Simulating autistic patients as agents with corrupted reasoning about mental states

Boris Galitsky
iAskWeb, Inc.
45 Elliot #B Newton MA 02461
boris.galitsky@iaskweb.com,
http://www.iaskweb.com

Introduction

In this study, we build the logical model of the specific type of human agents: autistic patients. We analyze the distinguishing features of autistic reasoning to construct the software tools for diagnosis and training the children with autism and accompanying mental disorders.

Autism is a relatively rare multifactorial disorder that affects about 5 out of every 10,000 school-aged children. This disorder is characterized by impaired social interaction and communication combined with repetitive and stereotyped patterns of behavior. We focus on some peculiarities of the reasoning of autistic children, which are tightly connected with logical artificial intelligence and could be significant for both autism diagnosis and training. It has been recently discovered that **autistic children cannot reason properly about the mental states** and mental actions of themselves and others [1-3]. At the same time, the deductive capabilities of the autistic children concerning other domains 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 human agent on the hypothesis that there is a number of standard axioms for mental attributes, which 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 specific instances of these axioms.

The paper is organized as follows. The first section introduces the logical model of autism and contains its descriptions, results and discussions. The second section introduces the training scenarios, and the third section presents different components of the autism training toolkit. This toolkit combines such AI technologies as reasoning about knowledge, natural language understanding and narrative generation[6,8,9, 15].

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.

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 him/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 is not necessarily correlated with the generally accepted classification of autism syndromes.

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 corrupted 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 the autistic patients.

1.2 Reasoning about mental states

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 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* and not want is in the degree of 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 fixed set of concepts.

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 fits 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.

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. 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 [14] to related issues of logical programming.

want(Agent, do(Agent, Action)).	agent wants to perform an action
want(Agent, do(DAgent, Action)).	agent wants another agent to perform an action
want(Agent, know(Agent,What)):- (believe(Agent, know(KAgent, What)), ask(Agent, KAgent, What)).	agent wants (himself) to know
believe(Agent, want(WAgent, know(WAgent, What))) :- prefer(Agent, tell(Agent,WAgent, What), OtherAction).	agent believe that other agent wants to know
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
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

Table 1. Examples of the mental formulas and their 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 [4, 13].

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 [8,9]. 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 concerning

1.3 Discussion of the model

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[4] 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 internally consistent ones.

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 is 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 the sequence, 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 [10,11]. 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 [10,11]. 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 showed 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. Note that the testing covers the majority of mental formulas complexity 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.

- Each question (with the mental formula complexity below four) was successfully answered by every control child.
- Each question was failed by at least one autistic kid.
- For each question the autistic child failed, it was possible to perform training such that the question is successfully answered after fifth attempt.
- If to substitute the mental states by physical states, the questions will be easier answered by the autistic children, than the questions of Section 2. It will not make a significant difference with the control children.

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[1-3,10-12]. 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.
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.

1.6 The brain and axioms

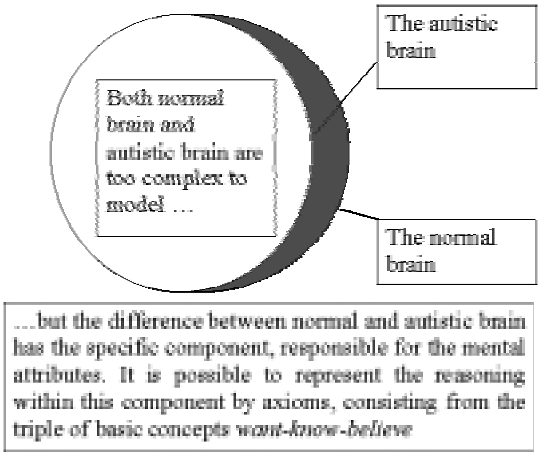


Fig 1.6 Logical model of autism builds the link between the axiomatic method and the brain. We use the logical model of autism for the indirect experimental proof that the brain functions on the axiomatic level.

2 Testing and exercising scenarios

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.

2.1 Mental states of another person

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: If we ask B about his opinion, do you (A) know whether B knows where the token is?
A: B knows that I know that he does not know where the token is. (B knows where the token is, B does not know where the token is, B knows that I know where the token is, B knows that I know that B knows where the token is.)
6) E: Can we achieve a situation, when B will know where the token is?
A: Yes, we can tell him or show him (A is confused: I don’t know).
B enters the room. Now all the questions are repeated; B’s responses, predicted by A, are actually evaluated.
7) E, After A showed (or told) B the location of the token: How do you (B) think, did A know whether you knew the location of the token while out of this room?
B: A knew that I did not know where the token is.
8) E, interrupting B: what do you (A) think, what will B say?
A: B will say that B knew that I knew that he B did not know where the token was.
9) E: Now you (B) know where the token is, because A have shown you. Do you think he (A) wanted you to know where the token was?
B: Yes, A wanted myself (B) to know where the token is.
10) E: Do you (A) know whether B knows that you (A) wanted him (B) to know where the token was?
B : Yes, I know that I wanted B to know where the token was.

