Science of Intelligence

White Paper

November 2011

Authors

This document is part of a rejected draft proposal for the second round of the German excellence initiative. The draft proposal was submitted to the DFG in September 2010. The official reviews by the DFG are appended. The following people contributed to the discussions that ultimately led to this White Paper:

Günter Abel	TU Berlin, Philosophy
Henry Brighton	MPI Berlin, Adaptive Behavior and Cognition
Oliver Brock	TU Berlin, Robotics & Computational Biology
Peter Frensch	HU Berlin, General Psychology
Gerd Gigerenzer	MPI Berlin, Adaptive Behavior and Cognition
Verena Hafner	HU Berlin, Cognitive Robotics
John-Dylan Haynes	Charité – Universitätsmedizin Berlin, BCCN
Ulman Lindenberger	MPI Berlin, Lifespan Psychology
Klaus-Robert Müller	TU Berlin, Machine Learning
Klaus Obermayer	TU Berlin, Neural Information Processing
Jörg Raisch	TU Berlin, Control Systems
Raul Rojas	FU Berlin, Artificial Intelligence
Marc Toussaint	FU Berlin, Machine Learning & Robotics
Elke van der Meer	HU Berlin, Cognitive Psychology
Felix Wichmann	Uni Tübingen, Neural Information Processing

We would like to acknowledge the financial support provided by TU Berlin for the preparation of the draft proposal.

Purpose and limitations of this document

Since this document developed as a proposal for a German Excellence Cluster there are inherent limitations. The presentation of the overall research agenda is exemplified through concrete projects that were part of the original proposal. Citations to our own work outnumbers citations of standard work in the literature. In a peer-reviewed scientific publication this would not be appropriate. We apologize in advance to all those who feel like their work should have been cited here.

Nevertheless, we think that this white paper is a valuable basis for a high-level discussion of research agendas which is essential to progress the field but typically out of the scope of scientific publications. Section 2 is a meta-level discussion of how to approach intelligence research in general. The reader interested in more concrete research agendas may jump directly to section 3 which—within the limits of this proposal format—discusses more specific research challenges and methodological approaches. Section 4 briefly outlines ideas on an educational curriculum to teach students an transdisciplinary approach to intelligence research.

If you have comments or suggestions, please feel free to contact Oliver Brock (oliver.brock@tu-berlin.de) or Marc Toussaint (marc.toussaint@fu-berlin.de).

1 Abstract of the Proposal

Understanding intelligence is one of the great scientific challenges of our time. Yet, in spite of extensive research efforts spanning many scientific disciplines, our understanding of intelligence remains fragmented and incomplete.

This white paper proposes a transdisciplinary research program to overcome the fragmentation of existing intelligence-related research, enabling progress towards a complete understanding of intelligence. A defining characteristic of the proposed Science of Intelligence is the synthetic approach to intelligence research. This approach requires that each insight, method, concept, and theory must demonstrate its merit by contributing to the generation of intelligent behavior in a synthetic artifact, such as a robot or a computer program. In doing so, the proposed Science of Intelligence overcomes the fragmentation of existing intelligence-related research by enforcing a truly transdisciplinary conceptual and physical integration. The synthetic approach also overcomes the dichotomy of basic science (concerned with the principles of physical and computational processes that enable intelligence) and applied science (studying and synthesizing specific behaviors in which intelligence manifests itself) because results of the two types of research must be integrated and validated in a single artifact.

Applying the proposed scientific strategies to a focused research program will fundamentally advance our understanding of intelligence, will unify theories, concepts, and insights from existing intelligence-related disciplines, will—through this unification—in turn catalyze progress in these disciplines, and—maybe most importantly—will advance our ability to construct intelligent technological artifacts for applications of societal importance.

2 Scientific and Structural Strategies

The proposed *Science of Intelligence* (SCIoI) aims to establish intelligence research as a unified scientific discipline. Today, intelligence research is highly fragmented and distributed across over a dozen disciplines, each leading to theories and empirical findings about specific aspects of intelligence. These theories and findings remain controversial and their synthesis into a unified understanding of intelligence represents a formidable and possibly insurmountable challenge. The scientific and structural strategies of this science provide a framework that avoids the tendency of today's science to become increasingly specialized and narrow—a trend actively opposed also in other scientific disciplines through inter- and transdisciplinary research efforts.

2.1 A Science of Intelligence

Today's Intelligence Research

Research on both human and machine intelligence in the recent past has been characterized by a diversity of theoretical perspectives and empirical approaches. In the area of human intelligence research, for example, theorists have not even been able to agree on a definition of what constitutes intelligence. A recent report published by the Board of Scientific Affairs of the American Psychological Association, the largest and most influential psychological organization worldwide, comes to the conclusion that "although considerable clarity has been achieved in some areas, no such conceptualization has yet answered all the important questions, and none commands universal assent. Indeed, when two dozen prominent theorists were recently asked to define intelligence, they gave two dozen, somewhat different, definitions." It is therefore not surprising that research on human intelligence has followed many disconnected paths. For example, much of the existing research has focused on generating and validating intelligence tests. These efforts have been entirely disconnected from empirical attempts to understand information-processing components

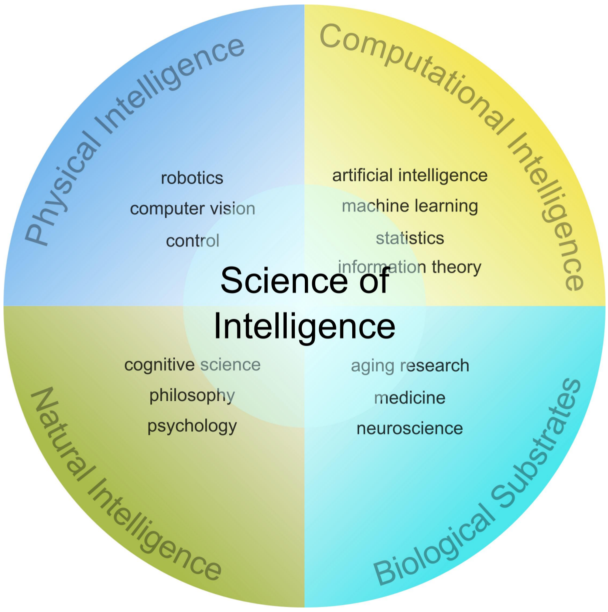

Figure 1: The proposed research unites four existing perspectives of intelligence research into a transdisciplinary "globe," representing a Science of Intelligence. The Northern Hemisphere represents synthetic approaches and the Southern Hemisphere analytic approaches.

of intelligence. As a result, today's understanding of human intelligence remains incomplete and fragmented. Research does not benefit from a common and consistent research strategy.

