Logical Thought Based on Word Presentations

Matthias Jakubec (jakubec@ict.tuwien.ac.at)
Klaus Doblhammer (doblhammer@ict.tuwien.ac.at)
Martin Fittner (fittner@ict.tuwien.ac.at)
Alexander Wendt (wendt@ict.tuwien.ac.at)
all
Vienna University of Technology, Institute of Computer Technology, Gußhausstr. 25-29
1040 Wien, Austria

Abstract

This paper is about the mechanisms of logical reasoning which would fit into a cognitive architecture based on the concepts of psychoanalytic metapsychology. As the results have to be based on a bionic approach beyond neurological findings a turn back to the roots of logic is required. According to psychoanalytical theory word presentations are the base of language on the one hand and connected to that the base of rational thought on the other. The paper shows, how a functional model of logical thought can be derived from basic metapsychological insights and how based on word presentations atomic propositions emerge from simple word presentations in a psyche. Finally it sketches briefly, how reasoning rules have to be handled based on the underlying model.

Keywords: cognitive architecture, SiMA, word presentations, reasoning, logic, rational thought

Introduction

For more than 15 years now the ambitious project of developing a human like cognitive architecture is observed at the Institute of Computer Technology of the Vienna University of Technology. The project called SiMA (Simulation of Mental Apparatus and Applications)[1] was set up to enable control units of performing adequate reactions in situations, where up to now only human beings show the requested flexibility and the necessary understanding for human cooperation partners, such as security control on airports or the interworking between human workers and robots.

To reach this goal, the engineers had to develop a functional model of the desired architecture, i.e. they had to identify its parts and to describe their interworking. The target to build a machine with humanoid behavior led to a bionic approach with the human brain as the blueprint for the design of the machine. If you want to understand the working of a computer application like Word you would not start investigating the computer's hardware, its circuits and the power supply, but rather would look for higher level functions, and so unlike in the Human Brain Project[2] the understanding of neurological basics was not the goal of the project team but primarily the upper layer of the brain's activities, the psyche respectively the mental apparatus.

Due to the bionic approach the team had to look for a holistic, functional description of the psyche as the template for their work and checking the different schools of psychology, finally the metapschological theory of Freud[3], the theoretical results of his psychoanalytical studies, turned out to be the only reliable scientific base for the intended quest.[4] According to metapsychology the work of the psyche is split into two major subparts: first in the primary process the input data (representations of the world provided by senses and drives) get treated in a kind of "quick and dirty" manner fully unconscious, then in the secondary process another step of processing the data makes them preconscious and conscious (Dietrich 2014).

While in the past 15 years the primary process had been elaborated down to a satisfying level of detail the secondary process remained rather underexposed. Artificial Intelligence research, computational intelligence studies, theoretical informatics, and mathematical logic have made enormous progress during the last 50 to 100 years in working out applicable models of logical reasoning, so it is no longer a severe problem for computer driven machines to play chess, proof theorems or even to win Jeopardy[5,6]. But there is more than justified suspense that these machines make use of a huge number of insights of the scientists in following their algorithms, algorithms which are not available to the human brain. So under the condition to find

[1] sima.ict.tuwien.ac.at, accessed 2015-02-18; up to January 2015 the name of the project had been ARS (Artificial Recognition System).

[2] www.humanbrainproject.eu, accessed 2015-02-18.

[3] *Metapsychology* is Freud's attempt to formulate a general, holistic, scientific model of the design and the functioning of the psyche, which he calls the psychic apparatus (comp. Freud, 1915; Freud, 1925). It has to be distinguished from (a) psychoanalysis as treatment for psychic disorder and (b) psychoanalysis as a method for the investigation of unconscious processes and contents (comp. Freud 1920).

[4] That of course doesn't mean that other findings in the fields of cognitive science or psychology are irrelevant, but psychoanalytic metapsychology at the time of the beginning of the project was identified as the only holistic theory suitable for the intended top-down-design of a model of the human psyche. For further arguments about psychoanalysis as base of the SiMA project see Dietrich, D., Bruckner, D., Zucker, G., et al. (2009).

[5] www.research.ibm.com/cognitive-computing/watson/index.shtml#fbid=lqOqQqGVuE0, accessed 2015-02-18.

[6] See e.g. Cohen & Lefebvre (2005); Gerla (2001); Goertzel, B., & Pennachin, C. (2007); Priest (2008).

a model, how the brain performs logical reasoning we need to step back again and turn to the roots. How can the human brain think logically based on an evolution which designed it as an optimized control unit of the human body, satisfying drives with minimum efforts?

