channel free from explicit reference (Holmqvist, Holsanova, 2007). Thus, similarly as in quotations, the gestural information which is not present in the spoken channel saves the speaker's face from the dangers involved in explicit spoken characterization. In case of disabilities, *e.g.* in aphasia, gestures and other non-verbal means of communication may compensate for the restricted use of language (Ahlsén, 2006).

The question remains how verbal and gestural communication becomes integrated by speakers and listeners. Do listeners look at the gestures produced by speakers in conversation? Goodwin suggests that speakers intentionally use gestures to attract the attention of their interlocutors. Cassell, McNeill, and MacCullough (1998) conclude from their mismatch studies that the speaker's gestural channel is indeed perceived by listeners and plays an important role for the understanding. Gullberg and Holmqvist (1999) found in an eye tracking study of gestures in interaction that auto-fixated gestures, *i.e.* gestures that the speakers themselves focus on, are attended more often by listeners than other gestures. This result confirms the assumption maintained by Streeck that gestures can be overtly marked as communicatively relevant by the speakers if they look at their own hands.

3. To tell and to draw

Apart from spontaneously using gestures, mimics, and body postures along with their utterances, speakers also draw pictures when they describe their experiences. Drawings – such as sketch maps of houses and areas, illustrative drawings of people and clothing, or explanatory diagrams – are often incrementally produced and modified as a part of a conversation. The use of iconic and pictorial representations is useful in communication, since it helps the speaker and the listener to interactively adjust their visualizations and achieve understanding. Pictures, maps, and sketches are visualizations that show how the described reality has been conceptualised (Tversky, 1999).

Drawings can represent (a) a concrete spatial domain, such as geography, sizes, shapes, spatial relations; (b) non-spatial domains, such as amounts of money, temporal relations; (c) abstract domains, such as intensity, contrast, and quality dimensions, and (d) dynamic processes, such as stages in a decision process, causal relations, development over time.

Apart from being a support for visualization and demonstration (specifying the location and shape of objects, spatial relations, and events), drawings also fulfil other, more abstract, functions in communication. They are a useful tool for the identification of "Where are we now?", serve as a storage of referents, as an external memory aid for the interlocutors, as an expressive way of underlining what is said, and as a representation of a whole problem discussed in the conversation (Holsanova, 2008).

Speakers' discourse contains concepts that appear as schematic representations and establish patterns of understanding. When speakers want to draw attention to a complex visual idea, *e.g.* a scene, a navigation route or an apartment layout, they have to organize the information so that their partner can understand it. By uttering ideas, speakers evoke images in the consciousness of the listeners, the minds of the speaker and the listeners get synchronized, and the listeners co-construct the meanings. The process of how the listener's internal image is constructed from spoken discourse and simultaneous drawing has been studied within image-oriented semantics (Holmqvist, 1993) where discourse understanding is described in terms of evolving mental images.

In spontaneous face-to-face conversation, this process has a more dynamic and cooperative character (Clark, 1996). The partners try to achieve joint attention, formulate complementary contributions, and interactively adjust their "visualizations". Quite often, the partners draw simultaneously with their verbal descriptions. Sketches and drawings are external representations that reflect the conceptualization of reality and serve as an aid for memory (Tversky, 1999). The utterances and non-verbal actions (such as drawing and gesturing) can be conceived of as instructions for the listeners on how to change the meaning, how something is perceived, how one thinks or feels, or what one wants to do with something that is currently in the conscious focus (Linell, 2005).

Using a delimited, common source of references makes it easier for the speaker and the listener to coordinate and to understand each other. Drawing as an external representation allows cognitive processes to be captured, shared, and elaborated. In a naturally occurring conversation, external visualizations help the partners to achieve a joint focus of attention and to coordinate and adjust their mental images during meaning making.

4. To tell and to imagine

In the previous sections, we mentioned mimetic devices that the communicative partners use when showing and re-enacting events and actions for the listeners. It has, however, been shown that that we use gestures even in a situation when our partners cannot see us. The question is thus whether we as speakers use mimetic devices and re-enactment for ourselves, in the form of mental imagery.

Mental imagery is "the mental invention or recreation of an experience that in at least some respects resembles the experience of actually perceiving an object or an event, either in conjunction with, or in the absence of, direct sensory stimulation" (Finke, 1989: 2). In popular terms, mental imagery is described as "visualising" or "seeing something in the mind's eye".

We use mental imagery when we mentally recreate personal experiences from the past, retrieve information about physical properties of objects or about physical relationships among objects, read novels, plan future events or anticipate possible future experiences, imagine transformations by mental rotation and mental animation and when we solve problems (Finke, 1989; Hegarty, 1992; Yoon, Narayanan, 2004). In other words, imagery plays an important role in memory, planning, and visual-spatial reasoning, and is considered a central component of our thinking.

Since mental images are closely connected to visual perception, this mental invention or re-creation of experience almost always results in observable eye movements that can be traced by new technology. Eye tracking methodology has become a very important tool in the study of human cognition, and current research has found a close relation between eye movements and mental imagery (Holsanova, 2001; 2006a). In order to verify the assumption that we use our ability to create pictures in our minds, we conducted a series of studies on mental imagery during picture description. The results of these studies contribute to our understanding of how speakers connect spoken discourse to mental imagery.

Already in our first eye tracking study (Holsanova *et al.*, 1998), we found some striking similarities between the participants' eye movement patterns when they looked at a complex picture and their eye movements when they later on looked at a white board and described the same picture from memory. We then conducted a number of new eye tracking studies where participants looked at a blank white board and visualized a scene they had previously either seen on a picture or heard as a spoken description (Johansson *et al.*, 2006; 2011; 2012; 2013).

Look at the following example showing a comparison of one person's eye movement patterns during picture viewing and during picture description from memory.

Fig. 1. The stimulus picture (Nordqvist, 1990)

Example 2: the stimulus picture (fig. 1) and one and the same participant's eye patterns after the viewing phase (fig. 2) and the description phase (fig. 3).

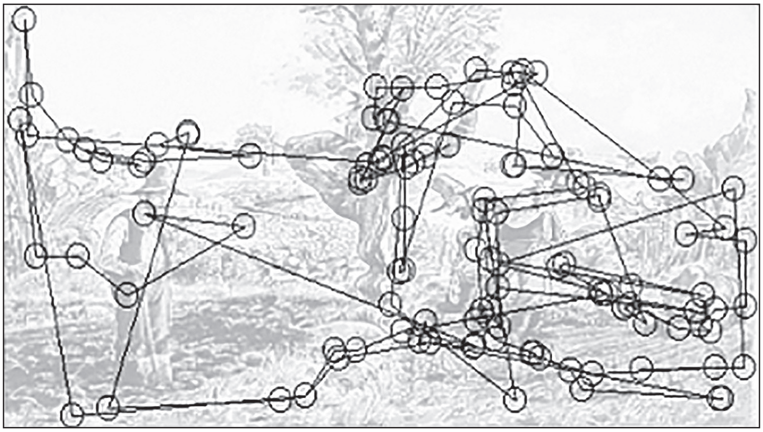

Fig. 2. Participant's eye patterns after the viewing phase

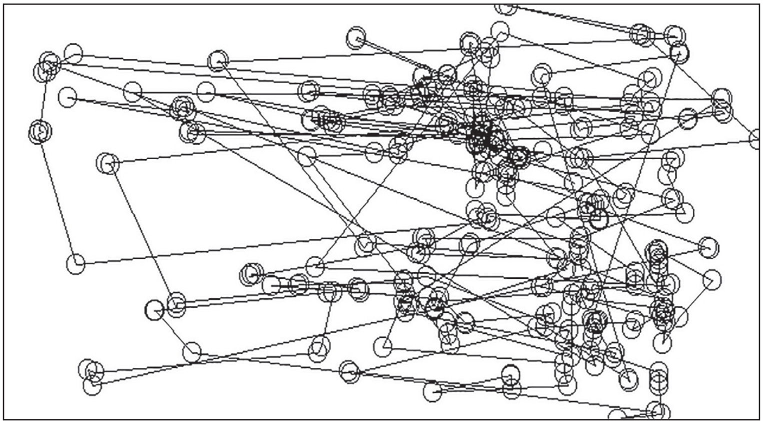

Fig. 3. Participant's eye patterns after the description phase

The results of our studies clearly showed that when describing a scene from memory the participants to a high degree moved their eyes in a pattern that "painted" the imagined scene on the white board in front of them. Additionally, it was found that the effect was equally strong irrespective of whether the original elicitation was spoken or visual, and that it was also present in complete darkness.

Our results support the hypothesis that mental scanning (Kosslyn, 1980) is used as an aid in recalling picture elements, especially when describing their visual and spatial properties. Mental imagery plays a functional role in our cognition and it seems to play an important role even for the speakers involved in discourse. We represent and re-enact our previous experience when we describe a scene from memory.

5. To show and to learn

The findings described above have decisive implications for learning. For speakers and listeners, both mental imagery and re-enactment of events can play an important role as a memory aid. For learning purposes, for instance in argumentation, quotations can be used to stage another person's talk or thought, to reinforce chosen opinions or to present a mental opponent whose opinions are to be questioned and undermined (Wästerfors, Holsanova, 2005).

In the context of collaborative learning, the use of sketches and drawings is advantageous. First, sketches and drawings show how the described reality has been conceptualized. Second, they allow revisions, regroupings, refinements, and reinterpretations and are therefore an important thinking tool (Suwa *et al.*, 2001). Third, they help the speaker and the listener to interactively adjust their visualizations and achieve understanding. In sum, drawings and sketches allow cognitive processes to be captured, shared, and elaborated.

In the context of textbook illustrations and instructional materials, it is important to outline visualizations according to the users' mental model, *i.e.* the way users conceptualize how everyday objects and situations are structured or how they work. Visualizations have world-like qualities resembling actual objects or events. By means of this analogy, they function as a substitute for the referent and evoke similar experience to the real-world referent. Visualizations in instructional materials can offer highly realistic impressions of objects and events, which might otherwise be too small, too large, too fast, too far away, or too dangerous to observe in reality. In that respect, visualizations do not only replace real-world experience, they may even improve this experience by providing information that would not have been accessible in the real world (Scheiter *et al.*, 2008).

Visualizations in their *representational function* depict objects and relations mentioned in a text in a way that the meaning of the text is made more easily accessible for learners by making a text more concrete. Visualizations with an *organization function* provide an organizational framework for a text (*e.g.*, how-to-do-it diagrams) and thereby make the content more coherent by highlighting the argumentative or organizational structure of the text. Accordingly, visualizations are often introduced in textbooks and multimedia instructions to clarify difficult passages and abstract concepts.

However, it should not be taken for granted that learners will extract the information from a visualization that was intended by the instructor. Rather, students need to be supported in extracting the relevant information from the visualization and guided as to how best to deploy their perceptual and cognitive resources. This support can be provided either by guiding learners' attention towards its relevant aspects (*e.g.*, highlighting) or by improving students' competencies in dealing with visualizations (Scheiter *et al.*, 2008).

To sum up, our communication is multimodal and embodied. By combining mimetic devices such as quotations, drawing and gestures, communicative partners depict, demonstrate, embody, and re-enact objects and actions in order to create vivid scenarios for each other. The usage of these devices evokes images in the mind of both the speaker and the listener, which in some respects resemble the experience of actually perceiving a scene or an event.

References

Adelsvärd, V., Holsanova, J., Wibeck, V. (2002). Virtual talk as a communicative resource: Explorations in the field of gene technology. *Sprachtheorie und germanistische Linguistik* 12 (1): 3–26.

Ahlsén, E. (2006). *Introduction to Neurolinguistics*. Amsterdam, Philadelphia: John Benjamins.

Allwood, J. (2002). Bodily communication dimensions of expression and content. In: B. Granström, D. Granström, I. Karlsson (eds.), *Multimodality in Language and Speech Systems* (pp. 7–26). Boston: Kluwer Academic Publishers.

Cassell, J., McNeill, D., McCullough, K.-E. (1998). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition* 7 (1): 1–34.

Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. (2004). Variations on a Ranarian Theme. In: Strömqvist, S.L. Verhoeven (eds.), *Relating Events in Narrative: Typological and Contextual Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Clark, H.H., Gerrig, R.J. (1990). Quotations as demonstrations. *Language* 66 (4): 764–805.

Finke, R.A. (1989). *Principles of Mental Imagery*. Massachusetts Institute of Technology: Bradford Books.

Frith, U. (1996). Cognitive explanations of autism. *Acta Paediactrica* 415: 63–68.

Goodwin, C. (2003). Pointing as a situated practice. In: S. Kita (ed.), *Pointing: Where Language, Culture and Cognition Meet* (pp. 217–241). Mahwah, NJ: Lawrence Erlbaum Associates.

Gullberg, M., Holmqvist, K. (1999). An Eye on Gesture. *Pragmatics and Cognition* 7 (1): 35–63.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition* 18: 1084–1102.

Holmqvist, K., Holsanova, J. (2007). Embodied communication and gestural contrast. In: E. Ahlsén, P.J. Henrichsen, R. Hirsch, J. Nivre, Å. Abelin, S. Strömqvist, S. Nicholson, B. Dorriots (eds.), *Communication – Action – Meaning: A Festschrift to Jens Allwood*. Göteborg: Göteborg University.

Holmqvist, K. (1993). *Implementing cognitive semantics: Image schemata, valence accommodation and valence suggestion for AI and computational linguistics*. Lund: Lund University Cognitive Studies 17.

Holsanova, J. (1998). The use of quotations in discourse about ethnicities. In: S. Cmejrkova *et al*. (eds.), *Dialoganalyse VI* (pp. 253–260). Tübingen: Niemeyer.

Holsanova, J. (2001). *Picture Viewing and Picture Description: Two Windows on the Mind*. Doctoral dissertation. Lund: Lund University Cognitive Studies 83.

Holsanova, J. (2006a). Dynamics of picture viewing and picture description. In: L. Albertazzi (ed.), *Visual thought: The depictive space of the mind*, Part 3: *Bridging perception and depiction of visual spaces* (pp. 233–254). Advances in Consciousness Research. Amsterdam–Philadelphia: Benjamins.

Holsanova, J. (2006b). Quotations as a vehicle for social positioning. In: H. Hausendorf, A. Bora (eds.), *Analysing Citizenship* (pp. 251–275). Talk Discourse Approaches to Politics, Society and Culture. Amsterdam–Philadelphia: Benjamins.

Holsanova, J. (2008). *Discourse, Vision and Cognition*. Amsterdam–Philadelphia: Benjamins.

Holsanova, J., Hedberg, B., Nilsson, N. (1998). Visual and verbal focus patterns when describing pictures. In: W. Becker, H. Deubel, T. Mergner (eds.), *Current Oculomotor Research: Physiological and Psychological Aspects.* New York–London–Moscow: Plenum.

Johansson, R., Holsanova, J., Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science* 30 (6): 1053–1079.

Johansson, R., Holsanova, J., Holmqvist, K. (2011). The dispersion of eye movements during visual imagery is related to individual differences in spatial imagery ability. In: L. Carlson, C. Hölscher, T. Shipley (eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Johansson, R., Holsanova, J., Holmqvist, K. (2013). Using eye movements and spoken discourse as windows to inner space. In: C. Paradis, J. Hudson, U. Magnusson (eds.).

The Construal of Spatial Meaning: Windows into Conceptual Space (pp. 9–28). Oxford: Oxford University Press.

Johansson, R., Holsanova, J., Dewhurst, R., Holmqvist, K. (2012). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance* 38 (5): 1289–1314.

Kita, S., Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48: 16–32.

Kosslyn, S. (1980). *Image and Mind*. Cambridge, MA–London: Harvard University Press.

Linell, P. (2005). En dialogisk grammatik? In: J. Anward, B. Nordberg (eds.), *Samtal och grammatik* (pp. 231–315). Lund: Studentlitteratur.