A similarly fragmented picture emerges in the context of Artificial Intelligence. After an initial focus on symbol processing, Artificial Intelligence specialized in diverse sub-disciplines like robotics, computer vision, machine learning, planning and reasoning, etc. Each of these fields today has led to significant scientific and commercial success in their own specific domain. For example, robotics has significant commercial relevance in automation and entertainment, and machine learning has become a versatile tool for data analysis in various scientific and commercial areas. However, despite (or maybe partially as a consequence of) their achievements in solving specialized problems, these fields remain disconnected and lost focus on the overall goal of achieving an integrated understanding of intelligence.

To summarize bluntly: Much of the research in intelligence-related disciplines has digressed from explaining the phenomenon of intelligence as a whole and resorted to explaining smaller and smaller aspects and components of it. This fragmentation is mirrored in the multitude of intelligence-related disciplines. They can be divided into four categories: those concerned with creating embodied agents, those that study computational aspects of intelligence, those that are concerned with natural intelligence, and those that study the biological substrates of intelligence (see Figure 1).

Today, intelligence research faces the challenge of integrating numerous findings, theories, and insights into a coherent and complete theory of intelligence. Ongoing research does not seem to be heading into this direction: Different fields of intelligence-related research continue to pursue a fragmented view of intelligence, make vastly divergent assumptions, communicate using inconsistent terminology, approach problems from irreconcilably different perspectives: a scientific Tower of Babel.

The Proposed Science of Intelligence

- The proposed Science of Intelligence (SCIoI) creates an understanding of intelligence informed by the views, findings, and theories of all intelligence-related disciplines. This consistency is a consequence of *involvement* of the disciplines in the *process* of creating this understanding—not a consequence of integration of existing theories.

- The understanding of intelligence developed by SCIoI will be constructive. By this we mean that every insight regarding intelligence must be demonstrated and validated by contributing to the intelligent behavior of a technological artifact. As Richard Feynman put it:

  > *What I cannot create, I do not understand.*
  > Richard Feynman (1918-1988)

- SCIoI aims to establish a global scientific community. This community will actively advance, evolve, and contribute to the proposed Science of Intelligence.
- SCIoI is synergistic with ongoing efforts in intelligence research; it is intended to complement the established disciplines—not to imply that other efforts have not been successful or insufficiently scientific.

The resulting science will provide new insights into intelligence and foster our ability to transfer these insights to machines, thereby advancing many application areas. We also expect that progress achieved in Science of Intelligence will serve as catalyst and accelerator for research within the established disciplines.

We acknowledge that the subject matter of intelligence cannot be studied exhaustively within a five-year period. To validate the strategies of SCIoI, we will conduct a focused transdisciplinary research program on specific and fundamental aspects of intelligence. This research program is described in section 3.

2.2 Scientific and Structural Strategies

Synthetic approach: The core scientific strategy of the proposed Science of Intelligence is to “create it”, following Feynman’s quote above. Each insight, method, concept, or theory developed as part of SCIoI must demonstrate its merit by contributing to the generation of intelligent behavior in a synthetic artifact. In this proposal, we refer to this as the synthetic approach to intelligence research. The synthetic approach prevents fragmentation as all disciplines have to produce results, theories, and insights that contribute to the behavior of a technological artifact. This enforces a truly transdisciplinary conceptual and physical integration among disciplines. Examples of the implications of this approach are given in Section 3.

From the perspective of analytic disciplines, the synthetic approach implies a shift from statistical modeling of data towards truly functional modeling of behavior. Instead of developing statistical methods to model experimental data, the synthetic approach aims to model the observed behavior by reproducing it in artificial systems.

Basic science and applied science: Intelligence organizes physical and computational processes into task-specific behavior. Intelligence research must therefore combine two perspectives: First, research concerned with the principles that govern the organization of physical and computational processes – which we call *basic science*. However, the appropriate organization of physical and computational processes—and therefore also the principles that govern this organization—depend on the specific task or challenge in an ecological niche. Consequently, the principles of intelligence can only be investigated in the context of a particular task, implying that a Science of Intelligence must also include research concerned with understanding the particular constraints and structure implicit in specific tasks, environments, and behaviors – which we call *applied science*. Intelligence research is therefore neither a basic science nor an applied science alone—it must be both at the same time.

