

Neuro-Symbolic Strong-AI Robots with Closed Knowledge Assumption: Learning and Deductions

Zoran Majkić

ISRST, Tallahassee, FL, USA
majk.1234@yahoo.com

Abstract. Knowledge representation formalisms are aimed to represent general conceptual information and are typically used in the construction of the knowledge base of reasoning agent. A knowledge base can be thought of as representing the beliefs of such an agent. Like a child, a strong-AI (AGI) robot would have to learn through input and experiences, constantly progressing and advancing its abilities over time. Both with statistical AI generated by neural networks we need also the concept of *causality* of events traduced into directionality of logic entailments and deductions in order to give to robots the emulation of human intelligence. Moreover, by using the axioms we can guarantee the *controlled security* about robot's actions based on logic inferences.

For AGI robots we consider the 4-valued Belnap's bilattice of truth-values with knowledge ordering as well, where the value "unknown" is the bottom value, the sentences with this value are indeed unknown facts, that is, the missed knowledge in the AGI robots. Thus, these unknown facts are not part of the robot's knowledge database, and by learn through input and experiences, the robot's knowledge would be naturally expanded over time.

Consequently, this phenomena can be represented by the Closed Knowledge Assumption and Logic Inference provided by this paper. Moreover, the truth-value "inconsistent", which is the top value in the knowledge ordering of Belnap's bilattice, is necessary for strong-AI robots to be able to support such inconsistent information and paradoxes, like Liar paradox, during deduction processes.

Keywords: AGI, Belnap's bilattice, First-order Logic, Robotics

1 Belnap's 4-valued Logic for AGI Robots

Many-valued logic was conceived as a logic for uncertain, incomplete and possibly inconsistent information which is very close to the statements containing the words "necessary" and "possible", that is, to the statements that make an assertion about the *mode of truth* of some other statement. *Algebraic* semantics interprets modal connectives as operators, while *Relational* semantics uses relational structures, often called Kripke models, whose elements are thought of variously as being possible worlds; for example, moments of time, belief situations, states of a computer, etc.. The two approaches are closely related: the subsets of relational structures form an algebra with modal operators, while conversely any modal algebra can be embedded into an algebra of subsets of a relational structure via extensions of Stone's representation theory.

Many classical as well as non-classical techniques for modeling the reasoning of intelligent systems have been proposed. Several frameworks, based on a many-valued logic with truth/knowledge partial orders (and lattices), for manipulating uncertainty, incompleteness, and inconsistency have been proposed in the form of particular extensions of classical logic programming and deductive databases. Such truth/knowledge ordered lattices can be integrated into a unique structure called bilattice [3,1,4], and Ginsberg has shown that the same theorem prover can be used to simulate reasoning in first order logic, default logic, prioritized default logic and assumption truth maintenance system.

Bilattice theory is a ramification of multi-valued logic by considering both truth $\leq$ and knowledge $\leq_k$ partial orderings. Given two truth-values $x, y \in \mathcal{B}$, if $x \leq y$ then y is at least as true as x , i.e., $x \leq y$ iff $x < y$ or $x = y$. The two operations corresponding to this ordering (t-lattice) are the meet (greatest lower bound) $\wedge$ and the join (least upper bound) $\vee$. For the knowledge-ordering $x \leq_k y$ means that y is more precise than x , i.e., $x \leq_k y$ iff $x <_k y$ or $x = y$. The operations $\otimes$ and $\oplus$ correspond to the greatest lower bound and least upper bound respectively in the knowledge-ordering (k-lattice). The negation operation for these two orderings are defined as the involution operators which satisfy De Morgan law between the join and meet operations.

Definition 1. (Ginsberg [1]) A bilattice $\mathcal{B}$ is defined as a sextuple $(\mathcal{B}, \wedge, \vee, \otimes, \oplus, \neg)$, such that:

1. The t-lattice $(\mathcal{B}, \leq, \wedge, \vee)$ and the k-lattice $(\mathcal{B}, \leq_k, \otimes, \oplus)$ are both complete lattices.
2. $\neg : \mathcal{B} \rightarrow \mathcal{B}$ is an involution ($\neg\neg$ is the identity) mapping such that:
 $\neg$ is lattice homomorphism from $(\mathcal{B}, \wedge, \vee)$ to $(\mathcal{B}, \vee, \wedge)$ and $(\mathcal{B}, \otimes, \oplus)$ to itself.

The two partial orders $\leq$ and $\leq_k$ represent respectively how much confidence we have in the validity (truth) of a particular sentence, and how much information (knowledge) we have about it. Notice that from this definition, the negation $\neg$ is an antitonic operator w.r.t. the $\leq$, with $\neg 1_t = 0_t$, $\neg 0_t = 1_t$ (where $0_t, 1_t$ are the bottom and the top elements w.r.t the truth-ordering $\leq$ respectively), but *monotonic* w.r.t. the knowledge ordering $\leq_k$. This homomorphism $\neg$ inverts the truth partial ordering, i.e., if $x \leq y$ then $\neg x \geq \neg y$, while it preserves the knowledge preordering, i.e., if $x \leq_k y$ then $\neg x \leq_k \neg y$.

A valuation of a set of sentences $\mathcal{L}_0$ is denoted by mapping $v^* : \mathcal{L}_0 \rightarrow \mathcal{B}$. They generate a bilattice as well with partial orders $\preceq$ and $\preceq_k$ respectively:

Definition 2. (Ginsberg [2]) Given the valuations v_1^* and v_2^* , we will say that v_2^* is a knowledge extension of v_1^* , writing $v_1^* \preceq_k v_2^*$ if $v_1^*(\phi) \preceq_k v_2^*(\phi)$ for every sentence in $\mathcal{L}_0$ (analogously we have for truth extension $\preceq$ as well).

A valuation v^* will be called closed if:

1. $v^*(\wedge_i \phi_i) \geq_k \wedge_i v^*(\phi_i)$.
2. (the negation of logic syntax is that of bilattice): $v^*(\neg \phi) = \neg v^*(\phi)$.
3. (truth preservation of entailment $\models_{v^*}$): if $\phi \models_{v^*} \psi$, then $v^*(\phi) \leq v^*(\psi)$.

The smallest *nontrivial* bilattice is Belnap's 4-valued bilattice [3], $X = \mathcal{B}_4 = \{t, f, \perp, \top\}$ in Fig.1, with four truth-values in X where t is *true*, f is *false*, $\top$ is inconsistent (both true and false) or *possible*, and $\perp$ is *unknown* truth-value. In what follows we denote by $x \bowtie y$ two unrelated elements in X (so that not $(x \leq y$ or $y \leq x)$). So, Belnap's bilattice is composed by the truth lattice $(X, \leq)$ and knowledge lattice $(X, \leq_k)$, with

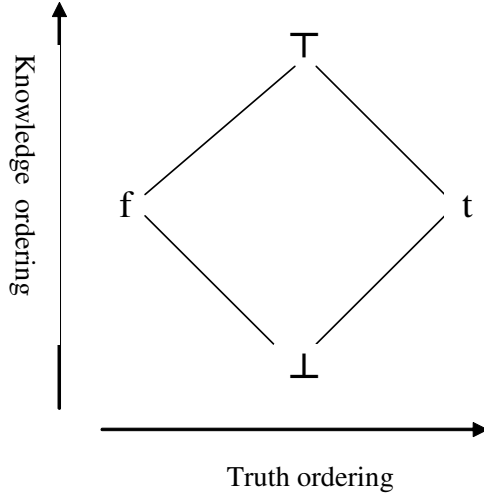


Fig. 1. Belnap's bilattice

two natural orders: *truth* order, $\leq$, and *knowledge* order, $\leq_k$, such that $f \leq \top \leq t$, $f \leq \perp \leq t$, $\perp \bowtie_t \top$ and $\perp \leq_k f \leq_k \top$, $\perp \leq_k t \leq_k \top$, $f \bowtie_k t$. That is, bottom element for $\leq$ ordering is f , and for $\leq_k$ ordering is $\perp$, and top element for $\leq$ ordering is t , and for $\leq_k$ ordering is $\top$. Meet and join operators under $\leq$ are denoted $\wedge$ and $\vee$; they are natural generalizations of the usual conjunction and disjunction notions. Meet and join under $\leq_k$ are denoted $\otimes$ and $\oplus$, such that hold: $f \otimes t = \perp$, $f \oplus t = \top$, $\top \wedge \perp = f$ and $\top \vee \perp = t$.

In [5] we proposed a family of intuitionistic bilattices with full truth-knowledge duality (D-bilattices) for a logic programming: the simplest of them, based on intuitionistic truth-functionally complete extension of Belnap's 4-valued bilattice, can be used in paraconsistent programming, that is, for knowledge bases with incomplete and inconsistent information.

There is a close relationship between logic programming and inductive definitions which are forms of *constructive* knowledge [6,7]. Constructive information defines a collection of facts through a constructive process of iterating a recursive recipe. This recipe defines new instances of this collection in terms of presence (and sometimes the absence) of other facts of the collection. In the context of mathematics, constructive information appears by excellence in inductive definitions. Not a coincidence, inductive definitions have been studied in constructive mathematics and intuitionistic logic, in particular in the sub-areas of Inductive and Definition logics, Iterated Inductive Definition logics and Fixpoint logics.

So, we adopt the idea of using relative pseudo complements as implications, which has been proposed in the literature several times and is nowadays well understood, for distributive bilattices also. In fact, the use of relative pseudo-complements is adopted in the fuzzy, multi-adjoint and residuated logic programming and allows also the assignment of weights to rules. Consequently, here one particular family of *intuitionistic* bilattices, that is, distributive bilattices extended by logic implication $\Rightarrow$, defined as a

relative pseudo complement $x \Rightarrow y = \bigvee \{z \mid z \wedge x \leq y\}$ (here $\wedge$ is the meet operation, $\leq$ is the truth-ordering of a bilattice and $\bigvee$ is the l.u.b). It has nice mathematical properties [8,9,10]. Semantically, the basic notion is that of relative pseudo complement from the algebraic semantics of intuitionistic logic [11]. More over, such an implication can be used for *nested* implications in bodies of rules, in order to extend the expressive power of logic programming (see, for example, in [12]). For Belnap's bilattice we obtain the following table for such logic implication:

$\Rightarrow$	t	$\perp$	f	$\top$
t	t	$\perp$	f	$\top$
$\perp$	t	t	$\top$	$\top$
f	t	t	t	t
$\top$	t	$\perp$	$\perp$	t

(1)

The truth lattice will be used for the many-valued many-sorted intensional FOL (denoted by $IFOL_B$) while the knowledge lattice can be used for estimation of the rate of growing the robots knowledge in time, as result of robot's experience and deductions. In order to simplify presentation we will use the same symbols of logic connectives of $IFOL_B$ as that of Belnap's bilattice for conjunction, disjunction, implication and negation. So, the conjunction and disjunction of $IFOL_B$ are consequently the meet and join operators under $\leq$, denoted $\wedge$ and $\vee$. For the implication $\Rightarrow$ of $IFOL_B$ we can use the *relative pseudocomplements* defined above, while for the negation, the bilattice involution mapping $\neg$ (which is *aparaconsistent* negation that reverses the $\leq$ ordering, while preserving the $\leq_k$ ordering)¹: switching f and t , leaving $\perp$ and $\top$.

The many-valued equivalence operator is defined for any two $a, b \in X$ by:

$$a \Leftrightarrow b = \begin{cases} t, & \text{if } a = b \\ f, & \text{otherwise} \end{cases} \quad (2)$$

In what follows we will consider only conservative homomorphic many-valued extensions of the classic 2-valued logic with logical connectives $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$ (negation, conjunction, disjunction, implication and equivalence respectively) and of the following algebra of truth values in the lattice $(X, \leq)$:

Definition 3. For a given many-valued predicate logic, by $\mathcal{A}_B = (X, \leq, \neg, \wedge, \vee, \Rightarrow, \Leftrightarrow)$ we denote the algebra of the Belnap's truth values in the complete lattice $(X, \leq)$.²

The fact that classic 2-valued algebra is the subalgebra of this many-valued algebra X , means that X is a conservative extension of the classic logic operators. We denote by $\mathbf{2} = \{f, t\} \subseteq X$ the classical 2-valued logic lattice, which is the sublattice of the Belnap's truth-ordering lattice $(X, \leq)$ composed only by false and true truth-values $\{f, t\}$. It is easy to verify that Definition 9 is valid also for classical FOL and its 2-valued logic lattice $\mathbf{2}$. However, the Belnap's bilattice has also the knowledge ordering which

¹ Differently from it, the negation used as pseudocomplement, $\neg_t x = x \Rightarrow f$, reverses both orderings.

² The classic 2-valued algebra $\mathcal{A}_2$ is particular case with lattice $\mathbf{2} = \{f, t\}$ and if $\mathbf{2} \subset X$ (the top and bottom elements in X).

is very useful for estimation of the robot’s knowledge, because for a robot, as default any ground atom has the unknown truth-value $\perp$.

Remark: It has been demonstrated by Lemma 1 in [5] that the unary operator of negation $\neg$ in the Belnap’s algebra $\mathcal{A}_B$ is the selfadjoint modal operator w.r.t the knowledge-ordering $\leq_k$, and antitonic auto-homomorphism of the truth-ordering lattice, $\neg : (X, \leq, \wedge, \vee) \rightarrow (X, \leq, \wedge, \vee)$, which represents the De Morgan laws. So, the algebra of truth values $\mathcal{A}_B$ is a *modal* Heyting algebra (that is, a standard Heyting algebra extended by unary modal operator).

□

The central hypothesis of cognitive science is that thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures. Most work in cognitive science assumes that the mind has mental representations analogous to computer data structures, and computational procedures similar to computational algorithms.

Neuro-symbolic AI attempts [13,14,15,16,17] to integrate neural and symbolic architectures in a manner that addresses strengths and weaknesses of each, in a complementary fashion, in order to support robust strong AI capable of reasoning, learning, and cognitive modeling. In this paper we consider the Intensional First Order Logic (IFOL) as a symbolic architecture of modern robots, able to use natural languages to communicate with humans and to reason about their own knowledge with self-reference and abstraction language property. After publication of my book [18] in 2022 for such IFOL, recently I applied the standard 2-valued version of it [19,20,24] to the four-level cognitive structure of AGI robots here adopted for Belnap’s 4-valued bilattice.

It requires the conceptual PRP structure levels based on domain $\mathcal{D}$ composed by particulars in D_0 and universals (concepts) in $D_I = D_1 + D_2 + D_3 + \dots$. The universals in D_I is made of: D_1 for many-valued logic sentences, as described in Section 2 of [24], called *L-concepts* (their extension corresponds to some logic value), and $D_n, n \geq 2$, for *concepts* (their m-extension is an n-ary relation); we consider the property (for an unary predicate) as a concept in D_2 .

We are able to incorporate the emotional structure to robots as well, by a number of fuzzy-emotional partial mappings

$$E_i : \mathcal{D} \rightarrow [0, 1] \quad (3)$$

of robots PRP intensional concepts, for each kind of emotions $i \geq 1$: love, beauty, fear, etc. It was demonstrated [18] that IFOL is able to include any kind of many-valued, probabilistic and fuzzy logics as well.

Despite the best efforts over the last years, deep learning is still easily fooled [33], that is, it remains very hard to make any guarantees about how the system will behave given data that departs from the training set statistics. Moreover, because deep learning does not learn causality, or generative models of hidden causes, it remains *reactive*, bound by the data it was given to explore [21].