Lodge, D. (1990). Narration with words. In: H. Barlow, C. Blakemore, M. Weston-Smith (eds.), *Images and understanding* (pp. 141–153). Cambridge: Cambridge University Press.

Nordqvist, S. (1990). *Kackel i grönsakslandet*, Bromma: Opal.

McNeill, D. (1992). *Hand and Mind*: *What the Hands Reveal about Thought*. Chicago: Chicago University Press.

Özyürek, A., Kita, S., (1999). Expressing manner and path in English and Turkish: Differences in speech, gesture, and conceptualization. In: M. Hahn, S.C. Stoness (eds.), *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society* (pp. 507–512). London: Erlbaum.

Peirce, C.S. (1960). *The Icons, Index, and Symbol (1902): Collected Papers*. Cambridge, MA: Harvard University Press.

Persson, T. (2008). *Pictorial primates: A search for iconic abilities in great apes.* Sweden: Lund University Cognitive Studies 136.

Redeker, G. (1991). Quotation in discourse. In: R. van Hoyut, E. Huls (eds.), *Artikelen van de eerste Sociolinguistische Conferentie* (pp. 341–355). Delft: Eburon.

Sanders, J., Redeker, G. (1996). The representation of speech and thought in narrative discourse. In: G. Fauconnier, E. Sweetser (eds.), *Spaces, Worlds and Grammar* (pp. 290–317). Chicago: Chicago University Press.

Scheiter, K., Wiebe, E., Holsanova, J. (2008). Theoretical and methodological aspects of learning with visualizations. In: R. Zheng, *Cognitive Effects of Multimedia Learning*. IGI Global, USA.

Streeck, J. (1996). How to do things with things. *Human Studies* 19: 365–384.

Suwa, M., Tversky, B., Gero, J., Purcell, T. (2001). Seeing into sketches: Regrouping parts encourages new interpretations. In: J.S. Gero, B. Tversky, T. Purcell (eds.), *Visual and Spatial Reasoning in Design II* (pp. 207–219). Sydney: Key Centre of Design Computing and Cognition.

Tannen, D. (1989). *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. Cambridge: Cambridge University Press.

Tversky, B. (1999). What does drawing reveal about thinking? In: J.S. Gero, B. Tversky (eds.), *Visual and spatial reasoning in design* (pp. 93–101). Sydney: Key Centre of Design Computing and Cognition.

Wästerfors, D., Holsanova, J. (2005). Examples as crucial rhetorical tools in discourse on "others". *Text* 25 (4): 519–554.

Yoon, D., Narayanan, N.H. (2004). Mental imagery in problem solving: An eye tracking study. In: A. Duchowski, R. Vertegaal (eds.), *Proceedings of the 3rd ACM Symposium on Eye Tracking Research and Applications* (pp. 77–83). New York: ACM.

Zlatev, J., Persson, T., Gärdenfors, P. (2005). *Bodily mimesis as "the missing link" in human cognitive evolution*. Lund: Lund University Cognitive Studies 121.

Aleksander Gemel
Bartosz Żukowski

SEMIOTICS, SIGNALING GAMES AND MEANING

1. Introduction

The 20th-century discussion on meaning was dominated by two distinct schools of thought – by that of naturalistic semantics and by that of structural semiotics. Although the two have traditionally been viewed as conflicting approaches, the main aim of this paper is to show that they can be seen as complementary to one another. In order to achieve this objective, we will examine the advantages and drawbacks of both approaches (represented, respectively, by Eco's semiotics and Skyrms' game-theoretical semantics). The results obtained this way provide a basis for further development of hybrid system by combining features of structural and naturalistic models.

2. Structural semiotics. Between signalization and signification

The distinctive feature of structural semiotics is a strict separation of two independent semiotic systems: signalization and signification. At the same time, in accordance with the basic premise of structural semiotics, it is only through both of these autonomic systems that a proper communication process can work successfully. Considered in itself, signalization is simply "the passage of a signal (not necessarily a sign) from a source (through a transmitter, along a channel) to a destination" (Eco, 1976: 8). The signals used for this process derive from a special kind of semiotic system called *s-code*, resulting from the division of some natural or artificial continuum. An essential feature of s-code is that all of the primary system components are isolated arbitrarily (*ibid.*: 77). Nevertheless, when selected, they are to form a well-organized, independent system, in which the identity of each unit is completely determined by reference to others, and hence specified by its position in the whole (*ibid.*: 38). The system established in this way is an internally structured matrix with certain combinatorial properties (determined by a set of combinatorial rules). As a result, it possesses also some informational potential. It is precisely (mathematically) defined by the intrinsic properties of the system (arrangement of elements and combinatorial rules), determining the amount of information possible to transmit. So configured, the system can operate autonomously – without any semantic reference: "s-codes are systems or 'structures' that can also subsist independently of any sort of significant or communicative purpose" (*ibid.*). Consequently, a purely signaling system can be considered as a communication system only in a purely mathematical sense – as a combinatorial syntactic structure, able to convey a certain amount of information, but devoid of any meaning in itself: "In a machine-to-machine process the signal has no power to signify insofar as it may determine the destination *sub specie stimuli*. In this case we have no signification, but we have the passage of some information" (*ibid.*: 8); "A signal is a pertinent unit of a system that may be an expression system ordered to a content, but could also be a physical system without any semiotic purpose; as such it is studied by information theory in the stricter sense of the term. A signal can be a stimulus that does not mean anything but causes or elicits something" (*ibid.*: 48). In effect, Eco's signalization is perfectly consistent with the transmission communication model proposed by Shannon and Weaver (1949) as a starting point for the mathematical theory of communication (Eco, 1976: 42–44).

A signaling system serves as a basis for a signifying system. Notwithstanding this, the assumption underlying the whole semiotic theory of language is that the two systems remain functionally independent of each other.[1] This central idea of structural semiotics seems to be an echo of the separation between semantics and syntax in logic and formal linguistics of the first half of the 20th century. The essence of signification is to establish a correlation of some units of a given s-code, considered as an expression system (plane), with the units of some other

---

[1] "A signification system is an autonomous semiotic construct that has an abstract mode of existence independent of any possible communicative act it makes possible" (Eco, 1976: 9).

s-code treated as a content system (plan) (*ibid.*: 50). The fundamental difference between signalization and signification is therefore that the latter involves an interpretative (decoding) response of the receiver, which is made possible by the previously established convention of correlating signals with contents: "When the destination [of a communicative process] is a human being, or 'addressee' [...] we are on the contrary witnessing a process of signification – provided that the signal is not merely a stimulus but arouses an interpretative response in the addressee. This process is made possible by the existence of a code" (*ibid.*: 8). By contrast with signalization, signification is a process within which signals refer to certain units of content and, thus, become units of meaning, which situates signification as a transmission of *signs*.

What is most important, however, is that both systems (plans) of expression and of content are structures, or systems of s-code type. What is being correlated here are, therefore, two independent and arbitrarily construed systems – two autonomous combinatorial matrices, organized internally as systems of positions and oppositions. And it is precisely the correlation between particular elements of such s-codes which is called "a sign-function" or simply "a sign". A sign (sign-function) is, therefore, a relationship linking two different s-codes, one acting as an expression plan and the other being a content plan. Furthermore, the method used to correlate both (*i.e.* to assign some content to a signal) is based on convention only. In other words, the set of correlating rules, called "a code", is purely conventional. It goes simply as follows: "When a code apportions the elements of conveying system to the elements of a conveyed system, the former becomes the expression of the latter and the latter becomes the content of the former" (*ibid.*: 48). All the observed complexity of structural semantics results from repeating and accumulating such correlations. For example, a particular sign-function (*i.e.* a particular relationship between an element of a system of expression and an element of a system of content), called denotation, can be easily correlated with an element of a third system (s-code), thereby forming a higher-order sign-function, called "connotation" (*ibid.*: 54–57). The resulting connotative code consists of two functives: a pre-existing sign function and some element of a third s-code. Repeated many times, this process gives rise to the formation of a desirable rich semiotic system. All this clearly proves the dependence of the theory of semiotics on structural linguistics.

There are, no doubt, some significant advantages of this approach. The first and foremost of them is substantial flexibility of the structural semiotics.

It provides a fertile ground for generation of new meanings and transformation of the current semantic field. At the same time, it seems to adequately reflect the dynamics of natural language, prone to constant reconfiguration of meanings. Thus, structural semiotics can rightly claim to be the 'logic of culture' (*ibid.*: 3). The aforementioned flexibility results from a combination of several features indicated above. Let us summarize them briefly as follows:

(a) Functional independence of signalization and signification – both systems are considered as autonomous combinatorial structures (s-codes).

(b) Arbitrary construction of s-codes of all types and levels. This feature is of special importance for the design of the plan of expression (*ibid.*: 77).

(c) Conventionality of the rules of linking the elements of the plan of expression with the elements of the plan of content.

(d) Rejection of the so-called 'extensional fallacy' (*ibid.*: 62–66). One of the most important features of structural semiotics not mentioned so far is rejection of referential semantics, correlating signs with some extra-semiotic objects: "From a semiotic point of view" the meaning of a term "can only be a *cultural unit*", understood as an element of some arbitrary construed matrix (*ibid.*: 67).

As a result of these assumptions, the semiotic system achieves the desired flexibility. It is easily susceptible to transformation – new units of content can be generated from within the system, by transformation of the existing semiotic infrastructure. In other words, the system has the capacity to freely evolve, just like culture itself.

Despite the abovementioned advantages of this system, it does still suffer from certain weaknesses, which should be overcome by partial naturalization. The first weakness has to do with the content plan – the problem is the absolute arbitrariness of the starting units. In fact, cognitive, anthropological and psychological studies revealed that there are some content universals common to all human cultures, which argues for the existence of natural determinants of content system (Bickerton, 2009; Bickerton, Szathmary, 2009; Brinck, Gärdenfors, 2003; Gärdenfors, 2004; Deacon, 1997; Ollera, Griebel, 2004). The second weakness is the avoidance of the question of the original sign correlations. Although Eco claims that the explanation of this issue requires a reference to the natural, pre-cultural conditions (Eco, 1976: 58–59, 77), he refrains from discussing this matter, probably because of the fear of falling into extensional fallacy (which, in our opinion, is excessive). Finally, the third problem with structural semiotics is its panlinguisticism. In order to fully reflect the actual functioning of language, structural semiotics must be supplemented with at least rudimentary external reference (correspondence to an extra-linguistic reality).

In conclusion, structural semiotics should be enriched with some components of naturalistic semantics, without, however, depriving it of its flexibility. Semiotics, for its part, can provide naturalistic models with generative/transformative potential, making them more flexible. A new hybrid system developed this way will feature the best advantages of both approaches. In the next section, we present an example of typically naturalistic semantics, which might serve as a starting point for such synthesis.

3. Signaling games

To illustrate the naturalistic approach to the question of coding conventions emergence and the fixation of meaning, we will use signaling games model, originally proposed by D.K. Lewis and developed by B. Skyrms (Lewis, 2002; Skyrms, 2010). This model provides a game theoretical instrument, which helps us to explain the genesis of such semiotic phenomena as code and meaning. Thanks to the game-theoretic framework, the concept of communication process in signaling games is recognized as a game between the sender and the receiver. The intention of Lewis' game is to provide a model of language and its semantic content genesis within a community devoid of any language system. Moreover, it is important to emphasize the fact, widely pointed out by Lewis, that a language constituted as a product of a sender-receiver game can only take a form of a fairly primitive, rudimentary proto-language (Lewis, 2002: 160). The mature form of a language system can potentially evolve from its proto-form, but Lewis does not deal with this issue. The mechanism of the mature language system formation is also not the subject of this paper, if only due to the degree of its complexity.

Lewis defines signaling game as a type of situation involving at least two agents (one in the role of a sender and the other as a receiver[2]). It meets four basic conditions. According to the first condition, at least one of several states of affairs occurs. States (S_1, …, S_n) are randomly picked by nature. In fact, random selection means that the occurrence of a particular state can be attributed with a certain probability. What is also important, in contrast to the sender, the receiver

---

[2] Lewis calls them communicator and audience (Lewis, 2002: 130–132).

occupies a privileged position to observe and correctly identify a given state of reality. According to the second rule of the game, the sender, having observed one of the states (S_1, …, S_m), takes one of several alternative measures (σ_1, …, σ_n) called signals. The set of signals must be greater or at least equal to the set of states ($n \geq m$). Taking action σ_i is equivalent to sending the signal to the receiver. The ontological nature of the signal is not pre-defined (it can be a sound, gesture, smell, etc.), since the signal is considered only in functional terms, *i.e.*, its essence is to evoke a specific reaction of the receiver. Moreover, sending the signal by the sender does not have to be intentional or even conscious. Similarly to the first condition, the receiver is in a good position to receive the sent signal. The channel is not noisy, although adding noise to the model is possible by manipulating the prior probability of the signal. According to the third rule, the receiver, after observing and identifying the signal, takes one of several alternative reactions (R_1, …, R_n). Importantly, the receiver takes action based only on the received signal without knowing anything about the state of the reality. The fourth signaling game condition is that both players have a set of strategies (contingency plans). The strategy of the sender consists in sending a specific signal according to a given state of affairs. In mathematical terms, this corresponds to function F_s associating one of the states in a set of states $\{S_i\}$ with a specific signal in a set of signals $\{\sigma_k\}$. On the other hand, you can specify the receiver's strategy as a function assigning the signal to the reaction. It is mathematically expressed as function F_r assigning function $\{\sigma_k\}$ to $\{R_j\}$. Lewis describes the combination of the sender's and receiver's strategies (F_s and F_r) which provides the relationship between the reaction of the receiver and the state of the affairs $<F_n, F_r>$ as the signaling system (Lewis, 2002: 130–132).

The essence of fixing a particular signaling convention is, therefore, the correlation between the reaction and the existing state of affairs, based on a proper signal. This correlation is effected through positive payoffs in the game. If the reaction of the receiver given a specific state of S_a leads to positive payoffs for both players, then signaling convention gets fixed. In general, reaction is considered to be proper and signal is said to be adequate if the receiver, after getting the signal, takes an action that he would take in the case of a direct experience of the state of affairs.

This model can be presented in a simplified version of two states, two signals, and two reactions in the following graphic form:

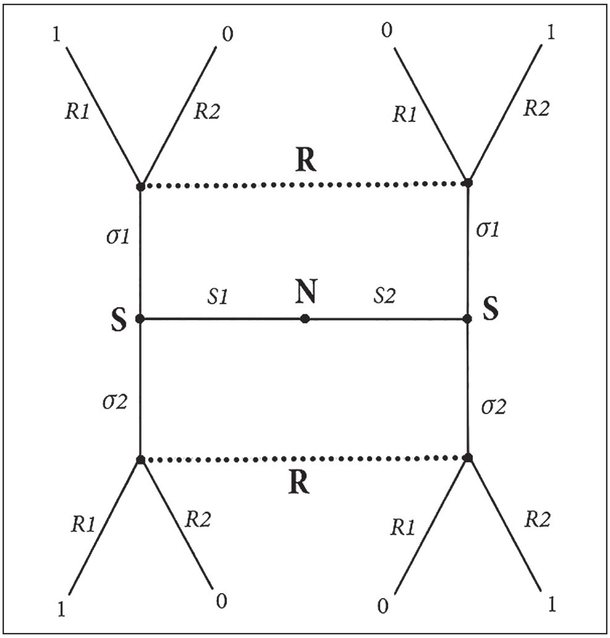

Fig. 1. A signaling system based on 2-states, 2-signals, 2-responses

The "nodes" of the tree represent the players (*i.e.*, **N**-nature, **S**-sender, **R**-receiver). The branches of the tree stand for events (*i.e.*, *S*1, *S*2 – occurring states of affairs; σ_1, σ_2 – sending the signals by the sender; *R*1, *R*2 – taking actions by the receiver).

