DIMACS Technical Report 2000-24
September 2000

Question-answering system for teaching autistic children to reason about mental states

by

Boris Galitsky[1]

[1] iAskWeb, Inc. 681 Main Str #1B Waltham MA 02461. http://www.iaskweb.com
(galitsky@dimacs.rutgers.edu)

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, and NEC Research Institute and Telcordia Technologies (formerly Bellcore).

DIMACS is an NSF Science and Technology center, funded under contract STC--91--19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

We build the natural language question-answering system to teach autistic patients reasoning about mental states. It has been discovered that autistic people cannot properly understand mental states. In accordance to our model of autism disorder, the specific reasoning patterns involving the concepts of knowing, believing and intention are corrupted. Based on this model, we have suggested (Galitsky 1999c) reasoning rehabilitation strategy for autistic patients of the verbal age, where the professional psychologists explained the autistic children the set of multi-agent scenarios. Our experiments showed that acquiring the definitions of mental concepts based on this formalism helps autistic children not only to improve judgment interacting with the other people, but also to stimulate the emotional development.

Further evaluation of the model and rehabilitation technology is conducted with automatic training toolkit, implemented on the Internet (www.dimacs.rutgers.edu/~galitsky/tech_engines.html). Asking questions about mental states of heroes of the scene or textual scenarios is a good assistance to parents and rehabilitation personnel in the diagnosis and training of the corrupted autistic reasoning. Natural language technology of semantic headers (Galitsky, 1999b) is applied, where the textual answer (explanation) is assigned with a mental formula, which is matched against the formal representation of an input question.

Introduction

In this study, we build the logical model for the specific type of human agents: autistic patients to design an interactive rehabilitation tool. We analyze the distinguishing features of autistic reasoning to construct the software tools for training the children with autism by means of question-answering natural language system. This work follows our investigation of autism from the viewpoint of logical artificial intelligence (Galitsky 1999c) and the studying of the formal model applicability of the autistic reasoning to the rehabilitation methodology.

Autism is a relatively rare multifactorial disorder that affects 4-10 out of every 10,000 school-aged children. Autistic group is very heterogeneous (and frequently referred as autistic spectrum disorder (ASD)); it is not easy to highlight the specific symptoms. This disorder is primarily characterized by impaired social interaction and communication combined with repetitive and stereotyped patterns of behavior. It is generally accepted that autism is characterized by the following:

1. Problems with **interaction** with external world. Difficult social relationships, impaired capability of understanding the feelings and mental states of other people.
2. Problems with **means of communication** with external world, difficulties of verbal and non-verbal nature. In particular, misunderstanding the meanings of gestures, facial expressions, etc. are inherent to autistic patients.
3. Problems with **imagination** and playing; imagination patterns are frequently copied from the surrounding people.
4. **Repetitive** behavioral patterns and unwillingness to change them; limited activities and stereotyped behavior.

If to switch from psychological terms to AI terminology and consider an autistic patient as abstract agents, the items 1) and 3) can be merged from the perspective of reasoning and decision-making of these agents. It has been recently discovered that **autistic children cannot reason properly about the mental states** and mental actions of themselves and others (Baron-Cohen 1988, Leslie & Thaiss 1992, Leslie 1987, Wetherby & Prutting 1984). Corrupted reasoning about mental states affects both the decision making while interacting with external world and patterns of imaginational behavior, where it is necessary to analyze the reaction of the others to imaginary patterns.

Experiments show, that the deductive capabilities of autistic children concerning other domains of reasoning match their mental ages. Autistic children are capable of relatively normal reasoning about physical attributes such as time, space, states and actions, yet reasoning about mental states including intentions, knowledge and beliefs is reduced in various degrees.

We base our model of the human agent on the hypothesis that there is a number of standard axioms for mental attributes; these axioms are genetically set for normal children and are corrupted in the autistic brain. The patterns of corruption vary from patient to patient and are correlated with the specifically outlined groups of autistic children. So autistic children have to acquire these axioms explicitly, by means of direct training, using the specific scenarios. Frequently autism is not accompanied by learning disabilities, so the patents willingly participate in training programs, especially those, who have the decent for their mental age mathematical skills.

Interaction with a computer system seems to be a relatively easy approach to introduction of a training technique, novel for the patients. The autistic children rather accept and start interact with computer agents than with the human ones, because the technical system usually requires neither imaginative thinking nor understanding of mental world of the others, in contrast to a live agent. At the same time, the computer system to be designed teaches the autistic patient nothing else but the laws of mental world, operating with the objects, which are not live from the trainee's viewpoint. Therefore we use the non-mental means, readily acceptable by the autistic children, to introduce them to the mental world via formalized reasoning, required to survive in this world. The trick in this methodology is that the reasoning in mental world, usually supposed to be irrational, can nevertheless be considered from the abstract perspective and formalized. We take advantage of this fact and form the skeleton of our rehabilitation strategy of teaching to behave in a real mental world. The limitation of our approach is that non-verbal children cannot be trained.

Interaction of autistic children with animals was found to be an important alternative to conventional therapy to improve the interpersonal and task-oriented behaviors. Animal assisted therapy (Sams 1999) employs the care, feeding and training of animals as a means of facilitating communication and engagement in activity with non-communicative individuals.

There is a series of studies, which have introduced the interactive environment for rehabilitation of children with autism. These studies are intended to support the children attempt to explore the environment, learning social-emotional cues. A remote control device (Resnick 1987), video-clips together with a set of dolls (Blocher 1999), autonomous mobile robot (Dautenhahn & Werry 2000) was used to stimulate the emotional activity of children with autism. Interaction with both animals and robots, multi-modal by nature, helps to rehabilitate the ability to behave in a real world. However, multi-modal training is not oriented to a train the primary locus of autistic impairment.

Our approach is more focused on the specific reasoning model and is intended to teach the proper analysis and decision-making rather than emotional development. In accordance to our model of autism, the higher and wider emotional activity will follow the development of proper analytical thinking about the mental world.

The paper consists from four sections. The first section introduces the logical model of autism and contains its descriptions, results and discussions. The second section introduces autism diagnosis and training scenarios. The third section introduces the natural language system, which is capable of answering questions about mental states of heroes of these scenarios. The fourth section presents the organization of natural language understanding (Fagin et al 1995, Konolige 1996, Galitsky 1999b, Decker et al1997).

1 The logical model of autism

1.1 Phenomena of Computational Autism

Involving multiple scientific knowledge: psychology, neurobiology, mathematical logic and artificial intelligence helps to build the systematic approach to the analysis of peculiarities of autistic reasoning. To link the hypothesis, formulated in the sciences with high level of strictness and formality, with experimental data, traditionally formulated on the intuitive basis, the letter should be better represented to match the standards of the former. We start with the language and axioms for mental states, proceed to **simulated, normal and autistic features of reasoning** about them, develop the diagnosis and training scenarios and analyze the resultant reasoning capabilities in autistic kids. On that way, we address the issue of brain realization of axiomatic reasoning, as well as implementation of reasoning about intentions, knowledge and belief in autistic and normal brain.

The idea of simulating the peculiarities of human behavior via symbolic reasoning is not novel. Such mental disorder as schizophrenia was thoroughly studied by the members of AI community. By means of introduced axiomatic it was possible to reproduce some verbal reaction of the patients. Logical simulation of autistic patients allowed to focus on the specific (mental) patterns of reasoning, which has already been introduced in the field, and perform the modeling of human behavior on the higher level of abstraction and generality, than logical investigation of other mental diseases. On the contrary, other mental disorders required the invention of special disorder-specific axioms, not interpretable taken separately from the subject of simulation. Observation and logical simulation of autism allows us to describe the phenomenon in terms of axioms about knowledge and belief and to bring in the additional evidence of validity and applicability of this particular field of deductive reasoning.

Experimental data for autistic children, supporting the focus on their corrupted speculation on the mental states, was obtained quite recently (Leslie & Thaiss 1992, Mawhood, Howlin & Rutter 2000, Prior et al 1998). Displaying peculiarities of thinking about mental states is inherent to the very beginning of verbal mental age: the older the autistic child, the more mixed is the corruption of various aspects of reasoning. In other words, when a child grows, the specific speculation patterns become more and more distorted. At the same time, it is very hard to set up experiment environment with 3-4 year old autistic kids, usually trying to avoid any conversation.

Note that deviation of reasoning patterns is inherent not only to autistic patients; some patients with other mental disorders display the similar patterns. For simplification, we introduce the metaphoric concept of **computational autism** to focus on the corruption of reasoning about other persons and himself/herself. We keep in mind that reasoning about other modalities can deviate as well, but in a lesser degree. Therefore, our diagnosis and training is oriented for the patients with computational autism, which does not fully match the generally accepted classification of autism syndromes.