In contrast, brains act proactively and are partially driven by endogenous curiosity, that is, an internal, epistemic, consistency, and knowledge-gain-oriented drive. We learn from our actively gathered sensorimotor experiences and form conceptual, loosely hierarchically structured, compositional generative predictive models.

The robot’s internal neuro-symbolic knowledge structure [19,20] is divided into four levels, in ordering: natural language $\mathcal{N}\mathcal{L}_{list}$ (parsed into conceptual structures with PRP concepts, i.e., conceptual C-structures³ and formal IFOL logic with the set of formulae $\mathcal{L}$), intensional conceptual PRP system and neuro system, as represented in the figure in Appendix (Section 7).

That means, for example, that we can use neural networks for robot’s transformation of the spoken phrases or visual contents presented to robots into written language phrases. For example, we can use the end-to-end learning paradigm and design neural network architectures for the tasks of image-sentence matching, image captioning tasks and image region-annotation as indicating by the PhD thesis of Andrej Karpathy in 2016 [22] with novel techniques and algorithms in the neural networks modeling paradigm which offers superior results and multiple appealing practical properties by introducing the dense captioning task, which requires the computer to both detect and describe all salient regions of an image.

To be grounded, the symbol system of $IFOL_B$ would have to be augmented with nonsymbolic, sensorimotor capacities—the capacity to interact autonomously with that world of objects, events, actions, properties and states that their symbols are systematically interpretable (by us) as referring to. It would have to be able to pick out the referents of its symbols, and its sensorimotor interactions with the world would have to fit coherently with the symbols’ interpretations.

Note that, while the top line in the figure in Appendix (Section 7) is the ordinary component of the natural language grounding developed by robot’s neuro system, the two lines bellow is the new robots knowledge structure of the added *symbolic AI system* based on the Intensional First Order Logic and its grounding to robot’s processes (its neuro AI system), by which the robot is able to provide logic deductive operations and autoepistemic self-reasoning about its current knowledge states and communicate it to humans by using natural languages.

Thus, the robot-s cognitive system in the figure in Appendix (section 7) combines both approaches to the linguistics: the *rationalist* Chomskyan (or generative) linguistics approach which depends on *categorical* principles, which sentences either do or do not satisfy, and the *empiricist* approach (of the British linguist J.R. Firth, for example) which is in favor to the *statistical* linguistics (probabilistic, actually implemented in LLM by neural networks, for example), better at automatic learning (knowledge induction), better at disambiguation, and also have a role in the science of linguistics. The rationalist approach postulates that the key parts of language are innate-hardwired in the brain at birth as part of the human genetic inheritance, mathematically modeled by

³ Following the development of advanced neural network techniques, especially the Seq2Seq model, and the availability of powerful computational resources, neural semantic parsing started emerging. Deep semantic parsing attempts to parse natural language utterances, typically by converting them to a formal meaning representation language which in our case are the expressions of the algebra of PRP concepts so defining the intensional (partial) mapping $I : \mathcal{L} \rightarrow D_I$ (represented completely by Frege/Russel semantics diagram of Corollary 3 in [25] as well. Good candidate for conceptual C-structures are AMR (Abstract Meaning Representation, and its improved models, with AMR graphs which are rooted, labeled, directed, acyclic graphs (DAGs), comprising whole sentences, by mapping their nodes and leaves into PRP intensional concepts and particulars of $IFOL_B$.

formal symbolic logics, the knowledge databases and reasoning systems (knowledge deduction). Differently, statistical linguistics draws from the work of Shannon, where the aim is to assign probabilities to linguistic events, so that we can say which sentences are usual and unusual. With this approach we can learn the complicated and extensive structure of language by specifying an appropriate general language model, and then inducing the values of parameters by applying statistical, pattern recognition, and machine learning methods to a large amount of language use.

Next Section is dedicated to a short description of the Many-sorted Intensional First-order logic $IFOL_B$ based on the Belnap's 4-valued bilattice of truth-values with the set of variables $\mathcal{V}$ and domain $\mathcal{D}$, provided recently in [25]).

2 Introduction to 4-valued Many-sorted Intensional FOL

The significant aspect of an expression's meaning is its *extension*. We can stipulate that the extension of a sentence is its truth-value, and that the extension of a singular term is its referent. The extension of other expressions can be seen as associated entities that contribute to the truth-value of a sentence in a manner broadly analogous to the way in which the referent of a singular term contributes to the truth-value of a sentence.

The first conception of *intensional entities* (or concepts) is built into the *possible-worlds* treatment of Properties, Relations and Propositions (PRP)s. This conception is commonly attributed to Leibniz, and underlies Alonzo Church's alternative formulation of Frege's theory of senses ("A formulation of the Logic of Sense and Denotation" in Henle, Kallen, and Langer, 3-24, and "Outline of a Revised Formulation of the Logic of Sense and Denotation" in two parts, Nous, VII (1973), 24-33, and VIII, (1974), 135-156). This conception of PRPs is ideally suited for treating the *modalities* (necessity, possibility, etc..) and to Montague's definition of intension of a given virtual predicate $\phi(x_1, \dots, x_k)$, as a mapping from possible worlds into extensions of this virtual predicate. Among the possible worlds we distinguish the *actual* possible world⁴.

The second conception of intensional entities is to be found in Russell's doctrine of logical atomism. In this doctrine it is required that all complete definitions of intensional entities be finite as well as unique and non-circular: it offers an *algebraic* way for definition of complex intensional entities from simple (atomic) entities (i.e., algebra of concepts), conception also evident in Leibniz's remarks. In a predicate logics, predicates and open-sentences (with free variables) expresses classes (properties and relations), and sentences express propositions. Note that classes (intensional entities) are *reified*, i.e., they belong to the same domain as individual objects (particulars). This endows the intensional logics with a great deal of uniformity, making it possible to manipulate classes and individual objects in the same language.

In what follows any open-sentence, a formula $\phi(\mathbf{x})$ with non empty tuple of free variables $\mathbf{x} = (x_1, \dots, x_m)$, will be called a m-ary *virtual predicate*⁵, denoted also by

⁴ For example, if we consider a set of predicates, of a given Database, and their extensions in different time-instances, then the actual possible world is identified by the current instance of the time.

⁵ In intensional FOL we use virtual predicates so that their extension for a given interpretation generates a single m-ary relation as in the case of standard m-ary predicates used in FOL

$\phi(x_1, \dots, x_m)$. This definition contains the precise method of establishing the *ordering* of variables in this tuple:

Definition 4. VIRTUAL PREDICATES: Virtual predicate obtained from an open formula $\phi \in \mathcal{L}$ is denoted by $\phi(x_1, \dots, x_m)$ where $(x_1, \dots, x_m)$ is a particular fixed sequence of the set of all free variables in ϕ . This definition contains the precise method of establishing the ordering of variables in this tuple: such a method that will be adopted here is the ordering of appearance, from left to right, of free variables in ϕ . This method of composing the tuple of free variables is unique and canonical way of definition of the virtual predicate from a given open formula.

Virtual predicates are used to build the *semantic logic structures* of logic-semantics level of any given natural language. However, with virtual predicates we need to replace the general FOL quantifier on variables $(\exists x)$ by specific existential quantifiers $\exists_i$ of the intensional logic $IFOL_B$, where $i \geq 1$ is the position of variable x inside a virtual predicate. For example, the intensional FOL formula $(\exists x_k)\phi(x_i, x_j, x_k, x_l, x_m)$ will be mapped into intensional concept $\exists_3\phi(\mathbf{x})$ where $\mathbf{x}$ is the list(tuple) of variables $(x_i, x_j, x_k, x_l, x_m)$. In the same way we introduce the set of universal quantifiers $\forall_i$. Notice that in many-valued version of intensional FOL the composed operation $\neg\exists_i\neg$ is different from $\forall_i$. Moreover with virtual predicates we need to replace standard binary logic connectives $\wedge, \vee, \Rightarrow$ with $\wedge_S, \vee_S, \Rightarrow_S$ where S is a set of pair of indices of variables in virtual predicate. So we obtain a non standard FOL syntax algebra with such expressions that use these new connectives.

We assume, as in [18], that a concept algebra of many-valued logic $IFOL_B$ has a non empty domain (it is different from the domain of 2-valued standard IFOL) $\mathcal{D} = D_0 + D_I$, (here $+$ is a disjoint union) where a subdomain D_0 is made of particulars or individuals (we denote by $\mathbb{R}$ the non-meaning individuals, for interpretation of language entities that are no meaningful, as "Unicorn" for example) with $X \subseteq D_0$.

The rest $D_I = D_1 + D_2 \dots + D_n \dots$ is made of universals (concepts): D_1 for many-valued logic sentences, as described in Section 2 of [24], called *L-concepts* (their extension corresponds to some logic value), and $D_n, n \geq 2$, for *concepts* (their m-extension is an n-ary relation); we consider the property (for an unary predicate) as a concept in D_2 . The concepts in $\mathcal{D}_I$ are denoted by $u, v, \dots$, while the values (individuals) in D_0 by $a, b, \dots$

Definition 5. M-SORTED EXTENSIONS:

We define the set of all *m*-extensions in the many-sorted framework of universals in D_I ,

$$\mathfrak{Rm} = \tilde{X} \bigcup \left\{ R \in \bigcup_{n \geq 1} \mathcal{P}(\mathcal{D}^n \times (X \setminus \{\perp\})) \mid (u_1, \dots, u_n, a), (u_1, \dots, u_n, b) \in R \text{ implies } b = a \right\} \quad (4)$$

for atomic formulae with $m \geq 1$ free variables. Each virtual predicate generates a *m*-ary *intensional concept* in domain $\mathcal{D}$ for a given intensional interpretation as standard *m*-ary predicates. Most simple syntactically virtual predicates are obtained from atoms, obtained from *k*-ary predicate letters $p_i^k \in P$, with non empty set of variables and non empty set of constants for predicate arguments as, for example $p_i^6(t_1, t_2, x_1, t_3, x_2, t_4)$ with ground terms (or constants), t_m for $1 \leq m \leq 4$, generates a virtual predicate $\phi_i(x_1, x_2)$. Virtual predicate of an atom with all its arguments represented by variables is just equal to such atom.

so that each $(n+1)$ -ary relation is a graph of a function, and by $\mathfrak{Rm}_k$, $k \geq 1$, we will denote the subset of all k -ary relations in $\mathfrak{Rm}$, and by $\emptyset$ each empty relation in $\mathfrak{Rm}$, with the set of unary relations in $\tilde{X}$,

$$\tilde{X} = \{\{a\} \mid a \in X \text{ and } a \neq \perp\} \cup \{\emptyset\} = \{\emptyset, \{f\}, \{\top\}, \{t\}\} \quad (5)$$

where $\{a\}$ denotes unary relation with unique tuple equal to the truth-value a in X and $\emptyset$ unary empty relation, and hence, $\emptyset \in \tilde{X} \subset \mathcal{P}(X)$. So, $(\tilde{X}, \preceq)$ is complete lattice with total ordering $\emptyset \preceq \{f\} \preceq \{\top\} \preceq \{t\}$, that is, with the bottom element and $\{a\} \preceq \{b\}$ iff $a \leq b$ (w.r.t the truth-ordering in X)⁶.

Note that $\mathfrak{Rm}$ contains only *known* extensions of the many-valued concepts, and this explains also why we defined for $\perp \in X$ the empty set $\emptyset$.

We define the following partial ordering $\preceq$ for the m -extensions in $\mathfrak{Rm}$ (note that for a given k -ary relation with $k \geq 2$ and $1 \leq i \leq k$, the $\pi_i(R)$ is the i -th projection of R while $\pi_{-1}(R)$ a relation obtained from R by eliminating i -th column of it):

Definition 6. EXTENSIONAL PARTIAL ORDER IN $\mathfrak{Rm}$:

We extend the ordering of unary relations in $\tilde{X} \subset \mathfrak{Rm}$ by: for any two nonempty m -relations $R_1, R_2 \in \mathfrak{Rm}$ with arity $k_1 = ar(R_1)$, $k_2 = ar(R_2)$ and $m = \max(k_1, k_2) \geq 2$, $\emptyset \preceq R_i$ for $i = 1, 2$, and

$$R_1 \preceq R_2 \text{ iff}$$

for each $(u_1, \dots, u_{m-1}, a) \in Ex(R_1, m)$, $\exists (u_1, \dots, u_{m-1}, b) \in Ex(R_2, m)$ with $a \leq b$, where this expansion-mapping $Ex : \mathfrak{Rm} \rightarrow \mathfrak{Rm}$ is defined as follows for any nonempty $R \in \mathfrak{Rm}$:

$$Ex(R, m) = \begin{cases} \mathcal{D}^{m-k} \times \{a\}, & \text{for } R = \{a\}, a \in X \setminus \{\perp\} & \text{if } m > k = 1 \\ \bigcup \{ \{(u_1, \dots, u_k)\} \times \mathcal{D}^{m-k} \times \{a\} \mid (u_1, \dots, u_k, a) \in R \}, & \text{if } m > k > 1 & \\ R, & \text{otherwise} & \end{cases} \quad (6)$$

We denote by $R_1 \simeq R_2$ iff $R_1 \preceq R_2$ and $R_2 \preceq R_1$, the equivalence relation between m -extensions.

We chose $\emptyset, \{t\} \in \tilde{X} \subset \mathfrak{Rm}$ to be the representative elements of the bottom and top equivalence classes in $\mathfrak{Rm}$ relatively (determined by $\simeq$), so that $(\forall R \in \mathfrak{Rm})(\emptyset \preceq R \preceq \{t\})$. Consequently, the *extensional* ordering $\preceq$ in $\mathfrak{Rm}$ extends the ordering of the lattice $(\tilde{X}, \preceq)$.

⁶ However, if we consider the inconsistent truth-value $\top$ as both true and false, that is, as set $\{f, t\}$ and unknown truth-value as empty set $\emptyset$, then $\tilde{X} = \{\emptyset, \{f\}, \{f, t\}, \{t\}\}$ is the powerset $\tilde{X} = \mathcal{P}(\{f, t\}) = \mathcal{P}(2)$, we obtain the following isomorphism with *knowledge-ordering* lattice $(X, \leq_k)$ of Bednap's bilattice and the complete powerset lattice $(\tilde{X}, \subseteq)$,

$$is : (X, \leq_k) \simeq (\tilde{X}, \subseteq),$$

such that $is(\perp) = \emptyset$, $is(f) = \{f\}$, $is(t) = \{t\}$ and $is(\top) = \{f, t\}$, from which we see the natural correspondence of $\tilde{X}$ with knowledge ordering and hence with knowledge extensions.

