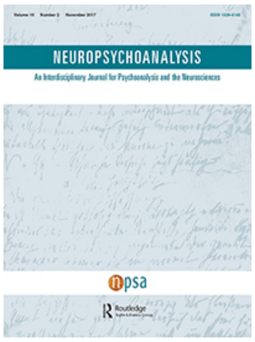

Routledge
Taylor & Francis Group

Neuropsychoanalysis

An Interdisciplinary Journal for Psychoanalysis and the Neurosciences

ISSN: 1529-4145 (Print) 2044-3978 (Online) Journal homepage: http://www.tandfonline.com/loi/rnpa20

On the way to bridging the gap between the mental apparatus and the neurobiological layer

Dietmar Dietrich, Gerhard Zucker & Klaus Doblhammer

To cite this article: Dietmar Dietrich, Gerhard Zucker & Klaus Doblhammer (2017) On the way to bridging the gap between the mental apparatus and the neurobiological layer, Neuropsychoanalysis, 19:2, 159-173, DOI: 10.1080/15294145.2017.1383180

To link to this article: http://dx.doi.org/10.1080/15294145.2017.1383180

Published online: 07 Nov 2017.

Neuropsychoanalysis, 2017
Vol. 19, No. 2, 159–173, https://doi.org/10.1080/15294145.2017.1383180

SOCIETY PROCEEDINGS

On the way to bridging the gap between the mental apparatus and the neurobiological layer

Dietmar Dietrich[a], Gerhard Zucker[b*] and Klaus Doblhammer[c]

[a]*TU Wien, Institute of Computer Technology, Vienna, Austria;* [b]*AIT Austrian Institute of Technology, Center for Energy, Vienna, Austria;* [c]*Psychoanalyst, Vienna, Austria*

(*Received 24 February 2017; accepted 8 September 2017*)

The mental apparatus – mind, psyche, soul – expresses itself in thought, play, love, creating, sex, planning, in fact in all human activities. Its foundation is the nervous system, as it is the central information system of the human being. For those interested in the connection between the neuronal level and the mental level, how is it possible to comprehend and model the connection between the neurobiological and the psychological? What does a natural scientific model of the mental apparatus look like? These questions led to the SiMA project (Simulation of the Mental Apparatus & Application). This article surveys the foundations of information theory and computer technology that underlie the SiMA project, and explains how the view of the neuropsychoanalytic world can be mapped into a model that can then be simulated and tested. The first results of several simulated case studies are described, suggesting that SiMA provides a way to validate theories of neuropsychoanalysis.

Keywords: Artificial intelligence; cognitive science; psychoanalytic theory; decision making; Mealy theory

1. Introduction

Although the origin of psychoanalysis had its roots in neurology, the psychoanalytic model of the mental apparatus evolved more or less independently from any considerations of the nervous system throughout most of the last century. This has led to a gap between both scientific worlds. For those interested in the neuropsychoanalytic project, or in the mind–body problem, an essential question must be: How can we close the gap between the neurological model and the model of the mental apparatus, to develop one consistent model? And if such a model were to be developed, would it be possible to validate this consistent model with natural science methods?

We propose that one productive way to work towards understanding how the mind works is to try to develop an artificial intelligence model of its internal dynamics that lead to external behavior. In doing so, researchers must grapple with specifying the actual "components" of the mind (in artificial intelligence terms, we call them "functions") and how they interact, in a way that is ecologically valid, and that can yield realistic outcomes when challenged in a simulation. This paper describes the efforts of our group to do just this, in a project we call the Simulation of the Mental Apparatus & Application (SiMA). In this project, a group of artificial intelligence and computer engineering experts are collaborating with psychoanalysts to sketch out a minimally viable model of the mind that might account for critical intrapsychic functions and that would generate realistic behavior when simulated.

Our paper thus has several aims. We provide an intellectual and historical context for our project, and then give an overview of some relevant aspects of information theory and artificial intelligence. We then describe some of the psychoanalytic functions modeled by SiMA, and summarize some of the first "use case" simulations. We hope this paper gives psychoanalytic readers a taste of the exciting generativity, as well as the significant challenges, in moving towards a fully realized psychoanalytic simulation of the mental apparatus. Some of the discussion that follows may seem difficult to follow for the psychoanalytic reader. We hope readers will persevere. We believe some of the challenging earlier sections are necessary for an appreciation of the final sections of the paper, and we very much look forward to future discussions with the psychoanalytically inclined as we pursue our project.

2. Intellectual context

2.1. Psychoanalytic precursors

A variety of positions have been taken on whether it is possible to bring together a psychoanalytic model of the mind that corresponds to, or can integrate with, other natural science, engineering, or information theory perspectives on the mind and its neurobiological substrate. Readers can see, for example, Holt, 1973, 1985; Gill, 1976, 1977; Peterfreund & Schwartz,

*Corresponding author. Email: gerhard.zucker@ait.ac.at

1971; Peterfreund, 1973. While these debates have explored important ideas, we think that recent developments in artificial intelligence now allow us to approach these questions more productively, as we hope to demonstrate in this paper.

2.2. Artificial intelligence

To help readers appreciate the potential contributions of the SiMA project, it is helpful to have a basic understanding of the history of artificial intelligence (AI), following the work of Palensky (2008). In the scope of this paper, we will have to simplify, but hope to cover some essential points. We also hope to demonstrate that SiMA represents a new development in contrast to traditional AI.

As a first point, it is essential for new readers to understand that in the realm of computer technology and communications engineering, we must distinguish between the laws of natural science that apply to physics, chemistry, physiology, for example, and the laws of information theory. Modeling the mind, with an ultimate goal of connecting it to its biological substrate, requires that we bring together both worlds, and it is critical that we understand that different laws apply to different levels of organization. Mixing both theories can lead to absurd solutions because information laws act in different areas than physical laws and cannot be compared.

Why is it that computer engineers, of all scientists, have been the first to state this so explicitly? Computer technology is the first science which has to handle explicitly the topic of manipulating huge amounts of data (information) by machines (physical objects), and thereby deal with the interface between these realms. In nature, of course, the nervous system was indeed the first information system, which handles such amounts of data by manipulating, memorizing and transferring them. But of course it has been difficult, if not impossible, for neuroscientists to experiment with simple or complex circuits using living neurons to learn their natural science laws. How could they possibly deduce all the laws of information theory which are necessary to understand the relationship between the hardware (in this case the nervous system) and the information layer (in this case the mental apparatus)? That was only possible with electronics.

With the interconnection of a few transistors (and later on also with simulated and emulated neurons (Lange, 1976)) engineers and scientists increased their understanding and their ability to differentiate between the hardware (physical part) and the information part. The computer became the second system in our world (after the nervous system) which was solely designed to handle huge amounts of data (information). Because of this it was essential for computer engineers – and we repeat it here with other words – to deal with the combination of the *physical* and the *information theory.* The first scientists in this field, like Konrad Zuse, John von Neumann, Claude E. Shannon, understood the problem and worked out different approaches, depending on whether they were mathematicians or engineers. Zuse and Neuman focused on exploring functions, algorithms, programing languages etc., while Shannon developed and composed the first information theory (independent of physics), which was significantly advanced over the last 70 years, for example by Mealy in 1955. One of his ideas is a central principle for SiMA and will be explained in Section 3.

From computer technology emerged artificial intelligence, then computer science, and finally robotics. Computer science focuses on algorithms (not the connection between hardware and software/information theory like in computer technology) which is not such an important topic for our work. The goal of robotics is machines that support humans by taking over automating tasks, which requires developing the right behavior for such machines. In contrast to this, the idea behind artificial intelligence is to explore machines with a specific "intelligence" – however scientists define this term –, which should be able to reason and make decisions. The founders were John McCarthy, Marvin Minsky, Nathaniel Rochester and again Claude E. Shannon. Their vision was based on logic, objectivity and – following the zeitgeist at that time – the independence of the mind from the body, with which Spinoza already has had his problems (Damasio, 2003, p. 12 and p. 150). In other words, we could say that their ideas were based on the assumption that thinking is an information process, which is independent of the body. Readers of this journal know that of course this was in opposition to Sigmund Freud's ideas.