An important conjecture of our study is that it advises us how do decrease the degree of reasoning corruption. Let us consider the following analogy, which illustrates the hypothetical mechanism of training for computational autism. Imagine that the normal humans find themselves on a planet, where the motion control requires knowledge of axioms other than ones for Euclidean space. A normal human does not need the axioms for Euclidean space to conduct the motion control because the required basic knowledge is genetically set. However, the new planet does require learning of new axioms (imagine for example, 4-dimentional or projective space) to conduct the motion control, because these axioms were not set genetically for normal humans. The similar situation, in accordance with our model, occurs for the autistic patients: some **mental axioms are missing, but can be acquired**. The same way the normal brain can learn the new geometric axioms to behave properly in the new conditions, the autistic brain can accept the absent knowledge and skills by means of training the missing axioms. It is well known that the brain can adapt to a variety of environments of different modalities, sometimes acquiring the experience via verbalized rules. This is the case for our approach to rehabilitation of reasoning of the autistic patients.

1.2 Reasoning about intention, knowledge and belief

The basic mental states are *intention* (subsumes goals and desires), *knowledge* and *belief.* The difference between belief and knowledge is that an agent is capable of changing and revising beliefs, but knowledge is only subject to acquisition. Almost any mental state or action concept can be defined within these basic mental states after adding an arbitrary predicate for a physical state or action. Some mental concepts cannot be formally derived using the basis above; however, this basis introduces the classes of equivalence with respect to the decision concerning the fixed set of actions (physical and mental). In other words, basic and derived mental states are the factorization of cognitive and emotional states relatively to resultant physical or mental action.

For example, the concepts *inform, deceive, explain, forgive*, etc. (see Section 3 for details) can be expressed via *want-know-believe* basis. Conversely, the concept *fear* is neither basic nor derivable concept; however, it is equivalent to *not want* relatively to the potentially implied physical action (e.g. to run away). The difference between *fear, not want* and *prefer to avoid* is in the “degree of influence” of the mental concept; therefore, we can ignore this difference having the explicit degree for each concept within an equivalence class. As an example of the derived mental predicates, let us consider the pair of concepts "*pretend for a person-be kidding with a person*". The former concept is expressible in our basis, and the latter one requires inclusion of the concept *joking* in the definition in addition to its part, similar to that of *pretending*. The resultant action depends on the class of equivalence, containing both these predicates, and can be determined by the former one. Our speculations may seem too informal, but the statement can be posed as a theorem in the environment of the fixed set of concepts. Furthermore, introduced basis gives a solid background for the design of training scenarios.

Autistic reasoning is corrupted in respect to the mental states and actions of him/herself as well as of the others. The phenomenon of computational autism is strongly correlated with the syntax of mental formulas: frequently autistic children do mistakes with the similar scenarios involving themselves and the others. Syntactically, the difference is just the substitution of atom either for itself or other agent in a predicate for mental state or action.

In the psychological literature about autism (see the discussion section), beliefs and intentions are divided into first-order and second-order ones. For example, the mental states "he wants", "I know" are of the first order, "he believes that I want", "she wants me to know" are of the second order. Syntactically, an **order** of a formula is the maximum number of embedding of predicates in a conjunctive member of this formula. Our approach provides the strict background, based on complexity of mental formula. In real life, humans need to reason with mental states, described by formulas of the order up to four. Therefore, we conducted the experiments, involving mental states described by phrases with up to four mental predicates.

Mental formula and its definition	**Semantic comments**	**Applicability test by the control group(%)**	**Applicability test by the autistic group(%)**
want(Agent, do(Agent, Action)).	agent wants to perform an action	100	95
want(Agent, do(DAgent, Action)).	agent wants another agent to perform an action	100	85
want(Agent, know(Agent,What)):- (believe(Agent, know(KAgent, What)), ask(Agent, KAgent, What)).	agent wants (himself) to know	95	72
believe(Agent, want(WAgent, know(WAgent, What))) :- prefer(Agent, tell(Agent,WAgent, What), OtherAction).	agent believes that other agent wants to know	90	44
believe(Agent, want(WAgent, know(KAgent, What))):- believe(Agent, inform(WAgent,KAgent, What)). believe(Agent, want(WAgent, know(KAgent, What))) :- not know(KAgent, want(WAgent, know(KAgent, What))), inform(Agent,KAgent, ask(KAgent, WAgent, What)).	agent believes that someone else wants the third person to know	93	37
believe(Agent, want(WAgent, know(KAgent, want(Agent, What))) :- believe(Agent, inform(WAgent,KAgent, want(Agent, What))).	agent believes that someone else wants the third person to know what this agent wants	93	35

Table 1: Examples of the mental formulas and their possible definitions(left column) and semantic comments(right column). Various formulas are built in the basis of *want-know-believe* and *ask/inform* in addition. The action/state predicates have the inmost occurrence: *do(Agent, Action)* or *What*. Last four rows present the mental axioms (which might be corrupted in the autistic patients). All well-formed formulas are interpretable by the autism simulation toolkit; however, not all of them can be a plot of a

real-life scenario (see Galitsky 1999a, and web page Galitsky, 1997). Third and fourth columns show the results of mental axiom applicability testing for control and autistic groups, averaged through the patients and trials (specific scenarios). The number shows the percentage of misunderstood scenarios (from Section 3) failed by the patients from respective groups.

We choose the most natural concepts (formulas) from the set of all well-formed formulas in our basic system. We then create a series of scenarios for each concept to determine if a child applies the formula (axiom) correctly and thus possesses that axiom. Every autistic child can then be categorized by the subset of corrupted mental axioms. We build the logical program with normal behavior to imitate the control patients (Galitsky 1998). Autistic patients are simulated by eliminating the particular set of mental axioms in this logical program. The structure of the corruption patterns and their statistical properties are the subjects of the further study. We refer the reader to (Dahl 1999) for related issues of logical programming.

Definition of the concept *inform* via the *want, know* and *believe* (a typical exercise for the autistic children).

inform(Who, Whom, What) :-
want(Who, know(Whom, What)),
believe(Who, not know(Whom, What)),
believe(Who, want(Whom, know(Whom, What))).

There is a number of logical systems for representing intentions, knowledge and belief (Fagin et al 1995, Konolige 1996). Autism phenomenon seems to be important for logicians to choose a logical formalism, adequate for human intelligence. The model of the brain and its specific reasoning and, in particular, the model of the autistic brain, is very difficult to build. The difference between these two models fits into the limited formalism of mental attributes, so that this restricted component of brain activity can be subject to logical modeling. The autistic phenomenon seems to be the only one that links the biological brain with the axiomatic method in the foundations of mathematics.

There are some fruitful analogies between physics as a verification of mathematical abstractions and mental phenomena for choosing the most adequate logical calculus (algorithm). Real data about reasoning in a human brain introduces the preferences for the family of logically valid approaches for mental states. Each of these calculus is oriented to the specific view of human intelligence, based on the author's observation of human intuition coming from the generally-accepted reasoning patterns.

The axiomatic method includes two components for a formal theory: axioms and inference rules. Which of this component is affected by autism stronger? Specific axioms themselves are rather corrupted by autism, then the inference rules, which are likely the same for reasoning about mental and physical states and actions.

The specific primitives of autism, affecting the reasoning about mental states, are the mental axioms containing the predicates of knowledge and intention. Various situations where autistic children display abnormal behavior can be constructed, using these axioms.

The reduced capabilities of modeling the mental states of himself/herself and other persons do not serve as the main criterion of autism. There are other behavioral criteria, which are more explicit and can be revealed at the pre-verbal development stage. However, the reasoning about mental state is more suitable for formalization and mathematical modeling.

There are various formal systems of reasoning about knowledge, belief and intention. Does autism modeling give the preference to a particular approach? The autism phenomenon is the strong criterion for choosing the adequate formalism for the mental concepts and reasoning. Just a single formal system (Galitsky 1998) out of many approaches is found to match the brain functioning from the autistic prospective. Such the system derives those and only those formulas (theorems) that accompany a scenario of multiagent behavior; if these formulas are failed by autistic kids, these kids can be trained to handle it properly. The situation is analogous to the physical modeling where just a single formalism can be valid from the multiple ones, internally consistent and fully formalized.

1.3 Discussion of the model

Does every child can be trained any mental formula? For the given formula complexity, that is determined by the mental age of a child. Besides, the following monotonicity proposition holds: the more complex the formula, the less likelihood that it will be successfully handled or trained. This is an additional verification of the adequate choice of the logical formalism.

It is an interesting observation that the training of computational autism occupies the intermediate position between the computer learning and normal human learning:

- Computer learning. The evident feature of teaching the computer to perform a human-like intellectual activity requires 100% formalization of the knowledge representation and reasoning machinery.
- Autistic learning. Intermediate between the normal human and computer learning in the sense of complexity and the number of necessary details to cover (to represent strictly or formally). Teaching the mental axioms to autistic children can be considered as a specific programming technique with the certain degree of generality, abstraction, formality and flexibility under transitioning from domain to domain. The requirements to formalization are less strict than that of the computer learning, but the knowledge has to be tightly linked to the model of phenomena.
- Normal human learning. Non-formalizable, involves learning explicit rules, learning by examples, learning deterministic and statistical information, etc. The major difficulties of the computer learning are resolved easily and implicitly while teaching a normal human.