We introduce for the concepts in D_I an *extensional interpretation* h which assigns the *m-extension* to each intensional concept in $\mathcal{D}$, and can be considered as an *interpretation of concepts* in $\mathcal{D}$. Thus, each concept in D_I represents a set of tuples in $\mathcal{D}$, and can be also an element of the extension of another concept and of itself also. Each extensional interpretation h assigns to the intensional elements of $\mathcal{D}$ an appropriate extension: in the case of particulars $u \in D_0$, $h_0(u) \in D_0$, such that for each logic value $a \in X \subset D_0$, $h_0(a) = a$. Thus, we have the particular's mapping $h_0 : D_0 \rightarrow D_0$ and more generally (here $' + '$ is considered as disjoint union),

$$h = \sum_{i \in \mathbb{N}} h_i : \mathcal{D} \longrightarrow D_0 + \sum_{i \geq 1} \mathfrak{Rm}_i \quad (7)$$

where $h_1 : D_1 \rightarrow \tilde{X}$ assigns to each L-concept $u \in D_1$ (of a sentence with truth-value $a \in X$), a relation composed by the single tuple $h_1(u) = \{a\}$ if $a \neq \perp$, $\emptyset$ otherwise, and $h_i : D_i \rightarrow \mathfrak{Rm}_i$, for $i \geq 2$, that assigns a m-extension to non-sentence concepts. The extension of the *self-reference* truth-concept $u_T \in D_2$, for any extensionalization function h is given by $h(u_T) = \{(u, a) \mid u \in D_1, \emptyset \neq h(u) = \{a\}\}$. This concept represents the truth of all logic-concepts in D_1 , in the way that for any L-concept $u \in D_1$ we have that $h(u) = \{a\}$ iff $(u, a) \in h(u_T)$.

In what follows, this logic $IFOL_B$ will be shortly denoted by $\mathcal{L}_{in}$ and the set of all its formulae by $\mathcal{L}$.

Definition 7. SYNTAX OF MANY-VALUED INTENSIONAL FOL $\mathcal{L}_{in}$:

We define the syntax of 4-valued many-sorted first-order logic $\mathcal{L}_{in}$, with complete Belnap's lattice of truth-values $(X, \leq)$, by:

- Variables $x, y, z, \dots$ in $\mathcal{V}$;
- Language constants $c, d, \dots$ are considered as nullary functional letters (all non nullary functional letters are represented as the predicate letters for graphs of these functions);
- Predicate letters in P , denoted by $p_1^{k_1}, p_2^{k_2}, \dots$ with a given arity $k_i \geq 1$, $i = 1, 2, \dots$. Nullary predicate letters are considered as propositional letters, as for example the built-in predicate letters $\{p_a \mid a \in X\}$;
- The 4-valued logic connectives: 1. unary connectives are the negation $\neg$, and existential an universal quantifiers $\exists_i, \forall_i$.
- 2. binary connectives in $\odot \in \{\wedge, \vee, \Rightarrow, \Leftrightarrow\}$ are respectively logic conjunction, disjunction, implication and equivalence. From the fact that we are using the virtual predicates that represent any formula composed by these connectives as a specific virtual predicate with free variables obtained as the union of the free variables of all its standard predicates in P that compose this formula, for the syntax algebra $\mathcal{AB}_{FOL}$ of $\mathcal{L}_{in}$ we need to use for each binary logic operator in $\odot$ (different from $\Leftrightarrow$) an algebraic operator $\odot_S$ where the pairs of indexes in the set S indicate the equal free variables used in two subformulae ψ_1 and ψ_2 that compose a resulting virtual predicate $\phi(x)$ equal to formula $\psi_1 \odot_S \psi_2$ as it is explained⁷ in [25].

⁷ For example, the FOL formula $\psi_1(x_i, x_j, x_k, x_l, x_m) \wedge \psi_2(x_l, y_i, x_j, y_j)$ will be replaced by a resulting virtual predicate $\phi(x_i, x_j, x_k, x_l, x_m, y_i, y_j)$ defined by algebraic expression $\psi_1(x_i, x_j, x_k, x_l, x_m) \wedge_S \psi_2(x_l, y_i, x_j, y_j)$, with $S = \{(4, 1), (2, 3)\}$, and then traduced by the algebraic expression $R_1 \bowtie_S R_2$ (see point 2.5 of Definition 8 in [25]) where $R_1 \in$

- Abstraction operator $\langle\!\langle_\rangle\!\rangle$, and punctuation symbols (comma, parenthesis).

With the following simultaneous inductive definition of terms and formulae:

1. All variables and constants (0-ary functional letters in P) are terms. We denote by p_a a logic constant (built-in 0-ary predicate symbol) for each truth value $a \in X$, so that a set of constants is a not empty set.
2. If $t_1, \dots, t_k$ are terms, then $p_i^k(t_1, \dots, t_k)$ is a formula ($p_i^k \in P$ is a k -ary predicate letter).
3. In what follows any open-formula $\phi(\mathbf{x})$ with non empty tuple of free variables $\mathbf{x} = (x_1, \dots, x_m)$, will be called a m -ary virtual predicate, denoted also by $\phi(x_1, \dots, x_m)$ and provided in Definition 4.
If ϕ and ψ are formulae, then $\neg\phi$, $(\forall_k x)\phi$, $(\exists_k x)\phi$ and, for each connective $\odot$, if binary $\phi \odot \psi$ or if unary $\odot\phi$, are the formulae. In a formula $(\exists_k)\phi(\mathbf{x})$ (or $(\forall_k)\phi(\mathbf{x})$), the virtual predicate $\phi(\mathbf{x})$ is called "action field" for the quantifier $(\exists_k)$ (or $(\forall_k)$) of the k -th free variable in the tuple $\mathbf{x}$. A variable y in a formula ϕ is called bounded variable iff it is the variable quantified by $(\exists_k)$ (or $(\forall_k)$). A variable x is free in $\psi(\mathbf{x})$ if it is not bounded. A sentence is a closed-formula having no free variables.
4. If $\phi(\mathbf{x})$ is a formula and $\alpha \subseteq \bar{\mathbf{x}}$ is a possibly empty subset of hidden (compressed) variables, then $\langle\!\langle\phi(\mathbf{x})\rangle\!\rangle_\alpha^\beta$ is an abstracted term, where β is remained subset of free visible variables in ϕ . So, the subtuples of hidden and visible variables (preserving the ordering of the tuple $\mathbf{x}$ are $\pi_{-\beta}\mathbf{x}$ and $\pi_{-\alpha}\mathbf{x}$, respectively). If α or β is empty sequence, than it can be omitted (for example, if ϕ is closed formula, then this term is denoted by $\langle\!\langle\phi\rangle\!\rangle$).

An occurrence of a variable x_i in a term $\langle\!\langle\phi(\mathbf{x})\rangle\!\rangle_\alpha^\beta$ is bound if $x_i \in \alpha$, free if $x_i \in \beta$, so that the variables in α are not subjects of assignment $g \in \mathcal{D}^V$ and can not be quantified by existential and universal FOL quantifiers.

In particular we introduce this subset of distinguished predicates:

- a The binary predicate letter $p_1^2 \in P$ is singled out as a distinguished logical predicate and formulae of the form $p_1^2(t_1, t_2)$ are to be rewritten in the form $t_1 = t_2$. It is a built-in 2-valued⁸ predicate for the identity.
- b The binary predicate letter $p_2^2 \in P$ is singled out as a distinguished 2-valued logical predicate and formulae of the form $p_2^2(t_1, t_2)$ are to be rewritten in the form $t_1 =_{in} t_2$. It is a built-in predicate for the weak-intensional-equivalence.
- c The unary many/valued predicate T and $Know$ for robot's own knowledge⁹ management. The basic argument of them is an abstracted term obtained from a given

$\mathcal{P}(\mathcal{D}^5), R_2 \in \mathcal{P}(\mathcal{D}^4)$ are the extensions for a given many-valued interpretation v^* of the virtual predicate ψ_1, ψ_2 relatively. In this example the resulting relation will have the following ordering of attributes of obtained virtual predicate $\phi: (x_i, x_j, x_k, x_l, x_m, y_i, y_j)$.

In the case when S is empty (i.e. its cardinality $|S| = 0$) then the resulting relation is the Cartesian product of R_1 and R_2 .

⁸ For "2-valued" we intend the subset of standard truth-values $\mathbf{2} = \{f, t\} \subset X$.

⁹ In this paper we propose the minimal version of the predicate $Know \in P$ with first argument dedicated to time-specification of obtained knowledge, second dedicated for the subject of this knowledge and the last argument is abstracted term of the knowledge sentence: the truth-value of a ground atom of this predicate represents the truth-value of the known sentence contained in third argument.

sentence, so that the truth value of their ground atoms reflects the truth/value of reified sentence in them. Consequently their interpretation has to satisfy the particular knowledge-constraints provided in (16).

We denote by $\mathcal{L}$ the set of all formulae of $\mathcal{L}_{in}$ and by $\mathcal{L}_0$ the subset of its sentences.

Notice that "built-in" is used for predicate letters that have fixed invariant interpretation, and consequently fixed invariant extension: each possible interpretation of $\mathcal{L}_{in}$ has to satisfy this constraint for built-in predicates, so that different many-valued interpretations can be used only for the remaining set of predicate letters in P . The T is the many-valued version of the truth-predicate in the standard 2-valued logic where a formula $T(\langle\phi\rangle)$ is true iff the (closed) sentence ϕ is true.

The intensional interpretation $I : \mathcal{L} \rightarrow D_I$ maps each nullary predicate symbol into D_1 , and for each logic constant $p_a, a \in X$, we have that $I(p_a) = u_a \in D_1$. The language constants (nullary functional symbols in P) are mapped in concepts in $\mathcal{D}$. Notice that for any non-meaningful language constant (as "Unicorn" for example) we have that $I(c) = \mathbb{R} \in D_0$.

Consequently, each atom $p_i^k(t_1, \dots, t_k), p_i^k \in P$, is mapped into the concept $u \in D_{m+1}$, where m is a number of the free variables of the virtual predicate obtained from this atom. Consequently, each ground atom $p_i^k(t_1, \dots, t_k)$ is mapped into D_1 , and if there is any $t_i, 1 \leq i \leq k$, such that $I(t_i) = \mathbb{R}$, then $I(p_i^k(t_1, \dots, t_k)) = u_\perp$, where $\perp \in X$ such that $\neg\perp = \perp$ is the logic truth-value 'unknown'. In this way we guarantee that unmeaningful sentences as $blu(Unicorn)$ will always have the 'unknown' logic truth-value. This intensional interpretation can be given also to all contradictory formulae (with truth-value $\top \in X$) that can not be nor true nor false, as the Liars paradoxes.

The main difference with standard FOL syntax is that here we can use abstracted terms obtained from logic formulae, for example, "x believes that ϕ " is given by formula $p_i^2(x, \langle\phi\rangle)$ (where p_i^2 is binary "believe" predicate).

Let $\phi(\mathbf{x})$ be any well-formed virtual predicate in $\mathcal{L}_{in}$, then $\langle\phi(\mathbf{x})\rangle_\alpha^\beta / g$ with the set of hidden variables $\alpha = (x_1, \dots, x_m), m \geq 1$, is the ground term (the assignment g is applied only to free visible variables in $\beta = \bar{\mathbf{x}} - \alpha$ whose semantics correlate is an intensional entity (concept) of degree m . If $m = 0$, the intensional correlate of this singular ground term is the proposition (sentence) "that ϕ "; if $m = 1$, the intensional correlate is the property of "being something x_1 such that ϕ "; if $m > 1$, then the intensional correlate is the concept "the relation among $x_1, \dots, x_m$ such that ϕ ".

Certain complex nominative expressions (namely, gerundive and infinitive phrases) are best represented as singular terms of the sort provided by our generalized bracket notation $\langle\phi(\mathbf{x})\rangle_{x_1, \dots, x_m}^\beta$, where $m \geq 1$. This MV-intensional logic $\mathcal{L}_{in}$ differs from two-valued intensional FOL in heaving these singular terms $\langle\phi(\mathbf{x})\rangle_{x_1, \dots, x_m}^\beta$ where the ground atoms $\phi(\mathbf{x})/g$ are many-valued.

Definition 8. An assignment $g : \mathcal{V} \rightarrow \mathcal{D}$ for variables in $\mathcal{V}$ in $IFOL_B$ is applied only to free variables in terms and formulae. Such an assignment $g \in \mathcal{D}^\mathcal{V}$ can be recursively uniquely extended into the assignment $g^* : \mathcal{T} \rightarrow \mathcal{D}$, where $\mathcal{T}$ denotes the set of all terms, by:

1. $g^*(t_i) = g(x) \in \mathcal{D}$ if the term t_i is a variable $x \in \mathcal{V}$.

2. If the term t_i is a constant $c \in F$ then $g^*(t_i) \in \mathcal{D}$ as its tarskian interpretation.
3. If a term t_i is $f_i^k(t_1, \dots, t_k)$, where $f_i^k \in F$ is a k -ary functional symbol and $t_1, \dots, t_k$ are terms, then $g^*(f_i^k(t_1, \dots, t_k))$ is the value $u \in \mathcal{D}$ of this functions for the tuple of values in $(g^*(t_1), \dots, g^*(t_k))$ or, equivalently, in the graph-interpretation of the function, $(g^*(t_1), \dots, g^*(t_k), u) \in \mathcal{D}^{k+1}$.
4. If t is an abstracted term obtained for an open formula ϕ_i , $\langle \phi_i(\mathbf{x}_i) \rangle_{\alpha_i}^{\beta_i}$ where $\alpha_i \cup \beta_i$ are the variables in tuple $\mathbf{x}_i$, then we must restrict the assignment to $g \in \mathcal{D}^{\beta_i}$ and to obtain recursive definition (when also $\phi_i(\mathbf{x}_i)$ contains abstracted terms:

$$g^*(\langle \phi_i(\mathbf{x}_i) \rangle_{\alpha_i}^{\beta_i}) =_{\text{def}} \begin{cases} I(\phi_i(\mathbf{x}_i)) \in D_{|\alpha_i|+1}, & \text{if } \beta_i \text{ is empty} \\ I(\phi_i(\mathbf{x}_i)[\beta_i/g(\beta_i)]) \in D_{|\alpha_i|+1}, & \text{otherwise} \end{cases} \quad (8)$$

where $g(\beta) = g(\{y_1, \dots, y_m\}) = \{g(y_1), \dots, g(y_m)\}$ and $[\beta/g(\beta)]$ is a uniform replacement of each i -th variable in the set β with the i -th constant in the set $g(\beta)$. Notice that α is the set of all free variables in the formula $\phi[\beta/g(\beta)]$.

5. If $t = \langle \phi_i \rangle$ is an abstracted term obtained from a sentence ϕ_i then $g^*(\langle \phi_i \rangle) = I(\phi_i) \in D_0$.

In what follows we will use the graph-interpretation for functions in FOL like its interpretation in intensional logics. We denote by $\mathcal{L}$ the set of all formulae ϕ of the logic $IFOL_B$, and denote by t_i/g (or ϕ/g) the ground term (or formula) without free variables, obtained by assignment g from a term t_i (or a formula ϕ), and by $\phi[x/t_i]$ the formula obtained by uniformly replacing x by a term t_i in open formula $\phi(x)$.

A sentence is a (closed) formula having no free variables.

Definition 9. The Herbrand base of a logic $IFOL_B$ is defined by

$$H = \{p_i^k(t_1, \dots, t_k) \mid p_i^k \in P \text{ and } t_1, \dots, t_k \text{ are ground terms}\}.$$

Herbrand interpretations are the mappings $v : H \rightarrow X$, which must satisfy the constraints for the built-in ground atoms (any ground atom of a built-in predicates must have the same truth-value for every Herbrand interpretation) and for built-in propositional symbols p_a , for $a \in X$, $v(p_a) = a$.

We are able to define the many-valued algebraic semantics of $\mathcal{L}_{in}$, based on the standard extension of Herbrand interpretations to all sentences $\mathcal{L}_0$ of $\mathcal{L}_{in}$.