4. Meaning in the signaling game framework

The signaling game theory provides the code creation model for Eco's semiotic theory. The other side of code creation is the emergence of meaning from the equilibrium of the signaling game. This process takes place as a result of the underlying behavioral-signaling mechanisms, *i.e.* connections of signals with appropriate responses to them. The basis of this approach is the thesis according to which the meanings are evolutionary established conventions of response to a signal. But this raises a question of what is meant by meaning in the signaling game theory. A comprehensive and original answer to this question can be found in Skyrms' *Signals. Evolution, Learning and Information* (2010). However, in order to provide answers to the question: what is the meaning of signal (or an information content as Skyrms puts it), we must first find the answer to two related questions, namely: "What is the quantity of information in a signal?" and "How should it be measured?".

4.1. Quantity of information

The concept of the quantity of information comes from the mathematical theory of information, which, as it is commonly known, does not deal with the informational content of a signal, since it understands the signaling process only in quantitative terms. According to mathematical theory of information, quantity is closely related to the value of probability that a given signal (stimulus) triggers a specific situation (reactions). As Skyrms puts it, the notion of information quantity can be easily applied to the signaling game model, since the occurrence of each of elements (events) within a game (*i.e.* state – signal – reaction) can be assigned to a certain probability (Skyrms, 2010: 34). Thanks to this feature, one can easily express the quantity of information carried by a signal as the ratio of the conditional probability of a particular state after sending a specific signal and the unconditional probability of this state (*i.e.* after sending the signal). This ratio gives an idea of how the probability of the state after sending the signal has changed with respect to the probability before the signal was sent. In formal terms, this may be expressed by the following equation (Skyrms, 2010):

$$\frac{P(S\,|\sigma)}{P(S)}$$

To illustrate this, let us assume that we have the simplest signaling game: 2-states, 2-signals, and 2-reactions, with an initial equal probability for each of them. This situation is a typical example of a state before reaching equilibrium (*i.e.* before the coding convention gets fixed). An increase in the level of the probability of a correct receiver's response to a specific signal corresponds to the achievement of equilibrium. In consequence, the signal becomes positively correlated with the state. A positive correlation increases also the probability of sending a specific signal (*e.g.*, σ_1) any other time when that certain state occurs (*e.g.*, S_1). Assume that the probability of this signal increases to a value of 0.9. The probability of σ_2 signal in the event of the state S_1 is thus reduced to a value of 0.1, since signals σ_1 and σ_2 are mutually exclusive events, and hence their total probability amounts to 1. Of course, changing the probability of a signal doesn't affect the value of prior probability of the state S_1, and thus, it is still equal to 0.5 (*i.e.* $P(S_1) = 0.5$). After reaching equilibrium, the conditional probability of signal σ_1 given state S_1 equals 0.9 (*i.e.* $P(\sigma_1|S_1) = 0.9$). The cumulative probability of signal σ_1 is the sum of two products. The first of them is the unconditional probability of state S_1 multiplied by the conditional probability of signal σ_1 given state S_1. The second product is the unconditional probability of state S_2 times the conditional probability of signal σ_1 given state S_2. In formal terms, $P(\sigma) = \mathrm{P}(\sigma_1 \cap S_1) + \mathrm{P}(\sigma_1 \cap S_2)$ and it is equal to 0.5. The overall probability distribution of our example is shown in fig. 2 in the form of a tree-diagram:

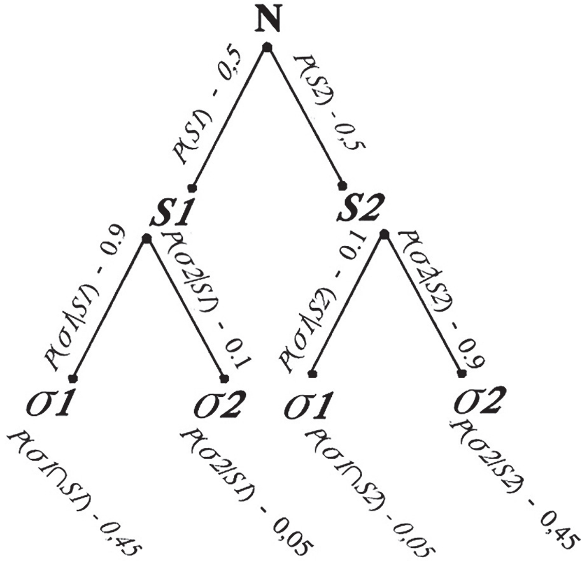

Fig. 2. Probability distribution of a signaling system of 2-states and 2-signals

By applying Bayes' theorem, we can easily calculate the value of the conditional probability of state S_1 after sending signal σ_1:

$$P(S_1|\sigma_1) = \frac{P(\sigma_1|S_1) \times P(S_1)}{\mathrm{P}(\sigma_1)}$$

$$P(S_1|\sigma_1) = \frac{0{,}9 \times 0{,}5}{0{,}5}$$

$$P(S_1|\sigma_1) = 0{,}9$$

Therefore, the ratio of $P(S_1|\sigma_1)$ to $P(S_1)$ gives a value of 1.8. This value expresses how many times the probability of state S_1 after sending signal σ_1 increases with respect to the prior probability of state S_1. It is the quantity of information carried by the signal. As Skyrms notices, we should take the logarithm of this ratio, since the quantity of information in the signal in present form does not allow us to express the situation when the signal does not convey any amount of information (*e.g.*, if the sender always gives the same signal, regardless of the situation) (Skyrms, 2010: 36). In this case, the ratio of the conditional probability of the state after sending the signal to the prior probability of the state equals 1, and not 0, as we would intuitively expect. Thus, the final form of the information quantity formula is as follows (*ibid.*):

$$Log_2 \frac{P\ (S_1|\sigma_1)}{P\ (S_1)}$$

Using the logarithm base 2 allows us to express the quantity of information in bits. This formula in its expanded form can be adapted to express information about the number of states, and to give a general measure of the information in the signal. Skyrms does this by using the Kullback-Leibler divergence formula, which expresses the average of the distribution of several probabilities (see: Kullback, Leibler, 1951; Skyrms, 2010):

$$I(\sigma_i|S_i) = \sum_i P(S_i|\sigma_i)\, log \left[\frac{P(S_i|\sigma_i)}{P(S_i)}\right]$$

The above formula is simply a weighted average, in which the weights are the conditional probabilities of states given signals.

In addition to the quantity of information in the signal on a given state, we can still distinguish the amount of information in the signal on a given act. It is measured in the same way, as the amount of state, and expresses the ratio of the conditional probability of response after sending a signal to the probability of reaction before it is sent. The informational content in the signal carrying probability on the act takes the following form (Skyrms, 2010):

$$I(\sigma_i|R_i) = \sum_i P(R_i|\sigma_i)\, log\left[\frac{P(R|\sigma_i)}{P(R_i)}\right]$$

Thus, the total amount of information in the signal consists of the two above values, *i.e.*

$$I(\sigma_i) = I(\sigma_i|S_i) + I(\sigma_i|R_i)$$

4.2. Informational content

Now let us turn to the issue of the information content of a signal. While the informational content in a signal is represented by a specified number (consisting of the sum of the amount of information about a state and about a reaction) expressed in bits, the information content of the signal is, according to Skyrms, represented by a vector whose components are the values of the informational content of states carried by the signal. The shape of the vector is related to a particular game. As Skyrms argues:

> Informational content must be a vector [...] within a given signaling game. It is implicit that this vector applies to the states or acts of this game. For a different game, the content vector shows how the signal moves probabilities of different states, or different acts. Content depends on the context of the signaling interaction. It is a modeling decision as to which game is best used to analyze a real situation (Skyrms, 2010: 40).

The original proposal of expressing the informational content, as suggested by Skyrms, allows us to reconcile the information theory with the logical interpretation of propositional content, understood as a set of possible situations. The shape of the informational content vector is determined by the components formed from the values of informational quantity carried by the signal, and has the following form:

$$\left\langle log\frac{P\ (S_1|\sigma_1)}{P\ (S_1)}, log\frac{P(S_2|\sigma_2)}{P\ (S_2)}, log\frac{P(S_3|\sigma_3)}{P\ (S_3)}, \ldots, log\frac{P(S_n|\sigma_n)}{P\ (S_n)}\right\rangle$$

A similar formula can be presented for the information content vector of actions carried by the signal.

$$\left\langle log\frac{P\ (R_1|\sigma_1)}{P\ (R_1)}, log\frac{P\ (R_2|\sigma_2)}{P\ (R_2)}, log\frac{P\ (R_3|\sigma_3)}{P\ (R_3)}, \ldots, log\frac{P(R_n|\sigma_n)}{P\ (R_n)}\right\rangle$$

To illustrate: suppose there are four states with an equal prior probability $P(S_n) = 0.25$ for $n = \{1, 2, 3, 4\}$, and that σ_2 signal is sent only when state S_2 occurs. This means that the value of $P(\sigma_2|S_2)$ is equal 1, and the value of $P(\sigma_2|S_1)$, $P(\sigma_2|S_3)$, and $P(\sigma_2|S_4)$ equals 0. Applying the Bayes' theorem, we get the value of the conditional probability of state S_2 after sending signal σ_2 equal to 1[3]. Thus, the probability of state S_2 after sending signal σ_2 increases four times with respect to the probability before the signal was sent:

$$\frac{P(S_2\ |\sigma_2)}{P(S_2)} = \frac{1}{0{,}25} = 4$$

Since signal σ_2 is not sent in states S_1, S_3 and S_4 (*i.e.* the conditional probability of signal σ_2 for these states is 0), it does not affect the probability of these states (*i.e.*, $P(S_1|\sigma_2)$, $P(S_3|\sigma_2)$ and $P(S_4|\sigma_2) = 0$). Therefore, an increase in the level of probability for states S_1, S_3 and S_4 after sending the signal is 0. If we take the logarithm of the values of all these probability ratios, we will obtain the informational content on all states in signal σ_2:

$$I(\sigma_2|S) = \langle -\infty;\ 2;\ -\infty;\ -\infty\rangle$$

Component $-\infty$ informs us that the probability of states 1, 3 and 4 approach 0. Value $-\infty$ is the result of the logarithm, and actually means that the signal does not carry any information about the state. But the question is: how to reconcile the above interpretation of the informational content with the logical in-

---

[3] $P(S_1|\sigma_1) = \frac{P\ (\sigma_1|S_1) \times P\ (S_1)}{P\ (\sigma_1)}$.

terpretation of the meaning understood as a proposition? Presenting his model of informational content, Skyrms answers this question. The proposition is, in fact, a set of possible worlds or situations, so in his opinion, the informational content of a proposition can be expressed as a set of states. A state belonging to the set is true when its informational quantity carried by the signal varies from $-\infty$. In other words, the informational content or meaning of the signal is the set of possible situations or states which are true or false. The state is said to be true when the signal carries some informational quantity about it, and is recognized as false when it does not carry any quantity of information. As Skyrms points it out, a true state can be defined by listing all false states from the set of all states (2010: 41).

We believe that it is especially important that the signaling games model is compatible with the basic assumptions of Eco's theory of semiotics. According to our hypothesis, the signaling games theory can provide a model of code creation for Eco's semiotic theory. In other words, we suppose that denotation in semiotic theory – correlating a signal with a specific meaning – is preceded by a signaling system, which by correlating a signal with a proper reaction, makes it possible for a specific denotational coding convention to emerge. In our opinion, the core of this process is based on an evolutionary-fixed behavioral-signaling mechanism, *i.e.* connections of signals with appropriate responses to them.

However, the naturalistic-evolutionary theory of language is not sufficient as a descriptive model of the processes of the broadly-understood cultural activity evolution. The first problem is interpretation within the game theoretical framework of the culturally understood utility. From the point of view of biological interpretation, the numeric quantities, which play a role analogous to "utility" in traditional game theory, correspond to the Darwinian fitness of individuals. However, the Darwinian concept of "fitness" in the cultural evolutionary interpretation is inconsistent. Consequently, the concept of fitness as the notion of utility used in traditional game theory cannot be simply moved to the game theory of cultural evolution. One must develop an alternate theory of utility/fitness that is sufficient to define a utility measure adequate for application of evolutionary game theory to cultural evolution.

The second weakness of naturalistic theories involves high generative rigidity of new semantic content in the system. The content cannot be generated from within the semantic system, as it is possible in the case of the theory of semiotics, but requires reference to an external reality (reactions and states of affairs).

The hybrid model proposed herein allows us to save the flexible potential of semiotic system, suitable for the explanation of cultural evolution on the one hand, and to root the semiotics in a natural order (by providing a model of emergence of evolutionary coding conventions) on the other. Such hybrid model requires, however, a formal tool to organize the semantic structure of the cultural system (in the signaling games framework, such tool is not needed, since the ordering of the meaning structure is carried by the values of informational quantity assigned to specific states). We will finish this paper by presenting a formal tool adequate for the task.

5. Content implication

The abovementioned tool is based on the propositional classical language extended by the binary intentional connective, called *content implication*, and represented by the sign of colon ":". It was proposed by Piotr Łukowski (1997, 2011).

The Classical Logic with Content Implication (*CLcont*), based on language $L = (For_U, \neg, \wedge, \vee, \rightarrow, \leftrightarrow, :)$, is a propositional logic given by an axiom set for classical propositional logic and the following formulas:

Ax_1. $((\alpha : \beta) \wedge (\beta : \delta)) \rightarrow (\alpha : \delta)$
Ax_2. $(\alpha \wedge \beta) : \alpha$
Ax_3. $(\alpha \wedge \beta) : (\beta \wedge \alpha)$
Ax_4. $\alpha : (\alpha \wedge \alpha)$
Ax_5. $((\alpha : \beta) \wedge (\beta : \alpha)) \rightarrow ((\neg\alpha : \neg\beta) \wedge (\neg\beta : \neg\alpha))$
Ax_6. $((\alpha : \beta) \wedge (\beta : \alpha) \wedge (\delta : \gamma) \wedge (\gamma : \delta)) \rightarrow (((\alpha \S \delta) : (\beta \S \gamma)) \wedge ((\beta \S \gamma) : (\alpha \S \delta)))$, for $\S \in \{\rightarrow, \leftrightarrow, :\}$
Ax_7. $((\alpha : \beta) \wedge (\delta : \gamma)) \rightarrow ((\alpha \S \delta) : (\beta \S \gamma))$, for $\S \in \{\wedge, \vee\}$
Ax_8. $(\alpha : \beta) \rightarrow (\alpha \rightarrow \beta)$

Modus Ponens (*MP*) $\{ \rightarrow \beta, \alpha\} \mid\!- \beta$ is the only inference rule of *CLcont*. One of the most important *CLcont*-theses is $\alpha : \alpha$, a trivial formula easily inferred by Ax_1, Ax_2 and Ax_4.

An adequate semantics for *CLcont* is the class of all so-called *CLcont*-models, *i.e.*, matrices $M = (A, D)$, such that $A = (A, -, \cap, \cup, \Rightarrow, \Leftrightarrow, \supset)$ is an algebra similar to L_U, D is a nonempty subset of A and for all $a, b \in A$,

1. $a = a \cap a$
2. $a \cap b = b \cap a$
3. $-a \in D$ iff $a \notin D$
4. $a \cap b \in D$ iff $a\ D$ and $b \in D$
5. $a \cup b \in D$ iff $a \in D$ or $b \in D$
6. $a \Rightarrow b \in D$ iff $a \notin D$ or $b \in D$
7. $a \supset b \in D$ iff $a = b \cap c$, for some $c \in A$

Semantic inference is defined in a standard way:

$X \models_{UCl} \alpha$ iff for any *CLcont*-model $M = (A, D)$ and $v \in Hom(L_U, A)$
$v(\alpha) \in D$, if for any $\beta \in X$, $v(\beta) \in D$.

According to the desired meaning of a new connective, $p : q$ is true if the content of sentence q is included in the content of sentence p. Thus, sentence $p : q$ is true if and only if the content of q is a part (not necessarily proper) of the content of p. In other words, $p : q$ is true, if sentence p says what is said by q. Of course, p can say something more than what is said by q. (Simultaneous truthfulness of $p : q$ and $q : p$ means that $p = q$ is true, and so p says what is said by q and q says what is said by p).