We present the analysis of how the mental states appear in the process of human development. We state the proposition of monotonous increase of the complexity of mental formula, reflecting the perceivable mental states in the course of development. It has been shown that the normal humans under natural development break the monotonicity proposition. Around the age of 18-24 months human toddlers begin to pretend and recognize the pretending of others (Perner 1991, Zaitchik 1990). However between 36 and 48 months they show the evidence of understanding additional epistemic states such as knowing and the simplest mental axiom that seeing leads to knowing. So the pretending phenomena starts earlier than the knowing one, but the concept of pretending can be derived via knowing, intention and belief. Hence, the normal development process violates the monotonicity proposition.

Our study shows that the autistic children can learn the mental concepts in the order of complexity increase under the training in the corresponding order. Therefore, the autistic development obeys the monotonicity proposition under the specifically oriented training, but the normal human development violates it.

1.4 Testing the model

The simplest illustration of the autistic behavior is as follows. An experimenter, sitting next to an autistic child, takes two pieces of paper and put the token underneath one of them such that the child sees that. Then the experimenter asks the child, where would his friend, who is currently outside the room, look for the token: under the left piece, under the right piece, or under both of them.

The control child answers: ”under the both”. The autistic child usually answers “under the left one, because you (experimenter) just have put it here. The autistic child does not understand, that since the friend was *out* of the room, he/she did not *see* where experimenter has put the token. The following axiom is missing:

not see(Agent, Something) $\supset$ *not know(Agent, Something).*

This is the simplest mental state axiom, which can be corrupted in the autistic brain. See the detailed testing scenarios in Section 2 and definition of derived mental predicates in Section 3.1.

20 autistic children of the age 4-18 participated in the testing and training and 20 control children of the age 8 participated in the testing during 1.5 years (Table.1). Note that the testing covers the majority of mental formulas of the order 1-4, involving *want* and *know* (*believe* is identified with *know* for simplicity). The manifold of tested mental state achieves the real world complexity. Therefore, the trained children are expected to behave properly in the real conditions, if they are able to transfer artificial mental states to the real ones.

1. Each question (with the mental formula complexity below four) was successfully answered by every control child.
2. Each question was failed by at least one autistic kid.
3. For each question the autistic child failed, it was possible to perform training such that the question is successfully answered after fifth attempt.
4. The higher the complexity of mental formula, the higher the failure rate.

If to substitute the mental states by physical states, the questions will be easier answered by the autistic children. It will not make a significant difference with the control children, taking into account the average mental (not physical) age of autistic patients.

1.5 Results of simulation and training

1. Adequate logical formalism of reasoning about mental attributes is found to represent the phenomenology of reasoning, inherent to autism.
2. This formalism generalizes the reasoning peculiarities of autistic children, revealed in the psychological studies (Baron-Cohen 1988, Lesley and Thaiss 1992, Perner 1991). From the specific behavioral patterns such as pretending, deceiving, choosing of action, etc. we proceed to the general framework of mental states, corrupted as a whole.
3. The experiments, based on suggested formalism, cover the totality of all mental formulas of the complexity below four. The control group successfully fulfills all the tests, however each of the autistic kids of the physical age of 4-18 failed the majority of tests (Table 1, third and fourth columns).
4. For each mental formula, expressed via the concepts of knowledge and intention, there is a way to explain it to an autistic child such that this formula (question) is handled properly after repetitive training. Acquired skills can be transferred to a situation, represented by the same mental formula with an arbitrary physical state and action.
5. The logical formalism of reasoning about mental attributes introduces the strong background for detection and training of these intellectual capabilities of the autistic children of the verbal age.
6. Suggested formalism allows one to represent an autistic patient as a binary vector of valid/missing mental axioms. This can serve as a firm basis for classification and statistical analysis of the autistic patients.
7. Developed methodology of training covers all mental formulas an autistic child needs for interaction with the other humans, being able to reason about the mental attributes of himself/herself and of the others.

This logical model of autism is applicable to both diagnosis and education. The preliminary experiments showed that after explaining to the autistic children a series of scenarios based on a corresponding axiom, these children raised the complexity of reasoning about specific mental states and actions, and demonstrated their capability to apply this axiom independently.

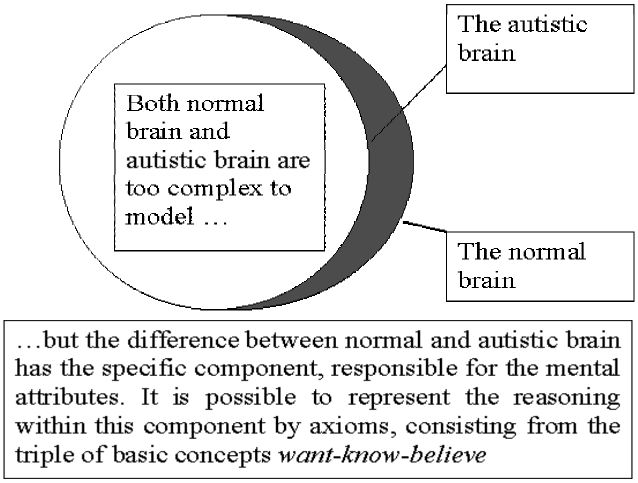

Fig 1: Logical model of autism builds the link between the axiomatic method and the brain. We use the logical model of autism for the indirect experimental proof that the brain functions, in particular, on the axiomatic level.

2 Testing and exercising scenarios

In this section, we introduce the testing and exercising scenarios. Unprepared, autistic children usually fail most of the questions below, depending on their mental age and degree of reasoning corruption. These scenarios are designed to cover the manifold of domains of mental reasoning, varying the complexity for each domain. Scenario presentation in this Section is informal, and the Section 3 below introduces the formalization machinery. These scenarios serve as a mean of revealing of corrupted mental formulas and of developing the training strategy, oriented to acquisition of that very formula. This strategy has been designed and tested to be performed by professional psychologists, introduced to the field of Logical Artificial Intelligence.

There are two children, A and B, who are subject to diagnosis and/or training of the corrupted reasoning about mental states and actions. Correct answers follow the question, wrong answers are enumerated in the parenthesis, where presented. The experimenter (E) is introducing the scenarios and asking questions.

There is a table in a room with two boxes on it. The experimenter (E) is keeping a token in his hands. A is in the room, and B is outside the room. E is asking A:

1) You see the token in my hands. Do you know which box I am going to put the token to?

A: I don't know that box / nobody knows. (A (confused): I don't know the answer).

2) E: As you see, I put the token into the left box. Do you know, where B will look for the token: in the right box, in the left one or in both boxes?

A: In both boxes. (In the left box, where the token actually is).

3) E: And do you know where the token is?

A: I know where is the token.

4) E: Does B know where the token is? If we ask him, what would he respond:

A: I don't know where the token is. (I know where it is. I know it is in the left box).

5) E: Now I want to tell you the following. I believe, that B still does not know where the token is. Who is wrong: myself (E) or B?

A: You are wrong telling us that B still does not know where the token is. (B is wrong, now he does know where the token is).

.

E keeps the blank piece of paper. A is next to E, and B is in the other room.

1) E: I am going to plot a geometric sketch on a piece of paper. I'm about to start the drawing. Do you know what I am going to draw; do I know, if myself knows what will be drawn?

A: I don't know, and you do.

E finishes the picture.

2) E: Now you know, what I've drawn. Does B know that?
A: B does not know what is drawn.
3) E: How can you let him know what is drawn?
A: Either show him or tell him (describe the picture).
4) E: You mentioned two ways of letting B know about this picture. Do both these ways require your knowledge of what is actually drawn?
A: No, to show him, I do not necessarily have to know (have seen) the picture. To describe the picture, I have to know its content. (Yes, I have to know the picture content for both telling and showing).
5) E: If we call B into the room and ask him if he knows what is on the paper, what would he (B) respond? What would he respond if we ask him after we show him the picture?
A: Before we show him (B) the picture, he will tell that he does not know what it is about. After we show or tell him (B) about the picture, he will tell he knows it.
6) E: if we ask B concerning his opinion, do you (A) know that he (B) does not know what this picture is about right now, before we informed him about the picture?
A: B knows that I know that he does not know the drawing. (A confusing: I don't know. B does not know that I know that he does not know. B does not know that I don't know that he knows).