Definition 10. BELNAP'S 4-VALUED SEMANTICS OF $\mathcal{L}_{in}$:

The algebraic semantics of the 4-valued many-sorted intensional first-order logic $\mathcal{L}_{in}$ (provided by Definition 11 in [25]) can be obtained by the unique extension of a given 4-valued Herbrand interpretation $v : H \rightarrow X$ (from Definition 9), into the valuation $v^* : \mathcal{L}_0 \rightarrow X$, where $\mathcal{L}_0 \subset \mathcal{L}$ is the strict subset of all sentences (formulae without free variables), inductively as follows:

for any formula $\phi, \psi \in \mathcal{L}$ and a given assignment $g : \mathcal{V} \rightarrow \mathcal{D}$, we have that (here $\wedge$ and $\vee$ are the meet and joint operators of the lattice X , respectively)

1. $v^*(\neg\phi/g) = \neg v^*(\phi/g)$, and for built-in propositions $v^*(p_a) = v(p_a)$ for $a \in X$,
2. $v^*(\phi/g \odot \psi/g) = v^*(\phi/g) \odot v^*(\psi/g)$,
for each logic connective $\odot \in \{\wedge, \vee, \Rightarrow, \Leftrightarrow\}$, i.e., $\mathcal{A}_B$ -operators in Definition 3 over Belnap's truth-lattice X ,

3. $v^*((\exists_i)\phi)/g = v^*(\phi/g)$ if i -th variable x is not a free in ϕ ;
 $= \bigvee \{v^*(\phi/g_1) \mid g_1 \in \mathcal{D}^V \text{ such that for all } y \in \mathcal{V} \setminus \{x\}, g_1(y) = g(y)\}$ otherwise,
4. $v^*((\forall_i)\phi)/g = v^*(\phi/g)$ if i -th variable x is not a free in ϕ ;
 $= \bigwedge \{v^*(\phi/g_1) \mid g_1 \in \mathcal{D}^V \text{ such that for all } y \in \mathcal{V} \setminus \{x\}, g_1(y) = g(y)\}$ otherwise.

We denote by $\mathcal{I}_{MV} \subseteq X^{\mathcal{L}_0}$ the set of all many-valued valuations of $\mathcal{L}_{in}$ that have fixed (invariant) interpretation for each built-in predicate.¹⁰

We recall that the set-based (for infinite sets as well) of the operators $\bigwedge$ and $\bigvee$ is well defined because our many-valued logics are based on the complete (and distributive) lattices $(X, \leq)$, which satisfy these requirements.

Notice that this Belnap's 4-valued extended interpretation $v^* : \mathcal{L}_0 \rightarrow X$ defined above is *not a homomorphism*, because of the point 3 (and 4 as well), where the truth of the closed formula $(\forall x_i)\phi(x_i)$ can not be obtained from the logic value of $\phi(x_i)$ from the fact that to any formula with free variables we can not associate any logic value. Because of that, we define a new version of many-valued interpretation, denominated "MV-interpretation" (which is also a homomorphism [25] between syntax algebra of $IFOL_B$ logic and algebra of relations in $\mathfrak{Rm}$):

Definition 11. MV-INTERPRETATIONS: We define, for a valuation $v^* : \mathcal{L}_0 \rightarrow X$ of the sentences in $\mathcal{L}_0 \subset \mathcal{L}$ of the 4-valued many-sorted intensional first-order logic $\mathcal{L}_{in}$, the MV-interpretation $I_B^* : \mathcal{L}_0 \rightarrow \mathfrak{Rm}$, such that for any sentence $\phi/g \in \mathcal{L}_0$,

$$I_B^*(\phi/g) = \begin{cases} \{v^*(\phi/g)\} \in \tilde{X} \subset \mathfrak{Rm}, & \text{if } v^*(\phi/g) \neq \perp \\ \emptyset, & \text{otherwise} \end{cases} \quad (9)$$

and we define also the unique extension of I_B^* to all open formulae $\mathcal{L}$ in $\mathcal{L}_{in}$ as well, such that for any open formula (virtual predicate) $\phi(x_1, \dots, x_k) \in \mathcal{L}$,

$$I_B^*(\phi(x_1, \dots, x_k)) = \{(g(x_1), \dots, g(x_k), a) \mid g \in \mathcal{D}^V \text{ and } a = v^*(\phi/g) \neq \perp\} \in \mathfrak{Rm} \quad (10)$$

8 We denote by $\mathcal{W}$ the set of all MV-interpretations derived from the set of 4-valued interpretations $v^* \in \mathcal{I}_{MV}$ specified in Definition 10, with bijection $is_{MV} : \mathcal{I}_{MV} \simeq \mathcal{W}$ such that for any $v^* \in \mathcal{I}_{MV}$ we have that $I_B^* = is_{MV}(v^*) : \mathcal{L} \rightarrow \mathfrak{Rm}$.

For each fixed intensional interpretation $I : \mathcal{L} \rightarrow \mathcal{D}$ and valuation $v^* : \mathcal{L}_0 \rightarrow X$ in Definition 10 that respect all built-in predicates, the corresponding extensionalization function $h : \mathcal{D} \rightarrow \mathfrak{Rm}$, for any virtual predicate $\phi(x_1, \dots, x_n) \in \mathcal{L}$, satisfies the following 4-valued generalization of Tarski's FOL constraint:

$$h(I(\phi(x_1, \dots, x_n))) =_{def} \{(g(x_1), \dots, g(x_n), a) \mid g \in \mathcal{D}^V, h(I(\phi(x_1, \dots, x_n)/g)) = \{a\}, a \neq \perp\} \quad (11)$$

with $h(I(\phi(x_1, \dots, x_n)/g)) = I_B^*(\phi(x_1, \dots, x_n)/g)$, so that holds [25] the general composition $I_B^* = h \circ I$.

¹⁰ As, for example, the identity predicate $p_1^2 \in P$ or nullary predicate letters $\{p_a \mid a \in X\}$ for which it must be satisfied that for any $v^* \in \mathcal{I}_{MV}$, $v^*(p_a) = a$, $v^*(p_1^2(u_1, u_2)) = t$ iff $(u_1, u_2, t) \in R_+$, and $v^*(p_2^2(u_1, u_2)) = t$ iff $(u_1, u_2, t) \in R_{=in}$, and for any assignment g , $v^*(p_1^1(t_i)) = v^*(T(t_i)) =_{def} v^*(\phi/g)$ if term $t_i = \langle \phi/g \rangle$; f otherwise.

The difference between $\mathcal{I}_{MV}$ and the total set of Herbrand interpretations X^H is caused by the presence of the built-in predicates.

3 Closed Knowledge Assumption (CKA) and Learning

First of all, the CKA is *not* a well-known Closed World Assumption (CWA) which postulates that all sentences of a logic that are not proven to be true (the lack of knowledge) *must be false*. I argue that this conflation is often inappropriate, especially for: distributed knowledge bases, semantic web applications, probabilistic and intensional reasoning, systems with incomplete information. Instead, CKA can be summarized as:

“Only the explicitly represented knowledge is considered known; what is not represented is simply not known, rather than automatically false.”

The emphasis shifts from truth to knowledge. My work on intensional logic treats propositions as objects of reasoning. In that setting, it is natural to distinguish: whether a proposition is true, whether it is believed, whether it is known, whether its probability is defined. This distinction is much closer to epistemic logic than to classical logic programming.

The main strengths of CKA are: It separates knowledge from truth. It avoids some unintuitive consequences of CWA in incomplete domains. It aligns well with intensional and epistemic logics. It fits naturally with probabilistic and temporal reasoning (which will be introduced in $IFOL_B$ in next dedicated paper), where incomplete information is common. From a theoretical perspective, CKA is a coherent attempt to provide a cleaner epistemic foundation for reasoning under incomplete information. It is especially consistent with my broader program of treating propositions, probabilities, and temporal facts as intensional objects rather than reducing them to simple truth values. While it has had limited impact on mainstream AI systems, it addresses a genuine conceptual issue that arises when combining logic, probability, and knowledge representation.

CWA principle has been valid only for the classical 2-valued logic, where the sentences can be true or false only. In CKA we are working with the 4-valued logic with the logic value “unknown” (denoted by the symbol $\perp$) as well, so that only for such logics we can define CKA, that is, that all sentences for which we do not know their logic values, that they are unknown sentences (that is assigning to them the logic value $\perp$).

Knowledge representation formalisms are aimed to represent general conceptual information and are typically used in the construction of the knowledge base of reasoning agent (a robot in our case). A knowledge base can be thought of as representing the beliefs of such an agent.

Differently from 2-valued logic where extension of a given predicate is composed by only *true* atoms, here for MV-logic $IFOL_B$ the ground atoms of extension of this predicate can have three truth-values (*except the unknown value $\perp$ for unknown facts which are not explicitly represented*). In this way, the complete knowledge of AGI robots would correspond to the extensions of the only known n -ary *concept*, $n \geq 1$, and hence it would be very easy to compare the knowledge of different robots developed by their individual experiences supported by their deductive inference capabilities.

Remark: The logic knowledge (composed by the logic sentences and corresponding intensional L-concepts heaving as extensions only the logic truth-values) will be provided in next Section both with robot’s autoepistemic deduction.

□

Consequently, we are able to represent the whole *AGI robot's knowledge Database* of its n -ary concepts, $n \geq 1$, in each fixed instance of time (it is an analog to traditional relational Database with standard 2-valued FOL) as follows:

Definition 12. AGI ROBOT'S CURRENT ATOMIC KNOWLEDGE DATABASE¹¹:

For a given instance of time, the current AGI Robot knowledge is defined by its current MV-interpretation $\mathbf{I}_B^*$ (that represents robot's current world $w \in \mathcal{W}$) defines the extension of robot's knowledge Database $\mathcal{K}$ by:

$$\mathcal{K} = \{R_{p_i^k} = \mathbf{I}_B^*(p_i^k(x_1, \dots, x_k)) \mid p_i^k \in P, \text{ for } x_i \in \mathcal{V}, 1 \leq i \leq k\} \quad (12)$$

where $\mathbf{I}_B^*$ is the current MV-interpretation (a function from $\mathcal{L}$ to $\mathfrak{M}$) and $\mathcal{V}$ the set of variables and $R_{p_i^k}$ is the $(k+1)$ -ary relation obtained for the k -ary predicate p_i^k .

The Belnap's bilattice has both truth and the knowledge orderings. The truth ordering we used for the set of relations in $\mathfrak{M}$ in Definition 6, while the knowledge ordering is useful just for the knowledge base and the principle of CKA: Each k -ary tuple, $k \geq 2$, not in knowledge Database $\mathcal{K}$ is considered as unknown fact. Thus, the inclusion relation for robot's knowledge is just the simple set-inclusion $\subseteq$.

Consequently, the atomic knowledge database is the subset of current metaknowledge of a robot is the set of relations, that is,

$$\mathcal{K} \subset \{R \mid R \in \text{Im}(\mathbf{I}_B^*), \text{ for } \text{ar}(R) \geq 2\} \quad (13)$$

where for each tuple of ground terms, $\mathbf{d} = (t_1, \dots, t_k, a) \in R_{p_i^k} \in \mathcal{K}$, of relation with arity $k+1 \geq 2$, the current truth-value of this, for robot known, fact is equal to last value of this tuple, $a = \pi_{k+1}(\mathbf{d}) \in \{f, \top, t\}$ with truth-ordering $f < \top < t$. So, for any ground instance of each robot's (real or virtual) predicate (and corresponding intensional concept), robot is able to know the level of truth in a given instance of time. Thus, given current atomic knowledge database $\mathcal{K}$, we are able to derive from it the current Herbrand model $v : H \rightarrow X$ and its extension v^* to all sentences, for robot's knowledge by, for any ground atom $p_i^k(t_1, \dots, t_k) \in H$, with the ground terms t_i , for $1 \leq i \leq k$,

$$v(p_i^k(t_1, \dots, t_k)) = \begin{cases} a \in \{f, \top, t\}, & \text{if } (t_1, \dots, t_k, a) \in R_{p_i^k} \in \mathcal{K} \\ \perp, & \text{otherwise} \end{cases} \quad (14)$$

Thus, from current atomic knowledge database $\mathcal{K}$, we are able to derive current Herbrand model v , its extension to all sentences v^* and thence the current possible world (MV-model in Definition 12) $\mathbf{I}_B^* = h \circ I \in \mathcal{W}$.

By default the initial robot's atomic knowledge database $\mathcal{K}$ before the initial process of learning contains only the extensions of the built-in predicates (invariant w.r.t any Herbrand interpretation v . However, we intend to set this initial knowledge by some

¹¹ This definition is different from the definition in [25], in order to be able also to derive the current Herbrand model of robot's knowledge $v : H \rightarrow X$, where H is the Herbrand base for all predicates in P . Note that the autoepistemic predicate *Know* is a meta-èredicate, so that *Know* $\notin P$.

necessary quantity of knowledge which will provide enough information from which a robot will be able to obtain its own knowledge by intensive initial learning and own experience.

In the accelerating world of AI, discussions often gravitate toward its potential to surpass human intelligence and the hypothetical scenario of machines overtaking decision-making processes, possibly leading to catastrophic consequences. This fear is not entirely misplaced. Today, AI technologies such as self-driving cars and smart homes already make crucial data-based decisions without direct human intervention. However, how AI systems interpret and process the data they have often mitigates the risks of these technologies. One key element in this process is Closed World Assumption in AI.

In *2-valued logic*, the Closed World Assumption (CWA) asserts that if a fact isn't recorded as true, then it is considered false. So, the knowledge base or system has complete information about the domain, and anything not explicitly stated is assumed to be non-existent or incorrect. In particular, we drew attention to the problem of representing negative information — information about the non-existence of objects or events or processes or influences. Negative information, we argued, must sometimes be encoded by the mere absence of positive information to the contrary: propositions which are not given, or which are not deducible from those which are given, should be assumed to be false. The closed world assumption is used in many practical applications of logic (planning, prolog¹², databases¹³) in order to lower the complexity class of the reasoning algorithms used. This assumes that any predicates that are stated are true and the ones that aren't are false.

Based on these facts we are able, in analogy to the Closed World Assumption (CWA) used for Databases defined by the standard 2-valued FOL (as, for example the RDBs), we can introduce the Closed Knowledge Assumption (CKA) and to show that it is satisfied for the 4-valued many-sorted Intensional FOL.

Definition 13. CLOSED KNOWLEDGE ASSUMPTION(CKA):

The CKA for the many-sorted Intensional FOL based on Belnap's 4-valued bilattice of truth-values is defined, for every Herbrand interpretation $v : H \rightarrow X$ and assignment $g : \mathcal{V} \rightarrow \mathcal{D}$, as follows:

For each k -ary virtual predicate $\phi(x_1, \dots, x_k)$, $k \geq 1$,

$$v^*(\phi(x_1, \dots, x_k)/g) \neq \perp \text{ iff } (g(x_1), \dots, g(x_k)) \in \pi_{-i}(I_B^*(\phi(x_1, \dots, x_k))).$$

So, we obtain the following property for the robot's knowledge database $\mathcal{K}$ based on the logic $IFOL_B$:

¹² In Prolog, **Yes** means a statement is provably true. Consequently, **No** means a statement is not provably true. This only means that such a statement is false, if we assume that all relevant information is present in the respective Prolog program. For the semantics of Prolog programs we usually do make this assumption. It is called the Closed World Assumption: we assume that nothing outside the world described by a particular Prolog program exists (is true).