Today we can call this first phase of artificial intelligence (AI) the *symbolic AI* (Lorenz & Barnard, 2007; Simon, 1965). The assumption – which is still accepted today – was that the nervous system perceives information over its sensors (Figure 1) and encodes the regarding data into micro-symbols and symbols until the mental apparatus interprets or produces symbols like *hot*, *hungry*, *oven* etc. (Dietrich, Fodor, Zucker, & Bruckner, 2009, p. 106). These symbols can be handled according to the laws of information theory. They can be manipulated, memorized, and transferred. Finally, the obtained symbols are decoded again to trigger electrical impulses that activate the muscles and glands.

In this phase of AI, we can say that the first theoretical bridge was built between the nervous system and the mental apparatus. However, it was not possible to

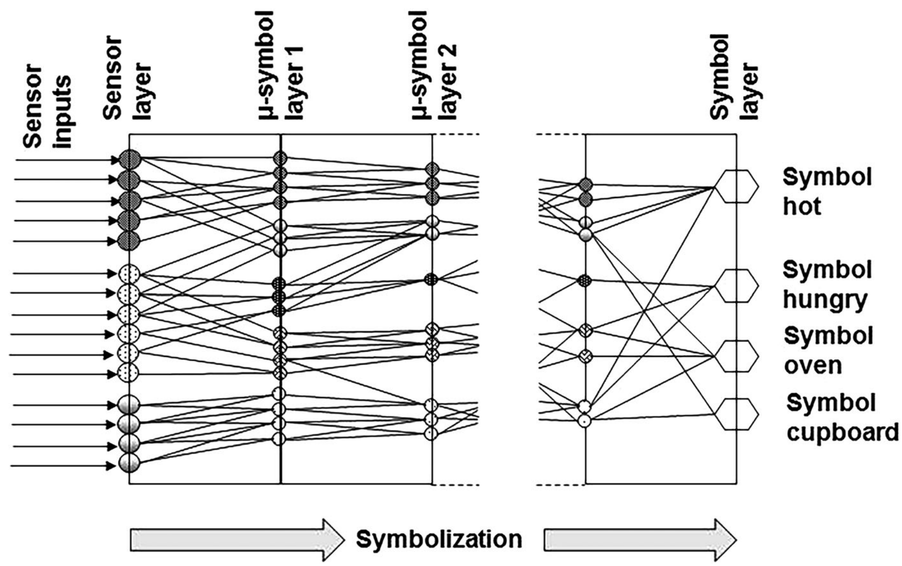

Figure 1. Symbolization of information, perceived over the sensors of the nervous system (see less abstract figures in Velik, 2008).

reach a solution in which the human mental apparatus could be explored in *one* model together with the electrical pulses that serve as inputs and outputs for this process. A continuous model was missing, which could explain how to follow the data flow from the electric pulses to the symbols to the mental apparatus part and back to the pulses in the actuators (the muscles and glands).

Scientists dealt with this aspect in the so called *statistical AI* (Minsky & Papert, 1969), the second generation of artificial AI. Using a bottom-up approach they tried to find the connection to the mental apparatus starting with the neural networks. Such developed artificial neural networks are still an important domain of scientific research and industrial development (Figure 2), but their solutions do not solve the gap problem between the neurobiological layer and the mental apparatus, because it is purely theoretical and statistical.

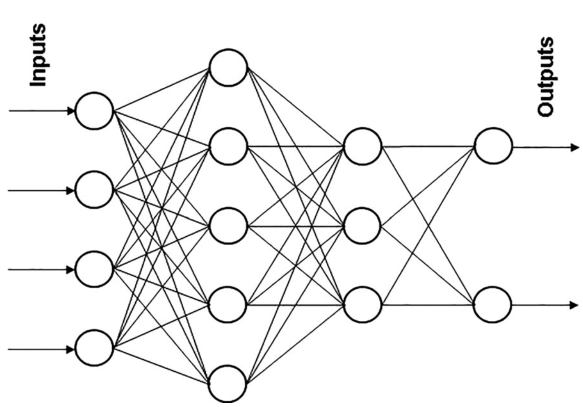

Figure 2. Artificial neural networks.

Then researchers decided that intelligence had to be developed in connection with the body. This was the third generation of AI, which is called behavior-based AI (Brooks, 1985). While this appears to be a step in the direction of truly modeling the mind, its name already shows a part of the problem, as scientists focused on the behavior of an intelligent process or system and not on any inner functions (the components of a system).

Humanoid robots are also counted among this behavior-based AI, although we see the humanoid aspect only in their outward appearance (see Figure 3 and e.g. http://videolectures.net/takashi_minato/; http://www.humanoid-robotics.org/). The "intelligence" of such robots is not human-like, because all aspects which represent a human being (as opposed to machines), are missing: this includes the mental apparatus, the unconscious and conscious activities, and the diverse evaluation mechanisms that are explained by the theories of psychoanalysis. Consequently, the fourth generation of AI emerged several years ago: the emotion-based AI (https://link.springer.com/chapter/10.1007/11550617_52; http://www.sony-aibo.com/aibo-models/sony-aibo-ers-7/; Canamero, 1997; Schaat, 2016). However, in our opinion, this leads to a dead end, because a commonly accepted definition of emotions or feelings does not yet exist.

Although these developments have not gotten AI close to actually modeling the mind yet, the reader should not be led to believe that AI is an absurd science. The opposite is the case. Internet without AI would not be thinkable. Numerous solutions of AI and cognitive science produce algorithms and architectures, which illustrate essential solutions for technical

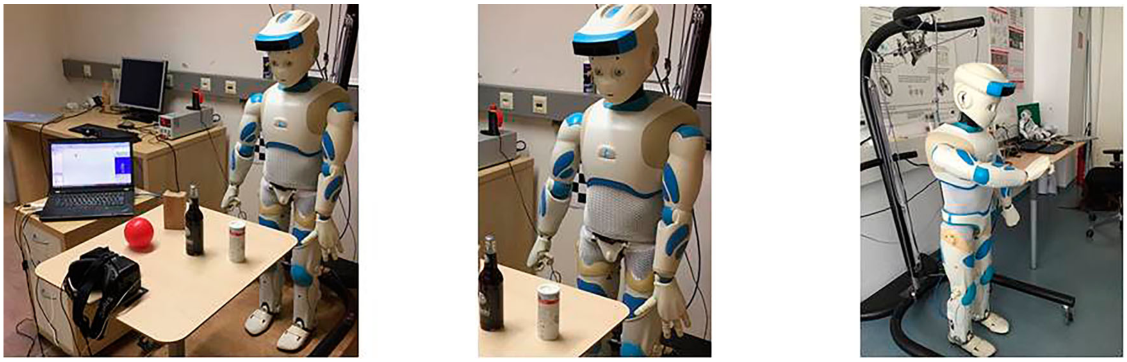

Figure 3. Robotics of today (Credit: ACIN, TU Wien).

and economical tasks. Their economic impact is enormous. However, the authors doubt that AI and cognitive science will efficiently contribute to explore the model of the nervous system in connection with the mental apparatus, on its current trajectory, because we believe that a more valid model of the mind must include internal motivation, value, agency, and subjectivity. That was exactly the reason to initiate the SiMA project, because these central processes or functions of mind are well-elaborated in psychoanalysis. We have pursued this project for the last 16 years. Our target is the fifth generation of AI.

3. Information theories

To be able to explain the model of SiMA and the results of our work to date, two fundamental theories of the modern information theory in computer technology should be understood, which we summarize in the following. Although the following section may be boring for psychoanalysts, we hope the reader will try to follow the train of thought.

3.1. Function and its behavior

Here we outline an essential differentiation that is necessary for understanding the SiMA project: the distinction between a *complicated* and a *complex* system. We think it is worth the effort to follow the argument, since we propose that the mind and nervous system are complex systems, not complicated ones, and this distinction dictates the principles that should be the foundation for designing functions that can be implemented in hardware or software to model how brain circuits can mediate complex mental behavior. As the reader will see, this goes far beyond a robot manifesting human-like behavior based on a relatively simple set of rules. A complicated process can be described by behavioral models e.g. specific algorithms (the dynamic of the procedure), which is mainly the field of computer science (less of computer technology[1]) or artificial intelligence.[2] Complex processes, in contrast, should be described as functional models, because it is impossible to consider all possible behavior phenomena.