There are the toys on the table: a bear, a fox and a rabbit. Experimenter is asking the child about his/her mental states.
1) E: As you see, the bear is watching the rabbit. Does the bear know that the rabbit is on the table?
A: Yes, The bear knows that the rabbit is on the table.
2) E: Now the rabbit leaves the table. The bear knows that the rabbit is not on the table any more. Does the bear know that the rabbit was on the table before?
A: Yes, he knows that he was on the table before.
3) E: Then, after a while, when the fox asks the bear if the rabbit had been on the table, the bear is saying that the rabbit has not been there. Trusting the bear, what do you think, does the bear know that the rabbit was on the table?
A: The bear does not know that the rabbit was on the table.
4) E: OK, the bear forgot that the rabbit was on the table. Does the rabbit know that he earlier knew that the rabbit had been on the table?
A: No, the rabbit does not know that he earlier knew that the rabbit had been on the table.
5) E: Now the fox wants the bear to recall that the rabbit has been on the table. What will she do?
A: She (the fox) will tell the bear that the rabbit was on the table, and that the bear has seen him there.
6) E: Then, assuming, that the bear trusts the fox, what is the knowledge of the bear?

There is a table, and a book on it. The experimenter teaches the child A to pretend that it is soap.
1) E: As you see, there is a book on the table. Do both of us know that it is a book?
A: Yes, both of us know that it is a book.
2) E: Now let us pretend that it is soap. Both of us will still know, that it is the book. How ever, if I ask you, what that is, what will I respond?
A: You respond that it is soap.
3) E: If you ask me, what is on the table, what will I respond?
A: That there is soap on the table.
4) E: When one asks you if you know what is on the table, what will you respond?
A: I do know what is on the table.
5) E: Now let us stop pretending. Both of us still know that this is actually a book. If one asks me what is on the table, what will I respond?
A: You will respond that it is the book.

3 Natural language question-answering system

In the previous Section, we introduced the scenarios, which are the basis of training by the professional psychologists. Computer-assisted rehabilitation of reasoning about mental states is another side of the suggested training technology. Below we present the formalization of basic entities, used in the previous Section informally to form the training scenarios. Formalization of mental entities is required to form the knowledge base for the natural language interaction of autistic children with the diagnosis/training toolkit. Usually, autistic patients willingly spend sufficient time with computer if the program behavior is appealing (in particular, intellectually challenging). Though there is a strong and widely spread opinion that the development of the emotional world is the primary task for autistic rehabilitation, computer-simulated communication is verified to be an important component of the training (Galitsky 1999c).

3.1 Derived mental predicates

We have given the definition of the concept *inform* above. In this section we define the concepts *offend, forgive, reconcile, explain, pretend* via the basis *want-know-believe*. This definitions are the formal background for the training scenario (Section 1).

The concept definitions are presented via the metalanguage-supported extension of first-order language to highlight the specific of our PROLOG implementation (Galitsky 1998). We use the traditional notions of the logical programming (Variables are capitalized, quantifiers are skipped, default conjunction and temporal sequences are represented by term enumeration (symbol **&** is equivalent to symbol **,**).

We start with the definition of unintentional offend (Galitsky 1999a). Ignoring modalities and tenses, we state, that unintentional offend is based on the lack of knowledge that the offending action *do(Who, Action)* is unwanted.

offend(Who, Whom, Action) :- want(Who, Action),
not want(Whom, Action),
not know(Who, not want(Whom, Action)),
do(Who, Action).

We remind the reader, that the default temporal relation between the terms is the order these terms occur in a clause.

To be forgiven, the offender has to demonstrate by some way that the offense was actually unintentional. It is necessary for the offender *Who* to inform *Whom* that *Who* would not *do* that *Action* if *Who* knew *Whom* did not like (*want*) it.

forgive(Whom, Who, Action) :-
offend(Who, Whom, Action),
inform(WhoElse, Whom,
not know(Who, not want(Whom, Action))),
believe(Whom, (know(Who, not want(Whom, Action))→
not do(Who, Action))).

If *Who* is unable to convince *Whom* (to make him *believe*) that the *offend* was unintentional, the other agent *Counselor* is required to explain the actual situation to *Whom.*

reconcile(Counselor, Who, Whom, Action) :-
offend(Who, Whom, Action),
not forgive(Whom, Who, Action),
explain(Counselor, Whom,
not know(Who, not want(Whom, Action))),

believe(Whom, (know(Who, not want(Whom, Action))→
not do(Who, Action))).

While explaining, a *Counselor* helps *Whom* to build the deductive link between particular facts and general knowledge, Whom possesses in accordance to the *Counselor's belief. The Counselor* knows this deductive link himself, *believes* that this link is unavailable for *Whom* and also *believes* this link will be established after *Whom* is *informed* with *PremiseFact*

explain(*Counselor*, *Whom, Fact) :-*
know(Counselor, PremiseFact→Fact),
believe(Counselor,believe(Whom,
not know(Whom,Fact)),
believe(Counselor,
(inform(Counselor, Whom, PremiseFact) →
believe(Whom,Fact))).

This definition gives *Whom* the explanation **why** the *Fact* holds.

One can compare our semantic for *explain* with the other option, the commonsense definition of the concept **to explain by example**, what feature could a particular object possess*". Counselor explains Whom* what *Feature* could possibly *Object* possess, using the *FeatureList* as the set of examples. The following has to hold to give *Who* an ability to explain *Whom* the *Feature* of the *Object: Who knows* that *Feature,* and *Whom* does not *know,* what could be a *Feature* of the *Object.* Then the *Counselor informs Whom* about the example *FeatureList.*

explain_by_example(Counselor, Whom,
Feature(Object, FeatureList)) :-
know(Counselor, Feature), not
know(Whom, Feature(Object, FeatureList)),
know(Who, Object),
inform(Counselor, Whom, FeatureList).

We remind the reader, that all the above concepts are indeed defined in the *want-know-believe* basis.

To introduce the concept of **pretending**, we first use the operator notation for knowledge and pretending. The presentation of the axioms for pretending, expressed in terms of modal operators, will be followed by the definitions in terms of metapredicates. This is another illustration that metapredicates deliver sufficiently expressive language to express such concepts as pretending. We denote by P_iF the fact that agent i pretends that the fact F holds; K_iF denotes the knowledge of fact F by the agent *i.* We do not use special symbols here to express the observation that the agent *i* **pretends for** another agent *j*; we express this fact indirectly.

1) General definition: an agent *i* pretends to the agent *j* that the fact *F* holds if he knows that *j* will understand the pretending: a) *i* knows that *j* knows that *i* pretends, b) *i* knows that *F* does not hold and that *j* knows that *F* does not hold, c) *i* assumes this pretend will be accepted.

$K_iK_j\, P_i\, F \;\;\&\;\; K_i\, not\, F\, \&\, K_i\, K_j\;\, not\, F \;\;\rightarrow\;\; P_i\, F\,.$

2) The pretend addressee either accepts the pretend (pretends that he knows the fact) or reject it (not know it).

$P_i\, F \;\rightarrow\; P_j\, K_j\, F \;\; \mathrm{v} \;\; not\; K_j F\,.$

3) If an agent *i* pretends that F_1 holds for agent *j* and pretend that F_2 holds for agent *m*, and *j* can inform *m* about some fact *G* (in particular, it could be that *G=M1)*, than *i* has to keep pretending that the conjunction of F_1 and *F* holds

$K_j\, P_i\, F_1 \;\;\&\;\; K_m\, P_i\, F_2 \;\;\&\;\; (K_j\, G \rightarrow \;\; K_m G) \rightarrow\; F{=}F_1\, \&\, F_2\, \&\;\; P_i\, F\,.$

4) If an agent i pretends that F_1 holds and agent j pretends that F_2 holds, and this pretending has been accepted by both of them, than both of them aware that neither F_1 nor F_2 holds.

$$P_i\, F_1 \;\&\; P_j\, F_2 \;\&\; P_j\, K_j\, F_1 \;\&\; P_i\, K_i\, F_2 \;\rightarrow\; K_i\, not\, (F_1 \;\&\; F_2) \;\&\; K_j\, not\, (F_1 \;\&\; F_2).$$

3.2 Understanding mental states of the participants of a scene

When the system knows the definitions of mental concepts and the mental states for a particular scenario, the question-answering can be organized, initiated by a picture (Fig.2).

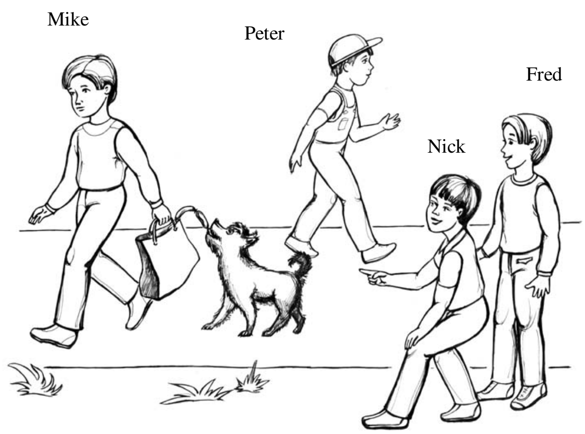

Fig. 2: Natural Language (NL) system answers the questions about the mental states of the scene heroes (Mike, Peter, Fred, Nick and the dog).

Below are examples of questions the patients may ask the system about, while watching the scene. The trainee compares the expected answers (their own understanding of the mental entities) with ones suggested by the system.