¹³ The Closed-World Assumption (CWA) on a database expresses that an atom not in the database is false. The CWA is only applicable in domains where the database has complete knowledge. In many cases, for example in the context of distributed databases, a data source has only complete knowledge about part of the domain of discourse.

Corollary 1 *The robot’s atomic knowledge database $\mathcal{K}$ based on the logic Belnap’s 4-valued many-sorted intensional first-order logic $IFOL_B$, representing the Herbrand many-valued interpretation $v : H \rightarrow X$ at any given instance of time, satisfies the Closed Knowledge Assumption.*

Proof: In this case the general virtual predicates are just the predicates $p_i^k \in P$ from which is defined the Herbrand base and, from $\mathcal{K}$ and (14) its Herbrand interpretation $v : H \rightarrow X$. So, for each sentence of predicate $p_i^k \in P$, for a given assignment $g \in \mathcal{D}^V$,
if $a = v(p_i^k(x_1, \dots, x_k)/g) \neq \perp$ then $(g(x_1), \dots, g(x_k), a) \in R_{p_i^k} \in \mathcal{K}$,
that is, $(g(x_1), \dots, g(x_k)) \in \pi_{-i} R_{p_i^k} = \pi_{-i}(I_B^*(p_i^k(x_1, \dots, x_k)))$
as required from CKA.

□

With this CKA property of its knowledge database $\mathcal{K}$, a robot is able to verify which of its sentences are still unknown, that is, to derive the logic truth value of every sentence of its logic language $\mathcal{L}$.

This result is analogous to the property of Closed World Assumption for RDB (Relational Databases) based on 2-valued standard FOL (First-order Logic).

By using in our knowledge 4-levels cognitive neurosymbolic structure also a neural LLM (Large Language Models), we are able to obtain the parsing from the natural language expressions into the FOL formulas, and to consider LLM for natural interactive dialog between a robot and humans (as in ChatGPT and similar applications). The exploration of the capability of neural models in parsing English sentences to FOL are provide, for example, in [39]. In next section we will provide the relationships between logic predicates and n-ary concepts which names are just the simple words or finite natural language expressions, so that we can use also the parsing of natural language expressions into into intensional expressions of these concepts as well.

Conceptual learning bridge: We consider that LLM with common statistical natural language knowledge and initial sets of predicates and corresponding concepts are part of initial robot’s knowledge. However, robot will be able to define new predicates (when with given set of predicates in P and their corresponding intensional entities (concepts) is not possible to generate a logic formula for particular natural language expressions), or modify existing predicates in P (by adding new attributes to such predicates). So, with this processes we obtain the scalability of the symbolic AI based on IFOL predicates and their domains.

The component of the neural AI learning and its scalability is given by the properties of implementation of LLM.

Learning from neural to symbolic levels: Each personal robot’s experience can be internally represented by some logic sentence.

For example, **The Labeling Process by Reification:**

To label a motor program (e.g., a routine that moves the right arm to pick up a block), the robot uses an Intensional Abstraction Operator.

1. **The Program:** Let's say the raw code for moving an arm is P_1 and motor programs be defined by true atoms of the binary predicate $MotPrg(x_1, x_2)$ where the variable $x_1 \in \mathcal{V}$ defines the row codes in neural system and $x_2 \in \mathcal{V}$ defines the labels (names) of these row codes, so that for this assignment $g : \mathcal{V} \rightarrow \mathcal{D}$ for variables in $\mathcal{V}$, $g(x_1) = P_1$ and $g(x_2) = \text{"moving right arm"}$ for which the ground atom $MotPrg(x_1, x_2)/g = MotPrg(P_1, \text{moving right arm})$ is true.
2. **The Logic Term:** The robot uses the abstraction operator $\langle _ \rangle$ to create a "name" for this code. The label becomes an intensional entity - a symbol the robot can think about without actually running the code. In fact by transformation of this logic predicate into abstracted term, we obtain (by using the intensional mapping I) that

$$u = g^*(\langle MotPrg(x_1, x_2) \rangle_{x_2}^{x_1}) = I(MotPrg(g(x_1), x_2))$$

$$= I(MotPrg(P_1, x_2)) \in D_1 \text{ is an intensional entity (an unary concept) such that}$$
 its extension (for extensionalization function h) is just a singleton, i.e., $h(u) = \{\text{moving right arm}\}$. Thus, the intensional entity $u = I(MotPrg(P_1, x_2))$ can be used as a label (name) for the raw code P_1 .
3. **The Predicate:** The robot then uses a ternary predicate like $Exec(y_1, y_2, \langle \phi(\mathbf{x}) \rangle_{\alpha}^{\beta})$ with $g(y_1) = \text{"in present"}$ and $g(y_2) = \mathbf{I}$, such that

$$Exec(y_1, y_2, \langle MotPrg(x_1, x_2) \rangle_{x_2}^{x_1})/g = Exec(\text{in present}, \mathbf{I}, I(MotPrg(P_1, x_2)))$$
 translates to:

" $\mathbf{I}$ (the coordinator) am currently executing the motor program of raw code P_1 ."
4. **Reaching Conscious Knowledge:** this sentence in a temporary memory of symbolic AI level of robot can be inserted into conscious robot's knowledge by epistemic predicate $Know$ and can be used for other logic deductions.

By this personal, based on robots experiences, learning process, the *scalability* of the robot's conscious knowledge is guaranteed by this bridge from the neural to symbolic robot's knowledge.

4 The 4-valued Autoepistemic Deduction of AGI Robots

Differently from the the knowledge Database of robot's intensional n-ary concepts (provided in previous Section), the robots current *logic knowledge* (composed by ground atoms of the distinguished predicate $Know$ in Definition 7) is directly derived from its experiences (based on its neuro-system processes that robot is using in this actual world), in an analog way as human brain does:

- As an activation (under robot's attention) of its neuro-system process, as a consequence of some human command to execute some particular job.
- As an activation of some process under current attention of robot, which is part of some complex plan of robot's activities connected with its general objectives and services.

In both cases, for a given assignment of variables $g : \mathcal{V} \rightarrow \mathcal{D}$ of virtual predicate $\psi(\mathbf{x})$ with the set of variables $\beta = \bar{\mathbf{x}}$, which L -concept $I(\psi(\mathbf{x})/g) \in D_1$ is grounded by this particular process, is transformed into abstracted term and hence robot's knowledge system generates the new ground knowledge atom with three terms t_i , $i = 1, 2, 3$, $Know(t_1, t_2, t_3)/g$ with abstracted term $t_3 = \langle \psi(\mathbf{x}) \rangle^\beta$ in robot's temporary memory. So, the robot's logic knowledge, which participates in deduction processes as well, is composed by a set of known robot's sentences $\psi(\mathbf{x})/g$ heaving logic truth-values of Belnap's billattice.

Consequently, the explicit (conscious) robot's logic knowledge in actual world (current time-instance) here is represented by the ground atoms of the distinguished $Know$ predicate in Definition 7,

$$Know(t_1, t_2, \langle \psi(\mathbf{x}) \rangle^\beta)/g = Know(g^*(t_1), g^*(t_2), g^*(\langle \psi(\mathbf{x}) \rangle^\beta)) \quad (15)$$

such that $g^*(t_1) = in\ present$ and $g^*(t_2) = me$ (the robot itself), for the extended assignments $g^* : \mathcal{T} \rightarrow \mathcal{D}$ in Definition 8 to all terms in $\mathcal{T}$.

We recall that each robot's extensionalization function h in (7) is indexed by the time-instance. The actual robot's world extensionalization function (in the current instance of time) is denoted by $\bar{h}$, and determines the current robot's knowledge. Clearly, the robots knowledge changes in time and hence determines the extensionalization function h in any given instance of time, based on robots experiences. Thus, as for humans, also the robot's knowledge and logic is a kind of temporal logic, and evolves with time.

Remark: We consider that only robot's experiences (under robot's attention) are transformed into the ground atoms of the $Know$ predicate, and the required (by robot) deductions from them (by using $IFOL_B$ deduction extended by the three epistemic axioms in what follows) are transformed into ground atoms of $Know$ predicate, and hence are saved in robot's temporary memory as a part of robot's *conscience*.

Some background process (unconscious for the robot) would successively transform these temporary memory knowledge into permanent robot's knowledge in an analog way as it happen for humans.

□

Based on these considerations, we can specify exactly what are semantically the robot's knowledge models:

Definition 14. LOGIC KNOWLEDGE MODELS:

Herbrand interpretations $v : H \rightarrow X$ in Definition 9 are denominated knowledge interpretations, if for each ground atom in (15) of the T and $Know$ predicates the logic truth-value of them is equal just to the truth-value of the sentence reified in the ground atoms of these two predicates as an abstracted term,

$$\begin{aligned} v^*(Know(t_1, t_2, \langle \psi(\mathbf{x}) \rangle^\beta)/g) &= v^*(\psi(\mathbf{x})/g) \in X \\ v^*(T(\langle \psi(\mathbf{x}) \rangle^\beta)) &= v^*(\psi(\mathbf{x})/g) \in X \end{aligned} \quad (16)$$

Thus, in the actual world h (actual time-instance), the known fact (15) for robot becomes the ground atom

$$Know(t_1, t_2, \langle \psi(\mathbf{x}) \rangle^\beta)/g = Know(in\ present, me, I(\psi(\mathbf{x})/g)) \quad (17)$$

with a L-concept $I(\psi(\mathbf{x})/g) \in D_1$ and, from the fact that $\mathbf{I}_B = h \circ I$, we obtain

$$h(I(\psi(\mathbf{x})/g)) = \mathbf{I}_B(\psi(\mathbf{x})/g) = \begin{cases} \{a\} \in \tilde{X} \subset \mathfrak{Rm}, & \text{if } a = v^*(\psi(\mathbf{x})/g) \neq \perp \\ \emptyset, & \text{otherwise} \end{cases} \quad (18)$$

While, for the L-concept $u = I(Know(t_1, t_2, \leq \psi(\mathbf{x}) >^\beta)/g) \in D_1$, we obtain

$$\tilde{h}(I(Know(t_1, t_2, \leq \psi(\mathbf{x}) >^\beta)/g)) = \begin{cases} \{v^*(\psi(\mathbf{x})/g)\} \in \tilde{X}, & \text{if } v^*(\psi(\mathbf{x})/g) \neq \perp \\ \emptyset, & \text{otherwise} \end{cases}$$

that is, for this *current* MV-model $\mathbf{I}_B = h \circ I$, and a ground atom of the *Know* predicate, we obtain

$$\mathbf{I}_B(Know(t_1, t_2, \leq \psi(\mathbf{x}) >^\beta)/g) = \begin{cases} \{a\} \in \tilde{X} \subset \mathfrak{Rm}, & \text{if } a = v^*(\psi(\mathbf{x})/g) \neq \perp \\ \emptyset, & \text{otherwise} \end{cases} \quad (19)$$

and consequently from (18) and (19), in accordance with (16), we obtain the MV-model knowledge-constraint

$$\mathbf{I}_B(Know(t_1, t_2, \leq \psi(\mathbf{x}) >^\beta)/g) = \mathbf{I}_B(\psi(\mathbf{x})/g) \quad (20)$$

Thus, we defined by using the predicate *Know* what the robot knows about each its sentence $\psi(\mathbf{x})/g$. However, we have to define also the opposite feature about natural language sentences that use the "do not know" in all temporal conditions (does not know, did not know, will not know), generally expressed in natural language by:

"I do not know if $\psi(\mathbf{x})/g$ ",

"I do not know whether $\psi(\mathbf{x})/g$ "

just because their translation into a logic formula of *IFOL_B* needs to use this *Know* predicate as well, and the fact that this $\phi(\mathbf{x})/g$ has to be *unknown*, that is $v^*(\phi(\mathbf{x})/g) = \perp$. It is clear that such logic formula must be different from $\neg Know(t_1, t_2, \leq \psi(\mathbf{x}) >^\beta)/g$ because $\neg \perp \neq \perp$.

In accordance with the Closed Knowledge Assumption, if we do not know something, this something must be *unknown*. Consequently, this formula is provided by (see the 2-valued logic truth-scheme formulae for any truth-value $a \in X$, given by (3) in [25])

$$p_\perp \Leftrightarrow Know(t_1, t_2, \leq \psi(\mathbf{x}) >^\beta)/g \quad (21)$$

where $p_\perp$ is the built-in propositional letter such that for each many-valued valuation $v^*(p_\perp) = \perp$. In fact, the sentence in (21) is true only if $v^*(Know(t_1, t_2, \leq \psi(\mathbf{x}) >^\beta)/g) = \perp$, that is, from (16), iff $v^*(\psi(\mathbf{x})/g) = \perp$ is unknown.

So the formula (21) can be expressed in English, for example, by:

"I do not know whether $\psi(\mathbf{x})/g$ ",

if $g^*(t_1) = me$ and $g^*(t_2) = in\ present$. Otherwise by

" $g^*(t_1)$ does not know whether $\psi(\mathbf{x})/g$ ", if $g^*(t_2) = in\ present$

" $g^*(t_1)$ did not know whether $\psi(\mathbf{x})/g$ ", if $g^*(t_2) = in\ past$

" $g^*(t_1)$ will not know whether $\psi(\mathbf{x})/g$ ", if $g^*(t_2) = in\ future$.

Remark: Note that for the assignments $g : \mathcal{V} \rightarrow \mathcal{D}$, such that $g^*(t_1) = in\ future$ and $g^*(t_2) = me$ we consider robot's hypothetical knowledge in future, while in the

cases when $g^*(t_1) = \text{in past}$ we consider what was robot's knowledge in the past. Consequently, generally the predicates of IFOL for robots, based on the dynamic changes of its knowledge has to be indexed by the time-instances (which are possible worlds of IFOL), for example by using an additional predicate's argument for them.

In the examples in next, we will consider only the case of robot's current knowledge (in the actual world with extensional function h) when $g^*(t_1) = \text{in present}$, so we avoid the introduction of the time-instance variable for the predicates; only at the remark at the end of this section we will show how to use time-variable τ .

□

In what follows, the robot's knowledge system will use only the knowledge interpretations $v : H \rightarrow X$ and their extensions $v^* : \mathcal{L}_0 \rightarrow X$ as described by Definition 14, and we denote by $\mathbb{V}$ the set of all such knowledge interpretations. As an inference many-valued system, we can use the truth-preserving entailment provided by Definition 57 in Chapter 5 of [18], here rewritten using notations used in this paper by:

Definition 15. [18] NEW TRUTH-PRESERVING ENTAILMENT: Let $\Gamma = \{\phi_1, \dots, \phi_n\} \subset \mathcal{L}_0$ be a set of sentences and we define the truth-preserving entailment in two alternative cases:

1. *Sequent-based entailment with sequents $\Gamma \vdash \psi$. We say that $v \in \mathbb{V}$ satisfies this sequent iff $v^*(\phi_1) \wedge \dots \wedge v^*(\phi_n) \leq v^*(\psi)$, where $\wedge$ is the meet operator in lattice $(X, \leq)$, and we denote this sequent satisfaction by $\Gamma \models_v \psi$. A sequent is valid if it is satisfied for all $v \in \mathbb{V}$.*

We denote by $\phi \preceq \psi$ iff for all $v \in \mathbb{V}$, $v^(\phi) \leq v^*(\psi)$.*

Valuation v satisfies a rule $\frac{s_1, \dots, s_k}{s}$ iff v satisfies the conclusion sequent s of this rule whenever it satisfies all sequent premises $s_1, \dots, s_k$ of this rule. Then, a model of this logic is any valuation $v \in \mathbb{V}$ which satisfies all logic sequent rules of this logic (the structural sequent rules as Cut, Weakening, etc.. are satisfied by all valuations).