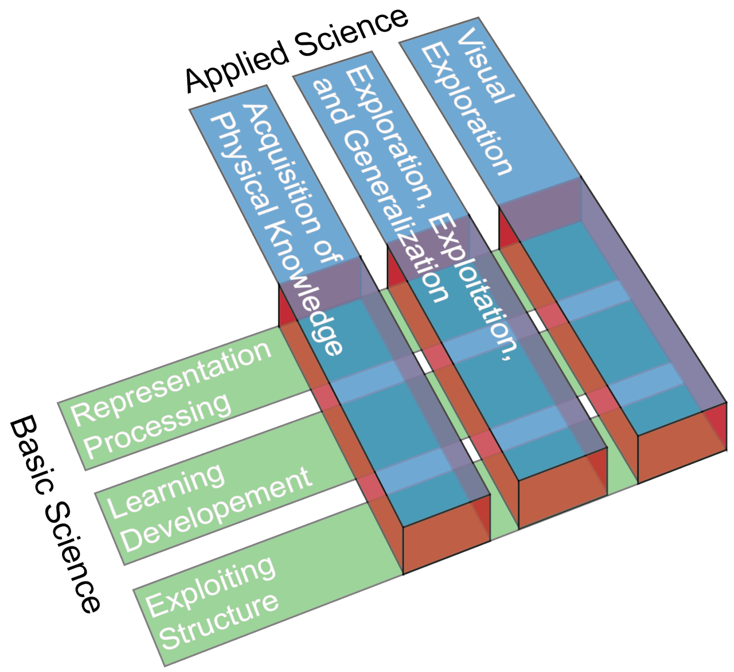

Figure 2: SCIoI links basic research (green) with applied (problem-oriented) research (blue) through the creation of synthetic artifacts (red). Basic research investigates how physical and computational processes can be organized to generate intelligent behavior. Applied research projects (three example projects shown) synthesize technological artifacts that replicate specific aspects of intelligent behavior. Both basic and applied projects must contribute to the creation of behavior in the same artifacts, effectively uniting aspects of basic and applied science.

A successful science of intelligence must account for this unique characteristic of intelligence research. A sole emphasis on basic (curiosity-driven) research will lead to theories that explain partial aspects of intelligence without furthering our understanding of intelligent behavior as a whole and without relevance to practical aspects. Research that is exclusively applied runs the risk of turning into an engineering exercise to solve an application-specific problem without uncovering generalizable principles.

The proposed Science of Intelligence unites basic research and applied research through the creation of synthetic artifacts. Basic research within SCIoI (green rectangles in Figure 2) investigates fundamental aspects of intelligence across different behavioral manifestations. Applied research projects within SCIoI (blue rectangles) recreate specific behaviors considered to be hallmarks of intelligence. These projects span and integrate the aspects of intelligence investigated within the basic research projects. This mutual “spanning” of basic and applied research is leveraged to link basic and applied research projects through synthetic artifacts (red boxes): basic science projects must contribute to the synthesizing of intelligent behavior in synthetic artifacts. These are the same artifacts synthesized in the applied research project. The result is a distinct and shared focus on the creation *and understanding* of intelligent behavior.

The synthetic approach will focus and direct research in the basic sciences. These sciences must now concentrate on developing theories and insights that can be transferred to scientific artifact. These theories must provide constructive models at the functional and behavioral level rather than statistical models over experimental data. At the same time, applied research projects are prevented from becoming engineering exercises as they must continuously incorporate, validate, and advance the theories of basic science.

2.3 What Science of Intelligence is Not

- *SCIoI is not artificial intelligence.* Artificial intelligence (AI) and SCIoI share the objective of creating intelligent artifacts. But even though AI has long moved beyond symbol manipulation and today draws inspirations from many other disciplines, SCIoI differs from AI through its distinct emphasis on transdisciplinarity, supported by a unifying research strategy.
- *SCIoI is not cognitive science.* Cognitive science studies intelligence and information processing in brains and machines in an interdisciplinary fashion. Cognitive science shares with SCIoI the objective to explain intelligence. SCIoI goes beyond this in its focus on synthesizing intelligence in technological artifacts. The synthetic approach puts research methods at the center of intelligence research that are not within the scope of cognitive science.
- *SCIoI is not product development.* Our aim is not to develop commercial machines. We use artifacts and machines to test and validate scientific hypotheses and to re-create intelligent behavior. Clearly, these artifacts must possess capabilities not realized before, but it is not our primary objective to render these artifacts immediately useful to specific commercial applications.

3 Research Areas

In this section we describe the proposed SCIoI research agenda that implements the strategies described above. It consists of basic research areas to study principles of intelligence relevant to many intelligent behaviors and applied research projects to synthesize specific behavior considered to be a hallmark of intelligence. Basic research areas and applied research projects interconnect and are linked through synthetic artifacts, as illustrated in Figure 2 on page 5.

We acknowledge that the subject matter of intelligence cannot be studied exhaustively within a five-year period. To validate the strategies of SCIoI, we will therefore focus on a specific and very fundamental aspect of intelligence: on intelligent physical behavior, i.e., intelligence pertaining to the physical interactions between intelligent agents and their environment. We view this “physical intelligence” as a necessary precondition for other aspects of intelligence, including language and social or societal aspects of intelligence. An understanding of intelligence as it manifests itself in physical interactions can serve as the starting point for investigations of these other aspects.

The basic research areas address three fundamental aspects of intelligence: *1)* representations & processing, *2)* learning & development, and *3)* leveraging structure. The aspect “representations & processing” investigates closed-loop perception-action systems that exhibit intelligent behavior. Research on “learning & development” examines how intelligent agents develop and adapt their representations and processing to improve performance. The area concerned with “leveraging structure” studies how structure inherent to the problem, the environment, or the agent is leveraged to find solutions in the first two areas.

Applied research projects synthesize robotic and software systems that exhibit behavior considered to be a hallmark of intelligence in animals or humans. Throughout the lifetime of the proposed research projects, new applied research projects will be created, targeting increasingly complex behaviors based on the artifacts created in earlier projects.

In the following sections, we describe the three basic research areas (Sections 3.1–3.3) and three selected applied research projects (Sections 3.4–3.6). The projects represent **examples** of the type of applied research projects to be conducted as part of SCIoI. Section 3.7 explains how basic research and applied projects interact and inter-connect through synthetic artifacts.

3.1 Basic Research Area: Representations and Processing

Intelligent agents rely on internal representations to learn, anticipate, reason, and generate behavior.[1] These representations reflect the hierarchies, abstractions, and stages of information processing the system employs to exploit the structure of the environment and the task.[2]

Research questions: What are good representations? What representations enable limited computational processes that lead to intelligent behavior? How can representations be found that reflect the structure (e.g., hierarchical or physical) of the environment? And do we need different computational processes for different tasks and representations?