In information theory, one foundation of computer technology is the strict separation between a function and its behavior. The term function is deduced from the (mathematical) description of a technical component and explains the structure of a component. The behavior is described by the data, running through the functions and generates the flow of data.

A system can be seen as a process when it is active. Then it shows behavior that is dependent on its inputs and its memories. This is the reason that a system can also be described by its behavior. The behavior describes actions and reactions.

If you look at results from cognitive science or artificial intelligence (Minsky, 1990), in which hardware – i.e. the connection to the physical world – does not play the major role, the scientists and engineers are mostly computer scientists (not computer engineers, see footnote 3) and use behavior models in terms of algorithms, which are normally economically efficient. In opposition to this, in computer technology the hardware usually plays a key role, therefore the way computer engineers think is another one. Their way of thinking starts from functional models (with the included components), with the link between hardware and software being a central aspect. The conclusion to be drawn from this fact is essential. Braitenberg explained this context in a book, which is fundamental today (Braitenberg, 1984). If function models become more advanced (complicated), the regarding behavior models can show complex behavior, which means not completely describable anymore. We can illustrate this further by describing Braitenberg's theoretical experiment.

The vehicle on the left side of Figure 4 has the functions "wheels" and "light sensors." The left wheel is connected with the left sensor. If the left sensor receives a lot of light, the wheel rotates fast; if the luminance is

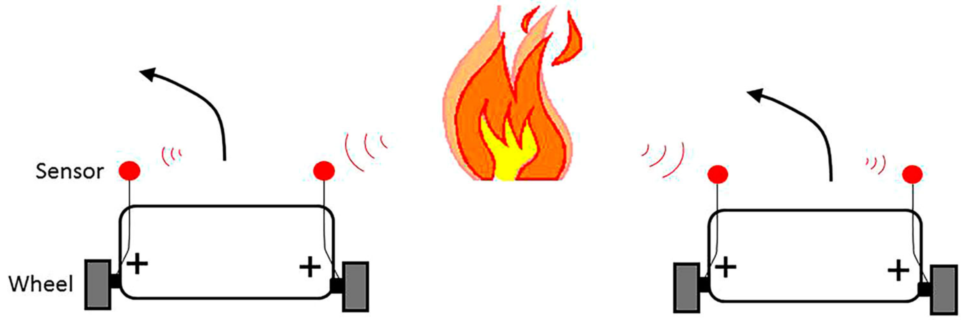

Figure 4. The simplest vehicles in Braitenberg's theoretical experiment.

low, the wheel rotates slowly. The same applies to the right wheel. Therefore, the left vehicle will drive to the left, *away* from the light. The right vehicle is identical to the left vehicle, but the connections between the sensors, motors, and wheels are crossed; therefore, if the light is to the left of the vehicle, the right wheel rotates faster and the vehicle will drive *towards* the light.

Braitenberg varies and extends his vehicles by changing, modifying and adding functions. He adds more sensors, modifies the connections (Figure 5), and suddenly the vehicle is able to follow complicated figures like a lying 8. If Braitenberg adds more and more sensors (which are in our understanding functions) and modifies the connections among the components (which means to modify the functions or even to increase the functionality of the vehicle), ultimately he gets an even more *complicated* functional description of the vehicle. At a certain point, its behavior becomes *complex*, which means: It is not describable (explainable) anymore, although Braitenberg is able to explain (describe) the functionality. So, the conclusion is: If the system is very complicated, it is still possible to describe/to develop a system on the base of functions, but not anymore to describe all behavioral phenomena, or – in engineering wording – to develop the behavior model. This essential relationship between describable functions and describable up to non-describable behavior is shown in Figure 6.

The conclusions to be drawn from these experiments are essential for analyzing and exploring complex processes. If a robot is programed on the base of a behavior model, it will exactly produce this behavior and nothing more. If engineers deduce the function and build it, they have to test and prove that the machine has the desired behavior. If the function becomes more sophisticated (complicated), the behavior can become complex. People, who explore computers on the base of functional models[3] always have to consider this aspect. However, in most industrial cases it is much more economically feasible to develop a smart system on behavior-based principles.

In SiMA the situation is the opposite. The nervous system is both a piece of hardware *and* an information system, which shows complex behavior, in Braitenberg's terms. Because it has a complex behavior, it is impossible to create a behavior model of the mental apparatus, but it is possible to compose a function model.

However, we have to admit that a function model of the nervous system cannot integrate all functions

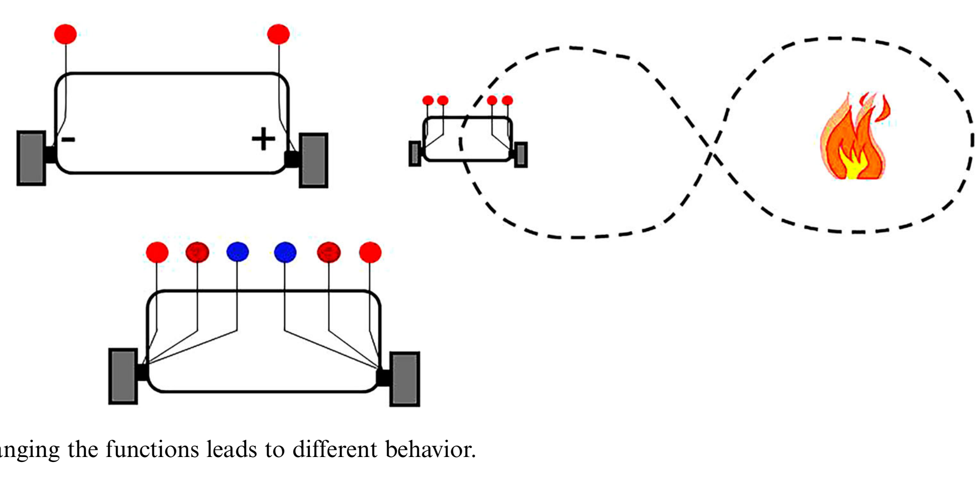

Figure 5. Changing the functions leads to different behavior.

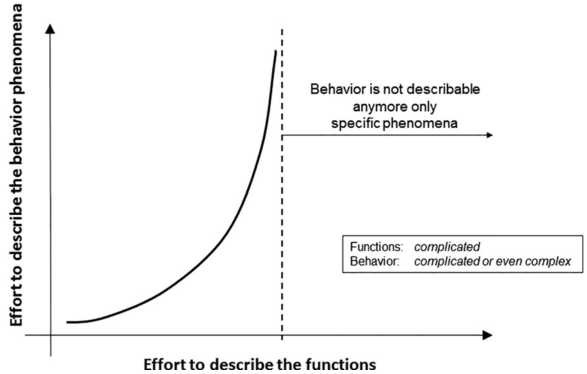

Figure 6. Relationship between the efforts to describe the function or its behavior.

and all function details. No institute has the capacity and money for it. The expert has to decide which details are necessary and which not and has to proof his decision by experiments. Then, over time, people can implement more and more functions and their details.

Back to the behavior model of a complex system: It only describes a specific detail of the whole system, but nothing more. For psychoanalysis this explanation becomes more comprehensible, if we think about unconscious procedures, which express their properties in conscious behavior: How would it be possible to describe all these procedures of behavior *completely* – which means in all details?

These considerations were crucial to use the theory of psychoanalysis as a basis for SiMA, because Sigmund Freud has created the prerequisites with his second model, the structural model, long time before computer engineers recognized that the connection between the hardware and information unit can only be bridged by a function model. This means – translated into the world of Freud – that the connection between the nervous system and the mental apparatus must base on function models. This clarifies that the computer engineers (not computer scientists) and psychoanalysts have the same way of thinking in this case.