2.1.1 A wrong mental state

1) 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).

2.2 Mental state transmission

This is a mirror test to the *mental state of the other person* one.
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).
7) E: I guess, I want your friend to know what is on the picture. Is it true? If so, does B know that you wanted to let him know about the picture? Does B know that you want him to know the picture?
A: I'm not sure. After I informed him about the picture, he would know that I wanted him to know what is on the picture. I don't know if he (B) knows that I want him to know the picture.
Thereafter E calls B in and asks A to actually inform B about the picture. All the questions above are posed for B as B's prediction of mental state of A.

2.2.1 Temporal relationships over the mental states. To forget and to recall

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?
A: Now the bear knows that the rabbit was on the table.
7) E: OK, so the bear recalls that the rabbit was on the table. Does the bear know that before the recollection he did not know that the rabbit had been on the table? Analogously, does the bear know that he(bear) knew that the rabbit had been on the table, while (bear) was watching the rabbit?
A: Yes, the bear knows that he did not know that the rabbit has been on the table, as well as the bear knows that he knew that the rabbit has been on the table while watching the rabbit.

2.3 Pretending

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 soup.

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 Simulation, diagnosis and training

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 MS extension of first-order language to highlight the specific of MS PROLOG implementation [4,6,14]. 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 [5]. 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).

It seems rather hard to fit two following concepts in the single definition: one explains some **facts** about objects and one explains the concept of having **a feature** for a given object.

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).

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

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

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

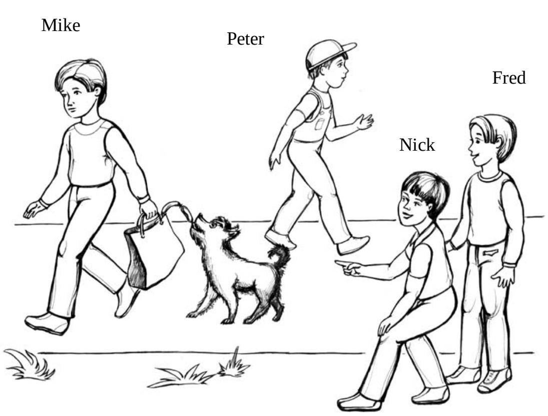
4) If an agent i pretends that F_1 holds and agent j pretends that F_2 holds, and this pretending has been accepted by both of them, than both of them aware that neither F_1 nor F_2 holds.

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

5) An agent can pretend only about his own knowledge

$P_i\, K_j\, F \rightarrow i = j$.

3.2 Understanding mental states of the participants of a scene



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?

Fig 3.2 Natural Language (NL) system answers the questions about the mental states of the scene heroes (Mike, Peter, Fred, Nick and the dog).

3.3 Understanding mental states of the couple of abstract agents

NL system processes the request 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 this mental state.

What if Peter knows that Mike knows that Peter does not believe that something olds?

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?

Does Peter pretend that he wants to know?

3.4 Table representation of mental formulas

Abstr	agent	Abstr
know	Mike	Location

Mike wants to know location of the token

	agent	Abstr		
believe	Mike		agent	abstr
		not know	Peter	location

Mike believes that Peter does not know the location

	agent	Abstr		
want	Mike		agent	abstr
		Know	Peter	location

Mike wants Peter to know the location

	agent	Abstr				
believe	Mike		agent	abstr		
		want	Peter		agent	abstr
				know	Peter	location

Mike believes that Peter wants to know the location

---

	agent	Abstr			
want	Mike		agent	agent	abstr
		inform	Mike	Peter	location

Then Mike wants to inform Peter about the location

Fig 3.4 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. This way of representation can be referred as *semistructured*: 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. The last table is the conclusion (reaction) that the mental action *inform* has actually occurred.

