Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
DOI : 10.47974/JIOS-2229

Modelling the process of agent cognition under conditions of incomplete or incorrect information

Alexander Kuznetsov *
Department of Computer Science and Computer Graphics
Tashkent State Transport University
1 Temiryolchilar Street
Mirobod District
Tashkent 100167
Republic of Uzbekistan

Dmytro Sytnikov †
Wessex Institute of Technology
Ashurst Lodge
Ashurst
Southampton SO40 7AA
United Kingdom

Shoxrux Sultonov §
Department of Computer Science and Computer Graphics
Tashkent State University of Transport
1 Temiryolchilar Street
Mirobod District
Tashkent 100167
Republic of Uzbekistan

Abstract

We present a formal framework for modelling agent cognition under incomplete or uncertain information using ontology graphs. An agent's knowledge is represented as a potentially flawed reflection graph of a true ontology. We introduce three core measures to evaluate epistemic states: a correctness measure for statements, an entropy-like uncertainty

* *ORCID-ID:* **0000-0002-0365-2077**
† *ORCID-ID:* **0000-0003-1240-7900**
§ *ORCID-ID:* **0000-0002-0462-8677**

* *E-mail:* **avkuz@bk.ru (Corresponding Author)**
† *E-mail:* **dsytnikov@googlemail.com**
§ *E-mail:* **sultonovsh@yandex.ru**

© Taru Publications

measure, and a graph dissimilarity metric as an ontological distance. Cognition is modelled as an optimisation process where the agent refines its graph by exploring neighbourhoods defined by these measures. The theory is extended to multi-agent competitive scenarios. This graph-theoretic approach provides a formal foundation for hybrid neuro-symbolic AI, offering precise criteria to validate generative model outputs and detect hallucinations by measuring their deviation from a knowledge base.

Subject Classification: *68T27, 03B48, 68T42, 91A27.*

Keywords: *Formal ontology, Epistemic logic, Graph theory, Uncertainty, Neuro-symbolic AI, Validation.*

1. Motivating Scenarios and Problem Statement

This article examines scenarios where an agent's behaviour appears logically inexplicable within a known relational framework, but becomes explicable within a broader, unknown framework. One of the goals of this article is to construct a theory that allows for assessing the proximity and truth of agent statements, including incorrect ones, within a given ontology, as well as the proximity of ontologies themselves. A further goal is the formal description of the process of cognition, particularly creative cognition, when the result is formulated during the process without a rigidly defined template. In short, we aim to formally describe from the inside all the uncertainty of the agent's world.

Consider several examples that demonstrate the manifestations of subjective probability and demagoguery in the interaction of several agents, at least one of which is not fully aware of the situation.

Model Example 1 (we are all sea urchins, in some sense):
Let there be two faces: *Il Dottore* and *Brighella*. *Brighella* says:

- (Premise) All people are brothers since they descended from Adam and Eve[i].
- (Premise) Brothers must share everything equally.
- (Premise) You are a man and I am a man.
- (Conclusion) So, you have to share equally with me.

Formally, this syllogism is sound; moreover, all premises can be shared by the recipient. In fact, if a *Il Dottore* is asked to share resources that *Brighella* had no role in obtaining, even the most ordinary *Il Dottore* will suspect a trick. This “inference” is an example of demagoguery. Unlike the healthy syllogism of philosophers, this sequence of words is used by swindlers to lull their victims' attention. The swindler uses assonances and fragments of seemingly true statements that resonate with the victim, but their goal is to illogically extract the victim's resources. The victim must temporarily become “convinced” (in a purely psychological sense, by finding fragments of shared values in the swindler's words) and voluntarily hand over these values. Below, we will develop a toolkit

that will allow, among other things, to formally describe not only “healthy” inferences but also such demagoguery.

Model Example 2 (deterministic non-determinism):

A corrupt judge is hearing a lawsuit between two parties, *Pantalone* and *Il Dottore*. *Il Dottore* submits all the necessary documents to support his position. The *Judge*'s decision is non-deterministic for *Il Dottore* - all the necessary input signals have been given, and *Il Dottore*, like any honest citizen, has no direct influence on the internal state of the *Judge*.

Pantalone gives the *Judge* money, and the *Judge* denies *Il Dottore*'s demands, finding a loophole in the laws. For *Pantalone*, the *Judge* is deterministic.

Model Example 3 (false probability):

There is a narrow box, at the bottom of which there are two balls, black on the right and white on the left. *Brighella* knows the location of the balls, but *Il Dottore* does not. *Il Dottore* thinks that he will get the white ball with a probability of 1/2, while *Brighella* knows he will get the white ball with probability 1 if he arrives first.

All examples have something in common: the exploitation through deception based on the different levels of understanding of reality by two sides. For instance, in these scenarios, a participant may perceive an outcome as probabilistic, while globally - and to a better-informed agent - the process is deterministic.

We need to do three things to construct a formal description of the situations in *Model Examples* 1-3:

- Introduce a distance between entities in ontologies to understand which statements should be accepted and which should be discarded as too far from the area under study,
- Gradate the concept of truth to separate simply true statements from those that are both true and useful to one or another agent,
- Account for the fact that, in tasks involving the concept of probability, neither events nor the probability measure are fixed; they depend on the agent in question and the time of the event.

2. Background and Related Work

This article aims to formally model scenarios where an agent's behaviour, which appears probabilistic or unformalisable within a known framework, becomes deterministic within a broader framework. This requires two components: (1) a probability-like measure quantifying system and statement uncertainty, and (2) a syntactic representation of typically informal semantics.

2.1 *Ways to Compare Ontologies*

The usefulness of a statement can be expressed in terms of the distance between concepts in an ontology (as ontology is understood by mathematicians[1]).

Quintero et al. [2] present the DIS-C approach for estimating the conceptual distance between concepts in an ontology. DIS-C is graph-based in the sense that the entire topology of the ontology is taken into account when calculating the weights of relationships between concepts. The methodology consists of two main steps. First, to take advantage of prior knowledge, the ontology expert assigns initial weights to each of the relations in the ontology. An automatic method for computing conceptual relations then refines the weights assigned to each relation until a stable state is reached. The DIS-C framework can incorporate a wide range of relationships between concepts such as meronymy, antonymy, functionality, and causation.

Moreover, it is possible to compare entire ontologies [3]. Distances between ontologies are useful for searching, matching, or visualising ontologies. A library of such distances is provided by the NeOn toolkit plugin, OntoSim.

One of the most important concepts when comparing ontologies is *semantic similarity* [4]. Ultimately, semantic similarity can be measured as graph similarity (if the ontology is represented by a graph) or via corpus-based measures that allow comparison of language units through analysis of unstructured or semi-structured texts. They are most often used to compare words, sentences or texts based on natural language processing techniques that most often rely only on statistical analysis of the use of words in texts, for example, by analysing word usage and their linguistic contexts.

In this work, the graph-based description of the ontology is, to a certain extent, close to description logic that describes the relationships of concepts, roles, and individuals [5]. The edges of ontology graphs correspond to atomic roles, and the vertices correspond to concepts of description logic. Note that identifying graphs and logical formulas is a well-known approach; see, for example, [6, 7].

Before moving on to the description of various probabilistic logical systems, we point out that ontologies can be compared not only via their topology but also via the attributes of their objects, using rough set theory [8]. This theory considers sets of objects that have attributes. These sets are divided into equivalence classes according to the P-indiscernibility relation, which identifies objects with attributes that coincide in some sense (for example, those for which the values of the function specified on the set of attributes coincide). The upper P-approximation of a set is the union of all equivalence classes that have a non-empty intersection with a given set. The lower P-approximation of a set is the union of all equivalence classes included in the given set. Any set can be approximated by a rough set, which is a pair of upper and lower P-approximations of the set.