It should be stressed here that the first model, the topographical model, which Freud could not bring together with the structural model, is in the wording of the information theory a behavior model; computer engineers also often say data model. Functions like the unconscious or conscious do not exist, only the information *is* un-, preconscious or conscious. The first model only describes the (behavioral) phenomena conscious and unconscious. Both models, the function model and the behavior (data) model, are in such a way the prerequisite for a complete specification of a system. They complement each other in the information theory and do not stand oppose each other.

3.2. Simulation of use cases

A theory, like a model of a system, needs a validation. In the engineering world, a system can be validated by simulation experiments. If an object, system, or process is theoretically developed – for example, a vehicle – the engineers and scientists first validate it in a simulation. Does it work? Does it fulfill the requested requirements? The basis for testing the result of the research is called a "*use case"*. In our project, we defined one or more individuals – in AI we call them "agents" – to handle certain situations. The agents contain the model (the functions) of the nervous system. It is essential to have these use cases, which have the necessary richness and ambiguity of a real-life situation, to specify which functions are necessary to navigate the situation. This is especially important for understanding, if the defined functions of the model are defined correctly and sufficiently, and if the function model fulfills the demands of psychoanalytic theory. Because of this, in the modeling of the mind, these use cases have to be worked out in SiMA by the psychoanalytical experts, because they do not only have the necessary knowledge and education in this topic, but also the longtime experience to explore and finally validate them on the base of their methods. The computer engineer, who develops the simulation program for this purpose, can then verify the results of the model on the basis of the use cases. This approach is – in the sense of computer technology – the necessary top-down design, starting with the desired process, the use case.

How does simulation work in SiMA? The first research question for the first use case was: Are the defined and developed basic functions of id, ego, and super-ego, and their sub-functions of the nervous system, worked out in the right way, and do they have enough functionality and fulfill the main requirements of psychoanalytic theory? To do this, psychoanalysts in SiMA developed the following use case: Two hungry people (agents), named Adam and Bodo, catch sight of a steak. Bodo happens to resemble Adam's brother. The narrative of the use case – following psychoanalytic knowledge – describes the procedure in the mental apparatus and how the agents behave. Do they want to share the steak? Which memories are activated regarding their siblings? What is the state of their emotions and feelings? By describing the use case in this way, the researchers are able to specify the parameters of the functions (not the behavior!) of the primary and secondary process.

Figure 7 shows the simulation procedure (iteration process) of the experiment principle. The starting point

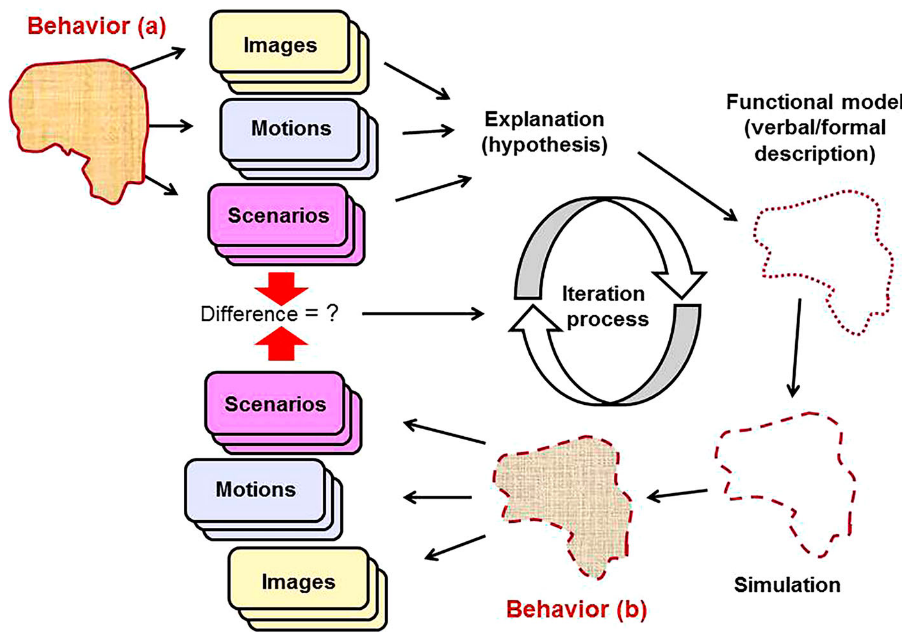

Figure 7. Simulation process.

in SiMA is the behavior (a), which is described in the use case. The behavior is described by values which are three different variables: drive representations, *images* and *motions*. Using neuropsychoanalytic models and explanations, this approach leads to the SiMA functional model of the agents (verbal/formal description). The use case, in which the agents operate, is simulated. The result is the behavior (b) of the model, which can be captured by drive representations, images and motions. The comparison of the variables and the scenarios of the behaviors in (a) and (b) also allow us to identify potential faults and inaccuracies in the model or the use cases. They indicate modifications which should be made in the model and use cases until all the faults are found and the inaccuracies of the experiment become reduced to an acceptable level.

A decisive advantage of a simulation is the simple way of testing all the process procedures (iteration process), which is the reason why the researchers in natural science have developed simulation programs in all areas. A psychoanalyst cannot have a view into the inside of the mental apparatus of his patient. He always depends on drawing the right inferences from the patient's behavior. In contrast to this disadvantage, the computer engineer can look in all corners of its simulated model because he integrates monitors in all functions and all places which the scientists want to observe. These monitors can record all variances that occur during operation. Figure 8 shows temporal sequences of several of these monitors in the described experiment: they record variables like elation, pleasure, drive, and perception, which can be compared with the expected perception (see Figure 7).

3.3. Axiomatics

A decisive requirement for simulation is to have precise definitions of all the terms to describe the functions of a model. Often people misunderstand this basic principle. We are not allowed to mix up the necessary, precise defined wording (the axioms) to describe the functions with the data which can deliver complete contradictory information contents in the model. This is a challenge in modeling or simulating any complex system, but SiMA is faced with two additional challenges. First, the diverse disciplines and schools of neurology, psychiatry, psychology and psychoanalysis have developed numerous, and sometimes contradictory definitions of a variety of terms. Second, which is one of the hardest challenges in SiMA, most psychoanalysts see themselves as humanists or social scientists and are concerned about over-simplifying or "operationalizing" the rich and individualized aspects of personal history; furthermore they do not want to give up the freedom to use terms in finely nuanced ways at different times and in different contexts, to be able to explain specific situations, objects etc. as well as possible. For reasons we hope are now clear, this does not work in the discipline of computer technology or computer science. We are not able – because of the limited length of this article – to explain this second and biggest problem of the

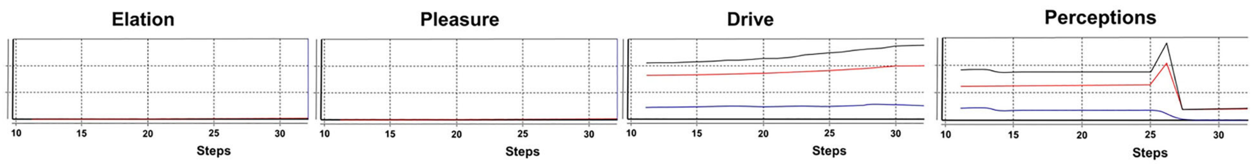

Figure 8. Monitors of the SiMA simulation program (Schaat, 2016, p. 138).

SiMA project in detail. The reader will find more detailed information in (Dietrich et al. 2015, p. 24).[4] However, we hope that psychoanalytic readers can understand the importance of precise and stable definitions of terms, and understand that our aim is to develop a model of the mind that can then account for, and allow us to simulate, moments and experiences of people that at least approximate the individualized, nuanced, ambivalent and conflicted mental life we each embody.

When this problem in SiMA was identified and understood, the team worked out specific axiomatics for the psychoanalytic wording. According to Bertrand Russell (Russell & Whitehead, 2009), a model needs a base of axioms. Axioms are not provable. They must be internally consistent and have to be defined unitary. Based on these definitions, rules and laws can be explored. In SiMA all the team members have worked on such a list of terms in English and German (see the left table in Figure 9), which we call the axiomatic glossary. All terms in it have to be thoroughly discussed again and again; when new terms are added, contradictions are recognized, or better wordings are found. Figure 9 shows an excerpt of the relationships between all the terms (the network of terms). The whole glossary of SiMA is published together with the scientific report (Dietrich et al., 2015, p. 150).