Methodological approach in SCIoI: Despite the successes of information theory, machine learning and neuroscience in defining theoretical, qualitative criteria for representations, (such as the preservation and compression of information and unsupervised learning),[3] many fundamental questions remain unanswered. These questions concern, for example, the automatic extraction of abstractions and hierarchies from data.[4] Most existing research addresses these questions in the context of static data processing, classification, regression, and compression problems. The approach of SCIoI is *1)* to shift focus on appropriate representations *for behavior*, that is, for organizing sensory-motor coupling and reasoning on various levels of abstraction and temporal scales, and *2)* to compare the representations (and the implied generalizations) used by humans to those gained from machine learning approaches (see applied project 3.5).[5] This approach benefits from SCIoI's focus on the transdisciplinary synthesis of behavioral systems.

Further, the issue of good representations is tightly related to the computational processes operating on them. When transferring existing methodologies of unsupervised learning to the realm of behavior, we need to investigate if generic computational principles exit across representations. Several hypotheses for such computational principles can be formulated.[6] Going beyond these very recent trends in the neurosciences and machine learning, the approach of SCIoI is to test such principles for the first time on real robotic systems and investigate whether they help to transfer unsupervised learning methods to the realm of behavior.

Example project: This project is based on the following hypothesis: Representations that support simple and efficient computational processes to generate a desired behavior reflect and leverage problem-specific structure of the world.[7] We test this hypothesis by building computational systems for decision making behavior. Recent trends in cognitive science relate computational

---

[1] G. **Abel**, Imagination und Kognition. Zur Funktion der Einbildungskraft in Wahrnehmung, Sprache und Repräsentation. In Th.S. Hoffmann & S. Majetschak (eds.), Denken der Individualität. Festschrift für Josef Simon zum 65. Geburtstag, pages 381-397, Berlin/New York: Walter de Gruyter 1995.

[2] T. Moor, J. Davoren & J. **Raisch**, Learning by doing – Systematic abstraction refinement for hybrid control synthesis. *IEE Proceedings Control Theory & Applications*, Special Issue on hybrid systems, 153(5):591-599, 2006.

[3] For instance with respect to visual representations in the cortex: E.P. Simoncelli & B.A. Olshausen, Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24: 1193-1216, 2001

[4] T.G. Dietterich, P. Domingos, L. Getoor, S. Muggleton & P. Tadepalli, Structured machine learning: The next ten years. *Machine Learning*, 73(1):3-23, 2008

[5] J.-D. **Haynes** & G. Rees, Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience* 7(7):523-34, 2006.

T. Kahnt, J. Heinzle, S.Q. Park & J.-D. **Haynes**, The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Science*, 107(13):6010-6015, 2010.

[6] For instance, probabilistic inference and the free energy principles: see K. Doya et al., Bayesian brain: Probabilistic approaches to neural coding. Cambridge (MA): MIT Press 2007.

M. **Toussaint** et al., Expectation-Maximization methods for solving (PO)MDPs. In: D. Barber et al. (eds.), Inference and learning in dynamic models, Cambridge: Cambridge University Press 2010.

K.J. Friston et al., Reinforcement learning or active inference? *PLoS ONE*, 2009.

[7] G. **Gigerenzer**, P.M. Todd & The ABC Research Group, Simple heuristics that make us smart. Oxford: Oxford University Press 1996.

cost to the number samples needed for Bayesian inference and use this to model human decision times.[8] In this project, we use the same measure for model selection, i.e., to optimize the structure of the model such that the decision becomes simpler. This contrasts existing models selection methods in statistics, which typically measure model complexity in terms of the internal degree of freedom or size of the function class. This project connects the discussion of “Knowing How instead of Knowing That” in philosophy[9] and “simple human heuristics” in cognitive science[10] to machine learning research and the synthesis of behavioral artifacts.

3.2 Basic Research Area: Learning and Development

Biological forms of intelligence learn from experiences and interactions with their environment. Such learning manifests itself in a modification of the representations and processing that form the basis of intelligent behavior.[11] In biological intelligent agents, the challenges mastered through learning vary with the developmental stages of the agent. Learning and development are therefore closely intertwined mechanisms through which the representations and processing of an agent change in order to address tasks and environmental challenges.[12]

Research questions: What are the differences between human-like and existing machine-like learning? Does active learning and exploration in machine learning relate to curiosity-driven exploration and learning in children? What priors underlie human learning that allow for the type of generalization and learning from few examples? Do developmental stages reflect problem decomposition? What are appropriate curricula of development in the natural world?

Methodological approach in SCIoI: Concerning learning, disciplines like statistics and machine learning proved very successful in the areas of statistical data analysis (pattern recognition, regression and classification). However, it has been argued that these types of learning lack many aspects of human learning,[13] in particular in terms of their dependence on a human expert to preprocess a given data set, design features and representations, and choose the appropriate learning algorithms and parameters. In SCIoI, we consider the existing theories of statistical learning as necessary foundation, but our approach is to develop them towards more autonomy and application in real behavioral systems. Our focus is not on scaling learning methods to even larger static data sets or addressing preprocessed static data sets; instead, we focus on autonomous exploratory behavior (see three applied projects). Further, the strategy of SCIoI is to synthesize systems that model human aspects of learning in terms of their exploratory preferences and abstract generalization based on few data (e.g., applied projects 3.4, 3.5).

---

[8] N. Chater, J.B. Tenenbaum & A. Yuille, Probabilistic models of cognition: Where next? *Trends in Cognitive Sciences*, 10(7):292-293, 2006.

[9] G. **Abel**: Knowing How. Eine scheinbar unergründliche Wissensform. In: J. Bromand & G. Kreis (eds.), Nicht-Propositionalität. Festschrift für Wolfgang Hogrebe. Berlin: Akademie-Verlag 2010 (to appear).

[10] G. **Gigerenzer** & H. **Brighton**: Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1(1):107-143, 2009.

[11] I. Wartenburger, H.R. Heekeren, F. Preusse, J. Kramer & E. **van der Meer**, Cerebral correlates of analogical processing and their modulation by training. *NeuroImage*, 48(1):291-302, 2009.