2.2 *Formalising Uncertainty and Partial Knowledge*

Our work is also related to one of the possible formalisations of the concept of subjective probability (in the sense understood by logicians, beginning with Ramsey [9]) and the theory of incomplete or imprecise knowledge associated with this formalisation. Subjective probability is usually understood as the degree of an agent's belief that a certain event will occur, in contrast to statistical probability, understood as the frequency with which an event occurs. Typically, scientists indignantly reject incomplete or imprecise knowledge as delusions, as well as events that occur once or twice and are not reproduced more than once. Here one might object that history, for example, consists entirely of isolated, irreproducible events. However, history as a science studies not these isolated, irreproducible events, but series of typologically similar events and objects. For example, a single archaeological culture, say, Sintashta, includes hundreds of virtually identical pots, very similar burials, very similar types of arrowheads, and so on. A single, unique arrowhead means nothing to a historian; such a find provides zero information, since it does not allow any conclusions to be drawn about the society that produced this object. However, in reality, isolated events occur, and from half-knowledge or knowledge tainted by error, it is entirely possible to extract pure, infallible knowledge, although there are few theories of this process.

The most systematic study of logical probability belongs to Carnap [10]. His formulation of logical probability [11] begins with the construction of a class of languages consisting of a finite number of logically independent monadic predicates (*naming properties*) applied to a countable set of individual constants (*naming individuals*) or variables, as well as ordinary logical connectives. The strongest (*coherent*) statements that can be made in a given language describe the universe in as much detail as the expressive power of the language allows. They are conjunctions of complete descriptions of each individual, each description itself being a conjunction containing exactly one occurrence (with or without negation) of each predicate of the language. These strongest statements are called *state descriptions*.

Any probability measure over state descriptions is automatically extended to a measure over all sentences, since each sentence is equivalent to a disjunction of state descriptions. We can define a probability measure to describe states so that descriptions encompassing several possible states of the universe would have a lower probability than states that correspond to only one state of the universe.

The article [12] shows that while logic provides an attractive formalism for representing knowledge and reasoning, there are many aspects of human reasoning that are not adequately handled by pure logic. For example, suppose a person is told that “all aliens are blue”. However, the person later also learns the fact that “Klingon-type aliens are not blue”. Although this new fact contradicts the original statement, the person will not have any difficulty in resolving the conflict. Most people will view Klingon-type aliens as a special case, and otherwise all other aliens will be considered blue. However, an automatic computer inference system cannot cope with such inference. The idea in the

article's approach is to assign a weight to each logical formula, which is interpreted as the "cost" of breaking the formula to handle inconsistencies and non-monotonic output patterns.

Systems of this type, one might say, "with fuzzy inference" should be distinguished from fuzzy logic systems, in which the concept of truth itself, rather than logical formulae, is graded. See, for example, the article [13], which provides a structured introductory overview of the landscape of weighted logics (in a general sense) that can be found in the artificial intelligence literature, highlighting their fundamental differences and their areas of application.

The idea of using probability theory in logic is quite old and can be seen in the work of Cox [14], describing the system in which the plausibility of a proposition is a real number and depends on the information we have about the proposition. Also, likelihoods should vary reasonably depending on the estimate of the likelihoods in the model, and if the plausibility of a proposition can be inferred in many ways, all results must be equal.

Related is the approach of evidence theory, or Dempster-Shafer theory (DST) [15]. DST is a generalisation of the Bayesian theory of subjective probability. Belief functions base degrees of belief for one issue on the subjective probabilities for a related issue. The degrees of belief themselves may or may not have the mathematical properties of probability; how different they are depends on how closely the two issues are related. Combination rules that reflect general assumptions about the data also contribute.

DST is based on two ideas: deriving degrees of belief for one question from subjective probabilities for a related question, and Dempster's rule for combining such degrees of belief when they are based on independent elements of evidence. Essentially, the degree of belief in a proposition depends primarily on the number of answers (to related questions) containing the proposition and the subjective probability of each answer.

Chkhartishvili [16] and colleagues examine the structure of agents' awareness in a reflexive game. The works introduce the concept of a set of states of nature and a set of possible worlds. There is a real world and, for each agent, worlds that he considers possible due to his awareness. Agents interact and observe the results of their actions, as a result of which the structure of their awareness changes. In this work, the focus is not on the interaction mechanism itself, but on the internal structure of the awareness function and its restructuring due to the agent's activity.

2.3 *Philosophical and Epistemological Foundation*

The second part of the problem under consideration is the development of formal semantics. Formal semantics allows us to treat statements in object languages as if they were ordinary formulas. Alfred Tarski pioneered this area in his article [17], which discusses expressing the concept of truth through the satisfiability of a propositional function. In the T-schema, the meaning of various parts of a statement is determined by recursively defining interpretation functions that map statements to predefined mathematical sets.

Unlike many well-known model-theoretic methods in formal semantics, our article assumes a virtually unambiguous interpretable object language, upon which languages that are simpler for practical use, but less unambiguous, can be built. The theory presented in this article is the result of psychological introspection and proceeds from the premise that cognition can be described by a sequence of layers of linguistic constructs, with the words of the lowest layer being words of the "language of nature" - primitives of colours, smells, tastes, and so on, evolved in the brain to describe the basic physical properties of the surrounding world. For example, the colour red, as it appears to humans, is also a word within the proposed system, albeit a specific one. The upper layers of descriptions will correspond to concepts and informal inferences. Strict formal logical inferences will reside somewhere in the middle layers of constructs. The deeper one must descend through these layers to reach a correct conclusion, the more complex, in a sense, the inference itself will be. When working with information and making decisions, it is always necessary to strike a balance between the detail of the model and the speed of decision-making. Sometimes the crudest model allows for a quick conclusion, while a detailed description only confuses the decision maker. Our work will be largely devoted to formalising this approach.

The article describes existence from the perspective of subjective idealism and reduces the world to atomic psychic phenomena, their interconnected groups, and their language, rather than to physical, objective reality. Reality itself is completely excluded, as something inaccessible to perception except through phenomena. In this respect, the theory is close to Indian philosophies (from Vaiśeṣika to Yogāchāra), with similar psychic atoms that also form a unique language [18, 19, 20]. Thus, the lower layers of ontologies will not contain the words or terms of some comprehensive textbook, like the "Dove Book" of Russian legends which contains all possible knowledge, but rather phenomena, the existentials of dasein, as understood by Heidegger [21], or dharmas, as understood by Dharmakīrti - that is, minimal units of mental experience. The atomic features themselves are established based on the applied problem based on the applied problem and the personal experience of the model's author. The history of empirical classification of phenomena spans hundreds of years. Even ancient philosophers were engaged in isolating such atoms from their own consciousness; it's enough to recall "A Clear Introduction to the 100 Dharmas" (Mahāyāna śatadharmā-prakāśamukha śāstra), by Vasubandhu. In this model, there are atoms of formal logical primitives such as generalisation and analysis of features, atoms of emotions, colours, shapes, sounds, and the like. Various relations of phenomena following each other, discrimination, generalisation, naming of phenomena, time, speed, etc. are also considered as a separate class of phenomena not directly observable by ordinary consciousness.

Since the article indirectly addresses the problem of human consciousness and cognition, the authors must also consider contemporary works on this topic. Numerous works have now been published that differ little in meaning from Julien de La Mettrie's "L'homme machine" (The Machine Man), who, back in

1747, defined man as a complex mechanism, like a watch, only more sophisticated. Proponents of all sorts of neurosciences and specialists in advanced mechanisms continue to reduce the diversity of mental phenomena to the work of purely material objects, such as neurons in the brain. The phenomenon of consciousness itself, i.e., being in existence, is declared an emergent phenomenon, as if such a declaration could explain anything.

The authors will not attempt to define the undefinable, but will simply settle for the fact that the symbols describing reality at the lowest levels will correspond to the features directly perceived by the agent, no matter what these features are.

Consequently, our theory does not attempt to define consciousness or reduce it to material substrates. Instead, it takes the agent's perceptual interface as its foundational given. This makes it well-suited for engineering autonomous systems (e.g., robots, UAVs), which have no direct access to an objective world but only to sensor-derived data streams. For instance, the generic IoT architecture proposed by Mageswari et al. [22] implements a pragmatic three-level hierarchy: local ontologies for individual sensors, global ontologies for device groups, and a central integrator ontology. This structure directly mirrors our formal layers, where lower levels consist of unambiguous perceptual primitives (sensor data) that are mapped and abstracted into higher-level operational concepts. Our framework provides the formal graph-theoretic foundation to describe, compare, and optimise such hierarchical knowledge integration processes.