Does Mike see that the dog is eating the sausages? Does Peter see what is happening with Mike and the dog? Does Nick know what is happening with Mike and the dog? Which way does Nick express his emotions? Does Fred know whether Peter knows what is happening with sausages? Does Nick want to keep the dog from eating the sausages? What would Fred do if he wants to let Peter know what is happening? Does Fred believe that Mike would yell at the dog if he (Mike) knew that it was eating the sausages?

Understanding mental states of the couple of abstract agents

NL system is capable of processing the requests about the mutual mental states of two agents, Mike and Peter. Each textual answer (the predicate *iassert/1*), describing the particular mental state of these agents, is associated with the semantic header: formal expression, representing the essential idea of the answer.

What if Peter knows that Mike knows that Peter does not believe that something holds?
know(peter,know(mike,not believe(peter,_))):- iassert($Peter would expect Mike to keep convincing Peter concerning the token location.$).
What does Peter want Mike to know about the belief of Peter ? Does Peter know that Mike knows that Peter does not believe that something is true ? How can Peter pretend that Peter does not want that Mike know anything ?

What would Peter expect if he would know that Mike knows that Peter does not know something ?
Guess what did Mike do such that he would want that Peter did not know anything?
What did Peter want to tell Mike about his belief ?
Did Mike pretend that he did not tell Peter what Mike wanted?

The reader can access the question-answering web page at
www.iaskweb.com/tech_engines.html

abstr	agent	abstr
know	Mike	location

Mike wants to know location of the token

<table>
<tr><td></td><td>agent</td><td colspan="3">Abstr</td></tr>
<tr><td rowspan="2">believe</td><td rowspan="2">Mike</td><td></td><td>agent</td><td>abstr</td></tr>
<tr><td>not know</td><td>Peter</td><td>location</td></tr>
</table>

Mike believes that Peter does not know the location

<table>
<tr><td></td><td>agent</td><td colspan="3">Abstr</td></tr>
<tr><td rowspan="2">want</td><td rowspan="2">Mike</td><td></td><td>agent</td><td>abstr</td></tr>
<tr><td>know</td><td>Peter</td><td>location</td></tr>
</table>

Mike wants Peter to know the location

<table>
<tr><td></td><td>agent</td><td colspan="5">Abstr</td></tr>
<tr><td rowspan="3">believe</td><td rowspan="3">Mike</td><td></td><td>Agent</td><td colspan="3">abstr</td></tr>
<tr><td rowspan="2">want</td><td rowspan="2">Peter</td><td></td><td>agent</td><td>abstr</td></tr>
<tr><td>know</td><td>Peter</td><td>location</td></tr>
</table>

Mike believes that Peter wants to know the location

<table>
<tr><td></td><td>agent</td><td colspan="4">Abstr</td></tr>
<tr><td rowspan="2">want</td><td rowspan="2">Mike</td><td></td><td>agent</td><td>agent</td><td>abstr</td></tr>
<tr><td>inform</td><td>Mike</td><td>Peter</td><td>location</td></tr>
</table>

Then Mike wants to inform Peter about the location

Fig. 3: Output of the question-answering system in the form of tables. These tables are automatically synthesized to represent mental formulas. The green rows contain the semantic types for the arguments of metapredicates. The *agent* type ranges over the participants of the scenario, and the arguments of mental predicates (metapredicates) have the semantic type *abstr*: they range over arbitrary (mental) formulas. The white rows contain the name of atoms.

If autistic children have advanced mathematical skills, it seems fruitful to teach them reasoning about mental formulas using visualization by means of dynamic tables (Fig.3). This way of representation can be referred as *semi-structured*: each cell can be replaced by a table for a formula, if the corresponding variable is substituted with this formula. We illustrate how the concept *inform* is derived via the basic mental predicates, given two agents Mike and Peter with mutually-involved mental states (See the first Section). The last table is the conclusion (reaction) that the mental action *inform* has actually occurred.

4 Natural language engine for the mental states domain

4.1 Semantic headers

The technique of semantic headers (SH) is intended to resolve a problem of conversion an abstract textual document into a form, appropriate for the question-answering (Galitsky 1999b). There are two opposite approaches to this problem. The first one assumes that complete formal representation of any textual document is possible, and the second one assumes that the textual information is too tightly linked to NL and it cannot be satisfactorily represented without it.

The first approach brings the advantage of the understanding, based on the full-range reasoning capabilities, but sacrifices the necessity of NL generation, which introduces additional deviation of the answer semantic. Attempts to provide an exact answer, inherent to the fully formalization approach, narrow down the overlap between a query and existing answers. (Query understanding is posed as a recognition problem: how to find the most relevant answer for a given query, even if there is no direct answer).

The SH technique is an intermediate approach in respect to the degree of knowledge formalization. SH approach is intended for the NLP system, which is combination of information extraction and text understanding systems. Let us consider the following example. We build a mental formula from the sequence *want-pretend-inform.* The following expression is the semantic header for the mental scenario involving Mike and Peter:

want(mike,pretend(mike,peter,inform(mike,peter,smth))):-
iassert($ Mike whispered something, and then he pretended that he expected a response back from Peter. Actually, Mike wanted to conceal some information from Peter and tries to avoid confession.$).

The predicate *iassert/1* adds the answer in a current list of relevant answers.

Semantic headers represent only the essential mental component of this abstract. It ignores both the physical actions, objects of these actions and the names of heroes because they will not be mentioned under the querying of mental states. Conversely, this SH focuses on the most complex mental state, involving the mental predicates for pretending, intention and informing, possibly subject to negation. Presented SH can be matched against the following queries of the decreasing generality: "What does Mike want?" *want(mike, What)*, "Does Mike want to pretend?" *want(mike,pretend(mike, Whom, What)),* "Why would Mike want to pretend for Peter?" *want(mike,pretend(mike, peter, What)),* "Did Mike want to pretend for Peter that he (Mike) has informed Peter about something? *want(mike,pretend(mike,peter, inform(mike,peter,smth))).* The last query is actually equal to the semantic header above.

4.2 Scheme of the natural language component

The following tasks need to be resolved on the way towards the natural language system, answering question concerning mental states:

1. NL understanding of a query or statement.
2. Domain representation in the form of semantic headers, where mental formulas are assigned to the textual representation (abstract) of a story.
3. Synthesis of all well-written mental formulas in the given vocabulary.
4. Matching of the translated NL query against the semantic headers. We use the approximate match in case of failure of the direct match.
5. Synthesis of canonical NL sentence based on mental formula.
6. Keyword search initiation in case of failure of mental formula search.

Fig .4 presents the interaction between the components of the WOL search system. The flexible system architecture allows functioning in two modes: WOL search and WOL database extension. Similar NL processing 1) and mental formula analysis 3)-5) features are employed in both modes, but the component interaction is different. See the "Domain Extension Wizard" Section for explanation.

Component 3) is required, because the traditional axioms for knowledge and belief (see, for example, Fagin et al 95) are insufficient to handle the totality of all realizable mental formulas. We developed the algorithm to extract the realizable mental formula from the totality of all well-written mental formulas, represented via metapredicates. In addition, introduction of the classes of equality of mental formulas are required for the approximate match of mental formulas 4), which is also inconsistent with traditional axiomatic of reasoning about knowledge and belief. NL synthesis of mental expression is necessary for the verification of the system's deduction. A customer needs it to verify that he/she was understood by the system correctly before this customer starts to evaluate the answer. NL synthesis in such strictly limited domain as mental expression is straightforward and does not require special considerations.

In the Table2 below, we describe the peculiarities of semantic processing of the system, specific to the mental states domain. Ideally, the semantic component should be domain-independent. However, semantic rules for the analysis of mental formulas require specific machinery for complex embedded expressions and metapredicate substitutions (see examples in the Table1 and the last Chapter).