[12] Y.L. Shing, M. Werkle-Bergner, S.C. Li & U. **Lindenberger**, Associative and strategic components of episodic memory: A lifespan dissociation. *Journal of Experimental Psychology: General*, 137(3):495-513, 2008.

I.E. Nagel, C. Preuschhof, S.C. Li, L. Nyberg, L. Bäckman, U. **Lindenberger** & H.R. Heekeren, Performance level modulates adult age differences in brain activation during spatial working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52):22552-22557, 2009.

[13] R. Douglas et al., Future challenges for the science and engineering of learning. Final NSF Workshop Report, 2008, http://www.cnl.salk.edu/Media/NSFWorkshopReport.v4.pdf.

Only few formal computational frameworks exist for understanding or realizing development.[14] The approach of SCIoI follows the hypothesis that stage-wise development reflects the problem decomposition adopted by the system—and thereby the inherent structure of the problem and the environment. To investigate this hypothesis, SCIoI will leverage the synthetic tools of machine learning and robotics to synthesize concepts and insights from developmental psychology. This will enable us to investigate the implications of development for problem decomposition, structuring representations and efficient learning.

Example project: In this project, we investigate the benefits of stage-wise development for the formation of visual representations in synthetic systems. The experimental setup mimics early child development where visual stimuli of natural scenes are first given in a myopic fashion, focusing on near objects with far objects out of focus. We use unsupervised learning methods (e.g., stacked auto encoders) for the extraction of visual representations. We compare such representations with those obtained from neural signals using machine learning techniques.[15] We test whether such representations can be formed more easily when first learning on myopic stimuli before relaxing the constraint to include all visible objects. The stimuli will be captured with cameras that allow us to control the focal length. The project benefits from SCIoI in that it requires the cooperation of experts in visual psycho-physics, robotics, machine learning, and developmental psychology. It follows SCIoI's research strategy by synthesizing concrete artifacts to test our hypotheses on the role of development.

3.3 Basic Research Area: Leveraging Structure

To exhibit intelligent behavior, humans and animals extensively leverage problem-specific structure. Structure focuses the limited resources available to an agent on those aspects of the environment relevant to the behavior. Structure is a consequence of the properties of the task, the embodiment of the agent, the environment, or the interactions between the agent and the environment.[16] Intelligent agents routinely use such structure to reduce the complexity of computational problems, to increase generalizability of acquired knowledge and skills, and to ensure the compositionality of learned actions and strategies. To synthesize intelligent behavior, we must understand how problem-specific structure can be identified and leveraged to enable intelligent behavior. This basic research area is closely connected with the first two, as representations & processing and learning & development must themselves leverage structure to be effective.

Important research questions: How can problem- and behavior-specific structure be identified? How can it be leveraged for representations & processing and for learning & development? Can limitations in biological intelligence (e.g. limited working memory, reduced visual range of babies, proximal to distal use of degrees of freedom in limbs, etc.) aid in discovering behavior-specific structure? What types of structure facilitates the generation of behaviors that generalize and can be composed into more complex behavior?

---

[14] Examples are: Y. Bengio, J. Louradour, R. Collobert & J. Weston, Curriculum Learning. International Conference on Machine Learning, 2009.

P.-Y. Oudeyer, F. Kaplan & V.V. **Hafner**, Intrinsic motivation systems for Autonomous Mental Development *IEEE Transactions on Evolutionary Computation*, Special Issue on Autonomous Mental Development, 11(2):265-286, 2007.

[15] J.-D. **Haynes**, Decoding visual consciousness from human brain signals. *Trends in Cognitive Science* 13(5):194-202, 2009.

J.-D. **Haynes** & G. Rees, Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience* 8(5):686-691, 2005

[16] D. Katz & O. **Brock**, A factorization approach to manipulation in unstructured environments. Robotics Research, Springer Tracts in Advanced Robotics, pages 1-16, Berlin/Heidelberg: Springer Verlag, to appear 2010.

U. Sassenberg & E. **van der Meer**, Do we really gesture more when it's more difficult? *Cognitive Science*, 34(4):643-664, 2010.

Methodological approach in SCIoI: The principles of how structure is leveraged to generate intelligent behavior can only be studied in the context of target behaviors and specific problems. The scientific strategy of SCIoI connects the investigation of leveraging structure to multiple applied projects through synthetic artifacts (see Figure 2 on page 5) thereby enabling the meaningful investigation of this aspect of intelligence. Projects in this research area will analyze what structure is leveraged and how to achieve specific target behaviors. As these target behaviors begin to require the use of previously synthesized behaviors, it is possible to study questions regarding the compositionality of synthesized behaviors.

This SCIoI methodology differs from that of most related work in synthetic intelligence research. Most often, the structure that must be exploited to exhibit a desired behavior is coded into the synthetic system by its designer. This means that an important component of exhibiting intelligence is not performed by the system itself but by humans, reducing the possible gain in understanding intelligence.

Example Project: With our description of this example project, we want to illustrate how basic research areas and applied research projects inter-connect, as illustrated in Figure 2. The applied research project Acquisition of Physical Knowledge, described in the next section, investigates the acquisition of knowledge about physical objects in the environment in human and robotic subjects. These subjects “play” with complex, jointed objects until they have discovered their full functionality.

The aim of this project is to understand what structure, when used to devise representations and learning methods, results in the most robust, generalizable, and efficient exploratory behavior to guide these acquisition of physical knowledge. To illustrate our approach, we describe a specific, very simple experimental procedure with robots. During the robot’s interactions with the environment, it gathers sensorimotor experiences. These experiences are the input for learning. To investigate the relationship between a particular type of structure and the effectiveness of representation and learning, we perform generate-and-test experiments. We automatically (possibly randomly) generate ways to structure the high-dimensional experience data, effectively reducing its dimensionality in accordance with the chosen structure. Now we test in experiments if this structuring of the data results in better performance. We can then analyze how the specific structure affects the robustness to uncertainty and the ability to generalize.