3. Graph-Theoretic Model of Agent Cognition

3.1 *Ontology Graphs and Agent Reflections*

Let us describe a method for constructing a set of levels of formal ontologies describing the world. The lower levels are described in a very primitive and unambiguously interpretable language in which the syntactic and semantic similarity of statements is one and the same. Upper-level ontologies operate with more expressive languages and, at the same time, define mappings from lower to upper layers, mapping the sets of primitive entities of lower-level ontologies to the entities operated by the agent. As a result, it will be possible to formalise and compare incorrect and poorly formulated agent statements by returning to lower ontology levels, and to assess the degree to which the agent's statements are “poor quality” based on the number of levels to descend. Although this article describes a theory related to mathematical logic, it will not discuss whether a variable has the value “true”, “false” or something in between, but rather whether it carries information about a particular phenomenon. Therefore, the article will introduce various measures of how much information is contained in a given statement. These measures will be called “correctness” and “reliability”, referring to the possibility of interpreting them in this way in certain specific cases.

Let a set of agents $Ag = \{ag_1, \dots, ag_m\}$ be given and an ontology $\mathbf{O}$ corresponding to the set of phenomena perceived by these agents, together with the time of perception. For simplicity, we will assume that the ontology corresponds to the ontology graph

$$\Gamma = (V, E, \Lambda_V, \Lambda_E, I, \varphi_V, \varphi_E),$$

whose vertices V correspond to instances and concepts, and the multiset of edges with the underlying set $E \subseteq V \times V$ corresponds to ontology relations. Here $I: V \times V \to \{\mathbf{0}, \mathbf{1}\}$ is the incidence function, $\Lambda_V \subseteq \mathbb{R}^p$, $\Lambda_E \subseteq \mathbb{R}^q$ are sets of attributes of vertices and edges, i.e., it is assumed that all attributes are encoded by vectors of real numbers, $\varphi_V: V \to \Lambda_V \subseteq \mathbb{R}^p$ is the vertex attribute function, $\varphi_E: E \to \Lambda_E$ is the edge attribute function , $p > 0$, $q > 0$. The symbol $\mathbf{0}$ denotes "False", the symbol $\mathbf{1}$ denotes "True". Sometimes, for simplicity, we will write the corresponding character string instead of a real number attribute, keeping in mind that characters can also be encoded numerically. We denote the set of all such ontology graphs $\mathcal{G}$.

We assume a set $\mathcal{P}$ of ontology subgraph selection operators is defined on $\mathcal{G}$. Each $\mathcal{P} \in \mathcal{P}$ maps $\Gamma \in \mathcal{G}$ to its subgraph $\mathcal{P}\Gamma \subseteq \Gamma$, and during the mapping the attribute vectors of vertices and edges may not go into $\mathcal{P}\Gamma$ completely, but only with some of their components. Furthermore, when we refer to an ontology graph Γ, we mean not only the universal ontology graph $\Upsilon \in \mathcal{G}$, (containing all possible world information) but also any fragment $\mathcal{P}\Upsilon$ thereof. The graphs under study are finite in all cases.

For each agent, a reflection $\mathbf{O}_{ag}$ of the ontology $\mathbf{O}$ is defined. It is the known (not necessary exact) to an agent ag part of $\mathbf{O}$. The corresponding reflection ontology graph is

$$\Gamma_{ag} = (V_{ag}, E_{ag}, \Lambda_{V_{ag}}, \Lambda_{E_{ag}}, I_{ag}, \varphi_{V_{ag}}, \varphi_{E_{ag}}) \in \mathcal{G},$$

where V_{ag} is the set of vertices Γ_{ag}, E_{ag} is the set of edges Γ_{ag}, I_{ag} is the incidence function of Γ_{ag}, the function of vertex attributes $\varphi_{V_{ag}}: V_{ag} \to \Lambda_{V_{ag}} \subseteq \mathbb{R}^p$, edge attribute function $\varphi_{E_{ag}}: E_{ag} \to \Lambda_{E_{ag}} \subseteq \mathbb{R}^q$.

Note that the agent's knowledge about the ontology may not correspond to the ontology itself and Γ_{ag} may not be a subgraph of Γ. We distinguish the agent's internal representation Γ_{ag} from the referential framework Γ, which may correspond to an external reality or a more complete model.

The ontology must include inference rules $\mathbf{R} = \{R_j\}$ and the corresponding graph $\Gamma_{\mathbf{R}} \subset \Gamma$, and they must be known to each agent $ag \in Ag$, i.e., $\Gamma_{\mathbf{R}} \subset \Gamma_{ag}$.

Let each agent $ag \in Ag$ have a set Prop_{ag} of possible statements about the entities of the ontology $\mathbf{O}$. These statements can be divided into false, true, meaningless statements of the first type, and meaningless statements of the second type. Only statements about vertices, edges, and incidence that match the actual graph Γ are true. It is also possible to make meaningless or incorrect statements about edges and vertices that do not exist in Γ.

True or false statements $\mathcal{TTF}$ about entities $\mathbf{O}$ include statements of the following types

$$\mathcal{TTF}_1(v_1, v_2): I_{ag}(v_1, v_2), v_1, v_2 \in V, \quad (1)$$

$$\mathcal{TTF}_2(v): \varphi_{V_{ag}}(v) \in \Lambda_{V_{ag}} \subseteq \mathbb{R}^p, v \in V, \quad (2)$$

$$\mathcal{TTF}_3(e): \varphi_{E_{ag}}(e) \in \Lambda_{E_{ag}} \subseteq \mathbb{R}^q, e \in E. \quad (3)$$

as well as logical formulae derived from the above and formulae derived from them using the inference rules **R**.

Atomic meaningless statements of the first type $\mathcal{TN}1$ include statements about vertices, edges and attributes that are not in the graph Γ. At the same time, the graph Γ can be embedded in some extension graph $\tilde{\Gamma}_{ag}$ in which such vertices, edges and attributes are possible, and, accordingly, for $\tilde{\Gamma}_{ag}$ these statements are true or false:

$$\mathcal{TN}1_1\colon I_{ag}(v_1, v_2), v_1 \in V_{ag}\backslash V \vee v_2 \in V_{ag}\backslash V, \tag{4}$$

$$\mathcal{TN}1_2\colon \varphi_{V_{ag}}(v) \in \Lambda_{V_{ag}}\backslash\Lambda_V \wedge v \in V, \tag{5}$$

$$\mathcal{TN}1_3\colon \varphi_{E_{ag}}(e) \in \Lambda_{E_{ag}}\backslash\Lambda_E \wedge e \in E, \tag{6}$$

$$\mathcal{TN}1_4\colon \varphi_{V_{ag}}(v) \in \Lambda_V \wedge v \in V_{ag}\backslash V, \tag{7}$$

$$\mathcal{TN}1_5\colon \varphi_{E_{ag}}(e) \in \Lambda_E \wedge e \in E_{ag}\backslash E, \tag{8}$$

$$\mathcal{TN}1_6\colon \varphi_{V_{ag}}(v) \in \Lambda_{V_{ag}}\backslash\Lambda_V \wedge v \in V_{ag}\backslash V, \tag{9}$$

$$\mathcal{TN}1_7\colon \varphi_{E_{ag}}(e) \in \Lambda_{E_{ag}}\backslash\Lambda_E \wedge e \in E_{ag}\backslash E. \tag{10}$$

Here, the type of statements (4) corresponds to the establishment of a non-existent relationship between entities; the types of statements (5), (6) correspond to the attribution of non-existent attributes to existing entities and relations; the types of statements (7), (8) correspond to the attribution of existing attributes to non-existent entities and relations; the types of statements (9), (10) correspond to the attribution of non-existent attributes to non-existent entities and relationships.

Logical formulae derived from formulae of types (4)-(10), as well as formulae derived from them using inference rules **R**, will also be classified as meaningless statements of the first type $\mathcal{TN}1$. Denote by $\mathcal{W}$ the set of graph expansion operators $\mathrm{W}_{ag}\colon \mathcal{G} \to \mathcal{G}$, $\mathrm{W}_{ag}\colon \Gamma \mapsto \tilde{\Gamma}_{ag}$.