3.4. Mealy machine

To proceed with our description of the SiMA project, we must give a summary of some additional underlying concepts or tools, which may be challenging for the non-specialist reader. First, we will describe a mathematical device called the "Mealy machine" which is central to our project.

Natural scientists sometimes describe phenomena and use explanations which are not imaginable, since they are based on pure mathematical deductions. In such cases we have to rely on mathematics. For example, the theory of general relativity is understandable, if we analyze the respective differential geometry equations, but it is not imaginable. Similarly, a black hole in astronomy is not imaginable. The Mealy machine, which belongs to the information theory of computer technology, is such an abstract construction; it describes a phenomenon which is also not

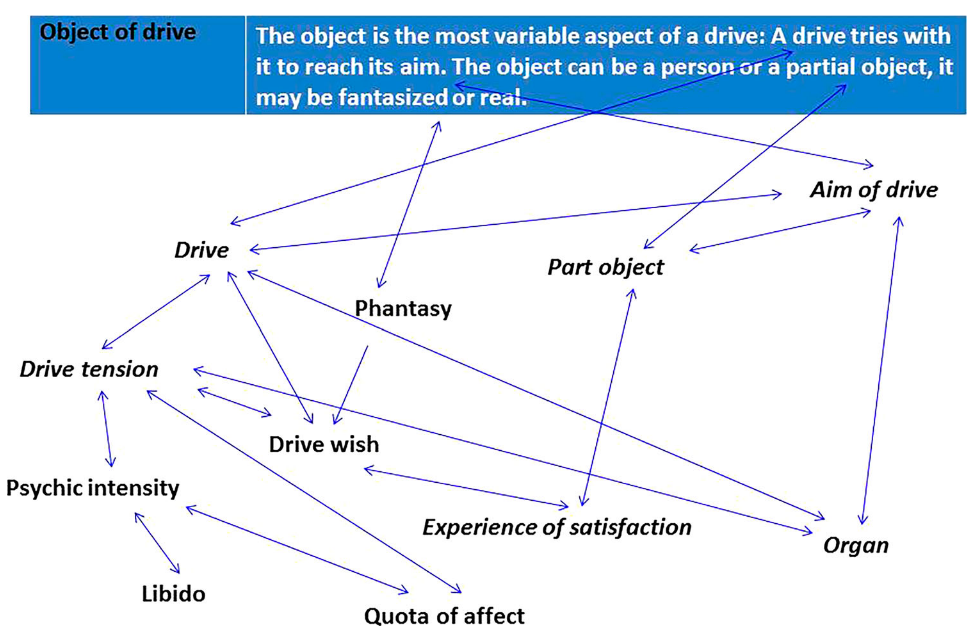

Figure 9. The neuropsychoanalytic axiomatics developed in the project (Dietrich et al., 2015, p. 150).

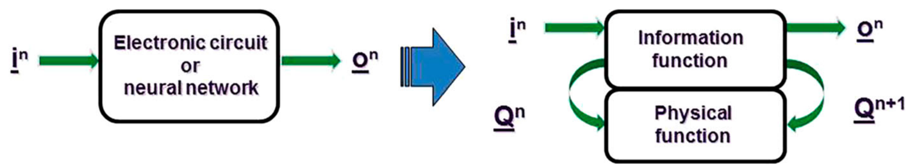

Figure 10. From the electronic design on the left hand side to the Mealy machine on the right hand side.

imaginable. However, to understand the SiMA model on the basis of the information theory, this machine is an essential precondition; therefore, here we describe how to apply the principle of the theory behind the Mealy machine, but we will not go into the deduction and the application in detail.

The theory of the Mealy machine allows computing all possible digitized processes. The basic idea is to divide the process in a *physical function* and an *information function*, which are explored separately (this procedure is only possible mathematically, but it is not imaginable). In other words, if we want to analyze or to compose a system, we have to map the left presentation in Figure 10 to the right presentation. The right representation shows the Mealy machine. It is essential that both functions of the Mealy machine are coupled by vectors, their interfaces. Again, we will not go into details; it is not necessary for a general comprehension.

In circuits (processes) – according to the Mealy machine – all physical parameters like time delays or temperature influences are only defined in the physical function block (lower block). Since it is a physical system, also the causality is defined in it. In the upper block (information function) only laws of information theory are applied; this is, for example, the algorithm, which connects information *a* with information *b* and leads to information *c*. Causal and temporal relationships do not apply in this function block.

It is obvious that in the model of the Mealy machine, the hardware (that is, the physical function) and information functions are strictly separated. The tasks of both functions are not allowed to be mixed. In contrast to this, it is not a problem to divide each function into several sub-functions. Engineers call functions and sub-functions in such a hierarchical system *layers* and *sub-layers* and the system itself a *layered system*.

Such a model is to be seen as a purely mathematical model and lies beyond our imagination. A human being cannot in any way imagine splitting a circuit, a system, an object, a process – however we call it – into a physical function (hardware) and an information function, coupled over information vectors. This is only possible using a mathematical representation. We can use this approach to extend our capabilities and analyze systems differently. It allows us to look at an information system/process like an electronic or a neuronal circuit and to describe and compute it as a model in all details. The reader will find many typical applications in internet or computer systems, in which the scientists developed one physical and six and more information layers (Wilamowski & Irwin, 2011, pp. 59–65). In SiMA we follow this principle.

4. SiMA model

Now that we have given a technical and theoretical context for our project, we can describe some of the core features of the SiMA project, in which we aim towards a psychodynamically valid model of the mind that can simulate specific moments or processes in the mental apparatus. In the following, we explain how we obtained the SiMA model based on what has been detailed so far, and we elaborate the structure of the model. Consequently, we also present and discuss the first results of the simulation experiments.

4.1. From the Mealy machine to the SiMA model

According to Mealy's principle (Figure 10) the nervous system is split into the *physical function* (layer 1) and the *information function*. The information function is again split into two further layers (Figure 11): the *neurosymbolic layer* (layer 2) and the *mental apparatus* (layer 3). The neurosymbolic layer prepares information coming from the hardware for the mental apparatus and coming from the mental apparatus going to the hardware (the (green) vectors in Figure 11).

The mental apparatus is the topic of psychoanalysis. Sigmund Freud differentiated this layer into the primary process and the secondary process (Freud, 1900). In his first model, the topographical model, Freud believed that the information in this layer (the lower sub-layer of layer 3 in Figure 11), the primary process, is unconscious, which means in the computer terminology that the topographical model is a behavior model.[5] However, in terms of Mark Solms' recent proposal of the "conscious id," this includes parts of the id

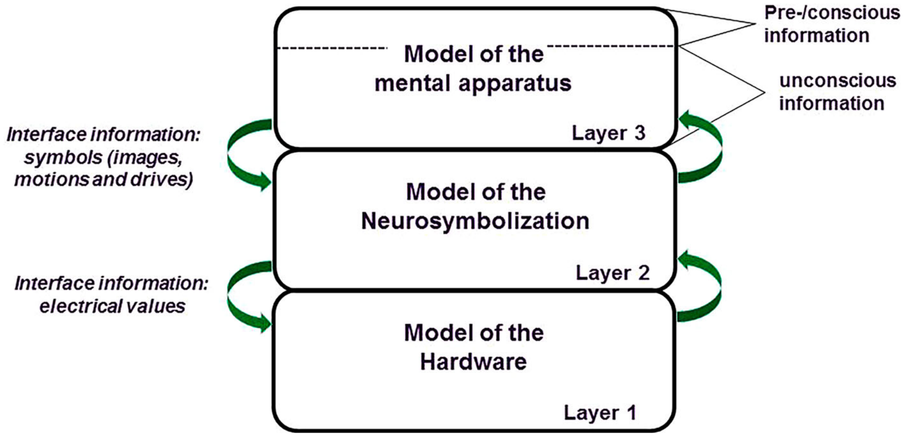

Figure 11. The layer model of the nervous system on the base of the Mealy theory.

which are conscious (Solms, 2013). In the secondary process most of the information is not unconscious.