Functional Model of SiMA

The SiMA project (Simulation of Mental Apparatus and Applications) was introduced under the name of ARS (Artificial Recognition System) by Dietrich, Fodor, Zucker, & Bruckner (2009). You can find brief discussions in Schaat, Wendt, Jakubec, et al. (2014) and Schaat, & Dietrich (2014).

As explained above, the SiMA model intends to provide a functional description of the psyche. The design process followed a top-down methodology. On the highest level (level 5) the psyche is described as the control unit of the human organism. In level 4 differentiations are made according to Freud's 2nd topological model: There are the functions of the Id, where bodily needs are treated. Competing with them Freud identified the Super-Ego which handles internalized social demands. Finally the Ego functions have to mediate between the different requests. Mind, that all this happens totally unconscious in the psyche. You usually are not aware in scenarios like the following, that you do not like a certain person because she reminds you of your sister which you have experienced to be a major competitor of yours in early days of childhood, and that this results in the reaction, to be especially kind towards this person. Only in the end of the process, a permitted subset of possible actions get presented to the psyche in a preconscious and conscious way by the Ego functions, so that they become part of rational decision making.

body, i.e. instructions for muscles and glands. There are four different input tracks. Two are signaling drives, i.e. sexual drives and self preservation drives, two more provide environment perception and body perception.

The input becomes psychical content if it gets cathected, that is some quota of affect is assigned to it, or to say it simple: the presented content is of interest for the psyche. Drives get represented by drive meshes and perceived contents get represented by thing presentation meshes[7]. The processing of drives and perception is mainly a task of sub-functions of the Id.

Before the psychic content can be forwarded to a rational treatment in the secondary process it needs to be filtered according to the question, whether it shows permitted thought or would be absolutely inacceptable for an individual as a member of a society. This filter task is performed by an interworking of Super-Ego functions, which provide super-ego rules according to which defense mechanisms have to handle the questioned thoughts, and some Ego functions which finally make decisions about permitted versus rejected content[8].

The contents which pass this filter, will then be transformed into preconscious thought by adding word presentations (WP, word presentation meshes in technical terms) to them (transformation track). Now they can be part of rational thought and planning in the reasoning track, and at the end a certain action gets selected which results in signals towards the actuators. Imaginary actions, whether actually executed or just fantasized, are fed back to the primary process to cause further associations and thus awake more psychic content (cathect it with some quota of affect).

An important factor throughout the whole process is the various measures of valuation. In principle they all are

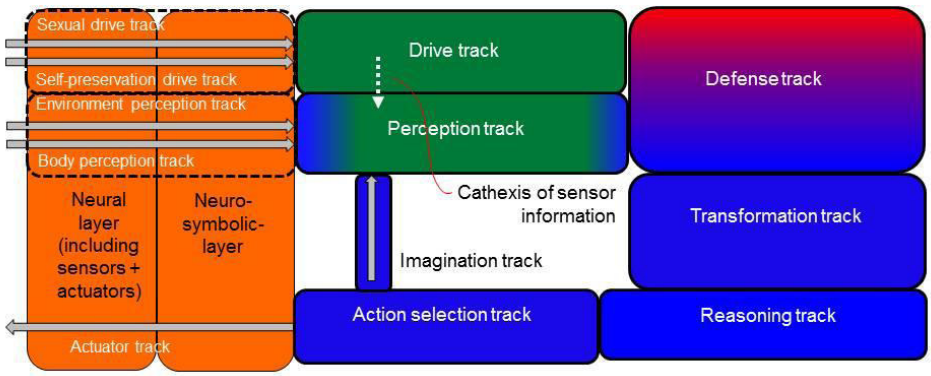

Figure 1: SiMA function model, level 3 (Dietrich, 2014).

Figure 1 shows level 3 of the function model. Leftmost two vertical columns are not part of the psyche but of the layers below: the neuronal layer of sensors and actuators and the neurosymbolic layer, where the translation of sensor data into neurosymbols is done and the other way round the translation of neurosymbols into actuator signals for the

[7] The technical term within psychoanalysis is *thing presentation*. The add-on *mesh* in the technical model indicates that drives and thing presentations never occur as isolated items but always are part of networks.