Meaningless statements of the second type $\mathcal{TN}2$ include statements that are made regarding objects that are not representable as vertices, edges and attributes of Γ and for which it is impossible (within the constraints that we initially establish when formalising the ontology) to construct a graph $\tilde{\Gamma}_{ag} \supset \Gamma$ for which these statements will be true or false. For example, for an approach in which all possible attributes in the world are already represented as numbers, this would be a statement like

$$I_{ag}(\text{“}Quack - quack''\text{,”}runcible'') = \mathbf{0}$$

and, moreover, incorrectly composed formulae like

$$I_{ag}(\text{“}Quack - quack''\text{,}\sqrt{2/}) = (,$$

since $\text{“}Quack - quack'' \notin \Lambda_V$, $\text{“}runcible'' \notin \Lambda_V$.

Letter combinations like “runcible” were taken by the authors intentionally and reflect the possibility of mathematical work with a text containing words that only resemble real ones. The source of inspiration was the famous “Glokaya Kuzdra” by academician L.V. Shcherba, an artificial phrase based on the Russian

language, in which all root morphemes are replaced with meaningless combinations of sounds [23].

We will further assume that agents do not make meaningless statements of the second type, although, in reality, they are quite common in informal reasoning and are not as useless as they might seem, since each meaningless statement of the second type $P \in \mathcal{TN}2$ generates a family of meaningless statements of the first type $\mathcal{TN}1$ or even true and false statements $\mathcal{TTF}$ close to P in some metric of formulae, compared as strings or as syntax trees of formulae (see, for example, [24]).

True or false statements $P \in \mathcal{TTF}$ generate the graph Γ^0_{ag}. For true statements of subtype (1) and their conjunctions, the generated graph Γ^0_{ag} can be embedded as a subgraph in the graph Γ, possibly ignoring the labelling of vertices and edges, $\Gamma^0_{ag} = \mathcal{P}_{ag}\Gamma$, $\mathcal{P}_{ag} \in \mathcal{P}$. Meaningless statements of the first type $P' \in \mathcal{TN}1$ also generate a graph Γ^1_{ag}, which can be embedded as a subgraph in the extension graph $\tilde{\Gamma}_{ag} = \mathrm{W}_{ag}\Gamma^1_{ag}$.

The agent, not knowing the ontology it is learning in full, and describing, in essence, not the entire ontology of Υ, but only its fragment $\Gamma = \mathcal{P}\Upsilon$, must take into account the possible need to go beyond the studied fragment into a wider graph $\Gamma' = \mathcal{P}'\Upsilon \supset \Gamma$. Knowing neither the names of vertices, nor possible edges, nor attributes in this wider graph, the agent must have a set of symbol vertices $V^S_{ag} = V_{ag}\backslash V$, symbol edges $E^S_{ag} = E_{ag}\backslash E$ and symbol attributes $\Lambda^S_{V_{ag}} = \Lambda_{V_{ag}}\backslash\Lambda_V$, $\Lambda^S_{E_{ag}} = \Lambda_{E_{ag}}\backslash\Lambda_E$, which will be used to designate entities beyond the current horizon of knowledge for the agent (for vertices and edges of $\Upsilon\backslash\Gamma$). Any statement about such symbols will be a meaningless statement of the first type. However, sometimes it is better to allow the agent to fantasise about the properties of the ontology than to strictly limit the possible attributes, vertices and edges, since otherwise we may end up with the inability to express the ontology in terms accessible to us and obtain meaningless statements of the second type. Symbols are analogues of variables in first-order logic. The number of possible $\mathcal{TN}1$ statements grows combinatorially with the size of the agent's symbolic sets V^S_{ag}, E^S_{ag}, etc.

3.2 *Illustrative Example: Grounding Unknown Terms*

Let us explain the above construction with an example. Consider the statement P_0 ="Ioannes is wearing a red jacket, white shirt, blue tie". This statement can be represented as a graph Γ with vertices and edges from Table 1.

In our ordinary world such a statement would not cause any surprise or difficulty; it could be true or false. Let us compare this with the statement P_1 ="Ioannes is wearing a red jacket, a white shirt, a blue tie and a gostak". We do not know what a "gostak" is (a statement of $\mathcal{TN}1_1$ type), so we (observer) cannot immediately assess the truth or falsity of the statement, but we can imagine a society in which a "gostak" is a wristwatch and obtain an extension graph $\tilde{\Gamma}_{\mathrm{We}}$, containing additional vertices: v_8 "gostak #1", v_9 "wristwatch", and edges (v_0, v_8) "Has an item of clothing", (v_8, v_9) "Belongs to class".

Table 1

Edges and vertices of the graph Γ

Vertex	Vertex Label
v_0	Ioannes #1
v_1	Jacket #1
v_2	Shirt #1
v_3	Tie #1
v_4	Red #1
v_5	White #1
v_6	Blue #1
v_7	Present #1

Edge	Edge Label
(v_0, v_1)	Has an item of clothing
(v_0, v_2)	Has an item of clothing
(v_0, v_3)	Has an item of clothing
(v_1, v_4)	Has colour
(v_2, v_5)	Has colour
(v_3, v_6)	Has colour
(v_7, v_0)	Time is

It is worth noting that checking the correspondence of $\tilde{\Gamma}_{\mathrm{We}}$ to reality requires an additional verification of the truth of the statement (v_8, v_9) "The gostak is a wristwatch", after which we can also check the statement (v_0, v_8) "Ioannes is wearing a gostak". Thus, the operation of expanding a meaningless statement of the first type P_1 up to a verifiably true or false one has the cost of checking one additional statement.

We can also make the statement P_2 ="Ioannes is wearing a red jacket, a white shirt, a blue tie, and a rath's hat" (and consider this as a statement of $\mathcal{TN}1_2$ type). We can also expand the assertion graph by adding the definition of the "rath's" attribute: "rath's hat is a green hat" and then we will need to check this statement in addition to checking the fact that Ioannes is really wearing a hat and that this hat is green.

3.3 *Correctness Measure for Agent Statements*

We come to the conclusion that for any expansion operator $\mathrm{W}_{ag} \in \mathcal{W}$ we can determine the cost of expansion $\mathrm{PW}_{ag}\mathrm{P}$ as the number of additionally introduced vertices and edges (or, what is the same, the number of additional statements, the truth or falsity of which the agent $ag \in Ag$ must check). In the above example, the cost of expansion is equal to 1. However, if we expand the statement P_1 by adding "Gostak is a sepulka, and a sepulka is a wristwatch", then the cost of the expansion will be more than 2, since it will be necessary to additionally check the truth of the statement "Gostak is a sepulka".

The atomic statements themselves can be divided as (1) - (3). However, since the statement $\mathcal{TTF}_1(v_1, v_2)$ is, in essence, the statement "The object v_1 is related to the object v_2", which is meaningless, it is advisable to consider instead of $\mathcal{TTF}_1(v_1, v_2)$ as atomic the statement $\mathcal{TTF}'_1(v_1, v_2; \beta) = \mathcal{TTF}_1(v_1, v_2) \wedge \mathcal{TTF}_3(v_1, v_2) = \beta$, i.e., "Object v_1 has a relationship of type β to object v_2".

Let

$$\mathcal{P}_{ag}^0 = \{\mathcal{TTF}'_1(v_1, v_2; \beta), \mathcal{TTF}_2(v), \mathcal{TTF}'_3(e) | v, v_1, v_2 \in V, e \in E, \beta \in \Lambda_{E_{ag}}\}$$

be the set of all possible statements $ag \in Ag$ of the above types.

Any statement of the form $f(P_1, \dots, P_q)$, f is a Boolean function of q variables, $P_i \in \mathcal{P}_{ag}$, $i = \overline{1, q}$ about the graph Γ can be represented as a perfect disjunctive normal form:

$$f(P_1, \dots, P_q) = \bigvee_{f(\sigma_1, \dots, \sigma_q)=\mathbf{1}} \bigwedge_{i=1}^{q} P_i^{\sigma_i}, \tag{11}$$

where

$$x^\sigma = \begin{cases} x, & \sigma = 1; \\ \overline{x}, & \sigma = 0. \end{cases}$$

We denote the set of forms of the form (11) as $\mathcal{PDNF}(q)$.