The second model, developed by Freud, the structural model, includes the three functions id, ego, and super-ego. These are the functions in which the information is handled unconsciously, preconsciously/consciously, and these three functions are distributed over the whole mental apparatus (layer 3 in Figure 11).

The hardware layer (i.e. the physical/physiological function in Figure 11) describes the physical and biochemical conditions of the neurons, sensors and actuators. Pertaining to the theory of Mealy on one hand, all three layers are connected through their interfaces/vectors (arrows in Figure 11), but on the other hand, the layers are to be seen as independent of each other. From the point of view of layer 3, it is irrelevant which kind of hardware is implemented in layer 1. It is only important that the interface to layer 2, and respectively the interface from layer 2 to layer 3, fulfills the desired demands (Loy, Dietrich, & Schweinzer, 2001, p. 6). The interfaces and layer 2 and layer 1 must be able to handle the information process demands precisely and quick enough. Otherwise layer 3, in our case the mental apparatus is not able to meet the necessary requirements.

For the first experiments in SiMA, the strategy was to develop, explore, specify, and simulate layer 3 as precisely as possible, before the researchers work out layer 2 and finally layer 1 in detail.

The task of layer 2 in Figure 11 is the mapping of the electrical signals to symbols, or in other words the generation of abstract representatives of the neuronal information. According to Velik, 2008, we call them neurosymbols. These neurosymbols are coded in different hierarchical stages (sub-layers), each generating new neurosymbols, and are finally transformed to three types of symbols: *images*, *motions* and *drive representatives* (see also Figure 7). They are transferred over the interface/vector to layer 3. If the symbols are transferred back from layer 3 over the interface/vector to layer 2, they must be mapped to the neurosymbols of layer 2 and finally to electrical impulses in layer 1. This hierarchical coding structure of the neurosymbols is the preconditions that layer 3, the mental apparatus, is able to compute the information on the base of symbols.[6] However, it must be emphasized that the interfaces/vectors and the three function layers and sub-functions are (abstract) model conceptions. The goal of this kind of description is the ability to formalize and compute the system. It would not make any sense to try to imagine how the information of layer 1 flows over layer 2 and then to layer 3, because layer 2 and 3 are only abstract, information models, which do not follow physical but information laws. In other words, developing the model of the mental apparatus is a pre-condition for any actual implementation, which is a completely separate process.

4.2. The structure of the SiMA model

One of the main reasons why it makes sense to choose the classical psychoanalysis of Sigmund Freud from all the different psychological schools, theories, and models, is that he used a functional model himself. When he created the structural model for the third layer (the mental apparatus), he thought functionally, i.e. he created a model based on functions (see layer 3 in Figure 12; ego: blue, Super-ego: red, id: green). This suits computer engineers, who are also used to thinking functionally. What does this mean in relation to the SiMA model?

As discussed earlier, Braitenberg use the term *function* for objects (units or components like s*ensor, engine drive, etc.*) which can produce, transfer or memorize information. So, the *sensor* function/component can initiate electrical impulses which can be interpreted

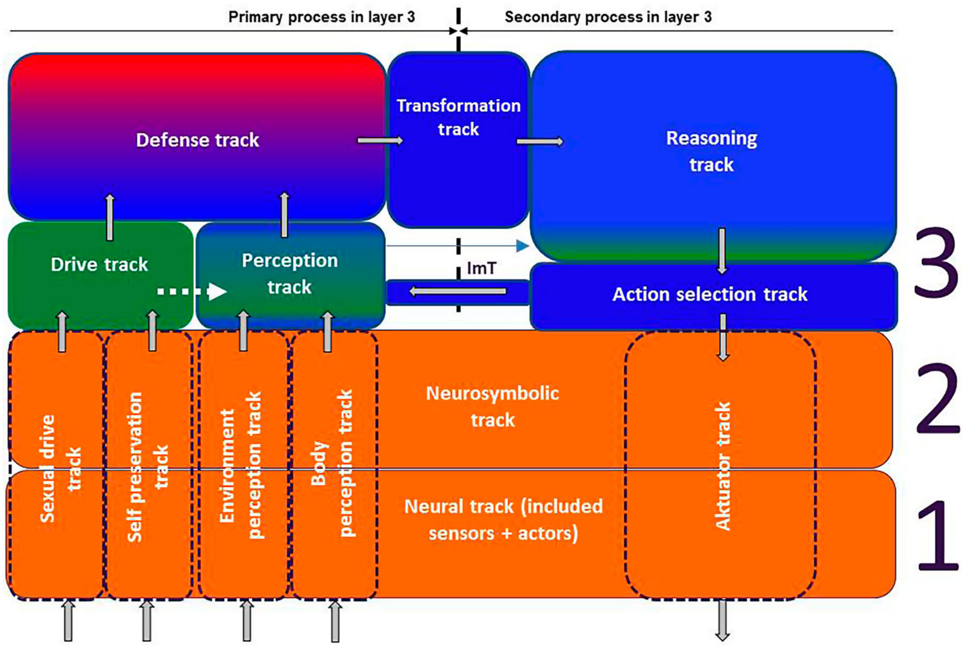

Figure 12. Functions of the 3 layers, but focusing on the mental apparatus (layer 3) (Dietrich et al., 2015, p. 73) (ImT: Imagination track).

as data and regarding a process as information. Such an object/component can be a specific network of transistors or neurons which can manipulate or memorize this data/information. In this sense, we can see the ego, super-ego, defenses, etc. as functions which compute the information in diverse ways, affected by memories and inputs from other functions or sensors. To point out that both unconscious and conscious information flow through the SiMA model – in Figure 12 from the bottom left hand side through layer 1, up to layer 2 and 3, going to the right part down and back to layer 1 – we use the term *track*.

We now will go into more detail about the essential structure of the model. The information inputs from the physical world, located on the lower left side of the figure, are subdivided into four information functions: the sexual drive track, the self-preservation drive track, the environment perception track, and the body perception track. In layer 1, the information consists of physical values (current, voltage, impulse frequency etc.), which are coded to neurosymbols in layer 2, and finally transferred to layer 3 as images, motions, and drive representations. The inner workings of layer 1 and 2 are not further elaborated here, since we want to focus on layer 3.

Figure 12 shows a step of the developed model in which one can get a good overview of the functionality and data flow of the whole system. Because of this, functions often include a mix of the three basic functions: ego, super-ego, id. The final development step of the SiMA model, in which it is broken down into smaller functions until they reach a size that can be implemented in programs, includes 46 functions in which each function is assigned to one of the three basic functions. These smallest defined functions are worked out in details and described especially by Muchitsch (2013), Gelbard (2015), Schaat (2016), and Wendt (2016). A detailed description, also of the least abstracted functions, is given in Dietrich et al. (2015).

Information is sent from the drive track to the perception track to evaluate it, which corresponds to the process of “cathexis” in psychoanalysis.

Figure 12 shows that the reasoning track in the secondary process contains an id-function, namely the perception of feeling, which corresponds to the conscious part of the id according to Solms (2013). This contradicts the classic theory of psychoanalysis, in which the information in the id is only unconscious in the primary process.

The different contents, which were generated in the drive track and the perception track by incoming data of layer 1 and layer 2, and also evaluated by quota of affects and emotions, can create conflicts with each other and with the demands of the super-ego (the red field in Figure 12). The task of the defense track is to retain or modify content using specific defense mechanisms, so that they become acceptable for the secondary process. In a next step the contents, which passed the defense track, are modified in the transformation track so that they obey the causal, rational and temporal laws and are able to become conscious. In the reasoning track, the contents are focused on the need that reached the highest evaluation. In psychoanalytic terms, we can say that the functions choose the need with the best possible feeling – with the highest amount of pleasure. The action selection track determines an action, which should be executed physiologically. The actuator track defines the moving patterns how the body should operate. In other words, this track determines which muscles and glands should be activated. Layer 1, the hardware, decodes the lowest level of symbols into electrical impulses which leads to physical actions. Activities which were not executed flow back over the imagination track into the primary process and influence the perception track.