The aim of construction of a new connective is simple: to express the fact that the content of one sentence is a part (not necessarily proper) of the content of another sentence. Therefore, the meaning of the new connective refers directly to the connective of conjunction. Truthfulness of sentence $p : q$ means that p is a conjunction, in which q is one of its conjuncts. In such a sense, the content of sentence q is a part of the content of p. This feature makes content implication an excellent tool for organizing the semantic structure of a given system of propositions. At the same time, content implication turns out to be the perfect complement to the hybrid model postulated in this paper.

6. Conclusion

The objective of this paper was to compare structural semiotics with naturalistic semantics. The advantages and disadvantages of each were examined, and both approaches were shown to be complementary to one another, thereby providing a basis for further development of a hybrid system by combining the strengths of the two models.

References

Bickerton, D. (2009). *Adam's Tongue: How Humans Made Language, How Language Made Humans*, New York: Hill & Wang.

Bickerton, D., Szathmary, E. (ed.), (2009). *Biological Foundations and Origin of Syntax*, Cambridge, MA: MIT Press.

Brinck, I., Gärdenfors, P. (2003). Co-operation and communication in apes and humans. *Mind & Language* 18 (5): 484–501.

Deacon, T.W. (1997). *The Symbolic Species: Co-evolution of Language and the Brain*. New York: W.W. Norton.

Eco, U. (1976). *A Theory of Semiotics*. Bloomington-London: Indiana University Press.

Gärdenfors, P. (2004). Cooperation and the evolution of symbolic communication. In: D.K. Ollera, U. Griebel, (eds.), *Evolution of Communication Systems* (pp. 237–256). Cambridge: Cambridge University Press.

Kullback, S., Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* 22 (1): 79–86.

Lewis, D.K. (2002). *Convention. A Philosopical Study*. Oxford: Blackwell Chicago Press.

Łukowski, P. (1997). *An approach to the liar paradox*. In: *New Aspects in Non-Classical Logics and Their Kripke Semantics*. RIMS: Kyoto University: 68–80.

Łukowski, P. (2011). *Paradoxes*. Dordrecht: Springer.

Ollera, D.K., Griebel, U. (ed.) (2004). *Evolution of Communication Systems*. Cambridge: Cambridge University Press.

Skyrms, B. (2010). *Signals. Evolution, Learning and Information*. New York: Oxford.

Shannon, C.E., Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

DOROTA RYBARKIEWICZ

OUT OF THE BOX THINKING

1. Introduction

If we accept that out of the box thinking is important, especially when hitherto existing solutions are applied to little avail, then understanding this process is also extremely important.

The primary observation here is that out of the box thinking is powerfully constrained by the very nature. In addressing this phenomenon we need to recognize that once the mind is furnished, it is difficult for it to entertain novelty.[1] Viewing the situation through the metaphor of canyon and plain not only allows for an insight into this question but also offers a method of inducing a change.

2. Canyon

Fancy that you are moving along some deep canyon and all you can see are its walls of stone. You feel safe inside and just follow the unique direction possible.

As much evidence shows, some part of human thinking may be compared to torrents flowing in deep canyons. And as in a real ravine the visual perspective is confined by its slopes and one can follow the only path, similarly, the train of thoughts is limited to the most easily accessible and immediate "pathways" without perceiving the possibility of choice or change. These are real thinking canyons, *i.e.* most deeply rooted and long-standing beliefs, attitudes, and opinions. They are the last to be changed. They play the role of a censor sometimes causing a cognitive dissonance, for example, a person who is asked to write an article

---

[1] I do not mean the "novelty" complying with canyons.

defending views contrary to their own and paid a negligible amount of money experiences cognitive dissonance and to avoid it starts to agree with what he/she has written. Thus, their prevailing canyon: *I am rational*, remains unaltered. Even Einstein succumbed to the canyons (he was convinced that the universe cannot change) and added a cosmological constant to stop the expansion (or contraction) predicted by his own relativity theory.[2]

The question about the source of such canyons has at least two complementary answers.

First, this feature of thinking seems to be strictly connected with the traits of our memory, but not, which should be emphasized, with any memory impairment. According to Eysenck, Keane (Eysenck, Keane, 2010: 272, 631, 636) declarative (explicit) memory registers events, rules, references and meanings, while non-declarative (implicit) one registers procedures and is connected with emotions, skeletal musculature, habituation, and sensitization (increase in the responsiveness to a stimulus as a result of its repeated presentation), and priming (increased sensitivity to certain stimuli due to prior experience) (Forster, Davis, 1984). The latter two supposedly explain the tenacity of canyons.

Moreover, the storage of memory and its persistence suggest the existence of engrams[3] by which memory traces are stored. Engram is thought of as a permanent impression left on protoplasm as the result of stimulus, *i.e.* any psychic experience. Neuroscientists have long sought the location of these memory traces. Recently (Liu, Ramirez, Pang, 2012) have identified the cells that make up part of an engram for a specific memory and reactivate it using a technology called optogenetics. Memories of experiences are encoded by chemical and physical changes in neurons, and by modifications to the connections between the neurons. The study also provides further evidence that memories are stored in networks of neurons that form memory traces for each experience we have.

Thinking, being naturally immersed in memory, shares its features. The differences between kinds of memory are reflected on neurological level, nevertheless, although separated, they often share neurological regions yielding kind of resonance, for example, *She is like lemon juice* apart from semantic regions activates sensual ones. (This is important for co-occurrence and reconciliation presented below).

---

[2] When Hubble's study of nearby galaxies showed that the universe was in fact expanding, Einstein regretted modifying his theory and viewed the cosmological constant as his "greatest mistake". However, now it is again seriously reconsidered.

[3] The term introduced by Richard Semon, German zoologist, in 1921.

When we gain a novel thought, a new engram appears and once it is formed it becomes part of memory/ies. The trace is left and cannot be wiped out. This must greatly influence the way people think, and especially what they think. Namely, they tend to think what they have already thought. Stereotypes are the most commonly recognized members of this group. In light of neuroscience canyonical thinking is biologically grounded.

Second, apart from biological, there are pragmatic reasons of this constancy. Our tendency to stay within/in the canyon seems to be ruled by Zipf's principle of least effort (PLE) (Zipf, 1949). The principle may be briefly spelled out as: the greatest effect with the least effort! It holds for all fields of human activity. The principle explains both: changes, for example, the linguistic preference for abbreviations rather than full forms as “flu” for “influenza”; as well as stagnation when its estimated value does not mean any gain as, for example, the refusal to replace the known computer software with a new one.

What is important, the principle seems to imply its conscious and free application assuming that people always strive to minimize the expenditure. How much freedom is actually realized is a different matter. The favorable balance may be disturbed by manifold factors, for instance, addictions. Sometimes the confrontation between evidence and reason turns into a battlefield on which the latter loses. The experiments conducted to investigate possible change of our opinions concerning others show that despite the fact that the evidence was sound, people chose not to adjust their beliefs to the new pieces of information, no matter how irrefutable these were. Once judgments have been formed, they have tendency to persevere even in face of totally discrediting data (Anderson *et al.*, 1980). This is possible due to the nature of brain (undoubtedly, it is difficult to leave a precipitous canyon) but also due to PLE because people see no reason to invest in such an unprofitable change. No additional psychological aspects that may block the change as rigid thinking, fears or compulsions have been considered, although they may pose an insurmountable obstacle to bear a new thought. For example, trauma makes canyons deeper and sometimes less reasonable, for a victim may even avoid people in the shoes similar to the ones the oppressor used to wear. Marginally, note that of special interests may be here the link between canyons, persuasion (psycho-manipulation), and PLE.

Canyons are carved by experience, either sensual, emotional, or discursive (information, persuasion, argumentation). The most intriguing, however, are canyons that enter by the back door. They often accompany “official visitors”, *i.e.* messages (or events) we are aware of. Such invisible guests are in fact canyon *occurrences* stemming from canyon *types* encoded in the language. In other words, some canyons are an inherent part of language and when language is used to communicate, canyons pass unnoticed but obviously they bear their content. Two such canyons can be specified: Hannah Arendt's *petrified analogies* and Lakoff and Johnson's *conceptual metaphors*. In cognitive terms, both are codified mappings of one domain (source domain) into another (target domain). This will be specified in the sequel.

Hannah Arendt (see: Arendt, 1991: 153) (following Kant) observed that abstract concepts are not referred to with the vocabulary indirectly related to our thinking abilities but by means of the words primarily connected with our common sensory experience. The process of borrowing the words is not arbitrary but based upon analogy. For Kant, analogy (metaphor) is the only way of being of pure reason. Metaphysical knowledge is possible due to analogy. According to Arendt all philosophical concepts are metaphors, which she calls *petrified analogies*. The function of metaphor is to direct the mind towards the world in order to connect mind's activity and the abstract outcomes of its activity with comprehensible words directly related to the world. One of the examples is the word *idea*. When Plato was introducing this word into philosophy he could hear it occasionally used in a pre-philosophical sense; it meant a pattern, model, image that an artisan must have before his eyes to make a concrete object like a vase, or a robe. Through such an etymological insight into words referring to abstract concepts there emerge petrified canyons with their dormant power of directing thoughts towards source domains, *i.e.* their original meaning.

Similar mechanism of viewing one thing in terms of another stands behind other canyons inherent to language, conceptual metaphors. Conceptual metaphors are glimpsed behind linguistic (dead) metaphors and are a cognitive phenomenon of mapping, as the tiny example below shows:

a. *Man consumes energy. Machine consumes energy.*
b. *Replace part of the body with… Replace part of the machine with…*

The conceptual metaphor here is: MAN IS A MACHINE.

Conceptual metaphors display systemic nature and show that we are more rational than it may have seemed so far. Lakoff and Johnson, having analyzed some basic conceptual metaphors like: LIFE IS A JOURNEY, LOVE IS WAR, TIME IS SPACE, CATEGORY IS A CONTAINER concluded that they pervade language and exert great influence on our thoughts and perception (Lakoff, Johnson, 1980). Nevertheless, their power in carving canyons is still underestimated. Strong connection between conceptual metaphor and models and paradigm is especially noteworthy in this context.

Canyons are a hidden power behind the throne. Concrete thoughts, decisions, and actions reveal their presence but not necessarily make them explicit in the form of clearly stated assumptions or reasons. Although they are detectable traces in different brains regions, the question how to elicit them remains open. The next question is how to change or replace them in order to initiate out of the box thinking. The strategy proposed is called *plaining*.

3. Plain

You have just managed to get upon the canyon edge. Standing on the cliff, you can see a plain all around with its fairy colors and shapes. New perspectives, almost unlimited number of directions, and finally… the necessity to choose one in order to continue the march.

Any kind of invention (but not only) is closely connected with a more general process that I call plaining.[4] One reaches the upper edge of the canyon and gains a double view, an old one on the canyon and a new one on the surrounding plains. Some fusion of the views is experienced and, in result, each looks different.

Plaining would consist in providing the system of thoughts (canyons) with a novel point of reference. On the one hand, it is true that everything is undergoing change in a continuous manner, so that we can enjoy novelty all the time. But on the other hand, we already have fixed engrams. Therefore, only in some cases, plaining may be a spontaneous response to the stimuli coming from the outer world. More often, it must be an intentional act performed with some effort. As the history of science shows, plaining that stands behind the eureka moments of the creative process, although is experienced as sudden and unexpected, is always provoked by the preceding thought labor. The question is how this crucial moment may be captured and turned into a more regular practice. This is not

---

[4] Plain – a large stretch of flat land, but also: clear, easy to see, hear or understand.

a new question and, for example, the creative use of metaphors is one of the answers. The present proposal is its natural extension. To this aim, next two steps can be made: *co-occurrence* and *reconciliation*.

4. Co-occurrence

You look and compare in order to choose the most suitable way or at least the one you like the most at the moment of choice.

Our brain tends to compare things. It also focuses on comparing things that are easily comparable and avoid difficult comparisons (Ariely, 2008: 7). We are always looking at things around us in relation to others, and not only physical things but states, emotions, attitudes, points of view, abstract concepts, anything. We always compare, and "analogy is the core of cognition" (Hofstadter, 2001). Most people even do not know what they want without comparison. Co-occurrence is a natural environment of comparison. Every co-occurrence of any items (experienced, perceived, thought of, imagined), is like reaching the edge of the canyon in the sense that two perspectives meet at the same point, some center of comparison. For example, the experience of passing through the shady canyon exerts influence on our attitude towards the sunny plains. Similarly, every thought and experience modify what is coming next, which may be called the extended priming effect. Even if the influence is subtle or not obvious, it seems that it always affects our ways to certain degree.

Most spectacular co-occurences procured by men in the fields of science (eureka effect), arts, and jokes (humor) have been described by Koestler as bisociation (Koestler, 1964). Koestler, investigating the secrets of creativity and inventions of various scientists and artists, noticed that their way of thinking is not associative but bisociative. He coined the term in order to make a distinction between the routine skills of thinking on a single plane, and the creative act, which always operates on more than one plane. The former can be called single-minded, the latter double-minded, "transitory state of unstable equilibrium where the balance of both emotion and thought is disturbed" (Koestler, 1964: 34–35). For Koestler, bisociation is the mixture of concepts taken from two contexts or categories of objects that are normally considered separate by the literal processes of the mind. It means to join unrelated, often conflicting, information in a new way. The pattern underlying the creative act is "the perceiving of a situation in two self-consistent but habitually incompatible frames of reference" (*ibid.*: 35). This has been justly objected by Hans Lenk who noticed that we frequently deal with more than two plains and, therefore, there are "multiple collisions, collusions (playing together), confounding phenomena, interconnections and inter-stimulations of many kinds and planes…" (Lenk, 2007: 307). More general term of co-occurrence adopted here, embraces an arbitrary number of elements compared. Moreover, it covers all cases of *interstimulations* and not only those characteristics for scientific and artistic achievements or jokes.

For Koestler there are two things making co-occurrence, and thus comparing, interesting. The first is pleasure found either in beauty of any kind (art) or in good humor (jokes). The other is the benefit of introducing a new model in science. Our point of departure is different, namely, the motivation to compare what co-occurs is a need to fill in gaps. Gaps arise in a wide variety of situations when, for instance, we face the unknown or feel lost, and are just a characteristic feature of our human condition. Gaps are byproducts of canyons and directly the effect of the way our brains work. The subsequent step, reconciliation, aims at gap filling. Davidson (see: Davidson, 1984) remarked, *a propos* the role of similarity in metaphorical meaning that everything is like everything else. So what is so interesting in comparing? I think that the chance for filling a gap, *i.e.* reconciliation, is.

5. Reconciliation

Reconciliation, like a puncture, forces us to stop and change…

Reconciliation is filling in gaps. Practically, it is a co-occurrence that induces change in thinking, *i.e.* a canyon shift which later may be manifest in action. But this change is always motivated by the discomfort of a gap. Any arbitrary co-occurrence may result in a kind of "puncture" in the way of thinking and resembles re-pumping of our mind. Reconciliation, by definition concerns only the switch of thinking and not the sways of moods nor cases of changes of cognition connected with physiology as, for example, the dependence of our sense of time flow on blood pressure. It must be admitted, however, that it often follows feelings[5]

---

[5] That affect is not necessarily post-cognitive was for the first time claimed by Robert B. Zajonc.

and is not so much the outcome of reasoning and argumentation as of mere comparing, as eight examples presented in the sequel show. It should be emphasized that a novel canyon introduced by it can transform perception, meaning, understanding, attitudes, and deeds.