Using this form, we can calculate the logical probabilities and other similar measures of the corresponding conjunctions and disjunctions in a similar way to [11]. The longer the elementary conjunction, the more accurately it describes the graph Γ, but the more elementary conjunctions, the greater the uncertainty in the description of the graph Γ. We denote $\overline{\sigma} = (\sigma_1, \dots, \sigma_q)$, $\mathrm{W}_{ag} P = \mathrm{Identity}\ P = P$ - the identity transformation of P if $P \in \mathcal{TTF}$ and define

$$\mathrm{Truth}\,(P, \mathrm{W}_{ag}) = \begin{cases} 0, P = \mathbf{0}, \\ 1, P = \mathbf{1}, \\ \dfrac{1}{1 + \mathrm{P}\mathrm{W}_{ag}\mathrm{P}}, \mathrm{W}_{ag} P = \mathbf{1}, P \in \mathcal{TN}1. \end{cases}$$

Let us denote $\mathrm{W}_{ag} = (\mathrm{W}_{ag}^1, \dots, \mathrm{W}_{ag}^q)$ and define the following function

$$correct(f(P_1, \dots, P_q), \mathrm{W}_{ag}) = \prod_{f(\overline{\sigma})=\mathbf{1}} \frac{1}{q} \sum_{i=1}^{q} \mathrm{Truth}(P_i^{\sigma_i}, \mathrm{W}_{ag}^i)$$

and, based on this function, a measure of the "correctness" of the statement:

Definition 1: Let us call the measure of the correctness of the statement $f(P_1, \dots, P_q)$ the number

$$\mathrm{correct}(f(P_1, \dots, P_q)) = \min_{\mathrm{W}_{ag} \in \mathcal{W}^q} correct(f(P_1, \dots, P_q), \mathrm{W}_{ag}). \tag{12}$$

This definition extends naturally from an SDNF of the form (11) to any DNF. The measure of "correctness" is needed to extend the measure Truth from atomic statements to disjunctions of conjunctions of atomic statements. The name "correctness" was chosen because a correct logical statement should not contain any ambiguity.

For true statements $f(P_1, \dots, P_q)$ it is true that $\text{correct}(f(P_1, \dots, P_q)) = 1$, for false $\text{correct}(f(P_1, \dots, P_q)) = 0$. The longer the statement and the more true terms $P_i^{\sigma_i}$ it contains, the closer $\text{correct}(f(P_1, \dots, P_q))$ is to 1, and vice versa, the shorter the statement and the fewer true terms $P_i^{\sigma_i}$ it contains, the $\text{correct}(f(P_1, \dots, P_q))$ is closer to 0.

By definition, $\text{correct}(f(P_1, \dots, P_q) \vee g(P_{q+1}, \dots, P_r)) = \text{correct}\,(f(P_1, \dots, P_q))$ $\text{correct}(g(P_{q+1}, \dots, P_r))$, if $f \in \mathcal{SDNF}(q)$, $g \in \mathcal{SDNF}(r)$ and $\{P_1, \dots, P_q\} \cap \{P_{q+1}, \dots, P_r\} = \emptyset$.

To assess the "correctness" of an elementary conjunction, the addition of the truth measures of the variables included in it was chosen, since the more information each variable of the conjunction carries, the more information the conjunction as a whole carries (the more is known about each variable, the easier it is to make a decision regarding the object characterized by these variables). It is impossible to multiply the truth measures of variables, since some variables may carry zero information (be "false"), moreover, in this case, the total amount of information in the conjunction will decrease if the Truth measure in its component variables is less than one, while the entropy type measure must be monotonic.

Between themselves, the measures for elementary conjunctions are multiplied, since the disjunction corresponds to the uncertainty of the description of the world (either one option or the other, or both at once) is true, and with the uncertainty of the description, the agent's awareness should decrease (it becomes less clear what decision to make). Thus, we can roughly say that our measure of "correctness" indicates, for example, how much useful information about the ontology is contained in the agent's statement. Next, we will formulate a similar measure in the language of graphs.

Let us illustrate the introduced measure with an example. Let the statement P ="Ioannes wear a red jacket, a white shirt, a blue tie and a gostak" be given and we expand this statement by adding that "the gostak" is a wristwatch. Let Ioannes really be wearing a jacket (P_1) of red colour (P_2), a shirt (P_3) of white colour (P_4), a tie (P_5) of blue colour (P_6) and a wristwatch (P_7). Then

$$\text{correct}(P) = \frac{1}{7}\left(1 + 1 + 1 + 1 + 1 + 1 + \frac{1}{1+1}\right) = \frac{13}{14}.$$

3.4 *Uncertainty and Probabilistic Error Measure*

The measure of the correctness of the statement from the previous section is useful, but has a limitation: the agent usually does not have complete knowledge and cannot assess how correct the statement was made. The one who evaluates the agents cannot evaluate this according to the measure (12). Let us denote as $\text{n}(P)$ the number of occurrences of the expression P in the right side of the expression (11). Next, for all P_i, $i = \overline{1, q}$, we determine the ratio of the difference in the number of occurrences of the expression P_i without negation and with its negation P_i^0 on the right side of the expression (11) to the total number of occurrences of P_i on the right side of the expression (11)

$$\mathrm{M}(P_i^1) = \frac{\mathrm{n}(P_i^1) - \mathrm{n}(P_i^0)}{\mathrm{n}(P_i)}.$$

Let a complete description of the ontology graph Γ require a conjunction r of statements $P_1 \wedge \ldots \wedge P_r$, but the agent has access only to information about the conjunction q of the required statements, $1 \leq q < r$.

Definition 2: Let us call the total uncertainty of the statement the number

$$\mathrm{N}(f(P_1, \ldots, P_q)) = 1 - \frac{1}{r}\sum_{i=1}^{q} |\mathrm{M}(P_i^1)|. \quad (13)$$

N can be considered both as a measure of the entropy type (how definitely the existing statements describe Γ) and as a measure of the probability type (how likely a possible error in the description of the graph is). Recall that each conjunction in DNF corresponds to some version of the graph description, and their disjunctions correspond to how many descriptions of the graph-ontology the agent assumes are possible.

Thus, N is a quantitative measure of how much we can extract from a statement of error-free information, given that the statement has mutually exclusive substatements. In a statement that has no mutually exclusive parts, all information is error-free, so N for such a statement is zero. However, $\mathrm{N}(f)$ is larger, the more mutually exclusive options there are in the statement f. If f contains all mutually exclusive options for describing the graph, then $\mathrm{N}(f) = 1$. Of course, if we build a description of the graph containing all possible options through "or", then some option will definitely turn out to be correct, and such a description will be reliable, but not informative. The fewer "or" with mutually exclusive options there are in the graph description, the smaller $\mathrm{N}(f)$ and the more informative the description, but the greater the likelihood that a real-life graph is not taken into account by the description.

Let us define N so that for a conjunction of 0 terms $P_\emptyset$ there is $\mathrm{N}(P_\emptyset) = 0$. Let us denote as $P_{All} \in \mathcal{PDNF}(q)$ a PDNF, in which for each conjunction $P_1^{\sigma_1} \ldots P_q^{\sigma_q}$ in its composition also occurs the conjunction $\overline{P_1^{\sigma_1}} \ldots \overline{P_q^{\sigma_q}}$. The statement P_{All} actually lists all possible variations of the graph Γ and says that at least one of them is possible. It corresponds to the maximum uncertainty of the world.

Statement 1: The function $N: \mathcal{PDNF}(q) \to [0,1]$ has the properties of an analogue of a measure and the $\sigma-$additive property under the disjunction operation:

- $\mathrm{N}(P_\emptyset) = \mathrm{N}(P_i) = 0,\ i = \overline{1,q}$,
- $\mathrm{N}(P_{All}) = 1$,
- $\mathrm{N}(\vee_{i=1}^{\infty} CP_i) = \sum_{i=1}^{\infty} \mathrm{N}(CP_i)$, $CP_i \in \mathcal{PDNF}(q)$, CP_i and CP_j do not contain common atoms $P_k,\ i \neq j,\ k = \overline{1,q}$.

The third property obviously follows from the definition (13).

For example, the statement $P =$ "*Pantalone* gave a bribe to the *Judge* (P_1), and *Il Dottore* provided all the evidence of his rightness in court (P_2) or *Pantalone* did not bribe the *Judge* ($\overline{P}_1$), and *Il Dottore* provided all the evidence of his rightness in court (P_2)" has in a naturally defined fragment of the ontology that describes these relationships

$$\mathrm{N}(P) = 1 - \frac{1}{2}\left(\frac{1-1}{2} + \frac{2-0}{2}\right) = \frac{1}{2}.$$

3.5 *Graph Dissimilarity as an Ontological Distance*

To quantify the divergence between an agent's reflection and the true ontology, or between different agents' reflections, we require a formal distance measure between ontology graphs. We employ a graph dissimilarity metric of the *graph edit distance* (GED) type, a special case of the Gromov-Hausdorff distance [25, 26]. This metric evaluates the minimal cost of transforming one ontology graph into another through a series of elementary edit operations (vertex/edge insertions/deletions, attribute changes) under an optimal mapping.