[8] Drive wishes for instance might be repressed or turned into its opposite.

based on the quota of affect which always reflects some drive tension. Concerning the active psychical contents there is the tension between the current and the desired state of the individual. Concerning memories the reduction of tension which had been the result of an action or a tile which played a role in it is stored with them. A significant amount of the sexual drive tension gets desexualized and that way under the term *neutralized intensity* is made available as a sort of fuel for secondary process activities. The state of the individual also gets rated by emotions, which directly may lead to bodily reactions (sweating, shortness of breath etc.) and thus is visible for others but the individual itself only can register it by observing these body reactions or does not detect its emotions at all. Finally the individual has feelings, i.e. emotions which get connected with word presentations and so become preconscious and if they are important enough also become conscious. So the individual is aware of its feelings, the others are not, if the individual does not talk about them.

Besides the input data two more categories of data play an important role: there is the huge mesh of memories, thing presentations and word presentations associated with each other, and a number of sub-functions make use of what we call personal factors. These personal factors are abbreviations of sets of memories or specific bodily reactions which result in certain behavior typical for the individual, such as the rate of sexual drive tension turned into neutralized intensity within a certain time span.

Word Presentations, Consciousness, Language, and Rational Thought

There is significant evidence that the following thesis is true:

Logical thought is always conscious.

Of course there are severe counter arguments about people who dream about the results of mathematical problems. Or let's think of the mathematician S. Ramanujan who came to a number of his results obviously merely by intuition (Ranganathan, 1967). But a lot of Ramanujan's results didn't hold. You wouldn't trust a logical result as long as you haven't checked it in full consciousness.

Psychoanalytical metapsychology shows, that consciousness and language belong together. Psychic content represented by thing presentations becomes preconscious in that moment when an associated word presentation gets activated. If the cathexis of the word presentation is strong enough, the content becomes conscious.

According to Freud a word presentation is the representation of a word of a natural language in the psyche. He has illustrated his idea about it in a famous sketch (figure 2). In his drawing things are represented by visual, acoustic and tactile associations and there is a strong association between the visual representation of the object and the sound-image of the word. There are other images of words as well, and they together build the word presentation.

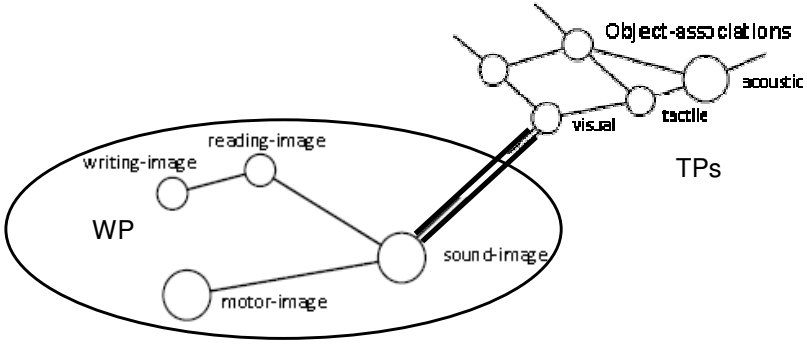

Figure 2: word presentation.
(WP … Word Presentation, TPs … Thing Presentations)
Drawing under usage of Deutsch (2011) based on Freud (1915).

A Functional Model of Logical Thought

According to the research program of SiMA in the following the functions needed to model logical thought will be developed step by step.

So the psychoanalytical finding is that rational thought, language and consciousness belong together. If we talk about language than usually its purpose for communication is seen in the first place, but that is not correct. There is no doubt that spoken language starts with the utterance of noises. We do this simply because it causes pleasure by itself. Just observe a baby when it starts experimenting with its first controlled sounds. She or he obviously enjoys the pleasure produced by the feelings around its oral areas. That explains, why we are speaking, but it doesn't explain why

speech is used as language. What happens is that individuals who live in social compounds make the experience, how certain utterances they make can have meaning to others in their group. They can also have meaning for themselves. One thing, the vocal utterance, can stand for another. It becomes a symbol. This is possible not only for acoustic objects but also for gestures or for external settings (writing). Today we understand that nearly everything might be a symbol, a representative, for something else. So Freud's drawing cannot be taken in its narrow sense. Otherwise it would mean that deaf people could not have word presentations (as they for sure have no sound-images of words) and thus would not have consciousness. This would be mere nonsense. What counts is that the object gets a semantic function which means its presentation in the psyche is associated with something else for which it is a representative. It has a meaning. The system of meanings is produced within a general social system (natural language).