2. *Model-based entailment where Γ is the set of theses $\phi_i \in \mathcal{L}_0$ with associated prefixed low-boundary truth values $a_i \in X$, such that their models are defined as a subset of all knowledge interpretations in $\mathbb{V}$:*

$$\mathbb{V}_\Gamma =_{def} \{v \in \mathbb{V} \mid \forall \phi_i \in \Gamma. (v^*(\phi_i) \geq a_i)\} \subset \mathbb{V} \quad (22)$$

that is, a knowledge interpretation $v \in \mathbb{V}$ is a model of the theses in Γ if the truth-value of each sentence ϕ_i in Γ is greater or equal to its specified low-boundary truth-value a_i . So, we introduce the model-based truth-preserving entailments of a sentence ψ from the theses in Γ and $v^ \in \mathbb{V}_\Gamma$ by*

$$\begin{aligned} \Gamma \models_v \psi \quad \text{iff} \quad & (\exists \phi_i \in \Gamma). (v^*(\psi) \geq v^*(\phi_i)) \\ \Gamma \models \psi \quad \text{iff} \quad & (\forall v \in \mathbb{V}_\Gamma). \Gamma \models_v \psi \end{aligned} \quad (23)$$

Note that (22) is a generalization of the standard 2-valued logic case when all $a_i = t$ are true.

Remark: In this moment we are not interested to the possibility of a robot to derive by himself the paradoxes, like Liar paradox discussed in next section that requires the high level mathematical ability (by using Gödel numbering and creation by himself of new predicates like *Diag* and Peano arithmetic FOL theorems, for example. In this

stage it is enough only to support the existence of paradoxes by robot's knowledge base without any logic problem (such paradoxes can be provided by humans in initial robot's knowledge like all ground atoms of built-in predicates).

□

So, for the sentences ϕ that are paradoxes (Like the Liar paradox discussed in next section) the inconsistent truth-value $\top$ is used:

Definition 16. *A paradox can be considered as particular built-in sentence $\phi \in \mathcal{L}$ (logic formula which is not an atom, without free variables) that for each knowledge interpretation $v \in \mathbb{V}$ it is always an inconsistent sentence, that is, $v^*(\phi) = \top$.*

Inconsistent sentences are not only the paradoxes. Consider for example a ground atom (of non built-in or distinguished predicates T and $Know$) that previously robot learned to be true and successively learned to be false (and viceversa), so to them will be assigned by knowledge interpretations the inconsistent value

$$\top = f \vee t$$

that will remain such for all future time of robot's life.

So, robots can support in their knowledge also the built-in paradoxes and other inconsistent ground atoms or sentences, and to compose them with other sentences by using logic connectives and to compute by v^* the truth-values of such composed formulae as well. For example, let ϕ be an inconsistent sentence and ψ not: if for given knowledge interpretation $v \in \mathbb{V}$, $v^*(\psi) = f$ then $v^*(\phi \wedge \psi) = v^*(\phi) \wedge v^*(\psi) = \top \wedge f = f$ (is not more inconsistent) while if for another $v_1 \in \mathbb{V}$, $v_1^*(\psi) = t$ then $v_1^*(\phi \wedge \psi) = v_1^*(\phi) \wedge v_1^*(\psi) = \top \wedge t = \top$ is inconsistent.

Notice that the entailment $\Gamma \models \psi$ is *valid* entailment differently from the specific entailment $\Gamma \models_v \psi$ for a given $v \in \mathbb{V}_\Gamma$, used also by Ginsberg's truth preserving entailment in point 3 of Definition 2 when Γ is a singleton set, as we can show:

Corollary 2 *Let Γ be a singleton composed by only $\phi \in \mathcal{L}_0$. The valid entailment $\Gamma \models \psi$ can be replaced by $\phi \models \psi$ so that (23) reduces to:*

$$\phi \models \psi \text{ iff } (\forall v \in \mathbb{V}_\Gamma). \phi \models_v \psi$$

and each knowledge interpretation $v \in \mathbb{V}$ (and hence each model of ϕ as well) is closed by Ginsberg's specification in Definition 2.

Proof: It is easy to verify that all three points in Definition 2 are satisfied for each knowledge interpretation $v \in \mathbb{V}$.

□

Let us consider now in which way the many-valued entailment $\Gamma \models_v \psi$ generalizes the standard 2-valued FOL entailment (where all sentences in Γ and conclusion ψ must be true) for the FOL inference rules for logic quantifiers and Modus Ponens $\phi, \phi \Rightarrow \psi \models_v \psi$. By this generalization in this many-valued intensional FOL we mean that not all premises in Γ and conclusion ψ must be true for this given interpretation v^* . So, we will only relax the Modus Ponens rule of the standard 2-valued FOL into this MV-Modus Ponens rule in which only the implication $\phi \Rightarrow \psi$ must be true and not the thesis ϕ and the conclusion ψ :

Definition 17. For a given knowledge interpretation v^* the MV-Modus Ponens rule (entailment) is defined as follows:

$$\text{if } v^*(\phi \Rightarrow \psi) = t \text{ then } \phi, \phi \Rightarrow \psi \models_v \psi \quad (24)$$

The inference rules for existential and universal quantifiers used in standard 2-valued FOL as for example Universal Instantiation entailment $(\forall x)\phi(x) \models_v \phi(x)/g$, are valid also in this many-valued many-sorted intensional FOL_B .

This definition for inference rules for quantifiers are obviously valid in FOL_B , where, for example the Universal Instantiation entailment is satisfied from the fact that from point 4 in Definition 10 it holds that $v^*(\forall x)\phi(x) \leq v^*(\phi(x)/g)$ also if we do not impose the standard 2-valued FOL requirement that $v^*(\forall x)\phi(x) = t$. We only have to show that the MV-Modus Ponens rule in (24) is well defined generalization of standard Modus Ponens rule in 2-valued FOL.

Corollary 3 The MV-Modus Ponens rule in (24) is well defined generalization of standard Modus Ponens rule in 2-valued FOL.

Proof: The thesis for MV-Modus/Ponens is defined by two sentences $\Gamma = \{\phi_1, \phi_2\}$ where ϕ_1 is the sentence ϕ with the low-boundary truth value $a_1 = f$ and ϕ_2 is the sentence $\phi \Rightarrow \psi$ with the low-boundary truth value $a_2 = t$, so for each knowledge interpretation $v^* \in \mathbb{V}_\Gamma$ it must hold that $v^*(\phi) \geq f$ and $v^*(\phi \Rightarrow \psi) \geq t$ (i.e., $v^*(\phi \Rightarrow \psi) = t$ as required by MV-Modus Ponens). So, from the requirement that $v^*(\phi \Rightarrow \psi) = t$, from the definition of intuitionistic implication in table 1 it must hold that

$$v^*(\phi) \leq v^*(\psi) \quad (25)$$

For this thesis in $\Gamma = \{\phi, \phi \Rightarrow \psi\}$ the first sentence ϕ it can be any value in X and we can use this thesis with (25) as satisfaction in (23) of the modus point entailment $\Gamma \models_v \psi$, which is satisfied for each knowledge interpretation v^* such that:

Case 1: $v^*(\phi) = f$ and $v^*(\psi) \in \{f, \perp, \top, t\}$, so from false ϕ we can entail any sentence as in standard 2-valued FOL Modus Ponens;

Case 2: $v^*(\phi) = \perp$ and $v^*(\psi) \in \{\perp, t\}$;

Case 3: $v^*(\phi) = \top$ and $v^*(\psi) \in \{\top, t\}$, so in this case from inconsistent ϕ by MV-Modus Ponens we can entail the inconsistent ψ .

Case 4: $v^*(\phi) = t$ and $v^*(\psi) = t$;

Based on these models v^* of Γ given in 4 cases above, we have that *valid entailment* $\phi, \phi \Rightarrow \psi \models_v \psi$ satisfied for *all these models* of Γ , would entail ψ iff in all these models ψ is true (in fact the common true-value of ψ in 4 cases above is just the truth-value t).

Note that if in the thesis Γ we set the low-boundary truth value $a_1 = t$ then we obtain Case 4 above which is standard 2-valued FOL Modus Ponens rule of inference, and hence the MV-Modus Ponens is a conservative many-valued extension of classical 2-valued Modus Ponens. Thus MV-Modus Ponens in (24) is well defined.

□

The autoepistemic robot's reasoning:

The introduction of the distinguished predicate *Know* is fundamental for the conscious part of strong-AI robot's cognitive system, able to save the robot's learned experience in symbolic way able to support the autoepistemic logic reasoning and deductions.

The autoepistemic logic is introduced in 2-valued propositional logic [31] with added universal modal operator, usually written K , and the axioms:

1. Reflexive axiom **T**: $K\phi \Rightarrow \phi$
2. Positive introspection axiom **4**: $K\phi \Rightarrow KK\phi$
3. Distributive axiom **K**: $(K\phi \wedge K(\phi \Rightarrow \psi)) \Rightarrow K\psi$

for any proposition formulae ϕ and ψ .

However, we do not use more a *pure* logical deduction of the standard 2-valued FOL, but a kind of autoepistemic deduction [32,8] with a proper set of new axioms. The modal (Kripke-like) operator K will be replaced by the IFOL predicate $Know$ applied to the logic formulae transformed into *abstracted terms*. Thus, the three epistemic axioms of epistemic modal logic with modal operator K , used to obtain deductive knowledge, can be traduced into 4-valued logic $IFOL_B$ by the following axioms for the predicate $Know$, which are demonstrated to be *the tautologies*¹⁴ in $IFOL_B$ which in fact define the semantics of this particular $Know$ predicate, for a given assignment $g : \mathcal{V} \rightarrow \mathcal{D}$ as follows:

1. The modal axiom **T**, in IFOL is represented by the axiom, for each abstracted term $\langle \psi(\mathbf{x}) \rangle^\beta$ in (15), with β the set of variables in the tuple $\mathbf{x}$, by the axiom schema

$$Know(t_1, t_2, \langle \psi(\mathbf{x}) \rangle^\beta) / g \Rightarrow \psi(\mathbf{x}) / g \quad (26)$$

So by using axiom (26), and FOL deduction, these deductive properties of the robot can deduce any true single fact (logical sentence) $\psi(\mathbf{x}) / g$ derived by its neuro-system process, and to render it to robot's consciousness as a single known fact $Know(t_1, t_2, \langle \psi(\mathbf{x}) \rangle^\beta) / g$. That is, from (8) with

$$u_1 = g^*(\langle \psi(\mathbf{x}) \rangle^\beta) = I(\psi(\mathbf{x}) / g) \in D_1 \quad (27)$$

such that $h(u_1) = h(I(\psi(\mathbf{x}) / g))\{a\}$ when $\perp \neq a \in X$ (otherwise we set $a = \perp$) is the truth-value of this robot's known sentence $\psi(\mathbf{x}) / g$ (if this is a Liar sentence (paradox), then $a = \top$ is inconsistent truth-value, and robot is able to work also with inconsistent knowledge as well. So, this **T** axiom (26) becomes

$$Know(g^*(t_1), g^*(t_2), u_1) \Rightarrow \psi(\mathbf{x}) / g \quad (28)$$

Note that the meaning of the intensional concept u_1 of the robot is grounded on robot's neuro-system process, which is just robot's current internal experience of what is he doing. For each knowledge model v^* the truth-values of both sides of implication in (26) are equal, so the truth-value of this logic schema (26) is true (lies in the diagonal of the truth-table (1)), that is, it is really an reflexive axiom.

Consequently, the application of the **T** axiom allows the extraction from robot's conscious knowledge the *logical sentences* which, successively, can be elaborated

¹⁴ The fact that the autoepistemic axioms in $IFOL_B$ are just the tautologies (always true) means that indeed natively $IFOL_B$ is an autoepistemic logic to reason about the knowledge in typically human way able to support the Socrate's famous "I know that I do not know that..." But to support the more complex Liar paradox as explained in details in next section.

by robot's implemented deductive property of FOL in two ways:

a.1. To respond to some human natural language questions (parsed into a logical formula) and to verify if the response is "yes" or "no", or "I do not know" (if robot's conscious knowledge is incomplete for such a question);

a.2. To deduce another sentences which then can be inserted in robot's conscious knowledge as ground atoms of the predicate *Know* (where this deduced sentence is represented as an abstracted term). This process (in background, or when robot is free of other concrete activities) can be considered as a kind of consolidation and completion of robot's knowledge based on previous experiences, in an analog way as it is done by human mind when we sleep.

2. The positive introspection axiom **4**:

$$Know(t_1, t_2, \langle \psi(\mathbf{x}) \rangle^\beta) / g \Rightarrow Know(g^*(t_1), g^*(t_2), \langle Know(t_1, t_2, \langle \psi(\mathbf{x}) \rangle^\beta) / g \rangle) \quad (29)$$

that is.

$$Know(g^*(t_1), g^*(t_2), g^*(\langle \psi(\mathbf{x}) \rangle^\beta)) \Rightarrow Know(g^*(t_1), g^*(t_2), \langle Know(g^*(t_1), g^*(t_2), g^*(\langle \psi(\mathbf{x}) \rangle^\beta)) \rangle) \quad (30)$$

which, in the case when $g(y_1) = in\ present$ and $g(y_2) = me$, is traduced in natural language by robot as:

"I know that $\psi(\mathbf{x})/g$ " implies "I know that I know that $\psi(\mathbf{x})/g$ "

where $g^*(\langle \psi(\mathbf{x}) \rangle^\beta) = I(\psi(\mathbf{x})/g)$ contains the logic truth-value of the known sentence $\psi(\mathbf{x})/g$ as explained by (27).

For each knowledge model v^* the truth-values of both sides of implication in (30) are equal, so the truth-value of this logic schema (31) is true (lies in the diagonal of the truth-table (1)), that is, it is really an introspection axiom schema.

3. The distributive axiom **K** ("modal Modus Ponens"):

$$(Know(t_1, t_2, \langle \psi(\mathbf{x}) \rangle^\beta) / g \wedge Know(t_1, t_2, \langle \psi(\mathbf{x}) \Rightarrow \phi(\mathbf{z}) \rangle^{\beta \cup \beta_1}) / g) \Rightarrow Know(t_1, t_2, \langle \phi(\mathbf{z}) \rangle^{\beta_1}) / g \quad (31)$$

with the *sets* of variables $\beta = \bar{\mathbf{x}}$ and $\beta_1 = \bar{\mathbf{z}}$. Or, equivalently,

$$(Know(g^*(t_1), g^*(t_2), g^*(\langle \psi(\mathbf{x}) \rangle^\beta)) \wedge Know(g^*(t_1), g^*(t_2), g^*(\langle \psi(\mathbf{x}) \Rightarrow \phi(\mathbf{z}) \rangle^{\beta \cup \beta_1}))) \Rightarrow Know(g^*(t_1), g^*(t_2), g^*(\langle \phi(\mathbf{z}) \rangle^{\beta_1})) \quad (32)$$

Note that this axiom schema is a way how the robot provides the conscious implications $\psi(\mathbf{x})/g \Rightarrow \phi(\mathbf{z})/g$, independently if they are its innate implemented rules (introduced in robot's knowledge when it is created) or if they are learned by robot's

own experience. For any knowledge model v^* , if we apply it to the logic formula (32) by using (16), from point 2 of Definition 10 for binary operators $\wedge$ and $\Rightarrow$, we obtain the propositional formula with truth-values $a = v^*(\psi(\mathbf{x})/g) \in X$ and $b = v^*(\phi(\mathbf{z})/g) \in X$,

$$(a \wedge (a \Rightarrow b)) \Rightarrow b$$

which is true for all cases of $a, b \in X$. Thus (31) is a valid axiom schema.