Let $\Gamma_0 = (V_0, E_0, \varphi_{V_0}, \varphi_{E_0})$ and $\Gamma_1 = (V_1, E_1, \varphi_{V_1}, \varphi_{E_1})$ be two ontology graphs. Let $\mathcal{X}$ be the set of admissible injective mappings $\chi: V_0 \circ V_1$ that preserve adjacency. For a given χ, vertices and edges in Γ_0 without a correspondent in Γ_1 (and vice versa) incur a unit penalty. For matched elements, their dissimilarity is computed via predefined, bounded functions: $\psi_V: \Lambda_V \times \Lambda_V \to [0,1]$ for vertex attributes and $\psi_E: \Lambda_E \times \Lambda_E \to [0,1]$ for edge attributes. The aggregate vertex and edge penalties under χ are denoted $\Phi_V(\chi)$ and $\Phi_E(\chi)$, respectively. The detailed construction ensuring metric properties follows the approach in [27].

The overall dissimilarity between Γ_0 and Γ_1 is defined as the minimal average cost over all such mappings.

Definition 1: (Graph Dissimilarity). Let $\alpha_V, \alpha_E > 0$ be weights balancing vertex and edge contributions. The graph dissimilarity is

$$diss(\Gamma_0, \Gamma_1) = \min_{\chi \in \mathcal{X}} \frac{\alpha_V \Phi_V(\chi) + \alpha_E \Phi_E(\chi)}{\alpha_V + \alpha_E}.$$

This measure serves as our formal *ontological distance*. In our layered model (see Section 2.3), the base layer of perceptual primitives ensures that syntactic graph similarity is a valid proxy for semantic similarity. For higher-level concepts, vertices must first be decomposed into their lower-layer composites before $diss$ metric can be meaningfully applied.

Computing $diss$ is NP-hard, related to the subgraph isomorphism problem. However, when the agent can establish vertex correspondences a priori, the calculation simplifies to a comparison of adjacency matrices and attribute vectors. In practical applications for large-scale ontologies, approximate methods based on graph embeddings [28] or efficient GED heuristics are employed.

Example 1: (Computing diss) Consider the two graphs Γ_0 and Γ_1 in Figure 1. Let $\psi_V(v_a, v_b) = 0.5$ for any similarly-shaped vertices, $\psi_V(v_a, v_b) = 1$ for any distinct-shaped vertices (circle vs. square), and $\psi_E \equiv 0$. With weights $\alpha_V =$

$\alpha_E = 1$ and the natural mapping $\chi: a_i \mapsto b_i$, the unmatched square vertex in Γ_1 contributes a penalty of 1, and the four matched but distinct-shaped vertex pairs contribute $4 \times 0.5 = 2$. With five vertices total, $\Phi_V(\chi) = (2 + 1)/5 = 0.6$. As there are no distinct edges, $diss(\Gamma_0, \Gamma_1) = 0.6/2 = 0.3$.

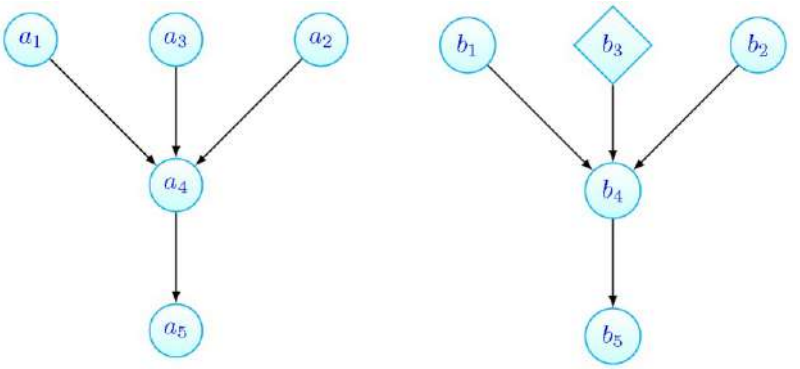


Figure 1
Finding the distance between graphs

Comparing Composite Vertices: To compare vertices that represent high-level concepts, one must recursively decompose them into their constituent subgraphs at lower ontology layers until a common level of granularity is reached. The distance is then computed between the resulting composite graphs. For instance, the statements "every person has a mother" and "every person has a head" may appear syntactically similar at a high layer, but their distances diverge significantly when their detailed biological and social constituent graphs are compared at a base layer of primitives.

3.6 *Information Utility and Truth-Seeking*

3.6.1 Quantifying and Reducing Epistemic Uncertainty

As follows from the previous section, each fragment of the Γ ontology generates a "cloud" of statements close to it $\mathcal{U}_\theta(\Gamma) = \{\Gamma' | \text{diss}(\Gamma, \Gamma') \leq \theta\}$. For an agent under conditions of insufficient information, each graph from this "cloud" may seem to describe objective reality equally correctly. The larger the size of this cloud $|\mathcal{U}_\theta(\Gamma)|$, the greater the probability of choosing from it a graph that describes reality incorrectly.

On the other hand, if agent $ag \in Ag$'s hypothesis about the world, expressed by the graph Γ_{ag}, is incorrect, then the smaller the size of the cloud $|\mathcal{U}_\theta(\Gamma_{ag})|$, the less likely it is to find a true description of the world Γ in it.

The union of all graphs from $\mathcal{U}_\theta(\Gamma_{ag})$ corresponds to a statement $f(P_1 \dots, P_q)$ about the constituent elements of these graphs, and each graph from $\mathcal{U}_\theta(\Gamma_{ag})$ corresponds to a conjunction of atomic statements about the graph, and each symbol of the union corresponds to a disjunction. For such a statement,

measures (12) and (13) are defined. Further, we will write $\mathrm{N}(f(P_1 \ldots, P_q)) = \mathrm{N}(\mathcal{U}_r(\Gamma_{ag}))$. Obviously, if $\mathcal{U}_r(\Gamma_{ag})$ has more than one graph, then $\mathrm{N}(\mathcal{U}_\theta(\Gamma_{ag})) > 0$. Thus, the agent needs, on the one hand, to increase θ so that $\mathcal{U}_\theta(\Gamma_{ag})$ includes the correct version of the description of the world, and on the other hand, to decrease $\mathrm{N}(\mathcal{U}_\theta(\Gamma_{ag}))$ so as not to have many contradictory description options. For example, one can select θ such that $\mathrm{N}(\mathcal{U}_\theta(\Gamma_{ag})) \leq 1/2$.

Definition 3: Let us say that a graph ontology Γ_{ag} has a degree of uncertainty θ and undefined with a logical probability p, if $N(\mathcal{U}_\theta(\Gamma_{ag})) \leq p$.

Thus, the neighbourhood $\mathcal{U}_\theta(\Gamma_{ag})$ is characterized both by its volume θ and the degree of heterogeneity p. We will further denote such neighbourhood as $\mathcal{U}_{\theta,p}(\Gamma_{ag})$.

3.6.2 Truth Approximation via Ontological Patterns

On the set $2^{\tilde{\Gamma}_{ag}}$ a family of relations of non-strict order (subjective "utility") $\circ_{ag}$ can be specified. We will say that information $\Gamma_1 \subseteq \tilde{\Gamma}_{ag}$ is more useful for agent $ag \in Ag$ than information $\Gamma_2 \subseteq \tilde{\Gamma}_{ag}$ if $\Gamma_2 \circ_{ag} \Gamma_1$. Such a relation must be introduced in order for the agent ag to view and select ontology graphs from $\mathcal{U}_\theta(\Gamma_{ag})$ in a certain order.