3.4 Applied Research Project: Acquisition of Physical Knowledge

Behavioral goal: One of the hallmarks of intelligence is the ability to “understand” the physical and psychological constraints and affordances of a given environment. Intelligent agents “know” which behaviors they can possibly perform in a particular surrounding to achieve a particular goal and have explicit expectations about the environmental changes that will result from their actions.

A first step toward understanding an environment is to generate a mental representation of the physical aspects of the surrounding world. In this project, we are concerned with how humans and artificial systems go about creating mental representations of some of the physical properties of their environments. More specifically, on the robotic side, we will develop an interactive, exploratory robot skill that determines the possible motions a previously unseen object is able to perform. Such a skill is the prerequisite for successful robot manipulation in everyday environments. A simple example: scissors consist of two parts, connected by a revolute degree of freedom, whereas drawers contain a prismatic degree of freedom. To cut with scissors or to open a drawer, the robot has to know the degrees of freedom of the object and to actuate them.

On the human side, our goal is to understand in which principled manner humans explore the possible motions of unknown objects and how unexpected motions of the object lead to a refinement of the acquired physical knowledge.[17] Ideally, the knowledge gained with human participants can be applied to and used in the development of the robotic skill; conversely, the application of knowledge gained from research with humans to the development of robots will uncover inconsistencies in theories of human behavior that will lead to additional and more specific research questions and efforts with human participants.

Experimental setup and procedure: The experimental task used in this research has been developed for use with robotic systems and will be adapted such that it can be used with human research participants as well.

In the robot experiments, the robot will be presented with an unknown, articulated, 3D object. The robot interacts with this object to incrementally build an internal representation of the kinematic structure of the object. Interactions consist of the robot pushing. If pushing results in a motion of parts of the object, the robot has "discovered" these parts of the object and can now track their motion in subsequent steps of the experiment. If the robot observes relative motion between two parts of the objects that can be explained by a revolute or prismatic joint, the robot adds this kinematic relationship to its internal model. When the robot has discovered all degrees of freedom of the object, an external observer ends the experiment.[18]

The robot experiments will be conducted on a bi-manual humanoid upper body. In addition, the experiments will be performed in simulation to enable the execution of many experiments in a short time period.

For use with human research participants, the robot task will be modified into a video game that is performed on a lab computer. Research participants take the role of the robots and are asked to discover the component structures as well as the possible motions of the same 3D objects used in the robot experiments. Participants are allowed to perform only the actions available to the robots; they are said to have "understood" the object when all objects components and possible movements have been generated and displayed on screen.

In order to investigate the neural coding of the mental representation of the physical objects, we will conduct a study using functional magnetic resonance imaging. Here again, the modified video game task will be used. Participants can use a joystick to explore the 3D objects used in the robot experiments. Then we will investigate neural responses to expectancy violations (objects behave deviate from their established behavior) and to learning load (realized as the number of critical object features). Specifically, this will help us assess which brain regions are involved in acquiring mental representations of novel physical systems. Of particular interest is to which degree it is possible to transfer learned information about an object to other similar objects.

Important research questions: One of the key questions addressed by both the human and the robotic research in this project concerns the strategy that humans and robots use to obtain a complete kinematic model of the object. On the human research side, it will be important to find out to what extent the strategic approach to "understand" the physical properties of a 3D object is principled and systematic, and whether it is affected by the complexity, difficulty, and perceptual similarity of the current object to previously encountered objects. In addition, it is as of yet unclear how expectations about the state of a new object are formed on the basis of previously viewed objects, which role these expectations play, and how the human mind and brain deal with failures

---

[17] H. Haider & P.A. **Frensch**, Conflicts between expected and actually performed behavior lead to verbal report of incidentally acquired sequential knowledge. *Psychological Research, Special Issue on conflicts and signals*, 73(6):817-834, 2009.

D. Rünger & P.A. **Frensch**, How incidental sequence learning creates reportable knowledge: The role of unexpected events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34(5):1011-1026, 2008.

[18] D. Katz, Y. Pyuro & O. **Brock**, Learning to manipulate articulated objects in unstructured environments using a grounded relational representation. Robotics: Science and Systems IV, pages 254-261, Cambridge (MA): MIT Press 2009.

of the expectations.

Similar research questions can be formulated for the corresponding research with robotic systems: The robot has to use its experience to determine the most appropriate action given its information about the environment. How can the experience be represented most effectively? How are new experiences incorporated into the robots knowledge base when the experience contradicts existing knowledge? What properties of objects are most relevant to obtain a compact and general knowledge base? How can the representation and learning procedure avoid overfitting and ensure generalizability? Is there a dependency between the order of experiences and the resulting representation? How can such a dependency be leveraged to devise a sequence of experiments that leads to a particularly good knowledge base? How do the answers to the questions vary if we vary the perceptual or manipulation capabilities of the robot?

Benefits of SCIoI approach: According to the synthetic approach, experimental strategies for human experiments must be transferred to robotic systems. As the exact same experimental tasks, procedures, 3D objects, and guiding theoretical questions are used with human participants as well as with robotic systems, results from robotic experiments will quickly uncover theoretical inconsistencies and lack of accuracy, and will pinpoint additional research efforts on the human research side. At the same time, the synthetic approach ensures that research on humans will be beneficial in guiding and constraining the large space of possible strategic approaches robots could adopt.

3.5 Applied Research Project: Exploration, Exploitation, and Generalization

Behavioral goal: Humans are typically curious to explore features of their environment—but they also lose interest once they are able to predict the outcome of their actions. When there is an overall goal to be solved, we expect a switch from a curiosity-driven phase to a goal-directed exploitation of the gained knowledge.[19] The interesting aspect about such behavioral switches from exploration (curiosity) to losing interest or exploitation is that it *uncovers the generalization* implicit in the construction of knowledge from experiences. Observing such behavioral switches therefore is an experimental means to analyze the structure of generalization assumed by humans, for instance whether they generalize experience with some objects to novel objects and therefore stop exploring them sooner. Our goal is to devise systems that perform switches between exploratory and exploitative behavior as humans do, which will require that they generalize in similar ways.