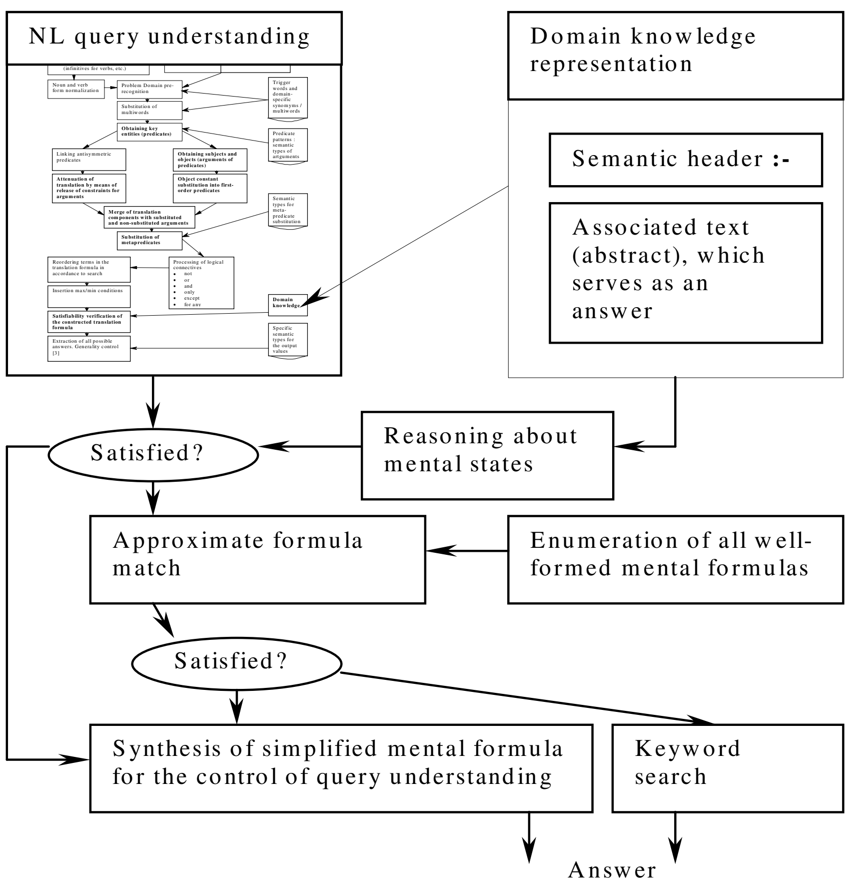

Fig.4: The scheme of the literature search system. NL query, containing the mental states and action of the heroes, is converted into mental formula. If there is no semantic header in the domain knowledge component, which satisfies the resultant translation formula for a query, the approximate match is initiated. Using the enumeration of all well-formed mental formulas, the system builds the approximation and deductive link between the translated query and existing semantic header. If even the

approximate match fails, the system generates the list of keywords to search through the texts of mental scenarios.

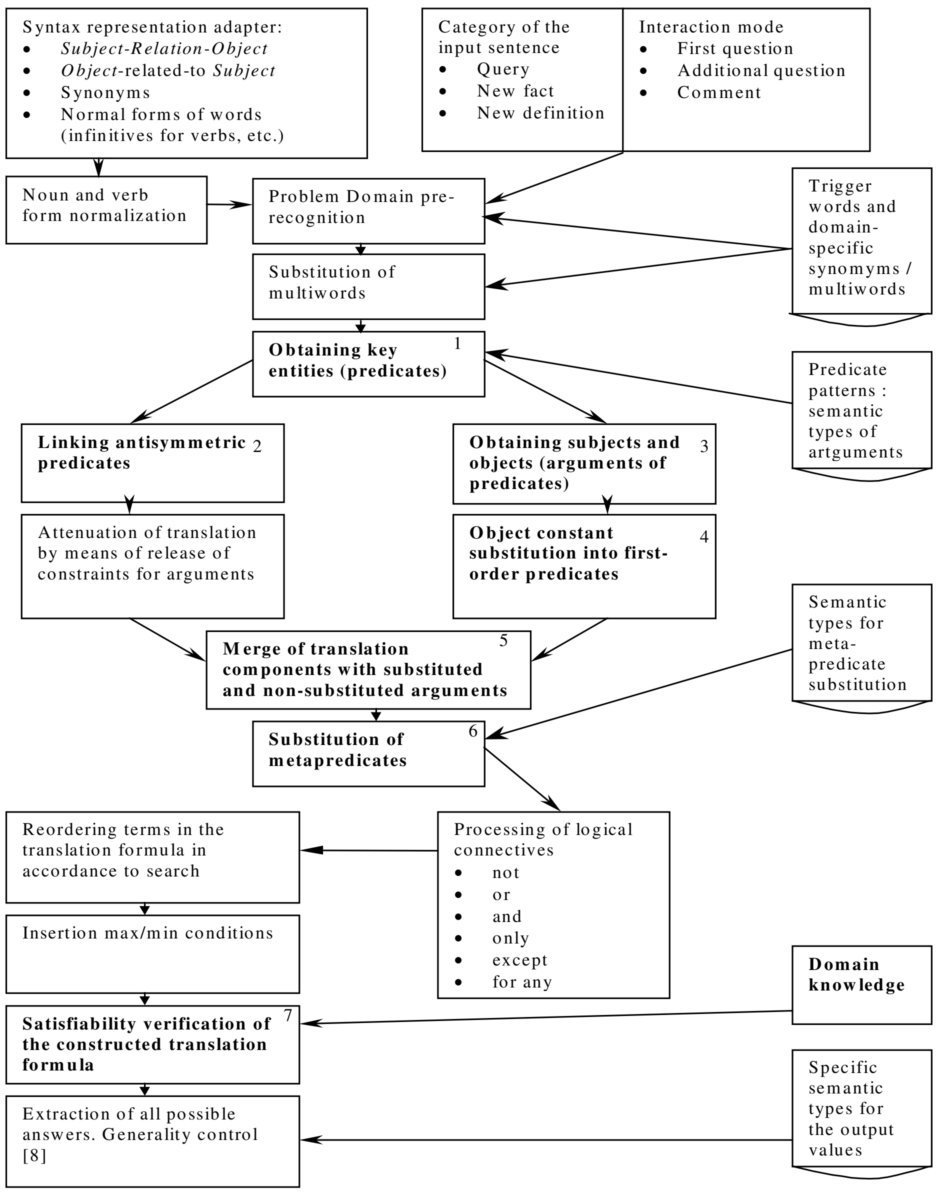

Fig.5 The general scheme of the semantic processor for the variety of problem domains of different nature, from semantic header to the fully formalized kind of domains. The units, essential for semantic processing in the mental states domain, are shown bold and presented in the Table 2 below. The semantic analysis implements the generality control (Galitsky 1999b), using the satisfiability verification of the translation formula as an additional criterion of application of semantic rules.

Syntax representation adapter (Fig 5) accepts the parsing tree, obtained by the external morphological and syntactic analysis subsystem, as well as the original NL query. The semantic processor allows various interactive modes (including iterative question-answering) as well as domain extensions. The system normalizes the parts of speech, pre-recognizes problem domains and substitutes multiwords. The further processing is presented in the Table 2 except attenuation of translation, which is standard for the mental states domain. If the resultant translation formula is found by the verification unit to be too general (too many answers) or too specific (no answers), then the family of syntactic transformations is applied to the translation to achieve an appropriate generality. Reordering terms, insertion logical connectives, etc. is not domain specific and does not require separate consideration for the mental states domain. Four units, storing the data for semantic representation, are shown on the right: domain-specific synonyms, which are required in addition to the general syntactic and semantic information, patterns of the predicates, enumeration of values for semantic types, and names for semantic types for the canonical XML-table output (see fig.3 for an example).

Unit name	**Unit function and implementation** (see http://www.dimacs.rutgers.edu/~galitsky/NL/unit.html for the rest of units and details)
1. Obtaining key concepts (predicates)	The normalized words and substituted multiwords of the sentence are matched against the set of all predicates for the domain. There are two classes of predicates in the mental state domains**: mental metapredicates** and **domain-specific predicates** of "physical" state and actions. In contrast to the physical predicates, the properties of mental metapredicates are embedded in the system; it is capable of reasoning about them. If a query includes the predicates from the other domains, we conclude that they do not contain information (no new knowledge, no additional constraints or links) within the given semantic model for the pre-recognized domain. If it is not the case, then we conclude that our current semantic model does not cover this query. Such a situation should not be considered as a misunderstanding. Either the user has to be notified that he/she is out of domain bounds, or the domain representation should be reevaluated. The metapredicates of mental states and actions are invariant in respect to a particular domain both in terms of NL processing and reasoning machinery. In this table, we consider the example of the following sentence: *"He pretends that she believes that he does not pretend that he is not a murderer of her husband"*. The system extracts the mental metapredicates *pretend0, believe, pretend1* and physical predicate *murder*, which does not participate in the further considerations.
2. Linking anti-symmetric predicates	Analysis of the anti-symmetric predicates is one of the most complex subtasks in the construction of the translation, involving the propositions, which sometimes link the anti-symmetric predicates together. We demonstrate the four possibilities of the links between the couples X_{s1},X_{s2} and X_{t1},X_{t2} of arguments for the pair of anti-symmetric predicates p_s and p_t : $p_s(X_{s1},X_{s2},\ldots,X_{sn})\&\ldots\& p_t(X_{t1},X_{t2},\ldots,X_{tn})$. These cases are: *1)* $X_{s1}=X_{t1}$ *&* $X_{s2}=X_{t2}$ *; 2)* $X_{s1}=X_{t1}$ *; 3)* $X_{s2}=X_{t1}$ *; 4)* $X_{s2}=X_{t2}$ *;* Note that in the cases 2)-4) other variables may or may not be equal. We present the following example: "*Who was informed about pretending*?" This unit will receive the formula $inform(X_{s1}, X_{s2}, pretend)\&pretend(X_{t1}, X_{t2},\ldots)$. In the first case, the first person informs another person that this first person is pretending to the second person about something. In the second case, the first person informs the second person about pretending for the third person. In the third case, a person is informed about his/her own pretending. In the fourth case, a person is informed that another person is pretending for this person. There are specific rules, which take the translation template, verify some constraints