Only the word presentations can get into relations with each other where they fulfil specific roles and thus build up a structure which forms a language. It is this ability to produce syntactical interdependencies which makes a set of words being a language which can be used as material for modelling the world. The first purpose of word presentations is to provide structures and orders between the things they denote, a capability which thing presentations alone do not have as they only stand in associative relations with each other, relations of the kind, "if I think of the one thing also the other comes into my mind". This primary characteristic of word presentations, the ability to build structure, is a result of the development of language as a social act, but the usability of language for communication purpose in that light appears as a second place feature behind its ability of bringing logical structure into the world.

Atomic Propositions

What is this magical step which enables word presentations to build up meaningful structures, while thing presentations only are connected via stronger or less strong associations with each other? Before we can perform any logical operations we need descriptions of the world, we need propositions as a foundation of any logic. The simplest propositions have the structure of a predicate expressed about an object. This is the structure of a pair, where the two constituents each play a specific role. Atomic propositions are ordered pairs.

The objection was made that you do not need ordered pairs for an axiomatic foundation of logical systems, as you can declare ordered pairs by a simple combination of unordered pairs (Gödel, 1931):

$$(a,b) = \{\{a\},\{a,b\}\}$$

But this objection overlooks, that the definition of the unordered pair itself makes use of the element relation, i.e. that an object is an element of a set or a class respectively, which already is expressed by ordered pairs.

$$a \in \{a\}$$
$$\{a\} \in \{\{a\},\{a,b\}\}$$

That is, you cannot declare any kind of formal relation without at least one kind of asymmetric relation, and thus it makes sense to use the ordered pair as a starting point from where you can derive any other kind of structure.

In Jakubec, Dönz, & Bruckner (2013) the fundamental role of ordered pairs in the light of psychoanalytic metapsychology as the constitutive structural element of the secondary process is described. Only in the secondary process there is

1. time dependent order
2. language
3. rational thought

where 3 needs 1 and 2 as preconditions. It must be observed as a real strange tradition, that formal logic distinguishes between the object-predicate relation (predicate-argument relation) on the one hand and the element-set (or -class) relation on the other. From an object oriented point of view both relations represent the same circumstances, that certain objects are instances of a certain class. The object in question belongs to the class of objects with exactly the predicate as defining property. There is no need to remain with this differentiation any longer and we can treat both kind of relations as one and the same.

There is a Boolean affinity between the propositional calculus and set algebra, where the implication corresponds with the subset relation:

$$((x \in A) \rightarrow (x \in B)) \Leftrightarrow (A \subseteq B)$$

Wang presents '→' as the symbol for the *inheritance relation* (Wang, 2007), which means that if A → B, A is a subcategory of B (e.g. *raven* → *bird*) and he reads it the way that A is a subject with the predicate B, which normally would be expressed by A ∈ B. It is obvious that the subset relation and the element relation are not the same but …

Big 'but'! Let's consider the following. An artificial agent Adam (e.g. based on the SiMA architecture) perceives his environment, where he, based on his memories, identifies several objects and among them another agent. (May his name be Bodo.) We now could model this situation in the way that there is a word presentation with the logical meaning of an object constant representing Bodo. If Adam now turns around he will lose perceptions of Bodo and if he then turns back again, he will again identify an object as Bodo. But can he really be sure, that both perceived objects are the same. The only thing what Adam could be sure about is, that there was an object with all the Bodo-attributes and seconds later, there again was such an object. Even if Adam does not lose contact with Bodo interim, there is no guarantee that the object identified in a certain moment is the same as the one in the next moment. The object continuity is a mere construction of the brain. (That's why we identify a jumping point of light on the screen as one and the same point jumping rather than different points highlighted consecutively.) But this would mean that there is no constant on object level. Instead of

$$P(O)$$

(predicate P holds for Object O), we rather would write

$$O(x) \wedge P(x)$$

(there is an object x which fits into the predicates O and P). If there are no object constants, we will end up with the following: if predicate P holds for an object O then the fact of an x being O implies that x fulfils P,

$$(O \in P) \Leftrightarrow (O(x) \rightarrow P(x))$$

So we face tight relationship if not to say equivalency between the concepts of predicate, subset and implication. 8 of the 16 binary operations of propositional logic are commutative. 4 select either the left or the right argument or their negation respectively, which means, that the respective other argument has no influence on the result but gets totally ignored. The remaining 4 are the implication operations ($a \rightarrow b$, $b \rightarrow a$) and their negations, where the order in the argument pair matters. Even though the implication itself is a static operation (as any other logical operation as well) it has a severe dynamical foundation: As soon as we have proofed the condition, we can assure the conclusion. In our imagination we handle it 'first a, then b'. In general we need time dependency to identify order, as time is the one Anschauungsform (form of intuition, Kant) which is directed. Any order is based on the asymmetry between first and next (Jakubec, Dönz, & Bruckner, 2013). It is our ability to recognise this directedness in the secondary process, the ability to handle temporal order, which also enables us to deal with object-predicate relations, the subset and element concepts, and implications.