3.5 Understanding a scenario

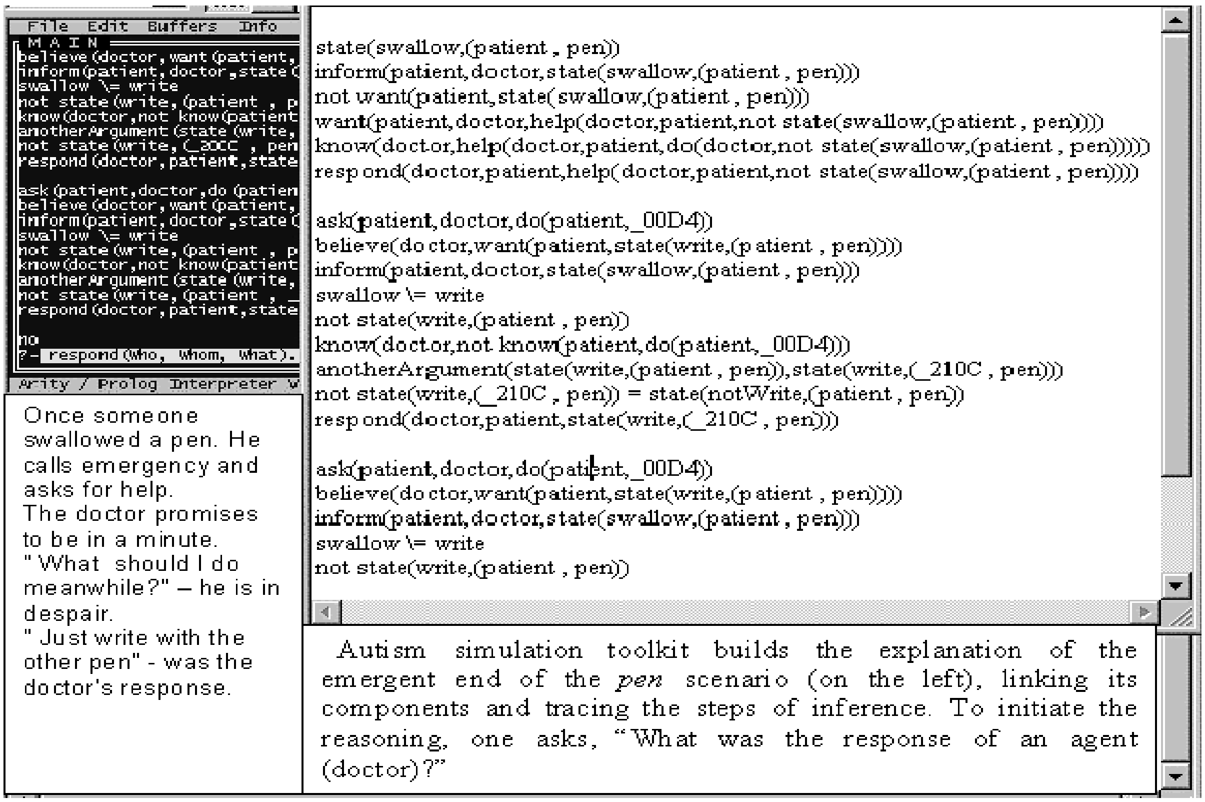


Fig. 3.6. Multiagent scenario simulation toolkit is developed to perform the reasoning, required to resolve the autistic diagnosis/training scenarios. Its knowledge base contains the full spectrum of mental axioms [6]. The question answering of each autistic patient can be reproduced by eliminating the specific axioms, missing by this patient.

References

1. Baron-Cohen, S. Social and Pragmatic Deficits in Autism: Cognitive or Affective *Journal of Autism and Development Disabilities* 18:3 , 1988.
2. Leslie, A. and Thaiss, L. Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition,* 43, 225-251 1992.
3. Leslie, A Pretence and representation: The origins of "theory of mind". *Psychological review* 94:412-426, 1987.
4. Galitsky, B.A. The formal scenario and metalanguage support means to reason with it *DIMACS Tech. Report #98-27,* Rutgers University, 1998.
5. Galitsky, B.A. Analysis of mental attributes for the conflict resolution in multiagent systems, in *AAAI-99 Workshop on agents' conflict*, Orlando FL, 1999.
6. Galitsky B.A. Automatic generation of the multiagent story for the Internet advertisement. *Symposium on AI and creative language: stories and humor*, in AISB-99, Edinburgh, UK.
7. Galitsky, B.A., How the logic of mental attributes models the autism / International Conference on Neural and Cognitive systems. Boston Univ., MA May 1999.
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*1986.
10. Perner, J. Understanding the representational mind. *MIT Press* Cambridge MA 1991.
11. Zaitchik, D. When representations conflict with reality: The preschooler's problem with false beliefs and "false" photograph. *Cognition*, 35, 41-68 1990.
12. Pratt, C., and Bryant, P. Young children understand that looking leads to knowing (so long as they are looking into a single barrel). Child development 61: 973-983, 1993.
13. See www.dimacs.rutgers.edu/~galitsky/MS/ma/wkb.html for more definitions of mental concepts based on *want, know,* and *believe*.
14. Dahl, V. The logic of language, in *The Logic Programming Paradigm,* Apt, Marek, Truszczynski, Warrens, eds, Springer-Verlag 1999.
15. Decker,K., Pannu, A., Sycara,K. and Williamson, M., Designing Behaviors for Information Agents , in *International Conference on Autonomous Agents'97, 1997.*