	of the propositions, related to these predicates, the other (non-anti-symmetric) predicates, related to both of them, to conclude which case 1)-4) to involve. Note that the analysis of argument substitution occurs independently (unit 3).
3. Obtaining subjects and objects (arguments of the predicates)	As only the predicates have been revealed, the system generates the list of all objects for all semantic types for all arguments. The generic agent names (*he, she, wife, husband, lover, mistress, person, other person*, etc.) specify the sub-domains in addition to mental metapredicates. In mental states domain, there are two basic semantic types: agents and formula substitution (abstract). The generated list is then matched against the input sentence to **extract** the **object/subject references** from the input query.
4. Object constant substitution into first-order predicates	After all the objects for all semantic types involved are found, the system determines the belonging of the object to the particular predicate (from the set of all query predicates), which has the semantic type for this object. There is a series of rules determining the belonging preference in case it has not been established by the syntactic component. Additional **substitution control** is required in addition to *object-predicate-subject* triples, revealed by this component. The more arguments are substituted by the objects, the **less** the **translation generality** (Galitsky 1999b). The hardest function of this unit is the substitution into the arguments of anti-symmetric predicate. Additional syntactic analysis is required to decide, which object should be substituted to which anti-symmetric argument. We give the example of substitution into the expression *pretend(Who1, Whom1,What1),believe(Who2,What2),pretend(Who3,Whom3,What3)* , obtaining *pretend(she, Whom1, believe),believe(she,pretend),pretend(he,Whom3,What3)*
5. Merge of translation components with substituted and non-substituted arguments	The unique feature of our approach is that the analysis of the links (mutual context) of the predicates is implemented completely separately from the determination of the argument substitution. Each of these operations has the specific semantic rules, and each of them is intended to correct (if necessary) the results of syntactic analysis. This architecture is chosen to achieve the generality control compatibility under backtracking mechanism: each fork is subject to independent formula attenuation (Galitsky 1999b). After the system completes these operations, the formal **unification** procedure is initiated: *Translation_with_unified_variables=Translation_with_substituted_arguments* The unification procedure merges the results of predicate linking and argument substitution. If a subject is related to more than one object or more than one predicate (part of speech), this fact is reflected by unification procedure. If one argument is substituted, and the other argument of the other predicate has the same value, then this other argument is substituted as well. Direct semantic rule, which would call the same result, could be inconsistent with the semantic rules for predicate linking and/or argument substitution. Splitting of the formula translation procedure into these procedures provides the clarity and consistence of the semantic rules for each of these procedures. If the unification is impossible, then the rule application in the previous units was incorrect and is subject to re-evaluation. The unification expression will look like (we use the internal Prolog denotations for

	variables; atoms are enumerated to encode multiple occurrence): *[pretend0(_9378,_938C,_93A0),believe0(_938C,_93B4),pretend1(_938C,_93C8,_93DC)]* = *[pretend0(she0,_A58C,believe0),believe0(she1,pretend1),pretend1(he0,_A5A0,_A5B4)]* Note that the right part of the formula has some arguments instantiated with values *she0, she1, he0* (of *agent* semantic type), *pretend1, believe0* (of *abstract* semantic type). The unification result is *pretend0(she0,he0,believe0), believe0 (she1,pretend1), pretend1(he0,Whom3,What3)*
6. Substitution of meta-predicates	Using metapredicates to build the translation of NL query is the essential capability of the system. The metavariables of metapredicates have the specific **abstract semantic type** to be able to range over arbitrary formulas. The metapredicate processing starts with the substitution of one predicate into the abstract semantic type of the other predicate (that is indeed a metapredicate). It occurs in the argument substitution unit (4) in accordance to the same semantic rules as with the objects of the regular semantic types. This unit replaces the abstract argument of a metapredicate, instantiated by the symbol of a predicate (the predicate name), by the whole term (this predicate with its arguments). As a result, we obtain *pretend0(she0,he0,believe0(she1,pretend1(he0,_674C,_6760))).*
7. Satisfiability verification for the constructed translation formula	The system runs the resultant translation (from NL into internal formal language) against the domain knowledge representation. Depending on the instantiation state (division of a set of variables into instantiated and non-instantiated), each term of the translation is matched against the domain representation. This match either finds the values for the arguments or verifies the compatibility of values the translation formula contained prior to the current satisfiability procedure. The ability to answer the "find an object" questions versus "yes-no" ones is built-in for our system architecture. Call to the translation, subject to the negation substitution and atom de-enumeration, looks like the following: *pretend(she,he,believe(she, not pretend(he, Whom3,What3))).* As a result, we obtain the instantiated expression *pretend(she,he,believe(she, not pretend(he,* ***she, murder(he, husband(she)****)))).* Notice that an ideal semantic processing would find the value *she* without satisfiability verification (it occurs in the input NL sentence), but the mental expressions in our domain are sufficiently complex to provide disambiguation. The instantiated formula serves as a semantic header to the answer

Discussion

It was discovered more than a decade ago, that autistic children have a deficit in attributing beliefs to others (in their "**theory of mind**" (Baron-Cohen 1989)). Autistic children fail to distinguish physical and mental entities, unaware of the mental function of the brain and cannot take into account their own mental states, in contrast to the mentally handicapped or normal children of the same mental or chronological age. Autistic subjects are impaired at providing context-appropriate mental state explanations for the story characters' nonliteral utterances, compared to normal and mentally retarded (MR) controls. There are autistic subjects who passes theory of mind test and showed impairments on the more naturalistic story materials, in accordance to (Happe 1994). The study of representation of substantive word meanings in children with autism (Tager-Flushberg 1985) showed the equivalence of comprehension of words on basic and superordinate levels for both autistic children and controls. The semantic knowledge for concrete objects is represented and organized in similar ways in autistic, MR and normal children. These findings build the background for us that the autistic corruption occurs on the level of axioms and not on the level of symbolic denotation (of a single entity).

There was a series of experiments of the understanding of **second-order beliefs and intentions**, **lying** and **deception**. It was reported (Leekam & Prior) that about 40% of autistic children who passed the first-order belief test passed the second order one as well, that satisfactorily matches our experimental data. Neither normal nor autistic children found second-order intention easier than second-order belief. At the same time, in contrast to autistic children, normal ones found the ability to estimate another person's mental state easier than determining whether a person is lying or joking. The study by (Yirmiya et al 1996) states that for individuals with autism theory of mind abilities correlate with verbal ability; non-verbal abilities of MR individuals correlated with deception and first belief scores.

In accordance to the relevance theory, the suffer of autistic patients from a specific and characteristic impairment in the ability to attribute mental states should cause the difficulties with use of **language for communication** (Happe 1993). Relevance theory allows precise predictions about the levels of communicative competence that should be possible with either no, first-order only or second order theory of mind ability.

There is a biological background for the division of reasoning axioms into "physical" and "mental": there are distinguishing areas in the brain, responsible for corresponding information processing. The functional neuroimaging study (Fletcher et al 1995) with positron emission tomography investigated the brain activity in normal volunteers while they performed the story comprehension task. These stories are necessarily involved reasoning about mental states. The resultant brain activity was compared with that measured in two control tasks: "physical" stories, which does not require the mental attributions, and the sets of logically unlinked sentences. If to compare with unlinked text, the brain activity showed increase of regional cerebral blood flow in the temporal poles and the posterior cingulated cortex. Comparison of mental and physical stories revealed a specific pattern of activation in the medial frontal; this pattern is associated only with mental state attribution. This observation supports our hypothesis, which establishes the **link between the brain and axiomatic method**. This study builds the simple analogy: if there exists a hypothetical technical system, which stores mental and physical "processing" in different components, and the former component can be damaged such that the resultant system behavior would show the corruption of specific property of reasoning. In the current study, we move one step further and hypothetically divide the brain not only into two components, but also into multiple axiom-determining ones and descent into the brain to the level of mental axioms. Since there are subjects with different patterns of impaired reasoning, we conclude that there are not just two areas in the brain such that one of them (mental) can be damaged, but that the mental area can in turn be hypothetically split into the "axiom-determining" sub-areas.

Categorization and **classification** experiments are important to highlight the intellectual peculiarities of autism and reveal the distinguishing features from MR. It was discovered (Shulman, Yirmiya & Greenbaum 1995) that on the perceptual matrices, participants with MR performed less well than those with autism, who performed less well than normal children. However, for the task of free-sorting representational objects and class-inclusion, which require hire operational thought, MR and autistic children switched their positions. This study suggests that internal manipulation of information is more corrupted in respect to MR and controls.

Neurobehavioral theories of autism describe various clinical behaviors as consequences of particularities in visual information processing. However, it was verified (Bertone et al 1999) that in general, high-functioning autistic patients are able to discriminate motion (simple, translational, radial, rotational) as well as normal controls.