Comparing is a vastly discussed topic. From Aristotle up till now a lot has been written on similarity, resemblance, analogy, common characteristics (Richards), associated commonplaces (Black), potential connotations (Beardsley), entrenched associations (Grady), or in more dynamic terms, mapping (Lakoff, Johnson), redescription (Black), and many other terms linked with metaphor and comparing. Although it is impossible to deal with this abundance in this paper, some traits of a comparing act must be specified to describe how and when it may result in reconciliation. It would be convenient to preserve the cognitivist terms for two domains compared (usually two are considered but I do not exclude more): source and target. Typically the target system is new or abstract – it is to be understood, explained, investigated, while the source system is the one in terms of which the target system is described, it is familiar and perhaps visualisable. Source and target correspond, respectively, to:

– the elements of metaphor structure *My dog is a donkey* (*dog* – TARGET; *donkey* – SOURCE);

– to scientific field of investigation (TARGET) and its model (SOURCE), *e.g.* sound waves – water waves;

– to a problem (TARGET) and its solution (SOURCE), *e.g.* MURDERER – a beast or a virus;

– to something unnamed (TARGET) and its proposed name (SOURCE) x – idea. The last one is the case of Black's catachresis.

As the authors of *Mental Leap* observed, similarity, structure, and aim constraint comparing (mapping, analogy) (Holyoak, Thagard, 1995). I claim that the predominant aim of comparing is to fill in the gaps and not mere interest or curiosity mentioned by the authors. Koestler's art and joke bisociations as not induced by gaps are not cases of reconciliations. The examples of reconciliations are:

1. Sensory illusion[6].
2. Metaphor.
3. Model in science.
4. Advert especially mind seducing.

---

[6] Exceptional character of a gap.

5. Some psycho-manipulation techniques.
6. Biblical parable.
7. Generative metaphor.
8. Psychotherapeutic story.

They all consist in seeing something as something else and are gap-oriented.

	Co-occurence	Reconciliation/ puncture	They
Sensory illusions	of two (or more) physical items	brain cognitive gap	cheat the senses
Metaphors (linguistic)	of two (or more) words, phrases, sentences	gap in meaning	express otherwise inexpressible[*]
Scientific models	of two domains (one theoretical)	gap in understanding why or how	elucidate
Seducing adverts Psychomanipulation	of two (or more) visual, auditory and linguistic items	gap in control	seduce to do something
Parables	of two layers of reality: physical and metaphysical	gap in access	reveal what is otherwise covered
Generative metaphors Psychotherapeutic story	of two phenomena: problematic and generative of two stories: problematic real life story and generative reference story	gap in solution	inspire to find solution

[*] It seems that even paraphrasable metaphors are created in the situation when the proper meaning is sought and not found in literary language. Besides, there are also gap-lacking metaphors used as pure decoration.

All the above examples are briefly described below.

1. Optical illusions boggle us but what we know for certain is the fact that they result from such a co-occurrence of elements that exposes gaps in our brain's fragile sense of reality. For example, two lines being actually perfectly straight and parallel look like they were bulging outward in the middle. Or if you hold two boxes, one large and the other small, of equal weight, you will perceive the larger box as lighter. One feels these illusions despite all of the information to the contrary.

That is the most tactile case of the power of co-occurrence and reconciliation revealing how helpless we may be while experiencing them.

2. However, such a feeling of helplessness seldom accompanies metaphors. Black initiated the change of paradigm concerning the metaphor (Black, 1962b). Since then metaphor has made a brilliant career leaving the exile of poetry, where it was admired but not treated seriously, and becoming the main mechanism of thinking. Metaphor may be described by means of three properties spelled out by Aristotle (*Poetics*, 1457b): transference, transformation (of meaning), and similarity/analogy (Rybarkiewicz, 1997). Notice that these properties are all known by various names in the philosophical, semantic, pragmatic, cognitive, literature. Aristotle's transference is a special co-occurrence of words of which one is used beyond its usual context, like "broken-winged bird" in:

Life is a broken-winged bird
That cannot fly.[7]

Broken-winged bird in this context is like a sudden sight of plains and as both landscapes are compared, so are the terms in the poem. The gap in meaning is filled by means of Aristotelian analogy. In the resulting reconciliation mind fixes the attention on the most accessible in a given context meaning and adheres to it. This accessibility is driven by experience so the interpretations may vary significantly. And just as only one direction of march across the plains is possible, comparing cannot be reversed (and thus metaphorical similarity is asymmetric):

(1) *His dog is a donkey.*

is not a synonym of

(2) *His donkey is a dog.*[8]

Black says that in metaphor we look at one thing through the lens of the other. Nevertheless, as more recent theories point out, both elements constituting metaphor influence each other and the meaning of the apparently neutral element – target – is also altered (for example, blending theory).

---

[7] From *Dreams* by Langston Hughes.

[8] Different features are the clue of interpretation: in (1) (possibly) obstinacy and in (2) (maybe) attachment.

The role of metaphorical co-occurrence has been thoroughly discussed in a great number of areas. Generally, it helps to convey complex information in brief time. It reserves blank space for individual interpretation, highlights and hides particular aspects of the presented information. Due to these advantages it has been adopted as a handy tool in many areas, some as remote from language as visual arts, music, dance, or architecture. Referring to these cases by means of *co-occurrence* and *reconciliation* permits to preserve traditional meaning of the word *metaphor*, which may be an additional advantage of the presented approach.

3. Scientific models seem to share the cognitive mechanism with metaphors. For this reason, they are also called metaphors. However, their sources and targets vary essentially, which justifies keeping separate terms for them. Many scientists who commented on the nature of their discoveries had the idea of a model and employed it in practice. For example, Huygens developed the wave theory of light making use of the known wave conception of sound, and Fourier's theory of heat distribution was based upon the analogy to the fluid dynamics laws. Scientists like Maxwell consciously applied models that were some known structures of abstract relations to construct their own theories. Black distinguishes scale, analogue, mathematical, and theoretical models and describes the conditions of their use (Black, 1962b). For example, analogue model reproduces the structure or web of relations of the original in some new medium. Its dominating principle is isomorphism. Theoretical model emphasizes the transfer of elements of one (secondary) system, which is "relatively unproblematic, more familiar, or better organized" (Black, 1962b: 230) (domain) into the primary system, which is the original field of investigation. There must be some "principle of assimilation" to start the transfer, and that principle was referred to in many various ways: analogy, intimations of similarity, framework through which one system is seen. Black calls this transfer the metaphorical redescription of the domain investigated. The transfer is a reconciliation resulting from the co-occurrence of two domains of which at least one is theoretical, the other may be "a serpent swallowing its tale".

4. Adverts are success pursuing co-occurrences directed to raising sale rate. The gap lies in the control of the market. Heath claims that the most effective are "mind seducers" (Heath, 2012: 4–6). He describes an astonishing case of advertising campaign: in 2001 a mobile network operator was launched under the name O2 and ran advertising campaign that showed blue water with bubbles bubbling through it and some lilting music in the background. The information was rather enigmatic: *O2, See what you can do.* Despite its cryptic character (blue water

and bubbles are hardly characteristics one might look for in a mobile phone), O2 had become market leader in four years. How is this possible? All schools of adverts sticking to a reliable information presented in an attractive way could be surprised by this absurd campaign. Heath suggests that the success may be attributed to the seduction of our consciousness. One of the possible explanations of how mind is seduced by such an advert is *mere exposure effect* (Zajonc, 1968). Mere exposure effect subsumes that the contact with a given thing alone suffices to change the attitude connected with this thing. It also happens in case of subliminal or peripheral exposure of less than 40 milliseconds. All such phenomena may be further explained by a Perceptual Fluency Model of R. Bornstein according to which perception without awareness “leads to inexplicable familiarity, which in turn raises favourability” (Bornstein, 1992, from: Heath, 2012: 81).

5. Closely connected with seducing are some methods of propaganda and psycho-manipulation. They also often resort to a straightforward co-occurrence. The most evident is the so called association technique already recommended in the Antiquity: Greek and Roman rhetoricians discovered that what we now call visualization (gr. *hypotyposis*) is more powerful in persuasion than any logical arguments. Some words as *flower, sun, lion, sea* immediately evoke an image thus activating various neural regions and have an overwhelming effect making us see, smell, hear and feel emotions. In brief, they bring synergy effect, which is difficult to control.

Especially Machiavellian is the co-occurrence of someone’s name and a lie. When consequently repeated, a lie like: *X is a thief* adheres to a person spoken about. Even if officially the audience declares misbelief, *X* is perceived as contaminated, which may be summed up: *X has something to do with a theft*. Blocking the canyon shift is almost beyond the will of the persuadee. Thanks to the mere exposure effect, reconciliation is flexibly adjusted to the purposes of the persuader. Notice that *X is a thief* is not a metaphor but, certainly, metaphors can also persuade with the same mechanism.

6. Parables serve not the control gap as adverts and psycho-manipulation do but are to fill in the gap in our access to otherwise inaccessible knowledge, experience, world. In the Bible they are to reveal the matters of the Kingdom of God to people. They are not mere moral instructions as some would like to think (McKenzie, 2005). In this sense a parable is a co-occurrence of two worlds: the known human, and the unknown divine. The analogies found may be so various that even the Apostles requested Jesus to explain the parables. Hermeneutics is of

much help in this case as the audience having no idea of the target (heavenly matters) has difficulty in the reconciliation; contrary to the scientists who match more known (a model) to a lesser known but known (theory they are just developing).

7. Generative metaphor is a practical strategy of problem setting that employs the typical co-occurrence aiming at reconciliation. It aims at finding a solution *via* transference of the problem somehow hulled out and set in a new context. The idea was introduced by Schön whose main field of interest was social policy (Schön, 1993). Schön noticed that dominant metaphors, *i.e.* co-occurrences within a discipline define the problems, *i.e.* canyons of this discipline and argued that the essential difficulties in social policy have more to do with problem setting than with problem solving. For Schön "the framing of problems often depends upon metaphors underlying the stories which generate problem setting and set the direction of problem solving" (*ibid.*:139).

As an example he explores the case of slum housing. If the underlying co-occurrence of a slum is a *blight* or *disease*, then this encourages a reconciliation governed by the corresponding medical remedies, including the surgery whereby the blight is removed. On the other hand, if the co-occurrence is that the slum is a *natural community*, then this orients a reconciliation in terms of enhancing the life of that community. The two co-occurrences and reconciliations are quite distinct and have quite different consequences in practice.

A variation of this kind of problem setting appears in the context of psychotherapy in a form of a story where the problematic real life story is replaced with a generative reference story seen either as a reinterpretation of the past or as a plan for future behavior. The psychological literature abounds in examples.[9]

6. Summary

1. Our thoughts flow in canyons, at neural level they tend to freeze and remain stable. For us, it means that once we have learnt something, we stay within this "knowledge" for two reasons: first, change would entail effort (which is regulated by PLE); second our brains are adjusted to survive, their engrams must be fixed to make us react quickly and properly if, for example, we come across a viper. Our tendency to stick to canyons is also encoded in language.

---

[9] See, for instance, Rosen (1982).

Petrified metaphors (Arendt) and conceptual metaphors (Lakoff and Johnson) are *petrified* proofs of this process. Usually we think, react, and act in a canyonical way and canyons are our basic assumptions.

2. Despite the resistance mentioned, we can leave a canyon. This is realized by means of a co-occurrence that presents us with a novel area to be compared, and gap-filling reconciliation which is an actual choice of the direction of the shift of canyon. I cannot think of any case of just leaving a canyon and remaining on its edge in a pre-decision state. Rather, our thinking tends to sink into another canyon, *horror vacui* at least on Earth, and either some external agents or we choose which. Two variations of the whole process are possible: aware or unaware.

3. An adequately procured co-occurrence and the resulting reconciliation is a more or less overt technique applied in manifold areas (examples above), but the main emphasis is laid upon turning it into a habitual thought pattern that not only enhances our creativity and rationality but also expands the sphere of canyons under (our own) control.

Natural field of further investigation will concern the effects of reconciliation: a canyon shift often influences categories, recalibrates assumptions, readjusts perception and touches emotions (not necessarily in this order). In consequence, purposes, decisions and concrete deeds are affected. Alas, often covertly. Then, discrete reconciliation should be another focus of inquiry.

References

Anderson, C.A., Lepper, M.R., Ross, L. (1980). Perseverance of social theories: The role of explanation in persistence of discredited information. *Journal of Personality and Social Psychology* 39: 1037–1049.

Arendt, H. (1991). *Myślenie*. Warszawa: Czytelnik.

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces that Shape our Decisions*. New York: HarperCollins Publishers.

Bornstein, R.E. (1992). Subliminal mere exposure effect. In: R.F. Bornstein, T.S. Pittman (eds.), *Percepcion without Awareness: Cognitive, Clinical and Social Perspectives* (pp. 191–210), New York: Guilford Press.

Black, M. (1962a). Metaphor. In: A. Ortony (ed.), *Models and Metaphors* (pp. 25–48). New York: Cornell University.

Black, M. (1962b). Models and Archetypes. In: A. Ortony (ed.), *Models and Metaphors* (pp. 219–243). New York: Cornell University Press.

Davidson, D. (1984). What Metaphors Mean. In: D. Davidson, *Inquires into Truth and Interpretation* (pp. 245–264). Oxford: Clarendon Press.

Eysenck, M., Keane, M. (2010). *Cognitive Psychology: A Student's Handbook*, 6th edition. Hove–New York: Psychology Press.

Forster, K., Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition* 10: 680–698.

Heath, R. (2012). *Seducing the Subconscious: The Psychology of Emotional Influence in Advertising*, Chichester: Wiley-Blackwell.

Hofstadter, D.R. (2001). *Analogy as the Core of Cognition*, Stanford Presidential Lectures in the Humanities and Arts, https://www.youtube.com/watch?v=n8m7lFQ3njk (accessed: 8.05.2014).

Holyoak, K., Thagard, P. (1995). *Mental Leaps: Analogy in Creative Thought*, Cambridge, MA: MIT Press.

Koestler, A. (1964). *The Act of Creation*, New York: Penguin Books.

Lakoff, G., Johnson, M. (1980). *Metaphors We Live By*, Chicago: University of Chicago Press.

Lenk, H. (2007). Creative ascent: Creativity through createaphors. In: F. Guttorm (ed.), *Aesthetics and Philosophy of Art*, vol. 9 (pp. 303–329). Berlin: Springer.

Liu, X., Ramirez, S., Pang (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* 484: 381–385.

McKenzie, A. (2005). Parables and proverbs. In: L. Jones (ed.), *Thomson-Gale Encyclopedia of Religion*, 2nd edition (pp. 69–76). Detroit: Thomson Gale.

Rosen S. (ed.) (1982). *My Voice Will Go with You: The Teaching Tales of Milton H. Erickson*, New York: Norton.

Rybarkiewicz, D. (1997). Three Aspects of Metaphor, *Logica Trianguli* 1: 93–103.

Schön, D.A. (1993). Generative metaphor: A perspective on problem-setting in social policy. In. A. Ortony (ed.), *Metaphor and Thought* (pp. 137–163), Cambridge: Cambridge University Press.

Zajonc, R.B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph* 9 (2): 1–27.

Zipf, G. (1949). *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press.

ANNIKA WALLIN

THE EVERYDAY OF DECISION-MAKING[1]

In this paper I will argue for the importance of studying the everyday of decision making, if we want to know more about what decision processes look like, how rational human beings are, and how we (if necessary) can improve decision making.

1. Decision making without the everyday

When issues such as rationality and decision making are discussed, we often go back to standard decision theoretic descriptions of decision making, leading to recommended strategies such as the *Principle of Maximizing Expected Utility.* In order to use this principle we have to have access to at least two courses of action that lead to different outcomes. Savage (1954) describes this situation by a chef cooking an omelet. At the time of the decision this person has broken five good eggs into a bowl, but a sixth one is available. For some reason, this egg has to be used now or thrown away (perhaps the kitchen is closing for the summer?). Savage describes the situation in the following way:

Act	State	
	good	rotten
Break into bowl	Six-egg omelet	No omelet, and five good eggs destroyed
Break into saucer	Six-egg omelet and a saucer to wash	Five-egg omelet and a saucer to wash
Throw away	Five-egg omelet and a good egg destroyed	Five-egg omelet

---

[1] This paper is partly based on the chapter “Decision making in everyday life” (Wallin, 2008), but has been updated and changed.