A human being is defined by the experiences he or she has had in the past. According to Solms and Turnbull (2002) human beings execute actions in rehearsal before they make a decision – and they pick the action that causes the most pleasure and the least unpleasure (Solms, 1997). To do so, human beings associate evaluated drive representations, images and motions that they take from their history with the current perception and the status of the drives. The result of this process is commonly phrased as "It feels like the right decision".

A simulation model of the psychic apparatus requires the concept of memories. This would make it necessary to recreate and implement all memories of an actual human being who has a history over years – which is an impossible task. This means that in the near future there will be no simulation of a single, specific person – but this is not our goal. Instead, we hope that SiMA will validate the findings of neuropsychoanalysis, which allows us to make a significant simplification. For the first experiments, we assume a human being who is about 30 years old and lived about 30,000 years ago. It receives an implementation of certain relevant memories that allow us to research behavior that are close to reality. The development and selection of the specific memories is the task of the experts and have to originate in psychoanalytic findings. The better the experts can formulate the memories, the better the simulation results will be.

5. Results of the experiments

To get optimized design results, a top-down approach is the standard method used in computer technology. Hence, first the top-most layer in Figure 12, i.e. the mental apparatus, has to be modeled and programed before we can model the other layers, i.e. the hardware. Therefore, in the first step of the SiMA project all functional requirements have to be defined in layer 3. Coming back to the use case from Section 3.2, where Adam likes to eat a steak, which also is perceived by Bodo, we ask: How did psychoanalysts specify Adam and analyze his behavior? Is this compliant with the simulation results?

In this use case, we can see the outcome of changing the strength of input data in combination with the characteristics of different functions. For instance, say that Adam has a high level of hunger. He does not necessarily take the whole steak and eat it himself: due to perceiving Bodo, who looks similar to his brother and therefore activates corresponding super-ego rules, different decisions are possible. We found, in one simulation, that high anger would lead Adam to share the steak with Bodo. Initially this sounds contradictory, but we are able to track the psychic processing that leads to that situation in our simulation. We showed (in Schaat, 2016) how undirected anger would lead to sharing in a personality with strong super-ego and the defense mechanism of reversal of affect. In such a case, anger is defended by guilt, which activates memories of sharing.

Thus, one advantage of simulation is to observe the interdependencies and interactions between the impact factors that determine behavior. For instance, high anger and a strong super-ego only lead to cooperative behavior if Adam has high drive representatives. In other words, there would only be cooperative behavior in this instance if Adam is very hungry, in which case his drive need would override the feelings or behavior generated by high anger. Of course, there is no direct influence of drive representatives, but in simulation we see how only sufficiently high drive representatives generate enough neutralized intensity for the defense mechanism of reversal of affect. However, too high drive representatives would lead Adam to eat the steak alone. Hence, in simulation we can observe and track the border between selfish and cooperative behavior. All these relations were programed, simulated and validated in the PhD theses of SiMA (see e.g.: Muchitsch 13; Hinterleitner, 2014; Gelbard, 2015; Schaat, 2016; Wendt, 2016).

Another easily comprehensible example of a simulation experiment is treated in Brandstätter, 2017. Dany plans a trip from Vienna to Salzburg, which takes about 3 h. He should give a scientific presentation at the University of Salzburg at 8:00 pm, but he cannot leave Vienna before 10:00 am. Dany can use various means of transport: car, bus, railway, hitchhiking, etc. Dany's mother lives a little distance off of the route. She would be glad to see him again. So, the agent has to find the right decision: Should he go directly to Salzburg to be able to rehearse his speech? Should he visit his mother, or should he try to save money? This decision depends on objective facts (money, time, comfort, etc.), but also on subjective mental states, wishes, and fantasies. The following subjective factors and parameters for the decision are perhaps essential: (1) the amount of the libido (hunger, sexual drive etc.) and his manner of handling it, (2) miscellaneous memories, which have different cathexes (regarding the mother, different speeches, times when he was not really good prepared etc.), and (3) abilities to sublimate, to desexualize, to think anticipatively, or other typical defense mechanisms, which Dany developed to solve conflicts. Finally, Dany will make a decision on the basis of the pleasure principle and reality principle of the primary and secondary processes, respectively.

In other words, the decision which has the biggest chance to become realized is that which promises the most pleasure and offers the least unpleasure. All of these relations and sequences can be visualized and validated in all details step by step in the simulation program of SiMA with the help of monitors, several of which are shown in Figure 7 as examples. Contradictions between theory and simulation sequences can thus become visible directly.

6. Technical applications

In 1999, the original goal of SiMA was to get knowledge about the functions of the mental apparatus (Dietrich, 2000), and to be able to apply it for industrial challenges – for example, to increase energy efficiency in all areas. However, SiMA focuses only on complex processes, because this is the strength of the mental apparatus. Any calculator can do mathematical operations much faster than humans can, but humans excel in handling complex processes. Anything else is better handled by algorithms that are developed in the domains of automation, computer technology, computer science, classical artificial intelligence, robotics or cognitive science using classical methods.

As a result of the SiMA project, two industrial projects were submitted and received a research grant: project CogMAS (Cognitive Multi-Agent System Supporting Marketing Strategies of Environmental-Friendly Energy Products; Schaat et al., 2015) and project KORE (Cognitive Control Strategy Optimization to Increase Energy Efficiency in Buildings; Zucker, Wendt, Siafara, & Schaat, 2016). Both have completely different application domains, but share one commonality: The systems show complex behavior that cannot be handled using a conventional approach. The evaluation of both projects shows that the underlying psychoanalytic model is applicable to technical problems, which is explained more detailed in Schaat (2016).

7. Open issues, limitations and possibilities

The SiMA model is the beginning of a new idea that builds the bridge between the neuronal layer and the psychological layer, brought together in one consistent model. This raises the question of where the limits of the approach are reached, both in technical and in societal regards. As for the technical aspects, we see memories, which are valuated by pleasure and unpleasure, as the defining factor of a human being; they build the foundation for making decisions. Until now we have only developed simplified use cases to validate the model and the underlying theories of neuropsychoanalysis, but this has to change and become more complex. A simulation of complex use cases requires a huge amount of memories, which have to be available to the simulation model. Consequently, through this process we can only support theories of psychoanalysis, but SiMA cannot be used for the complex dynamics and nuances of individual patient cases in therapy. First we have to create simulation tools, which allow us to acquire and process all available data. Then the model has to be engineered in detail, experiments have to be automated, and the evaluation process has to be revisited, since it is our opinion that this process is not sufficiently mature yet.

Looking at all the topics that have not been addressed in the model so far, we see the following main issues: in 2015 a sister project of SiMA worked out the *self* of the human being (Doblhammer, 2015). These results have to be integrated into SiMA. *Language* has not been thoroughly addressed, which is closely related to the concept of *consciousness,* and *unconscious, preconscious* and *conscious information*. Another topic is the integration of *hormones* as carriers of information, but this is not such a critical issue, although it may be quite a tedious task. A bigger challenge is *learning* and, similarly, *intelligence*. SiMA touched these topics. Learning cannot be defined generally but has to be mapped to each specific function in Figure 12. Many different learning approaches will have to be implemented for these different modules. All of these topics are topics of ongoing PhD work.

In conclusion, we suggest that a number of interesting avenues are opened by SiMA. Diverse questions in neuropsychoanalysis are based on hypotheses and logical deduction. The SiMA simulation tool can provide answers that are solely based on natural science (Dietrich et al., 2017).

One final aspect that shall be addressed here: SiMA has raised a lot of controversy in academia. There are not many engineers and scientists who are open to neuropsychoanalysis. Raising awareness for this field is a task for the future. But the other side faces similar challenges: psychoanalysis has to open the door for natural sciences and engineering in general or it will fall behind and become deprecated. This would result in insufficient funding to initiate and continue research projects like SiMA. We hope that our survey of the work so far will encourage readers to participate in this endeavor, and raise awareness about the potentials for exploring this interface between the mind and the nervous system.

Acknowledgement

The authors are thankful to Richard J. Kessler, D.O, for his valuable feedback in the course of writing this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is partly funded by the Austrian Research Promotion Agency (FFG) within project KORE [grant agreement number 848805].