Experimental setup and synthetic approach: Consider a physical setup with a number of buttons and knobs that have a priori unpredictable artificial effects (changing the state of the light or a display, turning on and off motors, generating sounds, etc). Similarly to the previous applied research project, we will combine human and robotic experiments. Humans will explore until they have uncovered the effects and underlying dependencies sufficient to achieve a set goal. We can count the number of manipulations and ask humans to predict effects. Starting points for building synthetic systems with comparable behavior are existing models of exploration in reinforcement learning. It is crucial to develop extensions of such methods to yield the same level of generalization as humans, possibly on the basis of relational representations.[20]

---

[19] E. **van der Meer**, R. Beyer, J. Horn, M. Foth, B. Bornemann, J. Ries, J. Kramer, E. Warmuth, H. Heekeren & I. Wartenburger, Resource allocation and fluid intelligence: Insights from pupillometry. *Psychophysiology*, 47(1):158-169, 2010.

[20] T. Lang, M. **Toussaint** & K. Kersting, Exploration in relational worlds. European Conference on Machine Learning (ECML 2010), 2010.

D. Katz, Y. Pyuro & O. **Brock**, Learning to manipulate articulated objects in unstructured environments using a grounded relational representation. Robotics: Science and Systems IV, pages 254-261, Cambridge (MA): MIT Press 2009.

Important research questions: Can we model human exploratory and generalization behavior using an artificial system trying to solve exactly the same task? Can we reproduce the statistics of trial numbers – and thereby implicitly the human generalization? What *representations* yield this type of generalization? What *computational processes* on these representations can generate exploratory and goal-directed behavior? What *learning* framework can leverage the problem *structure* to yield a type of generalization similar to humans?

Benefits of SCIoI approach: An appropriate strategy of balancing exploration and exploitation must leverage problem-relevant structure (see Section 3.3), as this structure leads to robust and general behavior. The SCIoI strategy addresses this requirement by enforcing close interactions between basic (leveraging structure) and applied (problem-specific exploitation) research. Furthermore, the emphasis on synthesizing artifacts, which is at the core of SCIoI, ensures the transferability of insights and results between human and robotic experiments. Open research questions that cannot easily be answered by designing experiments with humans might be easily tested and evaluated in a synthetic system, and vice versa.

3.6 Applied Research Project: Visual Exploration

Behavioral goal: The advanced capabilities of the human visual system are an important enabler of intelligent behavior. This project is concerned with the trade-off between visual data processing and visual data selection. In regards to the human systems this means: the foveated eye, by only providing high-resolution data for a narrow field of view, greatly reduces the amount of information that must be processed. At the same time, this information reduction necessitates active control of the field of view to scan the scene and acquire sufficient information to fulfill a perceptual task. The goal of this project is to synthesize an artifact that replicates and examines this tradeoff between information reduction and explorative strategy in the context of a specific visual task.

Experimental setup and procedure: To be able to make explicit the visual exploration strategy of humans, we will decouple the biological intention to perform eye movement from the eye motion itself. Human subjects will explore a visual scene with a robotic foveated camera. The camera is commanded either through a brain-computer interface (BCI) or with a joystick. By learning to use these new interfaces, humans re-learn and thereby make explicit their exploration strategies. BCI and functional MRI imaging provide valuable clues about the representations involved in the process and their temporal evolution.

Important research questions: Given a perceptual task, what is an appropriate visual exploration strategy? What representations of visual information and visual actions are well-suited to support such a strategy? What is the optimal tradeoff between information reduction and the complexity of the exploration strategy? How do representations, associated exploration strategies, and the tradeoff with information reduction depend on the visual task? How can exploration strategies be learned?

Benefits of SCIoI approach: In this project, the decoupling of the closed-loop visual system through a synthetic artifact renders the exploration strategy explicit, makes it observable, and replicatable in a technological artifact. At the same time, theoretical models of human exploration strategies can be validated on the synthetic system. The combination of measurable behavior and observational methods for mental states allow us to establish a direct link between internal representation and behavioral manifestation, thereby fully leveraging the scientific strategies of SCIoI.

3.7 Interactions Among Basic Research Areas and Applied Research Projects

Progress in the applied research projects depends on our ability to understand and leverage the problem-specific structure to devise appropriate representations, processing, and learning methods. The successful synthesis of a behavioral hallmark of intelligence thus depends on a close connection between applied research projects and the three basic research areas. As illustrated in Figure 2, the scientific strategy of SCIoI enforces exactly this.

At the same time, basic research areas are concerned with aspects of intelligence across multiple behavioral goals to identify general principles. These principles must contribute to the generation of intelligent behavior in multiple synthetic artifacts and towards multiple intelligent behaviors to be considered principles. Again, the scientific strategy of SCIoI enforces this.

Consequently, we believe that the scientific strategies of SCIoI are well-suited for the investigation and synthesis of intelligence.

4 MSc & PhD Training Program

The foundation of a new science must be accompanied by a supporting educational program. Also, the long term success of the proposed research agenda crucially depends on an appropriate training program for a new generation of researchers. Here is a starting point for a discussion regarding this.

The transdisciplinary nature of SCIoI research requires a curriculum that does not align with existing educational programs. To address this, SCIoI will establish a multi-university, integrated MSc & PhD training program. The goals of this program are *a)* to train students on a basic set of quantitative and intelligence research foundations covered by introductory courses, *b)* to provide students with an interdisciplinary perspective on their topic of specialization, and *c)* to establish an *internationally recognized profile* of students that graduate from our training program.