In order to decide upon a course of action the chef will have to decide how good or bad each possible outcome is. This is done by assessing the *utility* of a six-egg omelet, the same omelet with a saucer to wash, no good omelet and five good eggs destroyed, and so on. In addition the decision maker will have to have an idea of how probable it is that the egg is rotten. Once this is done, the *expected utility* for each possible action can be determined. For instance, the chef who thinks the probability of a rotten egg is low, and hates waste will probably break the egg directly into the bowl, whereas a cautious and hungry person will use the saucer. Bayesian decision theory states that a decision maker should always select the alternative with the highest expected utility. This is the *Principle of Maximizing Expected Utility*.

It is not entirely clear from the *Principle of Maximizing Expected Utility* how decision makers should proceed when making a decision. Should they make lists, like the one above, or do they have other options? The reason that Bayesian decision theory is relatively mute on the specifics of decision-making is that it is a normative theory. It describes what a good decision looks like, but does not presume anything about how decisions are made. Nevertheless, the most obvious option for someone attempting to maximize expected utility is to identify each possible action and state of the world, to assign each state a probability and each possible outcome an utility. But although the recommendation to do so seems very sensible, it is not altogether easy to follow. As a matter of fact, the theoretical framework developed by Savage requires an agent that has, among other things, full information and infinite sensitivity. This is obviously not true for any living decision maker, but the intention of models of ideal rational behaviour (such as Savage's) is not to be realistic, but rather to explore the nature of rational choice (*cf*. Sahlin *et al.*, 2010; Wallin, 2013).

2. Decision making without the everyday in judgment and decision-making

Interestingly, however, the ideal rational model (often named Economic man) has had a strong presence in the area of judgment and decision making despite the fact that it is obviously unrealistic. This is not entirely unsurprising: after all it is a model of rational decision making, and alternative models are scarce. Ward Edwards explicitly introduced Economic man into the field of judgment and decision-making in the 1950s:

> It is easy for a psychologist to point out that an Economic man who has the properties discussed above [complete information, infinite sensitivity, rationality] is very unlike a real man. In fact, it is so easy to point this out that psychologists have tended to reject out of hand the theories that result from these assumptions. This isn't fair. [...] The most useful thing to do with a theory is not to criticize its assumptions but rather to test its theorems. (Edwards, 1954: 382)

This approach was picked up by Amos Tversky and Daniel Kahneman's heuristics and biases research. Perhaps it culminated with a *Science* paper (Tversky, Kahneman, 1974), in which violations of some very fundamental "statistical rules" were summarized and described (*ibid.*: 1130). Among other things, participants did not seem sensitive to the prior probability of an outcome, to sample size, or to predictability. Through such deviations the authors argued for the existence of heuristics: simple decision rules described as being "highly economical and usually effective", but leading to "systematic and predictable errors" (*ibid.*: 1131). These, and similar results, have been famously described as having "bleak implications for human rationality" (Nisbett, Borgida, 1975: 935). How the results should be interpreted, whether participants violate fundamental statistical rules, and whether the implications for human rationality are truly bleak has been discussed since. The resulting debate has been named the rationality wars by Richard Samuels and colleagues (2002).

I will not focus on the rationality wars here, nor on the issue of how rationality has to be studied. I will simply point out that the research traditions that aim at studying how closely human behaviour approximates Economic man has a focus on using paper and pencil tasks where, for instance, sensitivity to prior probabilities is examined. This leads to famous and extremely well investigated tasks involving taxicabs, feminist bank tellers and Asian diseases that have little resemblance to the type of decisions and judgements we usually are engaged in. It is my strong belief that this research has to be complimented by studies of real life decision-making and this for three reasons. First, if our decision-making is shaped by anything, it is by common activities for which we receive feedback. Therefore, the decisions strategies shaped by, for instance, the tasks we accomplish in a supermarket, are more likely to influence our decision making regarding complex issues such as choice of medical treatments, voting behaviour and economic commitments than the other way around. Second, if we study familiar tasks we will get a more realistic, useful and likely positive view of human rationality. Thirdly, the study object – the tasks with which decision making is usually studied, is likely to have a great impact on what the resulting cognitive models are. In this I wholeheartedly agree with Ward Edwards when he says: "My own guess is that most successful models now available are successful exactly because of their success in describing tasks, not people" (Edwards, 1971: 640).

3. Why is it important to focus on the everyday?

Looking at real life decision situations it is obvious that decision makers do not compute the expected utility of each possible choice, as has been suggested by decision theoretical approaches to decision making. The important question, in my mind, then becomes how decisions are made – in real life settings – and in general.

One way in which my colleagues and I have tried to understand this is by investigating consumer behaviour in stores. By approaching people that do their everyday shopping we get a completely different decision task; a task that is familiar, where people are motivated, and where they over time receive feedback on the trade off between effort and decision outcomes.

In one of our studies we observed shoppers buying (among other things) jam in a supermarket. These consumers had a choice of 91 different types of jam, all differing in taste, size, price, land of origin etc. in up to at least 90 different attributes. Despite this plethora of information, the average participant spent only approximately half a minute contemplating his or her choice (Gidlöf *et al.*, 2013b). If Bayesian decision theory is interpreted as a recommendation to compute the expected utility of every possible jar of jam, and to select the one that best satisfies the decision makers' demands, grocery shopping would literally take forever. This state of affairs has been obvious to all researchers on decision-making, but regardless of this fact, the *Principle of Maximizing Expected Utility* was for a long time the *only* available description of how decisions are made. It was so ingrained in our way of thinking that researchers have spent decades trying to prove that it does not describe decisions in everyday life (see above). But if people do not list utilities and probabilities when they make a decision, what do they do? Can less taxing decision-making strategies still be reasonably successful? The answer is yes, but as we will see, decision quality then depends on the situation in which it is made. This is why I think that everyday decision-making is important.

So then, what do consumers do it if they do not maximize expected utility?

One possibility is that the consumers bought the same kind of jam they ordinarily prefer, not even trying to update their knowledge about the other available kinds. Such a strategy saves time, and given that the shoppers are happy with their default choice, it is bound to be reasonably successful. Note however that these consumers run the risk of ignoring an even better option – perhaps a jam similar to their preferred one in all respects except a lower price. Thus a shopper sticking with what he or she knows is willing to trade an optimal choice for a quick, sufficiently good one. Herbert Simon (1956) called this principle of trying to find something "good enough" *satisficing*.

There are many strategies of this type. Decision makers can, for instance, use the strategy *elimination by aspects*: first compare all jams with respect to one of their attributes, such as their price. If one jar is cheaper than the rest, pick it. If several jars remain, compare these (and only these) with respect to another attribute, such as how much fruit they contain. Continue searching through the attributes in this way until you are left with only one jar of jam (Tversky, 1972). It is of course possible that there is another jar of jam somewhere on the shelf that has a better combination of attributes than the one you picked. Thus, a consumer eliminating by aspects is also satisficing. They trade the cost of not finding the best possible option for the gain of not having to spend too much time and energy searching for the jam they buy. Such strategies are generally quicker and less demanding than comparing the utility of each option.

It is clear that people use relatively little time and information in most everyday decisions. To remain within the supermarket setting, Hoyer (1984) observed consumers buying detergent and found that they spent on average 13 seconds in the isle. When they were approached and interviewed about their decisions an overwhelming majority (90%) gave merely one reason for their choice such as its price, or their experience with the products' performance. That consumers pay little attention in such everyday decisions is further confirmed by the fact that if you approach consumers just after they have selected a product not even half of them can correctly estimate its price, and a fifth cannot even give a rough estimate (Dickson, Sawyer, 1990). It is highly unlikely that these decisions were made according to Savage's recommendations. The consumers' behaviour is far more similar to Simon's satisficers.

A decision maker using a relatively easily applied decision rule may also get additional advantages. There is some evidence that a decision that is too complex decreases in quality, both with respect to how satisfied and confident decision makers are with their choice and with respect to how likely they are to actually decide. One way in which this can happen is if decision makers are given too much information. This phenomenon has been given several different names over the years: information overload, choice deferral, the too much choice effect, and the paradox of choice. A particularly elegant demonstration of how complexity negatively affects decision makers is a study by the psychologists Sheena Iyengar and Mark Lepper (2000). They placed a tasting booth at an upscale supermarket where shoppers could stop and sample a variety of exotic jams. Some days the variety was larger, so that 24 types of jams were on offer. Other days only a subset of 6 of these 24 jams could be tasted (selected so that it contained the 2 most and least popular jams, and two jams of medium popularity). The shoppers were about twice as likely to try some of the 24 jams than they were to do so with the smaller set. But far more of those that stopped and sampled from the small selection actually bought a jar of jam.

One way this phenomenon has been explained is that it is too difficult to make a choice when the variety is large. A consumer choosing from a large variety will be confronted with more options, and someone trying to make an informed choice will thus have to consider more information. As a matter of fact, just telling people more about the properties of different options – such as the qualities of a particular jam – is enough to make choice more difficult and sometimes too difficult (Malhortra, 1982). Naturally the more information a decision rule requires, the more affected will the decision maker be when the amount of information increases. Thus it is possible that decision makers that rely on rules that require little information are less affected by the too much choice effect, that is, when they have to choose among many options, or are given lots of information about these options.

On the other hand it is clear that a decision maker that ignores information runs the risk of making a bad choice in the sense that s/he misses an even better opportunity. How worried should we be by this possibility? How much do decision makers loose by satisficing instead of optimizing? Even a satisficing decision maker may end up with the best possible option, however that is defined. To determine how damaging it is to ignore information we have to consider also the *environment* in which a decision is made. For instance, simulations reveal that a decision maker looking at a single most important attribute – say, price in the case of grocery shopping – is still highly likely to select one of the best available options (within 90% of the best possible value) if the environment is such that attractive features such as low price and a high content of fruit are positively correlated (Fasolo *et al.*, 2007). For decisions in which positive attributes co-vary, decision makers that ignore all information except that pertaining to a single subjectively important attribute will still make highly successful predictions. When positive attributes do not co-vary, ignoring information may lead to worse outcomes. Thus an important aspect of decision-making is to select strategies that fit the environments in which they are used. Research in judgment and decision making indicate that people's decision strategies vary with types of tasks and decision environments (see for instance: Payne *et al.*, 1993; Gigerenzer *et al.*, 1999), and we know from our research that consumers in supermarkets are moderately successful (Gidlöf *et al.*, 2013a). This leaves us with an important research questions: which strategies work well in which decision environments? (later on we will also ask how different decision strategies are activated).

4. Which strategies work well in which decision environments?

Decision strategies used in everyday life have to be frugal (*i.e.* require little information) and fast. It is impossible for a shopper facing almost a hundred types of jam to process all information carefully. Does this mean that we are bound to make bad choices? Actually not. Depending on the decision environment, fast and frugal decision strategies can be as successful as more complex strategies. They can even sometimes do better.

Imagine, for instance, a group of American and German students trying to determine for long lists of randomly selected pairs of cities, which is the larger one. Is Dortmund larger than Munich or vice versa? And what about San Diego and Sacramento? When actual students are given the same choices a remarkable thing happens: the American students make equally good predictions for German cities as they do for American ones (Gigerenzer, Goldstein, 1996; Goldstein, Gigerenzer, 1999; 2002). How can this be?

The answer appears to be that ignorance is not random. Students, and people in general, tend to recognize cities that are larger. Gigerenzer and Goldstein compared the number of mentions German cities got in the *New York Times*, and American cities in *Die Zeit*, and found a generally strong correlation between number of mentions and population size. If recognition is mediated by, for instance, mentions in the press, the *ecological validity* of recognition is, in this case, quite high. This means that a relatively ignorant student can utilize the non-random character of his or her ignorance when making a choice. Students that are familiar with both cities cannot, and have to rely on other information, such as the fact that Sacramento is the state capital of California. Such knowledge can be misleading (San Diego is actually bigger) and in such circumstances, knowing less can outperform knowing more. Ignorant people can make better choices – when the environment is right.

The recognition experiment demonstrates that it is sometimes an advantage to be ignorant. But there is other evidence that simple decision rules, such as taking the city you recognize (the recognition heuristic), may do well. Czerlinski and colleagues (1999) constructed 20 decision environments in which decision makers (in this case simulations following pre-defined decision rules) could search information about pairs of options and pick their winner, just as in the city example above. They were, for instance, required to predict dropout rates at high schools in Chicago given information about such things as their proportion of low-income students, non-white students and SAT scores. Another task was to predict the selling prices of houses in Pennsylvania based on information such as property taxes, garage space and total living space. These decision environments were based on statistics textbooks and reports, and were as true to real life as they could be. The authors constructed different rules that could be used to predict which house would be more expensive or which high school would have fewer dropouts. Some were relatively complicated such as multiple regressions, and some added the pros and cons without assigning them any weights (Dawes' rule). Other rules were even more simple, such as a rule that basically searched information in the order of its validity and made a decision as soon as one piece of information differentiated between the options. Since the rule searches information in order of its validity, that is, how accurate a decision based on this information will be, the rule was called take-the-best (note its similarity to elimination-by-aspects). The decision rules were then allowed to train on half of the available data sets (setting weights, or learning the validity of the different pieces of information as best as they could), and to make predictions for the other half. In this task the take-the-best-rule surprisingly outperformed both multiple regression, and Dawes' rule. A rule that ignored most of the available information thus made better predictions than rules that took all of it into account. How can this be? The most probable explanation is that the more complicated rules take information into consideration that does not help them make correct predictions. They take too much of the information in the training set into account – and "overfit" to circumstances that do not generalize (*cf*. Martignon, Hoffrage, 1999).

The conclusion to draw from the above is that sometimes using less information can be better – when the environment is such that it pays off. Nobel laureate Herbert Simon put it in the following way: "Human rational behavior is shaped by a scissors whose two blades are the structure of the task environment and the computational capabilities of the actor" (1990: 7). It is clear that most human endeavours (and in particular everyday decision-making) have to use limited time and resources. The fact that simple decision rules can perform so well thus gives us good reasons to assume that it is these types of strategies that people use in their day-to-day activities. Whether this intuition is true will have to be, and is currently, tested in experimental studies. They will not be covered here. Instead the next question we will turn to is how people recognize which environment they are in.

5. When are different decision rules activated?

Oppenheimer (2003) challenged Gigerenzer and Goldstein's recognition heuristic by constructing pairs of cities in which some where well-known and small (such as Chernobyl and Los Alamos) whereas others did not exist, but sounded like they were situated in densely inhabited areas (such as, Al Ahbahib, Weingshe, and Las Besas). Not surprisingly, participants were more likely to guess that the non-existing (and therefore not recognized) cities were larger. In a way this is not surprising, since these pairs of cities were explicitly constructed to recognitions' disadvantage. However it points to an important question: what determines whether a decision rule is activated or not?

One answer could lie in adaptation. With time and experience decision makers learn that a particular strategy is generally successful in a certain type of settings (*e.g.* Payne *et al.*, 1993; Gigerenzer *et al.*, 1999). Such an adaptation can occur at both an individual and a cultural level (Wallin, 2007). Cultural and individual adaptation can also explain how particular decision rules could function unproblematically in relatively unknown environments. For instance, decision makers use possible causal relations between cues and criteria (the available information, and the outcome to be predicted) as a proxy for the ecological validity of each cue (Garcia-Retarmero *et al.*, 2007). In this way, a rule such as take-the-best can be used also for environments in which cue validity is unknown.

Another answer is that we often have no choice. Given the type of situation we are in, the time, information and computational resources available for a decision may be small, and thus force us to use a relatively fast a frugal decision rule. A wonderful example of this is a study by Ap Dijksterhuis and colleagues (2006). They presented participants with more or less complex decisions, such as choosing a house or a car. One group of participants were given time to think about their choice, whereas others were prevented from conscious deliberation through a distractor task. Those that were given time to think about their choices were less likely to pick the best out of four cars when the choice was complex (*i.e.* the options had 24 attributes). When the choice was not complex (the options had only 4 attributes) those that had thought about their choice did better. Dijksterhuis and associates explain this by the advantages of unconscious processing. Another explanation, that better fits the theme of this chapter, is that even if the environment determines that little time will be spent on a decision, reducing the information that is consciously processed in order to reach it might improve decision-making.