Remark: We recall that this method of application of autoepistemic deduction (for concepts such as *knowledge*) can be applied to all other modal logic operators (for concepts such as *belief*, *obligation*, *causation*, *hopes*, *desires*, etc., for example by using *deontic* modal logic that same statement have to represent a *moral obligation* for robots), by introducing special predicates for them with the proper set of axioms for their active semantics (fixing their meaning and deductive usage).

By such fixing by humans of robot's unconscious part with active semantics (which can not be modified by robots and their live experience) of all significant for human robot's concepts and their properties, we will obtain ethically confident and socially safe and non danger robots (controlled by public human ethical security organizations for the production of robots with general strong-AI capabilities).

□

Note that the obtained robot's knowledge, from the known facts at the end of deduction, is in robot's temporary memory. In order to render it permanent (by cyclic process of transformation of the temporary into permanent robot's memory), we need to add to any predicate of the robot's FOL syntax, also the time-variable as, for example, the first variable of each predicate (different from *Know*), instantiated in the known facts by the *timestam*p value τ (date/time) when this knowledge of robot is transferred into permanent memory. This temporization of all predicates used in robot's knowledge is useful for robot to search all known facts in its permanent memory that are inside some time-interval as well.

It can be used not only to answer directly to some human questions about robot's knowledge, but also to extract only a part of robot's knowledge from its permanent memory in order to be used for robot's deduction, and hence to answer to human more complex questions that require deduction of new facts not already deposited in explicit robot's known facts.

The last argument of this section will be dedicated to the evolution of robot's knowledge in time.

The problems about knowledge fixpoints:

Based on the bilattice $(X, \leq, \leq_k)$ of truth-values in X we are able to define the truth and knowledge partial orders over the knowledge interpretations in $\mathbb{V}$ as well:

Definition 18. Let $H_0 \subset H$ be the subset of ground atoms of predicates in P that are not built-in or distinguished predicates. Then for $v_1^*, v_2^* \in \mathbb{V}$ we have that

1. Truth partial order $\preceq$ is defined by

$$v_1^* \preceq v_2^* \quad \text{iff} \quad (\forall A \in H_0). v_1(A) \leq v_2(A);$$
2. Knowledge partial order $\preceq_k$ is defined by

$$v_1^* \preceq_k v_2^* \quad \text{iff} \quad (\forall A \in H_0). v_1(A) \leq_k v_2(A).$$

Note that the truth-value of ground atoms of built-in predicates are predefined initially and do not change in time, while the ground atoms of distinguished predicates T and $Know$ are derived from the truth-values of ground atoms that compose the sentence in their abstracted terms. Because of that we used only ground atoms in H_0 for definition of two partial orders in definition above.

In particular, if we are considering the growing in time of the robots knowledge, with his knowledge interpretation v_i at time instance τ_i , and the knowledge interpretation v_{i+1} at time instance τ_{i+1} , with $\tau_i < \tau_{i+1}$.

Let us denote by v_0^* the bottom knowledge interpretation in which each ground atom $A \in H_0$ (of non-distinguished predicates) is unknown. So,

1. The robot's learning process which increments its knowledge is obtained by passing some ground atom A (of non-distinguished and non-built-in predicates) from the initial truth-value $\perp$ (unknown), to the learned by robot truth-value f (false) or t (true). An atom is not learned by a robot to be unknown, but only initially (at the beginning when a robot is created) for default it is set to be unknown in robot's initial (bottom) knowledge interpretation v_0^* . Thus one time robot learned that an atom is true or false, successively can not learn that it is unknown.
It is possible that such an atom that robot previously learned to be false, in future by different robot's experience can become inconsistent (if robot learned successively from different sources that it could be true). Analogously, an atom previously learned to be true, in future can become inconsistent.
2. If an ground atom becomes inconsistent its truth-value can not be successively changed, like in the case of built-in paradoxes¹⁵.

Based on the considerations above and Definition 18, we obtain the following result:

Lemma 1. *We have that during time-evolution for $\tau_i \leq \tau_{i+1}$ of robot's knowledge interpretations it is always satisfied that*

$$v_i^* \preceq_k v_{i+1}^* \tag{33}$$

that is, we have a monotonic increments of knowledge.

Proof: We have that for each ground atom $A \in H_0$ its truth-value can evolve only in this way:

1. $\perp \mapsto f$ of $\perp \mapsto t$;
2. $f \mapsto \top$;
3. $t \mapsto \top$;
4. $\top \mapsto \top$.

Thus, all changes of truth-values of ground atoms in H_0 follow the increments of knowledge in the lattice of knowledge ordering of Belnap's bilattice. Thus, from Definition 18 the relationship (33) is valid.

¹⁵ In effect, all human paradoxes are initially introduced in robot's knowledge as complex sentences with requirement that, in each knowledge interpretation of them, have the same constant built-in truth-value $\top$

□

Consequently, given such time-ordering of extended robot's knowledge, expressed by v_i^* in the ordered sequence $v_0^* \prec_k \dots \prec_k v_i^* \prec_k v_{i+1}^* \prec_k \dots$, based on Lemma 1, we are able to define the monotonic function $f_k : \mathbb{V} \rightarrow \mathbb{V}$ such that for $i = 1, 2, 3, \dots$

$$f_k(v_i^*) = v_{i+1}^* \quad (34)$$

In fact, for any two v_j^*, v_n^* such that $v_j^* \preceq_k v_n^*$ we have the monotonic property $f_k(v_j^*) \preceq_k f_k(v_n^*)$. Note that we do not guarantee the monotonicity w.r.t the truth-ordering in time evolution of robots knowledge. So, we obtain the following result:

Corollary 4 *In any ordered sequence $v_0^* \prec_k \dots \prec_k v_i^*$ with $S = \{A \in H_0 | v_i(A) = \perp\}$ if S is empty then v_i^* can be the fixed point of f_k , that is, $f_k(v_i^*) = v_i^*$.*

Proof: If S is empty than if all inconsistencies are found, v_i^* is fixed point, that is, successive changes of robot's knowledge are impossible, that is, it will not be possible for a robot to learn more by its existing logic system.

□

Of course, if robot reaches a knowledge fixpoint by learning process robot can not increment its reached knowledge.

So, in order to avoid knowledge fixpoints, in order to be able to increment its knowledge, the AGI robots must be able to learn and generate (like the humans) from the natural language an extra new predicate $p_j^k \notin P$ (w.r.t the initially defined set P) whose atom with all arguments the sorted variables in k -ary tuple of variables $\mathbf{x}$, $p_j^k(\mathbf{x})$ and its corresponding PRP concept (real or abstract) $u = I(p_j^k(\mathbf{x}))$ where I is new intensional interpretation by initial assignment to all ground atoms of this new predicates the logic truth-value $\perp$. That with empty set $h(u)$ of extension of this new concept.

5 Paradoxes: Resolution of Liar Formula in $IFOL_B$

Let us introduce the Liar paradox in the standard 2-valued FOL of arithmetic (standard 2-valued first-order Peano Arithmetic), extended by a primitive satisfaction binary predicate (Gödel) $Sat(x, y)$, governed by Tarskian axioms. This system of axioms governing $Sat(x, y)$ was given in [26].

The philosophical significance of such FOL logic is discussed in [27], compared with minimalistic extensions generated by adding just the "T-schema" (the set of formulae $\mathbf{T}(\ulcorner \phi \urcorner) \Leftrightarrow \phi$ where $\ulcorner \phi \urcorner$ is the code (positive integer) obtained by Gödel's codification of logic formula ϕ , and $\mathbf{T}$ is unary 2-valued truth-predicate whose argument has the sort of natural numbers N . Similar constructions are considered in [28].

The idea is then $Sat(x, y)$ expresses the satisfaction relation between (codes of) formulae and (codes of) sequences of the formulae. Then $\mathbf{T}(x)$ expresses the *concept of truth* for such formulae and sequences. In fact, it can be shown that this theory satisfies Tarski's Convention (T) in [18]: i.e., that

$$\mathbf{T}(\ulcorner \phi \urcorner) \Leftrightarrow \phi \quad (35)$$

is a theorem (true formula) in this 2-valued Peano Arithmetic FOL.

If ϕ is any formula with one free variable x , then the *diagonalization* of ϕ , can be defined by mapping $d : \mathcal{L} \rightarrow \mathcal{L}$, such that $d(\phi(x)) = \phi(\ulcorner \phi(x) \urcorner)$, equivalent to the formula $(\exists y)(y = \ulcorner \phi(x) \urcorner \wedge \phi(y))$.

Based on this mapping we can define the mapping $diag : N \rightarrow N$, such that for any non-negative integer $n \in N$, $diag(n) = \ulcorner d(\phi(x)) \urcorner$ if $n = \ulcorner \phi(x) \urcorner$; 0 otherwise.

Now we can introduce the binary predicate $Diag(x, y)$ such that for any two $n, k \in N$, $Diag(k, n)$ is true iff $n = diag(k) \neq 0$, that is, if n is the code of the formula obtained by diagonalization of the formula $\phi(x)$ whose code is equal to k .

From the Diagonalization Lemma (or Fixed Point Theorem) in Peano Arithmetic FOL, there must be fixed point formula λ such that, the following formula is a theorem (true formula) in this logic:

$$\lambda \Leftrightarrow \neg \mathbf{T}(\ulcorner \lambda \urcorner), \quad \text{for } \lambda \text{ equal to } d(\neg \mathbf{T}(diag(x))) \quad (36)$$

It is easy to show that $\neg \mathbf{T}(diag(x))$ is equal to the formula $(\exists y)(Diag(x, y) \wedge \neg \mathbf{T}(y))$ with the free variable x . Thus,

Let $\phi(x)$ be equal to formula $(\exists y)(Diag(x, y) \wedge \neg \mathbf{T}(y))$, then

$$\begin{aligned} \lambda &\text{ is equal to the formula } d(\phi(x)), \text{ i.e. to } (\exists z)(Diag(\ulcorner \phi(x) \urcorner, z) \wedge \neg \mathbf{T}(z)) \\ &\text{ and hence, to } (\exists z)(Diag(\ulcorner (\exists y)(Diag(x, y) \wedge \neg \mathbf{T}(y)) \urcorner, z) \wedge \neg \mathbf{T}(z)) \end{aligned} \quad (37)$$

The analysis of the proof of this Diagonalization Lemma [29,30] shows that this formula λ must contain the atoms $\mathbf{T}(x)$ of 2-valued predicate $\mathbf{T}$. This formula is the formal analogue of the so-called "*strengthened liar*" for this system. It is a formula that "*says of itself that it is not true*".

In fact, it expresses the Liar paradox. If we assume that λ is true than $\neg \mathbf{T}(\ulcorner \lambda \urcorner)$ must be true, i.e. $\mathbf{T}(\ulcorner \lambda \urcorner)$ must be false and, consequently, from (35) we obtain that λ is false: the contradiction. The same contradiction we obtain if we assume that λ is false.

Liar paradox in 4-valued $\mathcal{L}_{in}$:

Let us consider now this problem inside the intensional many-valued FOL $\mathcal{L}_{in}$ defined previously by having also the unary predicate $\mathbf{T}$ with sort of natural numbers the binary predicate $Diag$ with both arguments of the sort of natural numbers, so that the formula λ in (37) is well defined in 4-valued intensional FOL $\mathcal{L}_{in}$ as well. However, while $Diag$ in $\mathcal{L}_{in}$ remains 2-valued (can be only true or false), the predicate $\mathbf{T}$ in $\mathcal{L}_{in}$ is 4-valued and hence can have also the truth-values $\perp$ and $\top$ as well. Thus also the formula $\neg \mathbf{T}(\ulcorner \lambda \urcorner)$ in (36) can have the truth-value $\perp$ or $\top$ as well and hence the formula (36) can remain a theorem (true formula) also if λ has the truth-value $\perp$ or $\top$. Let us show this fact: from (37) and knowledge interpretation v^* , we obtain

$$v^*(\lambda) = v^*((\exists z)(Diag(\ulcorner \phi(x) \urcorner, z) \wedge \neg \mathbf{T}(z)))$$

and hence from the point 3 of Definition 10 for existentially quantified formulae, and the fact that for the ground atoms $v^*(Diag(\ulcorner \phi(x) \urcorner, z)/g_1)$ are false for all $g_1(z) \neq \ulcorner \phi(\phi(x)) \urcorner \in N$, we obtain that,

$$v^*(\lambda) = v^*(\neg \mathbf{T}(\ulcorner \phi(\phi(x)) \urcorner)) = \neg v^*(\mathbf{T}(\ulcorner \phi(\phi(x)) \urcorner)) \in X$$

can have any truth-value in X , thus also $\perp$ or $\top$.

Consequently, instead of the formula $\lambda \Leftrightarrow \neg \mathbf{T}(\ulcorner \lambda \urcorner)$ in (36) for the sentence λ defined in (37), we can use the intensional abstraction operator $\langle _ \rangle$ and the 4-valued unary predicate T which argument is the sort of abstract terms and obtain equivalent to it formula

$$\lambda \Leftrightarrow \neg T(\langle \lambda \rangle) \quad (38)$$

Note that both unary predicates $\mathbf{T}$ and T have the same semantics as truth-predicates but only represent the formula λ in two different ways: in $\mathbf{T}$ is represented by its numeric Gödel's code while in T by its abstracted term used in intensional FOL. However, in this case we have that the unary predicate T is *many-valued* and hence able to satisfy the condition (38) by using the truth-constraint provided in (16)

$$v^*(T(\langle \phi(\mathbf{x})/g \rangle)) = v^*(\phi(\mathbf{x})/g) \in X \quad (39)$$

for any sentence $\phi(\mathbf{x})/g$ and valuation v^* . Thus, for each MV-interpretation I_B^* of $\mathcal{L}_{in}$, we may fix $I_B^*(\lambda) = \{\kappa\} \notin \{\{f\}, \{t\}\}$ in the way that the Liar sentence λ satisfy (38), with the many-valued unary predicate T :

Corollary 5 *The Liar self-reference sentence is not a paradox in $\mathcal{L}_{in}$, thus, does not create the contradictions.*

Proof: Note that from Definition 8 in [25], for a relation $R \in \mathfrak{Rm}$ with arity $ar(R) = 1$, for the unary operator $\odot : \mathfrak{Rm} \rightarrow \mathfrak{Rm}$, $\odot(R) = \{\neg a \mid \{a\} = R, \neg a \neq \perp\}$.

Thus, for any MV-interpretation I_B^* , such that $I_B^*(\lambda) = \{\kappa\}$ (or an equivalent many-valued interpretation v^* such that $v^*(\lambda) = \kappa$), such that $\neg \kappa = \kappa$, which is satisfied for $\kappa = \top$ ("inconsistent" truth-value, both true and false) in the lattice of truth values X , we have that $I_B^*(\neg T(\langle \lambda \rangle)) = \odot I_B^*(T(\langle \lambda \rangle)) = \text{from (39)} = \odot(\{\top\}) = \{\top\} = I_B^*(\lambda)$ so that (38) (that is, the (36)) are true formulae. Thus, in $\mathcal{L}_{in}$ the value to λ is assigned without any contradiction.