What we need to do is to identify functions in the beginning of secondary process where atomic propositions are composed in the agent's psyche. In the transformation track (see figure 1) we can distinguish exactly three different relevant functions (Dietrich et al., 2014):

1. **F21: Transformation to secondary process (perception)** which activates word presentations (WPs) for perceived objects.
2. **F20: Composition of feelings** which creates feelings from current emotions and activates WPs for these feelings.
3. **F8: Transformation to secondary process (drive-wishes)** which identifies the wishes the agent wants to get satisfied currently and names them by activating associated WPs.

The result of F21 will be a list of propositions describing the agent's current perception. E.g. if the agent sees a Schnitzel and the thing presentation representing the Schnitzel is associated with the word presentation "Schnitzel" (it might be associated with other WPs as well, but these other associations are weaker) there will be the proposition Schnitzel(x), which simply means "there is something which is a Schnitzel". As stated above, there are no constant identifiers on object level, thus a WP for a variable gets activated for the perceived object. In natural language this would be something like "something" (which replaced "dada" in a late state of language acquisition). Our artificial agent may use predicate calculus as his natural language where "x" could be a valid WP for what he perceives.

At this point it has to be stated, that Hurford comes to the result that higher developed animals have some kind of proto-predicates at their disposal which they apply on object variables rather than on constant representations of objects. He gives

"$\text{LION}_{\text{baboon}}(x)$ & $\text{CROUCH}_{\text{baboon}}(x)$ & $\text{ROCK}_{\text{baboon}}(y)$"

as an example for a possible representation of a situation with a crouching lion and a rock in an animal brain, where the index 'baboon' means that these are the corresponding predicates of a baboon and not of a human (Hurford, 2007).

The result of F20 is a list of propositions, (one for each detected feeling) stating the word presentation of the feeling about the agent himself as the object (e.g. sad(Adam)). It might make sense, that in these cases the proposition's argument should be a constant identifier of the agent instead of a variable, so that the WP representing the agent would be the only constant WP on object level. But this question needs further investigation.

Finally F8 produces propositions in a similar way as F20, this time declaring the agents drive-wishes instead of his feelings (e.g. wants_to_eat(Adam)).

Reasoning Rules

Reasoning rules are ordered pairs of word presentation sequences. The first element contains a text (resp. corresponding WPs) which has to be replaced by the second element, if it occurs in the current thought (resp. the corresponding WPs). As we can see, also this element of rational thought relies on the fundamental role of ordered pairs and again they reflect a temporary order, a dynamic process: replace a currently available WP sequence by another one which then will be active at least in the step(s) of immediate future. There are a lot of open questions in how far such rules can be rule schemata instead, which have to be adapted in each concrete situation or whether such adaptation itself is part of the execution of an action plan. We expect more results concerning this in near future.

In the SiMA model two functions are identified in the action selection track (figure 1) to do the work. In **F26: Decision making** it is decided among other criteria based on the feelings identified in F20, which goal has to be reached next and at the same time which reasoning rules could be satisfactory if applied to the current thought. For selection such reasoning rules are available which are associated strongly enough with the currently active word presentations or word presentation sequences. In **F52: Generation of imaginary actions** the possible reasoning rules get tested, that is they get checked, in how far the to be replaced word presentation sequence appears in the active WP sequence – an activity which is performed in the lower layers of the model based on our innate ability of pattern recognition – and if a rule matches it gets applied, and the

result is checked, whether it leads closer towards the desired goal.