The agent may have a certain "ideal graph" Γ', possibly with symbol vertices (see Subsection 3.4), with which it compares the graphs $\Gamma \in \mathcal{U}_\theta(\Gamma_{ag})$. For simplicity, we can assume that the family of utility functions $u_{ag}: 2^{\tilde{\Gamma}_{ag}} \to \mathbb{R}_{\geq 0}$ is defined as follows: $\Gamma_2 \circ_{ag} \Gamma_1$ if and only if $u_{ag}(\Gamma_2) \leq u_{ag}(\Gamma_1)$, $u_{ag}(\Gamma') = 1 - \mathrm{diss}(\Gamma', \Gamma)$. Thus, the "ideal graph" serves as a pattern for the search for truth.

All the contents of this subsection are necessary in order to formalise the concepts "there is a grain of truth", "an almost correct statement", "despite errors, important information has been obtained", etc.

Let us illustrate the above with an example. Let the pattern graph be defined by the statement

$P^S =$ "[someone] is wearing [something #1] red, [something #2] blue, and [something #3] white",

where square brackets indicate symbol vertices.

The "cloud" $\mathcal{U}_\theta(\Gamma_{ag})$ consists of statements

$P_1 =$ "Ivan (v_0^1) is wearing a jacket (v_1^1) of red colour (v_2^1), a shirt (v_3^1) of brown colour (v_4^1), and trousers (v_5^1) of white colour (v_6^1)",

$P_2 =$ "Basilius (v_0^2) is dressed in a red jacket (v_1^2) and a drum (v_2^2)",

$P_3 =$ "Acacius (v_0^3) is wearing a jacket (v_1^3) of brown colour (v_2^3) and trousers (v_3) of white colour (v_4^3)".

The vertices of the corresponding graphs are indicated in parentheses. Intuitively, P^S is closest to P_1. If we assume that if the vertices represent the same colour, then the distance between their attributes is 0, otherwise it is 1. The

distance between the vertices and their corresponding symbol vertices is 0, then we can say that $\mathrm{diss}(P^S, P_1) = 1/6,\ \mathrm{diss}(P^S, P_2) = 3/6,\ \mathrm{diss}(P^S, P_3) = 3/6.$

The formation of a pattern (incomplete induction) occurs as follows. Let the fragments of the ontology graph with which the agent interacted up to a certain point be described as conjunctions

$$P_1^1 P_2 \dots P_q, \dots, P_1^m P_2 \dots P_q, P_i = \mathbf{1}, P_1^j = \mathbf{1},$$

$i = \overline{1,q}$, $j = \overline{1,m}$. Constructing a template looks like creating a meaningless statement of the first type $[P_1^S]P_2 \dots P_q$, in which $[P_1^S] \in \mathcal{TN}1$ is a symbol for the repeating atoms of P_1^j.

The application of a pattern (specialization) occurs due to the fact that when using the metric diss or another GED-type distance to compare the ontology fragment graph Γ_0 and the graph corresponding to the template Γ_T, the agent

1. Establishes that the dissimilarity between the symbol vertex or symbol edge α of the template graph Γ_T corresponding to the symbol $[P_1^S]$ and the corresponding vertex or edge β of any subgraph of the ontology graph Γ is equal to zero.
2. Searches for the graph $\Gamma_0 \subseteq \Gamma$ that is closest to Γ_T in the sense of the diss metric from the ontology fragment graphs available to the agent.
3. Constructs $\Gamma_R = \Gamma_T \cup \Gamma_0$, assuming $\beta = \alpha$ and considers the resulting reflection graph to be a true description of reality.

The agent breaks the pattern when it turns out that the statement P_R corresponding to the graph Γ_R is false or meaningless.

Let $Prop_{\Gamma - T}$ be the set of all atoms $P \in \mathcal{TTF}$ such that the graph corresponding to $PP_2 \dots P_q$ is not a subgraph of Γ. The statements in $Prop_{\Gamma - T}$ are not known to the agent in advance. The agent analyses statements one at a time and may not even get to the one that causes the pattern to break. The following statement about the probability of breaking the pattern at the analysis step is obvious.

Statement 2: The probability of the breaking the pattern by an agent is no more than $\frac{|Prop_{\Gamma - T}|}{|Prop_{\Gamma}|}$, where $Prop_{\Gamma}$ is the set of all atoms from $\mathcal{TTF}$ that completely describe the graph Γ.

The problem is that the agent does not know $|Prop_{\Gamma - T}|$ either. However, to calculate the subjective probability, the agent can use the estimate $|Prop_{\Gamma - T}| \leq |Prop_{\Gamma} \backslash Prop_{\Gamma + T}|$, where $Prop_{\Gamma + T}$ is the set of all atoms from $\mathcal{TTF}$ that completely describe the graph Γ, and studied by the agent at the current analysis step.

4. Cognitive Processes as Graph Optimization

4.1 *Single-Agent Cognitive Dynamics*

4.3.1 The Process of an Ontology Cognition as an Local Optimisation Problem

Suppose that the agent $ag \in Ag$ seeks to bring the ontology-reflection graph Γ_{ag} closer to the ontology graph Γ. This process can happen in different ways. In general, cognition can be described as the kind of local optimisation problem.

Let us denote as $\mathrm{correct}(\Gamma)$ the value of correct for the statement P corresponding to the graph Γ. The agent $ag \in Ag$, wanting to find the reflection of reality with correctness better than o, performs the following algorithm.

0. Set $t = 0$, the initial values of θ, p, o, the initial reflection Γ_{ag}, $\mathcal{U}(t) = \mathcal{U}_{\theta,p}(\Gamma_{ag})$.
1. Analyse the neighbourhood $\mathcal{U}(t)$ for the graph $\Gamma_{ag}(t)$ that most corresponds to reality,
$$\Gamma_{ag}(t) \in \mathrm{Arg}\max_{\Gamma' \in \mathcal{U}(t)} \mathrm{correct}(\Gamma').$$
2. If $\mathrm{correct}(\Gamma_{ag}(t)) \geq o$, go to Step 6 else go to Step 3.
3. Set $t := t + 1$.
4. Construct the neighbourhood $\mathcal{U}(t) = \mathcal{U}_{\theta,p}(\Gamma_{ag}(t))$.
5. Go to Step 1.
6. End.

If the value of $\mathrm{correct}(\Gamma_{ag}(t))$ is stuck and not increasing over the time t, we should increase θ and p until $\mathrm{correct}(\Gamma_{ag}(t))$ starts to increase. It will have the cost of the neighbourhood swelling and the analysis time increase.

This will lead to an increase in data volume and analysis time. In fact, the process of cognition is considered as a greedy search on graphs. It is unknown whether the process will converge, whether the agent will come to anything at all (the agent will either â€œget tiredâ€• or go through everything). Unfortunately, this is exactly how people function - their cognition may lead to nothing, and no goals will be achieved, or a completely suboptimal result will be obtained. Attributing illnesses to witchcraft or witch hunts are examples of such situations.

Additionally, an agent can use heuristics to analyse $\mathcal{U}(t)$. In practical applications, agents rarely need to know everything in the world, but rather only need to accurately reflect the portion of the graph Γ that corresponds to their interests. Let us give a few examples of such heuristics.

Heuristic 1: An agent sees some kind of information as more attractive than others, and he will look for those fragments of the ontology graph that correspond to a certain pattern, as in Subsubsection 3.6.2. Let us specify an agent which prefers to search for spies using a pattern

P^S =“[Spy] ([v_0^S]) wears black (v_2^S) glasses (v_1^S) and a cloak (v_3^S)”.

There are two statements:

$P_1 =$ "Basilius (v_0^1), who is a spy, does not wear black (v_2^1) glasses (v_1^1) and wears a green (v_4^1) T-shirt (v_3^1)" and

$P_2 =$ "Acacius (v_0^2), who is not a spy, wears black (v_1^2) glasses (v_2^2) and a cloak (v_3^1)".

The part of reality unknown to the agent is indicated in curly brackets; there are no other possible spies except Acacius and Basilius. As before, the symbol vertices are indicated in square brackets, and the names of the vertices corresponding to the statements (minus the unknown part in curly brackets) P^S, P_1, P_2 of the graphs Γ^S, Γ_1, Γ_2 are in round brackets.