Notes

1. The requirements in computer technology are mainly to research and develop hardware and software in combination (closing the gap between both areas). The area of computer scientists is mainly algorithms.
2. Only the dynamic (dataflow) is in the focus, not the functions (the entities, the structure etc.) of a system, which determines the dataflow.
3. In computer technology processes are often abstracted, so that engineers can use behavior models in the sense of Braitenberg which are much easier to program with specific algorithms, because the market offers highly efficient tools for this task.
4. One of the authors, Dietmar Dietrich, worked in such international bodies for over 20 years, which was essential for the work in SiMA.
5. We have to consider that in the information theory of computer technology the functional models are described by their functions (see again (Braitenberg, 1984) and behavioral models by information, the data which are memorized, and manipulated by the functions.
6. Here it must indicate that the term symbol is used in the sense of the computer information theory. It is nothing else as an abstract representation of something. In psychoanalysis a symbol has a different meaning.

ORCID

Gerhard Zucker http://orcid.org/0000-0003-4258-2849

References

Braitenberg, V. (1984). *Vehicles - experiments in synthetic psychology.* London: The MIT Press Cambridge.

Brandstätter, C. (expected end of 2017). *Brain-like location awareness and navigation for psychoanalytically inspired autonomous agents* (Ph.D. thesis at the Institute of Computer Technology). Technische Universität, Wien, Austria.

Brooks, R. (1985, September). *A robust layered control system for a mobile robot* (MIT AI Lab Memo 864).

Canamero, L. (1997). Modeling motivations and emotions as a basis for intelligent behavior. In W. L. Johnson (Hrsg.), *Proceeding of the first international conference on autonomous agents* (pp. 148–155). New York: The ACM Press.

Damasio, A. (2003). *Looking for Spinoza*. London: Vintage Books.

Dietrich, D. (2000). *Evolution potentials for fieldbus systems.* IEEE international workshop on factory communication systems WFCS 2000. Instituto Superior de Engenharia do Porto, Portugal, pp. 343–350.

Dietrich, D., Brandstätter, C., Doblhammer, K., Fittner, M., Fodor, G., Gelbard, F., … Widholm, R. (2015). *Natural scientific, psychoanalytical model of the psyche for simulation and emulation* (Scientific Report). Wien, Austria: Technische Universität. Retrieved from http://publik.tuwien.ac.at/files/PubDat_240983.pdf

Dietrich, D., Fodor, G., Zucker, G., & Bruckner, D. (2009). *Simulating the mind – a technical neuropsychoanalytical approach*. Wien: Springer.

Dietrich, D., Jakubec, M., Schaat, S., Doblhammer, K., Fodor, F., & Brandstätter, C. (2017). The fourth outrage of man - is the turing-test still up to date? *Journal of Computers*, *12*(2), 116–126. (Commitment for publishing).

Doblhammer, K. (2015). *Das Selbst eines Roboters.* Online unter. Retrieved from http://sima.ict.tuwien.ac.at/new/wp-content/uploads/2015/02/Das-Selbst-eines-Roboters-abschlussbericht_final_150115.pdf (Stand: 30.09.2016)

Freud, S. (1900). Die Traumdeutung; GW II/III, S. 607; Fischer Verlag, 1999.

Gelbard, F. (2015). *Psychoanalytic defense mechanisms in artificial intelligence exemplified implementation in complex technical systems* (Ph.D. thesis at the Institute of Computer Technology). Technische Universität, Wien, Austria.

Gill, M. M. (1976). Metapsychology is not Psychology, in Gill/Holzmann: Psychology is Metapsychology: Essays in Memory of George Klein, New York.

Gill, M. M. (1977). Psychic energy reconsidered: Discussion. *Journal of the American Psychoanalytic Association*, *25*, 581–597.

Hinterleitner, I. (2014). *Modelling of inner and outer speech by means of situational context* (Ph.D. thesis at the Institute of Computer Technology). Technische Universität, Wien, Austria.

Holt, R. R. (1973). The crisis of metapsychology. Ansprache in der American academy of psychoanalysis. Page 500 in: Rangell, L. (1975). Psychoanalyse und Veränderung: Ein Versuch über Vergangenheit, Gegenwart und Zukunft. *Psyche – Z Psychoanalytic*, *29*, 481–502.

Holt, R. R. (1985). The current status of psychoanalytic theory. *Psychoanalytic Psychology*, *2*, 289–315.

Lange, D. (1976). *Aufbau zweier Neuronenmodelle mit modernen Bauelementen nach Küpfmüller-Jenik zur Nachbildung gegenseitiger Erregungs- und Hemmvorgänge bei neuronaler Verkopplung über exzitatorische beziehungsweise inhibitorische Synapsen* (diploma thesis). Technical University of Karlsruhe, Institute of Cybernetic.

Lorenz, B., & Barnard, E. (2007). *A brief overview of artificial intelligence focusing on computational models of emotions.* Proceedings of ENF 2007 – 1st International Engineering & Neuro-Psychoanalysis Forum, 1–12, July 23, Vienna, 2007.

Loy, D., Dietrich, D., & Schweinzer, H.-J. (2001). Open Control Networks, LonWorks/EIA 709 Technology; Kluwer Academic Publishers.

Minsky, M. (1990). Logical vs. Analogical or Symbolic vs. Connectionist or Neat vs. Scruffy. In Artificial Intelligence at MIT, Expanding Frontiers 1.

Minsky, M., & Papert, S. (1969). *Perceptrons, an introduction to computational geometry.* Cambridge: MIT Press.

Muchitsch, C. (2013). *Human-like perception for psychoanalytically inspired reasoning units* (Ph.D. thesis at the Institute of Computer Technology). Technische Universität, Wien, Austria.

Palensky, B. (2008). *Introducing neuro-psychoanalysis towards the design of cognitive and affective automation systems* (Ph.D. thesis at the Institute of Computer Technology). Technische Universität, Wien, Austria.

Peterfreund, E. (1973). On information-processing models for mental phenomena - a reply to Lawrence Friedman's critical review of information, systems, and psychoanalysis by Emanuel Peterfreund. *The International Journal of Psychoanalysis*, *54*, 351–357.

Peterfreund, E., & Schwartz, J. (1971). *Information, systems, and psychoanalysis. An evolutionary biological approach to psychoanalytic theory.* New York: International University Press.

Russell, B., & Whitehead, A. N. (2009). *Principia Mathematica, 3 volumes, Merchant Books.*

Schaat, S. (2016). *Simulation of foundational human - information-processing in social context* (Ph.D. thesis at the Institute of Computer Technology). Technische Universität, Wien, Austria.

Schaat, S., Dickert, S., Geveze, E., Miladinovic, A., Wilker, S., & Gruber, V. (2015). *Modelling emotion and social norms for consumer simulations exemplified in social media*. International conference on affective computing and intelligent interaction, pp. 851–856, Xian.

Simon, H. A. (1965). *The New Science of Management Decision: The Shape of Automation for Men and Management*. New York: Harper Torch Books.

Solms, M. (1997). Descartes' error: Emotion, reason, and the human brain. (By Antonio Damasio. New York: Putnam, 1994). *Journal of the American Psychoanalytic Association*, *45*, 959–964.

Solms, M. (2013). The conscious id. *Neuropsychoanalysis*, *15*, 5–19.

Solms, M., & Turnbull, O. (2002). *The brain and the inner world*. London: Karnac/Other Press, Cathy Miller Foreign Rights Agency.

Velik, R. (2008). *A bionic model for human-like machine perception* (PhD thesis at the Institute of Computer Technology). Technische Universität, Wien, Austria.

Wendt, A. (2016). *Experience-based decision-making in a cognitive architecture* (PhD thesis at the Institute of Computer Technology). Technische Universität, Wien, Austria.

Wilamowski, B. M., & Irwin, J. D. (2011). *The industrial electronics handbook, industrial communication systems.* Boca Raton, FL: CRC Taylor & Francis.

Zucker, G., Wendt, A., Siafara, L. C., & Schaat, S. (2016). *A cognitive architecture for building automation*. Proceedings of the 42nd annual conference of IEEE industrial electronics society (IECON 2016), Firenze, Italy.