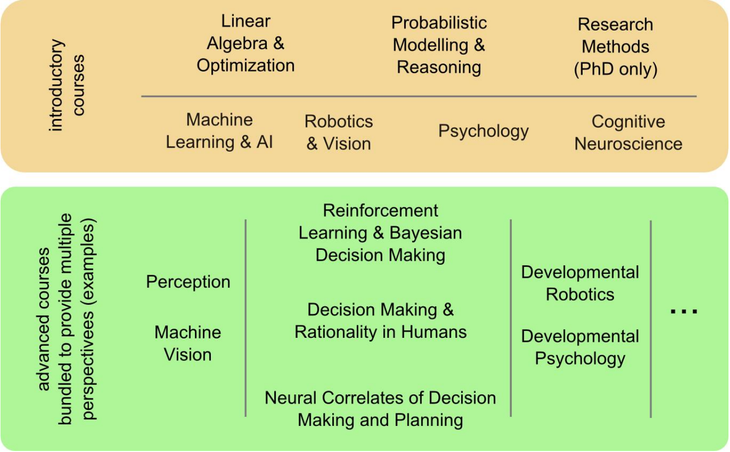

Figure 3: Outline of the course curriculum in the MSc & PhD training program.

MSc program: Figure 3 outlines a possible course curriculum. The introductory courses ensure a solid mathematical grounding for research and an introduction to the general scope of the disciplines. The curriculum of a particular student will be tailored to complement her or his background. For instance, students with a cognitive science background will have to focus on the quantitative courses; students with a computational background will have to focus on the cognitive sciences courses.

The advanced courses are bundled to cover multiple perspectives of the same aspect of intelligence (e.g., development, perception, or decision making). Students specializing on development take a course on developmental psychology and on developmental robotics; students specializing on visual perception will learn equally about classical Gestalt concepts as about modern computer vision methods. This bundling of courses provides an interdisciplinary perspective on the research areas to avoid a narrow focus on a certain discipline or methodology. The concept of bundling allows us to offer a full curriculum from day one, leveraging the full spectrum of established courses in all disciplines.

PhD program: PhD students must initially cover the introductory courses of the MS program, either by taking the classes or by showing sufficient prior experience, as evaluated by the adviser. Every PhD student has to take a class on research methods, covering the scientific method, proper scientific conduct, literature search, empirical evaluation methods, statistics, experiment design, writing research papers, and giving academic presentations. In addition, the PhD program includes the following:

- yearly reporting and evaluation of each student,
- lab rotations between the four quadrants of Figure 1 on page 3,
- supervisors from two different quadrants, and
- funding for a three-month visit to an international research lab, for example that of an External Investigator (EI).

Begutachtung der DFG:

Technische Universität Berlin

EXC 1045, Science of Intelligence

Koordinator: Oliver Brock

Das Konzept zur Etablierung und Neuausrichtung einer ganzheitlichen Intelligenzforschung stellt einen ambitionierten umfassenden Leitgedanken dar und wird getragen von einer Gruppe von Wissenschaftlerinnen und Wissenschaftlern mit herausragender Expertise auf ihren jeweiligen Feldern. Ein starker und glaubwürdiger interdisziplinärer Ansatz vereint Personen aus den Fachgebieten der Psychologie, Robotik, Neurowissenschaften, Künstlichen Intelligenz und dem Maschinenlernen. Das Forschungsprogramm ist in einen stark grundlagenwissenschaftlich orientierten und einen eher anwendungsorientierten Teil sinnvoll gegliedert. Der erste Teil ist sehr visionär und global aufgesetzt. Doch liegt es nahe, die adressierten Fragen, die sämtlich von zentraler Bedeutung sind, unvoreingenommen zum jetzigen Zeitpunkt – wie geplant – ganz neu und in überfachlicher Kompetenz anzugehen, da auf den beteiligten Fachgebieten im letzten Jahrzehnt relevante neue Ergebnisse erzielt wurden, die das Fundament des geplanten Clusters darstellen. Hier will man sich dem komplexen Kern der Intelligenzforschung widmen. Lediglich eine etwas stärkere Auseinandersetzung mit der Entwicklung der Intelligenzforschung über die Zeit wurde vermisst.

Die Vorstellung einiger beispielhafter, sehr konkret dargelegter Projekte macht das in Teilen noch eher vage Konzept verständlich und glaubwürdig. Gut gefallen auch Ideen zum Thema „Bewusstsein“ sowie die relevanten Fragen zur Modellierung.

Die Qualität der Beteiligten ist durchweg hervorragend. Die Gruppe vereint die richtigen Wissenschaftlerinnen und Wissenschaftler aus den notwendigen Disziplinen. Das Konsortium aus 15 Personen ist mit Expertise aus den Berliner Hochschulen TU (7 Personen), HU (3 Personen) und FU (1 Person) äußerst passend besetzt und wird durch drei Wissenschaftler aus dem Berliner MPI für Bildungsforschung (3 Personen) sowie einen Forscher aus dem *Bernstein Center for Computational Neuroscience* an der Charité ergänzt. Darunter sind zwei Wissenschaftlerinnen.

Auch vor dem Hintergrund der bereits vorhandenen Graduiertenschule „Berlin School of Mind and Brain“ und des Exzellenzclusters „Neurocure“ ist Berlin ein sehr geeigneter Standort für die Thematik. Erfreulich konkret und überzeugend ist auch die geplante Einbindung von externen internationalen Wissenschaftlerinnen und Wissenschaftlern, die als Gastforscher für eine Teilnahme am Exzellenzcluster gewonnen werden sollen. Die bestehende Vernetzung mit internationalen Wissenschaftlerinnen und Wissenschaftlern sowie Partnern aus der Industrie überzeugt. Auch die

Ideen zur Öffentlichkeitsarbeit gefallen. Die Konzeptidee, der Standort und das wissenschaftlich attraktive Umfeld werden zweifelsfrei viele qualifizierte Nachwuchsgruppen begeistern können. Die Ansätze zur Berufung weiterer Wissenschaftlerinnen sind wohl durchdacht und realistisch, zumal es bei eventuell auftretenden Schwierigkeiten auch alternative Vorstellungen zur Rekrutierung von Nachwuchswissenschaftlerinnen gibt.