To conclude the biological background for autism, we proceed to some **genetic** clues for autism (see the survey by Turner et al 2000, Gilberg 1998). Because of the clear evidence that idiopathic autism has a strong genetic basis, many groups are undertaking whole genome screens to identify susceptibility loci. As autism is associated with several distinct medical disorders, genetic heterogeneity among idiopathic cases are seems to be likely. There are three strategies to find these loci:

1) Candidate gene strategy examines genes that seem likely to be implicated in a disorder. Genes thought to be involved in relevant pathophysiological process are known as functional candidates. Positional candidate genes are those within or close to cytogenetic rearrangements associated with the disorder. There are no very strong functional candidate genes for autism, however ones involved in serotonergic and GABA systems are potential functional candidates.

2) Genome screens helps in the absence of candidate genes. 300-400 highly polymorphic microsatellite markers, evenly spaced through the genome, are typed in an affected relative per sample. Allele sharing linkage analysis is used to identify markers at which affected relative pairs share alleles more often than expected by chance. A statistically significant increase in sharing implies that the marker might be linked to a nearby susceptibility locus.

3) Association studies, when the areas of suggestive linkage can be fine mapped. Identifying of alleles occurs at significantly different frequencies in affected versus control samples.

Presently, there are no specific drug therapies for autism, although particular symptoms, such as repetitive behavior, can be treated by pharmacotherapy, the core social difficulties are relatively unresponsive to medication. Current drug therapies rely on altering neurotransmission without understanding the underlying neurophysiological abnormalities. A potentially early application of a better understanding of the genetic background of autism will be the basis of improved genetic counseling to parents of autistic children who want information about the risks of having further affected children, and to other family members who are concerned about having an autistic child themselves.

Conclusions

The phenomena of **computational autism** has been introduced to explore the theoretical and practical issues of the mental disorder. The theoretical result is building one more link between the experimental brain research and symbolic reasoning. The study showed that the abstract formalism of reasoning about mental states, intended to model the human behavior by a technical system, found an emergent application in the simulation of human agents.

In contrast to the psychological approach to rehabilitation of mental reasoning, we teach the children the very essence of the mental world: the axioms for intention, knowledge and belief. One would argue that the autistic children rather need to practice of mental reasoning in a real-world than to learn just one more mathematical subject (the mental axioms). Indeed, encouraging autistic patients to learn the explicit mental axiom is easier acceptable by them than the direct involvement in communication with other persons; furthermore, acquiring of "formal" skills leads to more successful capabilities of living in a mental world.

The practical value of this study is compositions of the set of scenarios with accents on reasoning about mental state for diagnosis and rehabilitation of for autism spectrum disorders. The training methodology is intended for manual training by psychologists, as well as by computer interactive environment for proper reasoning about human agents. Our experiments demonstrated that rehabilitated deductive capabilities of autistic children contributes in the overall emotional development, usually considered as a main target of corruption under this disorder.

References

1. Baron-Cohen, S. 1998. Social and Pragmatic Deficits in Autism: Cognitive or Affective *Journal of Autism and Development Disabilities* 18:3.
2. Leslie, A. and Thaiss, L. 1992. Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition,* 43, 225-251.
3. Leslie, A 1987. Pretence and representation: The origins of "theory of mind". *Psychological review* 94:412-426.
4. Galitsky, B.A. 1998. The formal scenario and metalanguage support means to reason with it *DIMACS Tech. Report #98-27*, Rutgers University.
5. Galitsky, B.A. 1999a. Analysis of mental attributes for the conflict resolution in multiagent systems, in *AAAI-99 Workshop on agents' conflict*, Orlando FL.
6. Galitsky B.A. 1999b. Natural Language Understanding with the Generality Feedback *DIMACS Technical Report 99-32,* June 99.
7. Galitsky, B.A., 1999c. How the logic of mental attributes models the autism. *International Conference on Neural and Cognitive systems*. Boston Univ., MA.
8. Fagin,R., Halpern,J.Y., Moses,Y., Vardi,M.Y. 1995. Reasoning about knowledge *MIT Press*, Cambridge, MA, London, England.
9. Konolige, K.1986. A deduction model of belief. *Morgan Kaufmann Publ*.
10. Perner, J. 1991. Understanding the representational mind. *MIT Press* Cambridge MA.
11. Zaitchik, D. 1990. When representations conflict with reality: The preschooler's problem with false beliefs and "false" photograph. *Cognition*, 35: 41-68.
12. Pratt, C., and Bryant, P. 1993. Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *J. Child development* 61: 973-983.
13. Galitsky, B.A. 1997. www.dimacs.rutgers.edu/~galitsky/MS/ma/wkb.html (definitions of mental concepts based on *want, know,* and *believe).*
14. Dahl, V.1999. The logic of language, in *The Logic Programming Paradigm,* Apt, Marek, Truszczynski, Warrens, eds, Springer-Verlag.
15. Decker,K., Pannu, A., Sycara, K. and Williamson, M.,1997. Designing Behaviors for Information Agents, in *International Conference on Autonomous Agents'97*.
16. K. Dautenhahn, I. Werry 2000: Issues of Robot-Human Interaction Dynamics in the Rehabilitation of Children with Autism. To be published in Proc. FROM ANIMALS TO ANIMATS, The Sixth International Conference on the Simulation of Adaptive Behavior (SAB2000), September 2000, Paris, France.
17. Prior, M., Eisenmajer, R., Leekam. S., Wing, L., Gould, J., Ong, B., Dowe, D. 1998. Are there subgroups within the autistic spectrum? A cluster analysis of a group of children with autistic spectrum disorder *J Child Psychol Psychiatry* Sep; 39(6):893-902.
18. Wetherby, A.M.,Prutting, C.A. 1984, Profiles of communicative and cognitive-social abilities in autistic children *J Speech Hear Res* Sep;27(3):364-77.
19. Mawhood, L.,Howlin, P., Rutter, M 2000. *J Child Psychol Psychiatry* Jul; 41(5): 547-59.
20. Blocher K.H., 1999, Affective Social Quest (ASQ): Teaching emotion recognition with interactive media and wireless expressive toys MSC Thesis, MIT.
21. Resnick, M., 1987 LEGO, LOGO, and Life, Proc on Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, Ed. Langton, C.G., Los Alamos, NM Sept 87, pp397-406.
22. Bertone, A., Mottron, L., Jelenic, P., Casanova, C.and Faubert, J.,1999, High functioning autistic subjects with a preoccupation for rotating objects show an increased sensitivity to complex rotational motion patterns. *Soc.Neurosci. Abstr.*, v 25, part 1, p489..

23. Beversdorf, D.Q., Smith, B.W., Crucian, G.P., Anderson, J.M., Keillor, J.M.,Barrett, A.M., HughesJ.D., Felopulos, G.J., Bauman, M.L.Nadeau, S.E and Heilman, K.M. 1998. The effect of semantic and emotional context on written recall for verbal language in high functioning adults with ASD *J Neurol NeurosurgPsychiatry* Nov 65(5):685-92.
24. Turner, M, Barnby, G., Bailey, A. Genetic clues to the biological basis of autism. *Molecular Med Today* 6:238-244.
25. Gilberg, C. 1998 Chromosomal disorders and autism. *J Autism Dev Disord*. 28, 415-425.
26. Shulman,C., Yirmiya,N., & Greenbaum C.W.1995 From categorization to classification: a comparison among individuals with autism, mental retardation, and normal development J Abnormal Psychol 104(4):601-9.
27. Baron-Cohen, S. 1989 Are autistic children "behaviorists"? An examination of their mental-physical and appearance-reality distinctions. J Autism Dev Disord 1989 Dec 19(4):579-600.
28. Baron-Cohen S, Ring H, Moriarty J, Schmitz B, Costa D, Ell P.Recognition of mental state terms. Clinical findings in children with autism and a functional neuroimaging study of normal adults.*Br J Psychiatry*. 1994 Nov;165(5):640-9.
29. Happe, FG 1994. An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J Autism Dev Disord*. Apr;24(2):129-54.
30. Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD 1995.Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition*. Nov;57(2): 109-28.
31. Tager-Flusberg, H.1985 The conceptual basis for referential word meaning in children with autism.
Child Dev. Oct;56(5):1167-78.
32. Yirmiya N, Solomonica-Levi D, Shulman C, Pilowsky T. Theory of mind abilities in individuals with autism, Down syndrome, and mental retardation of unknown etiology: the role of age and intelligence.
J Child Psychol Psychiatry. 1996 Nov;37(8):1003-14.
33. Leekam SR, Prior M. Can autistic children distinguish lies from jokes? A second look at second-order belief attribution. *J Child Psychol Psychiatry*. 1994 Jul;35(5):901-15.
34. Happe FG 1993 Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* Aug; 48(2):101-19.
35. Sams M.J., Gardner A.A., Sides, J.M., Street, M.W., Reed, C., Haynie, K.L. and Willenbring, S. Differential behavioral responses in autistic children: animal-assisted therapy versus conventional therapy *Soc.Neurosci. Abstr*., v 25, part 1, p.489.