The result of the function always has to be an action as its name indicates, an action for the immediate next step. That means in word presentations it is an imperative for the agent himself. In form of a proposition – as the result of a logical derivation can only be a proposition (whether atomic or combined) – it is a predicate expressing that the agent is expected to do something[9]. Nevertheless the requested action can also be to do more reasoning in the next steps of processing and not to do anything else at the moment.[10]

What Remains to be done

Besides the open points already mentioned above the most important question is of course the content and quality of the agent's memory database. Which word presentations and reasoning rules are associated with which perceived situations? The contents of this database can be attained in two different ways. They can be designed by hand or they can be the result of learning processes. Currently learning is too wide a field as to expect substantial results concerning our problem, thus for the moment our choice was for the first option. Step by step the agent's memory has to get extended by relevant knowledge about the world. Later on, when the principles are understood well enough, there might be tools to acquire more information automatically from the internet as already sketched in Jakubec, Dönz, & Bruckner (2013).

A closer problem and current topic of the research is the adequate embedding of the abstract concepts of logic in the agent's memory which is radically based on bodily experiences. In this question we again expect some useful support from psychoanalysis and its know-how at grounding adult behaviour in suppressed early childhood experiences. Research in this direction is continued.

References

Cohen H., & Lefebvre C. (Eds.) (2005). *Handbook of Categorization in Cognitive Science*. Amsterdam, Netherlands: Elsevier.

Deutsch, T. (2011). *Human Bionically Inspired Autonomous Agents*. Doctoral dissertation. Institute of Computer Technology, University of Technology, Vienna, Austria.

Dietrich, D., et al. (2014). *Naturwissenschaftliches, psychoanalytisches Modell der Psyche* (Technical Report II), Wien, Austria: Vienna University of Technology, Institute of Computer Technology.

Dietrich, D., Bruckner, D., Zucker, G., et al. (2009, September). *Psychoanalytical model for automation and robotics*. In *AFRICON, 2009. AFRICON'09*. (pp. 1-8). IEEE.

Dietrich, D., Fodor, G., Zucker, G., & Bruckner, D. (Eds.) (2009). *Simulating the Mind.* Wien, Austria: Springer.

Freud, S. (1915). *The Unconscious*. In: *On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works*. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, volume XIV (1914-1916). London, United Kingdom: Vintage.

Freud, S. (1920). *Two Encyclopaedia Articles. The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Volume XVIII (1920-1922) (pp. 233-260). London, United Kingdom: Vintage.

Freud, S. (1925). *An Autobiographical Study. The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Volume XX (1925-1926) (pp. 1-74). London, United Kingdom: Vintage.

Gerla, G. (2001). *Fuzzy logic: mathematical tools for approximate reasoning* (Vol. 11). Springer Science & Business Media.

Goertzel, B., & Pennachin, C. (Eds.) (2007). *Artificial General Intelligence*. Berlin, Heidelberg, Germany: Springer.

Gödel, K. (1931). *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I.* In *Monatshefte für Mathematik und Physik, 38* (pp. 173-198). English *On formally undecidable propositions of Principia Mathematica and related systems*. Courier Corporation.

Höpp, G. (1970). *Evolution der Sprache und Vernunft*. Berlin, Heidelberg, Germany: Springer.

Hurford, J. R. (2007). *The Origins of Meaning*. New York, NY, USA: Oxford University Press.

Jakubec, M., Dönz, B., & Bruckner, D. (2013, November). *Grounding of relations and abstract symbols in the decision unit of an embodied agent*. In *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE* (pp. 6660-6665). IEEE.

Priest, G. (2008). *An Introduction to Non-classical Logic: from if to is*. Cambridge University Press.

Ranganathan, S. R. (1967). *Ramanujan: the Man and the Mathematician*. UBS Publishers Distributors.

Schaat, S., & Dietrich, D. (2014). *Case-Driven Agent-Based Simulation for the Development and Evaluation of Cognitive Architectures*. In *26th Benelux Conference on Artificial Intelligence* (pp. 73-80).

Schaat, S., Wendt, A., Jakubec, M., et al. (2014). *ARS: An AGI Agent Architecture*. In B. Goertzel, L. Orseau, & J. Snaider (Eds.). *Artificial General Intelligence. Proceedings of the 7th International Conference, AGI* (pp. 155-164). New York, USA: Springer.

Wang P. (2007). *The Logic of Iintelligence.* In Goertzel, B., & Pennachin, C. (Eds.). *Artificial General Intelligence*. (pp. 31-62) Berlin, Heidelberg, Germany: Springer.

[9] The eligible relation between imperative and indicative is not yet determined. There is the serious idea, that imperative might have been the earlier language mode (Höpp, 1970), which might have an impact on the details of the data in the SiMA model.

[10] Pleasure principle and reality principle will be observed in the decision making, but are the task of other function modules than F52.