Naturally there are also differences between people in how much time they spend on a decision, or which decision rules they use. There have been attempts to measure this as a personality trait with questionnaires such as "need for cognition". Dijksterhuis and colleagues complemented their laboratory studies with one in real life. They looked a choices made in two different actual stores. One of them offers relatively complex choices (IKEA) whereas the other sells more simple products such as clothing accessories (Bijenkorf). Costumers were approached as they exited the store and were asked to fill in a questionnaire aimed at identifying consumers who had thought a lot, or relatively little about their choice. Some weeks later the customers were contacted and asked how happy they were with their choice. Non-thinking consumers were generally more happy with their complex choices, than thinking consumers were, and the opposite effect was found for the less complex products.

6. Summary

The goal of this chapter has been to show that everyday decision making matters. It is not enough to construct formal, normative, models of how decisions should be made, because these models are difficult to apply to real life decision tasks. Furthermore, the environments in which decisions are made to a large extent determine the success of different decision rules. Given the right environment, ignorance can even outperform knowledge. There are many, highly useful, shortcuts to good decision making, such as exploiting ignorance, imitating others, or ignoring less important information. Such decision strategies can only be fully understood if we study the environments of everyday reasoning.

References

Czerlinski, J., Gigerenzer, G., Goldstein, D. (1999). The recognition heuristic: How ignorance makes us smart. In: G. Gigerenzer, P.M. Todd, The ABC Group, *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

Dickson, P.R., Sawyer, A.G. (1990). The price knowledge and search of supermarket shoppers. *Journal of Marketing* 54: 42–53.

Edwards, W. (1954). The theory of decision-making. *Psychological Bulletin* 51 (4): 380–417.

Edwards, W. (1971). Bayesian and regression models of human information processing – a myopic perspective. *Organizational Behavior and Human Performance* 6: 639–648.

Dijksterhuis, A., Bos, M.W., Nordgren, L.F., van Baaren, R.B. (2006). On making the right choice: The deliberation-without-attention effect. *Science* 311: 1005–1007.

Fasolo, B., McClelland, G., Todd, P.M. (2007). Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory* 7: 13–26.

Garcia-Retarmero, R., Wallin, A., Diekmann, A. (2007). Does causal knowledge help us be faster and more frugal in our decisions? *Memory and Cognition* 35 (6): 1399–1409.

Gidlöf, K., Wallin, A., Dewhurst, R., Holmqvist, K. (2013a). Using eye-tracking to trace a cognitive process: Gaze behavior during decision making in a natural environment. *Journal of Eye Movement Research* 6 (1): 1–14.

Gidlöf, K., Wallin, A., Mögelvang-Hansen, P., Holmqvist, K. (2013b). Material distortion of economic behavior and everyday decision making. *Journal of Consumer Policy* 36 (4): 389–402.

Gigerenzer, G., Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 98: 648–704.

Gigerenzer, G., Todd, P.M., The ABC Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Goldstein, D., Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In: G. Gigerenzer, P.M. Todd, The ABC Group, *Simple heuristics that make us smart* (pp. 37–58). New York: Oxford University Press.

Goldstein, D., Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review* 109: 75–90.

Hoyer, W.D. (1984). An examination of consumer decision making for a common repeat purchase product. *Journal of Consumer Research* 11, 822–829.

Iyengar, S.S., Lepper, M.R. (2000). When choice is demotivating: Can one desire too much of a good thing. *Journal of Personality and Social Psychology* 79, 995–1006.

Malhortra, N.K. (1982). Information load and consumer decision making. *Journal of Consumer Research* 8: 419–430.

Martignon, L., Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In: G. Gigerenzer, P.M. Todd, The ABC Group, *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.

Nisbett, R.E., Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology* 32 (5): 932–943.

Oppenheimer, D.M. (2003). Not so fast! (and not so frugal!) Rethinking the recognition heuristic. *Cognition* 90: B1–B9.

Payne, J.W., Bettman, J.R., Johnson, E.J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.

Sahlin, N.-E., Wallin, A., Persson, J. (2010). Decision science: From Ramsey to dual process theories. *Synthese* 172: 129–143.

Samuels, R., Stich, S., Bishop, M. (2002). Ending the rationality wars: How to make disputes about human rationality disappear. In: R. Elio (ed.), *Common sense, reasoning and rationality* (pp. 236–268). Oxford: Oxford University Press.

Savage, L.J. (1954). *The foundations of statistics*. New York: Dover Publications Inc.

Simon, H. (1956). A behavioral model of rational choice. *Quarterly Journal of Economics* 69: 99–118.

Simon, H. (1990). Invariants of human behavior. *Annual Review of Psychology* 41: 1–19.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review* 79: 281–299.

Tversky, A., Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.

Wallin, A. (2007). Explanation and environment: The case of psychology. In: J. Persson, P. Ylikoski (eds.), *Rethinking Explanation* (pp. 163–176). Boston Studies in the Philosophy of Science. Dordrecht: Springer Verlag.

Wallin, A. (2008). Decision making in everyday life. In: P. Gärdenfors, A. Wallin (eds.). *A Smorgasbord of Cognitive Science* (pp. 249–264). Nora: Nya Doxa.

Wallin, A. (2013). A peace treatise for the rationality wars? External validity and its relation to normative and descriptive theories of rationality. *Theory and Psychology* 23 (4): 458–478.

Magdalena Grothe[1]
Bartosz Żukowski

SHORT- AND LONG-TERM SOCIAL INTERACTIONS FROM THE GAME THEORETICAL PERSPECTIVE: A COGNITIVE APPROACH

1. Introduction

Game theory is a growing field of analytical modelling of interactions between agents. It has found many applications in economics and business administration, as well as in social and cognitive sciences. Among various forms of one-off and multi-period games, the probably most popular one is known as a prisoner's dilemma. This paper analyses the social interactions from the perspective of the equilibria resulting from the prisoner's dilemma models with a short and indefinite time horizon.

A one-period interaction modelled in the prisoner's dilemma setting results in an equilibrium point where the agents are maximising their short-term profit and thus, interacting, they reach a suboptimal point of overall welfare. However, in a model with an indefinite time framework, it is likely to reach an equilibrium which is consistent with the cooperative behaviour over a longer period of time. In result, an optimal point from the perspective of welfare is reached. In the following analysis, we relate these model results to social interactions and discuss their possible applications in cognitive science.

The paper is related to the literature on game theory and its applications in social interactions. Research in game theory dates back to the nineteenth century, when, among others, the theories of strategic interaction in oligopolies have been

---

[1] Views expressed in this paper do not necessarily represent the views of the European Central Bank.

developed (see, *e.g.* Bertrand, 1883, or Varian, 2006 for an overview). In the first half of the 20th century, Neumann and Morgenstern (1944) applied game theory to the economic interactions among agents who maximise their utility (see also Neumann, 1928 or Copeland, 1945 for the review of Neumann, Morgenstern, 1944). Following this, the field grew intensively, with the range of underlying concepts developed and extensive applications to many areas of economic research (see, *e.g.* Shapiro, 1989; Kreps, 1990 and Selten, 1999). In particular, in 1994 John C. Harsanyi, John F. Nash Jr. and Reinhard Selten received the Nobel Prize in Economic Sciences "for their pioneering analysis of equilibria in the theory of non-cooperative games". Following this work, game theoretical results have been validated from the behavioural perspective, as well as by experiments (see, *e.g.* Smith, 1992 or Crawford, 1997),[2] while the framework to analyse conflict and cooperation of economic agents has been developed further.[3] In cognitive science, game theory has been also widely used for a variety of applications and research questions (see, *e.g.* Carlsson, 1998; Skyrms, 2010). The contribution of this paper is to add some new aspects to this discussion.

The remainder of the paper is structured as follows. Section 2 outlines a basic model of the prisoner's dilemma, with focus on a one-period and multi-period interactions. Section 3 analyses social interactions from the perspective of the equilibria of prisoner's dilemma game, as derived in section 2. Section 4 concludes.

2. Prisoner's dilemma

The most popular model in the game theory is the prisoner's dilemma. The game stems originally from the research conducted for military purposes in the middle of the 20th century. The following section discusses, first, the one-period interaction, outlining the game's set-up, the Nash equilibrium and the welfare calculation. The second part of the section presents the iterated prisoner's dilemma, discussing possible strategies and their dominance.

---

[2] 2002 Nobel Prize in Economic Sciences for Daniel Kahneman "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty" and for Vernon L. Smith "for having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms".

[3] 2005 Nobel Prize in Economic Sciences for Robert J. Aumann and Thomas C. Schelling "for having enhanced our understanding of conflict and cooperation through game-theory analysis".

2.1. One-period interaction

Prisoner's dilemma game is a standard example of a cooperation problem. The game in its simple set-up has two parties, who interact in a one-off meeting. Originally, the game formulated by M. Flood and M. Dresher in the context of military-related research was formalised in Kuhn and Tucker (1950). The game is described as a problem of two prisoners, accused of committing jointly a crime, and facing an interrogation. Each of them can decide, without consulting with the other one, to admit the crime or not. If they both refuse to confess, they will both get only a mild conviction – an outcome resulting in a maximal joint welfare. In case one of the prisoners confesses, while the other will deny the crime, the latter will suffer a long conviction, while the confessing prisoner will be set free – an outcome implying a maximum profit for one of the parties and a maximal loss for another party.

In a more general setting, each of the players faces a simple choice: to "cooperate" or "defect". The outcomes of the game for each of the players can be summarised as follows:

		Player II	
		cooperate	defect
Player I	cooperate	(A,A)	(B,C)
	defect	(C,B)	(D,D)

Fig. 1. Matrix of outcomes for one-period game

where $C > A > D > B$, and $2A > (B + C) > 2D$. In a numerical example, the above setting would correspond to the following:

		Player II	
		cooperate	defect
Player I	cooperate	(100,100)	(0,150)
	defect	(150,0)	(25,25)

Fig. 2. Matrix of payoffs for one-period game

The outcome always denotes the payoff in units of each of the players, depending on their action as well as the action of the counterparty (*i.e.* units received by player I, units received by player II). The contingencies involve a choice between "cooperate" and "defect" for each player. This would correspond to "refuse to confess" and "confess" in an original example of the game.

Analysing the possible outcomes from the overall welfare perspective, it is generally optimal that both players "cooperate" with each other, as they would maximise the joint payoff and minimise the inequality of the payoff distribution. However, from the perspective of a strategy, which is an optimal response to the other players' strategies, *i.e.* from the perspective of Nash equilibrium, the optimal strategy is quite different. Nash equilibrium can be defined as follows. Theorem as in Nash (1950):

Let $(P, R)_{m\times n}$ be the payoffs of a bimatrix game.[4] Then there exists a mixed strategy $x^* = (x_1{}^*, x_2{}^*, \dots, x_m{}^*)$ for player I and a mixed strategy $y^* = (y_1{}^*, y_2{}^*, \dots, y_m{}^*)$ for player II such that for any mixed strategy $x = (x_1, x_2, \dots x_m)$ for player I and for any mixed strategy $y = (y_1, y_2, \dots y_m)$ for player II,

$$\left\langle x^*, Py^*\right\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}x_i^*y_j^* \geq \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}x_i\,y_j^* = \left\langle x, Py^*\right\rangle,$$

and

$$\left\langle x^*, Ry^*\right\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij}x_i^*y_j^* \geq \sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij}x_i^*y_j = \left\langle x^*, Ry\right\rangle.$$

For a broader discussion, see also Raghavan (2002) and Hillas and Kohlberg (2002), as well as Khan and Sun (2002), all in Aumann and Hart (2002).

Nash equilibrium of the prisoner's dilemma game implies the outcome, where both players defect and reach the jointly minimal payoff. The Nash equilibrium is reached by the players following their individual profit maximising strategy, which implies that they would prefer the choice yielding their personal maximum payoff. In particular, the decision faced by each player is that the choice of the strategy "cooperate" is less productive than the choice of the strategy "defect", *i.e.* C > A in the example above. Knowing this, each player anticipates the strategy of the counterparty, and tries to maximise his profit, given the strat-

---

[4] A bimatrix game is a finite non-cooperative game with two players I and II, with payoff matrices *P* and *R*, respectively.

egy, *i.e.* max((B,D)|counterparty strategy "defect"). Chammach and Rapoport (1965) show that the analogous result for the Nash equilibrium holds for several available strategies.

Another way to present the same problem is the diagram of choices of one of the players, depending on the unknown decision of another player:

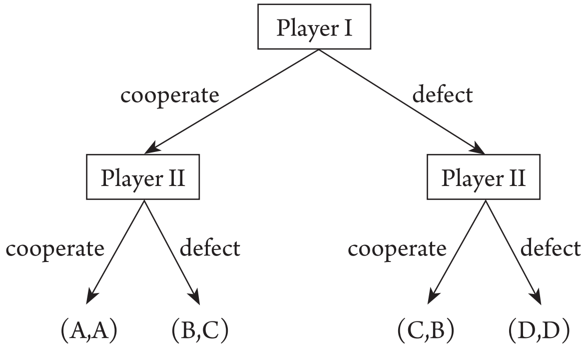

Fig. 3. Diagram of choices for one-period game

In this form, a solution can be found by dividing the game into parts, so-called subgames, with subgame perfect Nash equilibria (see also Harsanyi, Selten, 1988). In the example above, the overall game would consist of two subgames played by player II, with the choice of player I set as exogenous in each of the subgames. The Nash equilibrium derived in this way is also ("defect", "defect").

Overall, the one-period prisoner's dilemma game implies that if players follow a rational expectation of anticipating the counterparty's move, the equilibrium will always be suboptimal form the perspective of (i) individual payoffs and (ii) the overall welfare.

2.2. Multi-period interactions

In a repeated prisoner's dilemma game, the uncooperative equilibrium is not straightforward any more. The cooperation can be theoretically supported either in an infinitely repeated game (see, *e.g.* Rubinstein, 1982), or in a game with a finite number of iterations and an uncertainty about the counterparty's strategy (see, *e.g.* Kreps *et al.*, 1982).

As a starting point, consider a two-period game of a prisoner's dilemma with a payoff structure as in section 2.1. The cumulative payoffs of players I and II, depending on possible sequences of choices between "cooperate" and "defect" strategies, are the following:

Iteration 1	Iteration 2	Sum payoff
(c,c)	(c,c)	(200,200)
	(c,d)	(100,250)
	(d,c)	(250,100)
	(d,d)	(125,125)
(c,d)	(c,c)	(100,250)
	(c,d)	(0,300)
	(d,c)	(150,150)
	(d,d)	(25,175)
(d,c)	(c,c)	(250,100)
	(c,d)	(150,150)
	(d,c)	(300,0)
	(d,d)	(175,25)
(d,d)	(c,c)	(125,125)
	(c,d)	(25,175)
	(d,c)	(175,25)
	(d,d)	(50,50)

Fig. 4. Matrix of payoffs for multi-period game

The solution of this game can be found by identifying subgame perfect Nash equilibria and using backward induction (see, *e.g.* Harsanyi, Selten, 1988). In the final iteration of the game, each player decides between their two strategies, contingent on the outcome of the previous iterations, *i.e.* starting in the respective point in the decision tree. The above calculation shows that for any outcome of iteration 1, the strategy "defect" is optimal, in the sense of Nash equilibrium, for each player. As a result, through backward induction, the players would have to decide on their moves for iteration 1, knowing that the outcome of iteration 2 will be ("defect", "defect"). Comparing the payoffs from rows with (d, d) in iteration 2, again, a Nash equilibrium would be ("defect", "defect") in iteration 1.