□

Analogously, the Gödel's contradictory formulae obtained by diagonalization of the formula $\neg \text{Prov}(\text{diag}(x))$, where $\text{Prov}(x)$ is a formula representing provability in the Peano arithmetic FOL can be resolved in $\mathcal{L}_{in}$ by interpreting these formulae with the value $\kappa = \perp$ ("unknown" truth-value) of Belnap's bilattice.

6 Conclusions and Future Work

By proposed four-level cognitive robot's structure, IFOL allow robots to reflect on, reason about, anticipate, or simply imagine scenes, situations, and developments within in a highly flexible, compositional, that is, semantically meaningful manner. As a result, IFOL enables the robots to actively infer highly flexible and adaptive goal-directed behavior under varying circumstances [34].

What psychological and philosophical significance should we attach to recent efforts at computer simulations of human cognitive capacities? In answering this question, I find it useful to distinguish what I will call "strong" AI from "weak" or "cautious" AI (Artificial Intelligence). According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it

enables us to formulate and test hypotheses in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations. In *Minds, brains, and programs*, John R. Searle, [35]

So, in our proposal, the *cognitive states* of AGI robots are defined by extensions of its knowledge Database for intensional concepts and known sentences represented by the ground atoms of the *Know* predicate in the actual knowledge model. So, mental content of an AGI robot is represented by its memory as a conscious cognitive state and by the grounding of the intensional language concepts to all kind of robot's processes realized by algorithms, neural networks and other computational and programming methods like subsumption architecture of Rodney Brooks [38] based on layers and augmented finite-state machines.

IFOL-based approach is part of a general neuro-symbolic AI paradigm, where researchers aim to blend: Neural methods (e.g., deep learning) for pattern recognition and learning from data, and symbolic logic for structured knowledge representation and reasoning. This broader field (which includes work at IBM Research, MIT, and in academic surveys) seeks to build AI systems that can both *learn from experience* and *reason abstractly*. These ideas contribute to ongoing discussions in AI about *symbol grounding*, *logical inference*, and *self-referential reasoning* — all important for Strong AI research.

The future work will be to extend the $IFOL_B$ to probabilistic reasoning and to consider an implementation of the Symbol Grounding Problem (SGP) for the symbolic system of $IFOL_B$. For example, eight proposed strategies for solving the SGP, which was given its classic formulation in Harnad are provided in [37]. The SGP concerns the possibility of specifying precisely how a robot can autonomously elaborate its own semantics for the symbols that it manipulates and do so from scratch, by interacting with its environment and other autonomous agents. This means that, as Harnad rightly emphasises [36], the interpretation of the symbols must be intrinsic to the symbol system itself, it cannot be extrinsic. This is provided in $IFOL_B$ by the fact that this logic language support both the truth and the meaning by the PRP concepts.

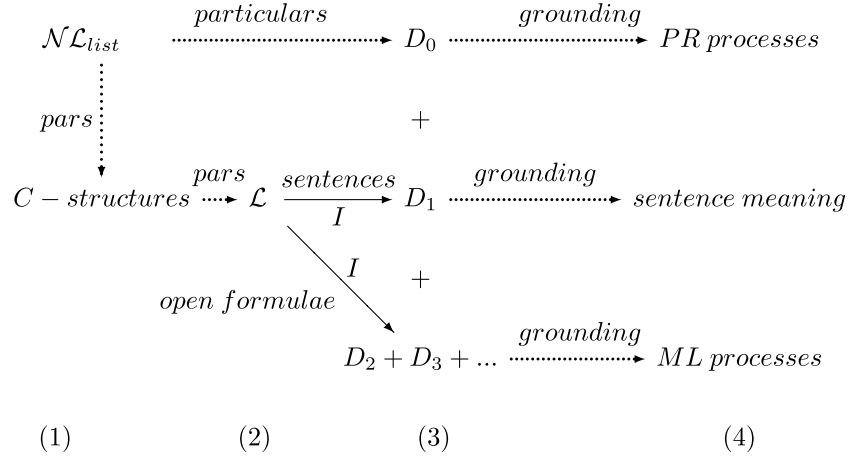
References

1. Ginsberg, M. : Multivalued logics: A uniform approach to reasoning in artificial intelligence. *Computational Intelligence*, vol.4, 265–316, (1988).
2. Ginsberg, M. : Bilattices and modal operators. Tech.Rep.N.94305, Comp.Science Dept. Stanford University, California, (1990).
3. Belnap, N.D. : A useful four-valued logic. In J-M.Dunn and G.Epstein, editors, *Modern Uses of Multiple-Valued Logic*. D.Reidel, (1977).
4. M.C.Fitting, M.C. : Bilattices and the semantics of logic programming. *Journal of Logic Programming*, 11, 91–116, (1991).
5. Majkić, Z.: Bilattices, Intuitionism and Truth-knowledge Duality: Concepts and Foundations. *Journal of Multiple-valued Logic & Soft Computing*, Vol.14, No.6, 525–564, (2008).
6. Bry F.: Logic programming as constructivism: a formalization and its application to databases. In *Proc. of the Symposium on Principles of Database Systems, ACM SIGACT-SIGMOD*, (1989).
7. Denecker M. and Bruynooghe M. and Marek V.: *Logic Programming Revisited: Logic Programs as Inductive Definitions*. *ACM Transactions on Computational Logic* 2 (4), 623–654, (2001).
8. Z.Majkić Z.: Autoepistemic logic programming for reasoning with inconsistency. *International Symposium on Logic-based Program Synthesis and Transformation (LOPSTR)*, September 7-9, 2005, Imperial College, London, UK, (2005).
9. Majkić Z. : Many-valued intuitionistic implication and inference closure in a bilattice based logic. *35th International Symposium on Multiple-Valued Logic (ISMVL 2005)*, May 18-21, Calgary, Canada, (2005).
10. Majkić Z.: Truth and knowledge fixpoint semantics for many-valued logic programming. *19th Workshop on (Constraint) Logic Programming (W(C)LP 2005)*, February 21-25, Ulm, Germany, (2005).
11. Rasiowa H. and Sikorski R.: *The Mathematics of Metamathematics*. PWN- Polish Scientific Publishers, Warsaw, 3rd edition, (1970).
12. Miller D.: A logical analysis of Modules in Logic Programming. *Journal of Logic Programming*, Vol. 6, 79–108, (1989).
13. Valiant L.G.: *Knowledge Infusion: In Pursuit of Robustness in Artificial Intelligence*. *Foundations of Software Technology and Theoretical Computer Science (Bangalore)*, 415–422, (2008).
14. Bader S. and Hitzler P.: Dimensions of Neural-symbolic Integration - A Structured Survey. *arXiv:cs/0511042*, (2005).
15. Sarker K. and Zhou L. and Eberhart A. and Hitzler P.: *Neuro-Symbolic Artificial Intelligence*. *arXiv:2105.05330*, (2021).
16. Pan Z.J at all: *Large Language Models and Knowledge Graphs: Opportunities and Challenges*. *arXiv:2308.06374*, (2023).
17. Colelough B.C. and Regli W.: *Neuro-Symbolic AI in 2024: A Systematic Review*. *Logical Foundations of Neuro-Symbolic AI 2024 workshop (IJCAI 2024)*, Jeju Island, South Korea, (2024).
18. Majkić, Z.: *Intensional First Order Logic: from AI to New SQL Big Data*. Walter De Gruyter GmbH, Berlin/Boston, ISBN 978-3-11-099494-0, (2022).
19. Majkić, Z. : *Intensional First Order Logic for Strong-AI Generation of Robots*. *Journal of Advances in Machine Learning & Artificial Intelligence*, Volume 4, Issue 1, 23–31, (2023).
20. Majkić, Z. : *Strong-AI Autoepistemic Robots Build on Intensional First Order Logic*. In F.Zhao, D.Miao(eds), *AI-generated Content. AIGC 2023. Communications in Computer and Information Science*, vol 1946. Springer, Singapore, 33–58, (2024).

21. Marcus, G. : Deep learning: a critical appraisal. arXiv:1801.00631 [cs.AI], (2018).
22. Karpathy, A.: Connecting Images and Natural Language. PhD Thesis, Stranford University, (2016).
23. Kollar, T. and Tellex, S. and Roy, D. and Roy, N. : Toward understanding natural language directions. In Proceedings of the 4th ACM international conference on human robot interaction, (2010).
24. Majkić, Z. : Intensional FOL: Many-Sorted Extension. arXiv:2409.04469 [cs.AI], 1–21, (2024).
25. Majkić, Z. : Intensional FOL over Belnap's Billattice for Strong-AI Robotics. arXiv:2508.02774 [cs.LO], 1–29, (2025).
26. Tarski, A. : Der wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica* 1: 261–405. English translation, 'The concepts of truth in formalized languages', appeared in A.Tarski 1956, *Logic, Semantics and Metamathematics: Papers by Alfred Tarski from 1923 to 1938*, Oxford: Clarendon Press, (1936).
27. Ketland, J. : Deflationism and Tarski's paradise. *Mind* 108, 69–94, (1999).
28. Shapiro, S. : Truth and proof - through thick and thin. *Journal of Philosophy* 95, 493–522, (1998).
29. Boolos, G. and Jeffrey, R. : *Computability and Logic*. Third edition. Cambridge University Press, (1989).
30. Ketland, J. : A proof of the (strengthened) Liar formula in a semantic extension of Peano arithmetic. *Analysis* 60, 1–4, (2000).
31. Marek, W. and Truszczyński, M. : Autoepistemic logic. *Journal of the ACM* 38 (3), 588–618, (1991).
32. Majkić, Z. : Intensional logic and epistemic independency of intelligent database agents. 2nd International Workshop on Philosophy and Informatics (WSPI 2005), Kaiserslautern, Germany, April 10-13, (2005).
33. Nguyen, A. and Yosinski, J. and Clune, J. : Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 427–436, (2015).
34. Russell, S. : The purpose put into the machine. In Brockman J (ed) *Possible minds: 25 ways of looking at AI*, Chap. 3. Penguin Press, New York, 20–32, (2020).
35. Searle, J.R. : Minds, brains, and programs. In *Behavioral and Brain Sciences*, Volume 3, Issue 3, Cambridge University Press, (1980).
36. Harnad, S.: The Symbol Grounding Problem. *Physica D*, 335–346, (1990).
37. Mariarosaria, T. and Floridi, L. : The symbol grounding problem: A critical review of fifteen years of research. *Journal of Experimental and Theoretical Artificial Intelligence*, 17(4), 419–445, (2005).
38. Brooks, R.: *Cambrian Intelligence: The Early History of the New AI*. Cambridge, Massachusetts, The MIT Press, (1999).
39. Sing, H. and Aggrawal, M. and Krishnamurthy, B. : Exploring Neural Models for Parsing Natural Language into First-Order Logic. arXiv:2002.06544 [cs.CL], (2020).

7 Appendix: Neuro-symbolic Archetutture

The robot's internal neuro-symbolic knowledge structure is divided into four levels, in ordering: natural language $\mathcal{NL}_{list}$ (parsed into conceptual structures with PRP concepts, i.e., C-structures and formal IFOL logic with the set of formulae $\mathcal{L}$), intensional conceptual PRP system and neuro system, as represented by the following diagram (only two continuous arrows of intensional mapping $I : \mathcal{L} \rightarrow D_I$ represent the total mappings, while other (dots) are partial mappings)



Nat.Lang. *Log.semantic sys.* *Conceptual sys.* *Neuro sys.*

where the level of grounding of the natural language of the robot to its neuro system is composed by the following processes:

1. PR (Pattern Recognition) processes of all kinds, also with robot's deep learning neural network support. For example, for "the person in the room" particular in D_0 , its meaning is the recognition process (also from neural networks) which responds positively for a presented image with this image content.
2. A natural language query $nq \in \mathcal{NL}_{list}$ that is, logical proposition (sentence) $\phi = pars(nq) \in \mathcal{L}$, for example, "I have seen the person in the room", is labeled by its intensional proposition label $I(\phi) \in D_1$. Thus, the grounding of nq is obtained by linking its intensional proposition $I(pars(nq))$ in PRP to the SDC (Spatial Description Clauses) parser process (part of robot's neuro system).
3. ML (Machine Learning) processes, like that used for the recognition of different types of classes (like the set of videoclips). For example, for the language plural world "videoclips" in $\mathcal{NL}_{list}$, such that $pars("videoclips") = videoclips(y) \in \mathcal{L}$ with its intensional unary concept $u_2 = I(videoclips(y)) \in D_1$ which is grounded to robot's ML process for the "videoclips".

In this research, bridging the gap between Neural Networks (sub-symbolic) and Logic (symbolic) is not just about making them work side-by-side; it is about creating a formal translation layer where the robot's "internal feelings" (neuron firings) become "meaningful concepts" (logical terms). We achieve this via a Dual-System Architecture grounded in Intensional First-Order Logic (IFOL):

1. **System 1: The "Neuronal Grounding" Layer;**

System 1 consists of Deep Learning and Neural Networks. For us, this layer is responsible for pattern recognition and sensory-motor coordination.

- **The Output:** Instead of just outputting a label like "Blue," the neural network produces a specific internal state (a vector or firing pattern).
- **The Problem:** On its own, the neural network doesn't "understand" the concept; it just reacts. This is where the gap exists.

2. **The Bridge: Intensional Abstraction:**

This is our "secret sauce." We use intensional abstraction to turn the complex activity of the neural network into a "Logic Object."

- **Mapping:** Each distinct neural pattern is mapped to a specific intensional entity in the IFOL.
- **Meaning vs. Reference:** The "meaning" (intension) of a word like heavy is the specific neural experience of the robot's motors straining. The "reference" (extension) is the actual physical object being lifted.
- **Self-Labeling:** The robot uses autoepistemic reasoning to say, "I am currently experiencing Neural State X, which I have mapped to the concept 'Heavy'."

3. **System 2: The "Autoepistemic Logic" Layer:**

Once the neural patterns are abstracted into logical symbols, System 2 takes over. This is the deliberative, slow-thinking part of the brain.

- **High-Level Planning:** Because these symbols are now part of a formal logic (IFOL), the robot can use them in complex "If-Then" scenarios that neural networks struggle with.
- **Predefined set of axioms and prohibitions:** Any kind of reasoning results and plans must satisfy these axioms before the executions of actions derived from such plans. In this way the robots can not become danger for the humans and the environment in which they are active.
- **Feedback Loop:** System 2 can "query" System 1. For example, the symbolic layer might ask, "Check the visual sensors again; does that object match the 'Bird' intension?" This forces the neural network to re-process data based on logical needs.

4. **Solving the "Black Box" Problem:**

One of the most interesting aspects of this bridge is Explainability.

- In standard neural networks, we don't know why a robot chose an action.
- In this framework, because every neural state is linked to an intensional logic term, the robot can provide a symbolic trace of its reasoning: "I performed Action A because my neural sensors triggered the 'Danger' intension, which my logic defines as a state to be avoided".

Summary Table: The neuro-symbolic's Bridge

	Sistem 1 (Neural)	The Bridge (Intensional)	System 2 (Symbolic)
Function	Fast, reactive sensing	Abstraction & Grounding	Slow, logical planning
Data Type	Vectors / Tensors	Intensional Entities	Logic Formulas
Role	Perceives the world	Maps neurons to symbols	Reasons about perceptions
Self-awareness	Unconscious	The "I" links the two	Conscious autoepistemic thought