The same logic can be applied to finitely repeated games. The players anticipate the last round, where it is optimal (in the Nash sense) to defect, and defect in the previous round, going along the decision tree back to the first round.

However, when the horizon is infinite or the end point is distant and uncertain, other equilibria may occur. One of the possibilities is a strategy, which starts with a cooperative move and always reflects the move of the opponent in the previous round, the so-called "tit-for-tat" strategy. Kreps, Milgrom, Roberts, and Wilson (1982) show that in a prisoner's dilemma game with a small probability of a small fraction of players having a "tit-for-tat" strategy, this strategy starts to dominate and other players converge to it as well. As an outcome, players may as well reach a welfare-optimal equilibrium of cooperation in each round. Confirming this empirically, Axelrod (1984) reports on a contest of an N-step prisonner's dilemma, where participants needed to submit the strategies. In fact, the "tit-for-tat" strategy, submitted by A. Rapoport, appeared to be the dominating strategy in the contest (see also, *e.g.* Chammach, Rapoport 1965, and Rapoport 1970).

To summarise, this section illustrates how the same set-up of interactions can accommodate materially different equilibria, only depending on the horizon of the interactions. In particular, in a game with a one-off interaction, an uncooperative equilibrium is achieved, where agents receive their second-worst pay-offs and overall welfare is suboptimal. A similar result is achieved in a multi-step setting with the end-point of interaction being commonly known and not distant. In contrast, a setting of multi-period interactions with indefinite (or very distant) end-point, may likely result in a cooperative behaviour with higher individual pay-offs and, consequently, a higher overall welfare. All these results are based on rational maximisation of profits from the individual perspective.

In the next section, the theoretic equilibria of one-period and multi-period prisoner's dilemma are related to social interactions.

3. Social interactions in the equilibrium of prisoner's dilemma – cognitive applications

The results discussed in section 2 can be used in social sciences in two distinct but complementary ways. First, they can serve as theoretical explanatory model to understand different attitudes of social agents towards cooperative behaviours, depending on whether the horizon of social interaction is finite or not. Second, from the point of view of cognitive science, they provide a successful computational framework to model complex interactions between intelligent agents.

The explanatory power of the above analysis may seem rather obvious. The correlation between the anticipated horizon of interaction and the increase of cooperative attitudes appears to be fully compatible with everyday experience. In case of both commercial and social interactions, individuals seem to be more likely to cooperate depending on the expected horizon of coexistence. The same is supposed to be true even for international relations.[5] What is less obvious is how this observation can be used as a persuasive or negotiating tool to obtain optimal point from the perspective of individual payoffs or overall welfare.

The most intriguing aspect of the results presented above is that they can provide a sufficient mathematical explanation for the whole complex of phenomena documented in social psychology. We mean the well-documented fact that the willingness to cooperate increases if agents are faced with potential infinity and inevitability of interaction (see, *e.g.* Aronson 2011, Darley, Berscheid 1967). For example, it has been observed that xenophobic attitudes decrease as a result of the interaction between people of different ethnic groups, provided that it was seen as unavoidable for the future coexistence (Aronson 2011).

Conventionally, these socio-psychological phenomena have been explained in purely psychological terms. They can, for example, be interpreted within behavioural learning paradigm (as one of many instances of the attitudes shaped by the behaviours), or, alternatively, regarded as an example of cognitive-dissonance reduction (Aronson 2011). Sometimes they are simply considered to be self-explanatory, or at least explainable by common-sense reasons.

While being intuitively appealing, conventional approaches seem insufficient for a precise and formal analysis of social interactions. Also, predictions about likely human behaviours need a more formal underlying model. Consequently, conventional approaches to understanding socio-psychological phenomena are of limited use for the purposes of the operationally-oriented cognitive science. This leads us to the second point mentioned above, *i.e.* the use of computational framework to model complex interactions between intelligent agents. In contrast to the socio-psychological explanations, the mathematical framework presented in section 2 provides sufficient formal theory for computational modelling of attitude changes in short-term and longer-term social interactions.

---

[5] However, the issue becomes more complex when more sophisticated models are considered. In at least some cases the results considered here can lead to counterintuitive conclusions. Consider, for instance, arms race games or war games with the conventional/nuclear distinction as a determinant of interaction horizon.

While the theory is very appealing, there still remains a methodological problem of adequacy, *i.e.*, the correspondence of theory to reality. In other words, it seems important to ask whether the theory is true. While this issue is essential if the realistic explanatory theory is needed (*i.e.*, if you want to describe what is really going on in agents' minds), it seems less important form the instrumentalist point of view. In fact, if the proposed formalism is to be applied in the operationally-oriented cognitive science, the issue of adequacy becomes irrelevant. The reason for this is that the model simulating human behavior must not necessarily work the same way as human mind. In the following, we leave it open for discussion, whether the above should be interpreted in a realistic or instrumentalistic manner (even though we are inclined to the latter opinion). This is, in fact, part of a wider problem of the epistemological status of a game-theoretic approach as a whole, which, nevertheless, has been successfully applied from the very beginning of cognitive science.

4. Conclusion

In this paper, we discuss the social reactions from a perspective of the game theoretical model of prisoner's dilemma. The model results in largely different equilibria, depending on the length of the interaction between agents. In a short-lived interaction, the model results in a suboptimal point of overall welfare due to the fact that agents maximise their short-term profit and act in an uncooperative manner. In an interaction of long indefinite duration, the model results in an optimal point of overall welfare due to the fact that agents, maximising their long-term overall profit, act in a cooperative manner. Overall, this paper presents theoretically founded explanation of the strategically different behaviour in short-term and long-term social interactions, providing successful computational framework for cognitive modelling.

References

Aumann, R.J., Hart, S. (eds.) (2002). *Handbook of Game Theory with Economic Applications*, vol. 3. Amsterdam: Elsevier Science Publishers.

Aumann, R.J., Heifetz, A. (2002). Incomplete information. In: R.J. Aumann, S. Hart (eds.), *Handbook of Game Theory with Economic Applications*, vol. 3 (pp. 1665–1686). Amsterdam: Elsevier Science Publishers.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Aronson, E. (2011). *The Social Animal*. New York: Worth Publishers.

Bertrand, J.L. F. (1883). Review of Théorie mathematique de la richesse sociale and of Recherches sur les principles mathematiques de la théorie des richesses. *Journal des Savants* 67: 499–508.

Camerer, C.F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.

Capraro, V. (2013). A model of human cooperation in social dilemmas. *PLoS ONE* 8 (8): 1–6.

Carlsson, B. (1998). *Evolutionary Models in Multi-Agent Systems*. Lund University Cognitive Studies 72.

Chammach, A.M., Rapoport, A. (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor: University of Michigan Press.

Copeland, A.H. (1945). Review: Theory of games and economic behavior by John von Neumann and Oskar Morgenstern. *Bulletin of the American Mathematical Society* 51 (7): 498–504.

Crawford, V.P. (1997). Theory and experiment in the analysis of strategic interaction. In: D. Kreps, K. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications*, *Seventh World Congress*, vol. 1 (pp. 206–242), Cambridge: Cambridge University Press.

Darley, J.M., Berscheid, E. (1967). Increased liking as a result of the anticipation of personal contact. *Human Relations* 20: 29–40.

Debreu, G. (1970). Economies with a finite set of equilibria. *Econometrica* 38 (3): 387–392.

Fehr, E., Fischbacher, U. (2003). The nature of human altruism. *Nature* 425 (6960): 785–791.

Varian, H.R. (2006). *Intermediate microeconomics: A modern approach*. New York–London: Norton.

Harsanyi, J.C., Selten R. (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Hillas, J., Kohlberg, E. (2002). Foundations of strategic equilibrium. In: R.J. Aumann, S. Hart (eds.), *Handbook of Game Theory with Economic Applications*, vol. 3 (pp. 1597–1663), Amsterdam: Elsevier Science Publishers.

Khan, M.A., Sun, Y. (2002). Non-cooperative games with many players. In: R.J. Aumann, S. Hart (eds.), *Handbook of Game Theory with Economic Applications*, vol. 3 (pp. 1761–1808), Amsterdam: Elsevier Science Publishers.

Kreps, D.M. (1990). *Game Theory and Economic Modelling.* Oxford–New York: Oxford University Press.

Kreps, D.M., Wilson, R. (1982). Sequential equilibria. *Econometrica* 50 (4): 863–894.

Kreps, D.M., Milgrom, P.R., Roberts J., Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory* 27 (2): 245–252.

Kuhn, H.W., Tucker, A.W. (eds.) (1950). *Contributions to the Theory of Games*, vol. 1. Princeton: Princeton University Press.

Milgrom, P.R. (1984). Axelrod's "The Evolution of Cooperation". *The Rand Journal of Economics* 15 (2): 305–309.

Nash, J.F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences* USA 36: 48–49.

Nash, J.F. (1951). Non-cooperative games. *The Annals of Mathematics* 54 (2): 286–295.

Neumann, J. von (1928). Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen* 100 (1): 295–320.

Neumann, J. von, Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Raghavan, T.E. S. (2002). Non-zero-sum two-person games. In: R.J. Aumann, S. Hart (eds.), *Handbook of Game Theory with Economic Applications*, vol. 3 (pp. 1687–1721), Amsterdam: Elsevier Science Publishers.

Rapoport, A. (1970). *N-Person Game Theory. Concepts and Applications*. Ann Arbor: University of Michigan Press.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica* 50 (1): 97–110.

Osborne, M.J. (2003). *An Introduction to Game Theory*. Oxford: Oxford University Press.

Selten, R. (1999). *Game Theory and Economic Behavior: Selected Essays*, vol. 2. Cheltenham-Northampton: Edward Elgar Publishing.

Shapiro, C. (1989). The theory of business strategy. *The Rand Journal of Economics* 20 (1): 125–137.

Skyrms, B. (2010). *Signals. Evolution, Learning and Information*, New York: Oxford University Press.

Smith, V.L. (1992). Game Theory and experimental economics: Beginnings and early influences. *History of Political Economy* 24 (suppl.): 241–282.

Tversky, A., Shafir, E. (eds.) (2004). *Preference, Belief, and Similarity. Selected Writings*. Cambridge, MA: MIT Press.

NOTES ABOUT AUTHORS

Peter Gärdenfors is Professor of Cognitive Science at Lund University since 1988. His main research interests concern concept formation, concept learning, the semantics of natural language and the evolution of cognition. His main books are *Knowledge in Flux*, Bradford Books, MIT Press, 1988 (second edition 2008, College Publications). *Conceptual Spaces*, Bradford Books, MIT Press, 2000, *How Homo Became Sapiens*, Oxford University Press, 2003, and *The Geometry of Meaning*, MIT Press, 2014.

Aleksander Gemel is Assistant Professor at the Department of Cognitive Science, Institute of Psychology, Faculty of Educational Sciences, University of Łódź. His main research interests concern the semantics of natural language and its evolution and acquisition, formal methods of conceptual structure modeling, and metaphorical theory of language. He is author of *Lingual Model of Cognition. Cognitive Components in Continental Philosophy of Language* (2015, in Polish) devoted to the multifaceted relations between contemporary cognitive science and continental philosophy of language.

Magdalena Grothe (born Malinowska) works in the Directorate Monetary Policy of the European Central Bank. She holds PhD from the University of Cologne (2008, scholarship of Deutsche Forschungsgemeinschaft). Between 1999 and 2005 she studied at the University of Lodz (M.Sc. in Economics), University of Mannheim (Erasmus scholarship) and University of Cologne (Dipl.-Volksw., Mummert scholarship). Her research focuses on liquidity, risk pricing, market microstructure and inflation expectations. Her work at the ECB involves on analysing euro area financial markets from the monetary policy perspective. She contributes to ECB publications and is involved in the process of monetary policy preparation and implementation.

Kenneth Holmqvist is Professor in Psychology at Lund University, and founder of Humanities Laboratory. He works in a large variety of eye-tracking-based research stretching from reading research and scene perception, over to newspaper reading, advertisement studies and gesture recognition in face-to-face interaction. He is the main author of the international standard *Eye Tracking: A Comprehensive Guide to*

Methods and Measures, Oxford University Press 2011 and chairperson of the initiative for standardization of data quality assessment.

Jana Holsanova is Associate Professor in Cognitive Science at Lund University. Her research focuses on cognition and visual communication, image perception, the interplay between language and images, the role of images for learning, visual thinking, and users' interaction with various media. Hers main books are *Discourse, Vision, and Cognition. Human Cognitive Processing*, John Benjamins Publishing Company, 2008, and *Picture Viewing and Picture Description: Two Windows on the Mind*, Lund University Cognitive Studies 2001.

Roger Johannson is former employee at Department of Cognitive Science Lund University, currently Department of Psychology Lund University. His research mainly focuses on the relationship between eye movements, attention, memory and mental imagery, and eye-tracking methods of investigation. He is author of *Tracking the Mind's Eye: Eye Movements During Mental Imagery and Memory Retrieval*, Lund University 2013.

Piotr Łukowski is Professor at the Department of Cognitive Science, Institute of Psychology, Faculty of Educational Sciences, University of Łódź. His research focuses on theory of persuasion, cognitive aspects of communication and thinking, and also natural language semantics. His last books are *Paradoxes*, Springer 2011, and *Practical Logic with Elements of the Knowledge on Psycho-manipulation* (in Polish), Wolters-Kluwer 2012.

Konrad Rudnicki is psychology student. Member of seminar at the Cognitive Science Department, Institute of Psychology, University of Łódź. Collaborator at the Department of Neurobiology, University of Łódź. His main research interest concern fuzzy sets theory, neurobiological aspects of cognition, formal-logical tools in psychology, neurochemistry of catecholamine metabolism.

Dorota Rybarkiewicz is Assistant Professor (on leave) at the Department of Logic, Institute of Philosophy, University of Łódź. Member of seminars at Cognitive Science Department, Institute of Psychology, University of Łódź. Her main research interest concern speech acts theory, theory of metaphor, theory of persuasion, cognitive aspects of communication, thinking, and decision-making.

Annika Wallin is Associate Professor in Cognitive Science at Lund University. She is fascinated by everyday decision-making, and how it is best understood, from both a normative and a descriptive point of view. She has, among other things, studied the visual attention of consumers doing their everyday shopping and is currently working on how scientific evidence and experience are allowed to influence medical decisions.

Paula Quinon is researcher at the Departments of Philosophy and Cognitive Science, Lund University. Her studies include, in particular, conceptual analysis of the relation between the concept of natural number and other concepts such as "computability" or "cardinality". Recently, in collaboration with developmental psychologists and cognitive scientists, she studies the emergence of the concept of natural number in early childhood. On this basis she develops a revised version of naturalized epistemology.

Frank Zenker is researcher at the Departments of Philosophy and Cognitive Science, Lund University, and researcher at the Department of Philosophy, University of Konstanz. His postdoc project (2008–2010) applied Gärdenfors' Conceptual Spaces and Olsson's model of Agenda-Driven Change to theory-dynamical processes in the empirical sciences. He works in Social Epistemology and Philosophy of Science. He is author of *Ceteris Paribus in Conservative Epistemic Change*, Peter Lang 2009.

Bartosz Żukowski is Assistant Professor at the Department of Cognitive Science, Institute of Psychology, Faculty of Educational Sciences, University of Łódź. His research focuses on the semantics of natural language, theory of truth, philosophy of mind and early modern thought. He is author of *Esse est percipi? George Berkeley's Metaphysics of Idea* (2012, in Polish).

REVIEWER
Jacek Malinowski

INITIATING EDITOR
Urszula Dzieciątkowska

EDITOR OF ŁÓDŹ UNIVERSITY PRESS
Katarzyna Gorzkowska

TYPESETTING
Oficyna Wydawnicza Edytor. org
Lidia Ciecierska
AGENT PR

COVER DESIGN
Łukasz Orzechowski

COVER IMAGE
© Depositphotos.com/agsandrew

First Edition. W.06997.15.0.K

Publisher's sheets 10.5; printing sheets 12.25