At the initial moment of time, the agent's reflection graph is defined by the pattern $\Gamma_{ag}(0) = \Gamma^S$ and $\mathcal{U}_\theta(\Gamma_{ag}(0)) = \{\Gamma^S, \Gamma_1, \Gamma_2\}$. Due to the agent's preferences, the innocent Acacius will be more likely to be suspected of espionage at time $t_1 > 0$ and the innocent Acacius will be checked first, since $\mathrm{diss}(\Gamma_1, \Gamma^S) > \mathrm{diss}(\Gamma_2, \Gamma^S)$. At this moment, the agent's reflection of reality will become $\Gamma_{ag}(t_1) = \tilde{\Gamma}_1$, where $\tilde{\Gamma}_1$ corresponds to $\tilde{P}_2 =$ "Acacius (v_0^2), who is a spy ($\tilde{v}_4$), wears black (v_1^2) glasses (v_2^2) and a raincoat (v_3^1)", as a derivation (essentially a union of graphs $\Gamma^S \cup \Gamma_2$ in which the symbol v_0^S "Spy" is considered one vertex with v_0^2 "Acacius") from Γ^S and Γ_2. After checking and finding out at time $t_2 > t_1$ that Acacius is not a spy, the agent breaks the pattern, the agent assumes $\Gamma_{ag}(t_2) = \Gamma_1$ and establishes the truth.

Heuristic 2: Let us denote by $\rho_{\mathbf{O}}: \mathbf{O} \times \mathbf{O} \to \mathbb{R}_{\geq 0}$ the semantic distance between entities in the ontology $\mathbf{O}$. The graph Γ corresponds $\mathbf{O}$ as before and l_Γ is a distance on the graph Γ. We will assume that the $\rho_{\mathbf{O}}$ distance and the ontology have the following property. The farther vertices, corresponding to entities of the ontology, on the graph Γ in the sense of the l_Γ distance, the greater the semantic distance between them in the ontology and vice versa.

Let $ag \in Ag$ be given a set of things and relations $\Gamma_{ag}^+ \subset \Gamma$ that are of interest to it. The agent estimates the value $c(\Gamma_{ag}(t))$ of information $\Gamma_{ag}(t) \subset \Gamma$ by the proximity of $\Gamma_{ag}(t)$ to Γ_{ag}^+, for example as

$$c(\Gamma_{ag}(t)) = \max_{\gamma_1 \in \Gamma_{ag}(t), \gamma_2 \in \Gamma_{ag}^+} l_\Gamma(\gamma_1, \gamma_2)$$

and constructs the neighbourhood $\mathcal{U}_{\theta,p}$ only from graphs Γ' with $c(\Gamma') < \varepsilon$.

It should be kept in mind that the agent (and the external evaluator) in the general case does not have complete information about the external world and receives knowledge about $\mathrm{correct}(\Gamma_{ag}(t))$ retroactively. However, in some applied problems, such as supervised learning problems, the external estimator has complete information, so $\mathrm{correct}(\Gamma_{ag}(t))$ can be calculated. Moreover, the agent always knows its approximation $\mathrm{correct}(\Gamma_{ag}(t))$, in which, for example, the measure of the truth of conjunctions about which nothing is known is determined arbitrarily, and then refined at the next step.

4.1.2 The Process of an Ontology Cognition as an Global Optimisation Problem

Let the external time $t \in \mathbb{Z}_{\geq 0}$ be given and at each moment t the agent ag analyses some subgraph $\Gamma_{ag}^{O}(t) \subseteq \Gamma$, $\Gamma_{ag}^{O}(t) = \mathcal{P}_{ag}(t)\Gamma$, $\mathcal{P}_{ag}(t) \in \mathcal{P}$ is an selection operator. The agent must make a decision $D_{ag}(t)$ with respect to $\Gamma_{ag}^{O}(t)$. This decision, in its most general form, corresponds to studying $\Gamma_{ag}^{O}(t)$ with more or less care. The more carefully the graph is studied, the more accurate a reflection of the graph is obtained and the more time it takes. Making and executing a decision will take some time $T(D_{ag}(t))$, proportional to the power of sets of edges, vertices, and their attributes in $\Gamma_{ag}^{O}(t)$. The agent can both clarify information regarding the part of the graph known to it and receive information about a previously unknown part of the graph. The result of the inference should be a new reflection $\Gamma_{ag}(t)$ corresponding to $\Gamma_{ag}^{O}(t)$.

Thus, the sequence of decisions $D_{ag}(t_1), \dots, D_{ag}(t_q)$ corresponds to the final reflection of the graph $\Gamma_{ag}^{O}(t) = \Gamma_{ag}^{O}(D_{ag}(t_1), \dots, D_{ag}(t_q))$, $t > t_q + T(D_{ag}(t_q))$. We denote the set of all possible decisions of the agent as $\mathcal{D}$.

The condition for stopping the process of studying the graph Γ by the agent can be, for example, the achievement of specified values of uncertainty θ and logical probability N or the absence of changes in their values during the prolonged study of the graph (at $t \geq T_0$ for some large T_0).

The accuracy of information (even in a very narrow area) is a measure of correspondence between the real $\Gamma_{ag}^{O}(t)$ and reflected ontology $\Gamma_{ag}(t)$ at a given moment t of time for the analysed fragment of the graph:

$$\text{accuracy}(t; D_{ag}(t_1), \dots, D_{ag}(t_q)) = 1 - \text{diss}(\Gamma_{ag}(t), \Gamma_{ag}^{O}(t)).$$

Each agent at time t has its own inaccurate copy $\Gamma_{ag}^{C}(t)$ of the entire graph Γ. The closer $\Gamma_{ag}(t)$ is to Γ, the higher the awareness of the agent $S_{ag}(\Gamma)$ about the entire ontology, a kind of volume of his knowledge:

$$\text{awareness}(t; D_{ag}(t_1), \dots, D_{ag}(t_q)) = 1 - \text{diss}(\Gamma_{ag}^{C}(t), \Gamma).$$

So, when iteratively constructing a reflection ontology, the agent must be guided by three goals: minimizing the probability of making a false or meaningless statement of the first type, which depends on the accuracy of the correspondence of the agentâ€™s ontology-reflection with the real ontology and on the share of the real ontology known to the agent, and minimizing the time spent on each iteration-clarification, which is achieved by incomplete research of the ontology, taking into account the value of information.

In essence, the agent solves the following optimisation problem:

$$\text{awareness}(t; D_{ag}(t_1), \dots, D_{ag}(t_q)) \to \max,$$
$$\text{accuracy}(t; D_{ag}(t_1), \dots, D_{ag}(t_q)) \to \max,$$
$$\sum_{k=1}^{q} T(D_{ag}(t_k)) \to \min,$$
$$D_{ag}(t_1), \dots, D_{ag}(t_q) \in \mathcal{D}.$$

Finally, an agent can fantasise and, instead of analysing the part $\Gamma_{ag}^{O}(t) \subset \Gamma$ available to him, express meaningless statements of the first or even second type (i.e., generate symbols of edges, vertices and their markings), perhaps similar in the sense of the metric diss to the part Γ already known to him, and then look for the closest constructions to them in the graph Γ or construct an extension graph $\tilde{\Gamma}$ so that the statements it makes are meaningful (i.e., true or false). This corresponds to a change in the scientific paradigm (see Kuhn [29]) and intuitive "brilliant insights", rather than the routine accumulation of knowledge.

4.1.3 Learning Strategies: Refinement, Expansion, and Exploration

Let us model the process of improving the ontology reflection graph by the agent from the previous subsection. The agent can consider already known or new fragments of the graph, selecting them randomly or selecting them purposefully. Let us list examples of possible actions of the agent.

D_1: The vertices $\Gamma_{ag}(t_k) \subseteq \Gamma$ are selected randomly from the vertices Γ every time t_k. In the second case, the vertices $\Gamma_{ag}(t_k) \subseteq \Gamma$ are chosen as the closest (in some sense, based on the distance l_Γ on the graph) to the set of things and relations Γ_{ag}^{+} that the agent is interested in, or as the closest to the already known vertices $\Gamma_{ag}(t_{k-1})$.

Next, we can introduce the probability of constructing an erroneous reflection of $\Gamma_{ag}^{O}(t)$ in Γ_{ag}, namely, the erroneous addition or exclusion of vertices, edges, or attributes.

D_2: The agent can improve the quality of the existing reflection by repeatedly learning $\Gamma_{ag}^{O}(t_{k-1})$ or

D_3: study a new subgraph $\Gamma_{ag}^{O}(t)$. Sometimes it is reasonable to assume that upon repeated study, the reflection and the reflected subgraph will coincide, i.e., if $\Gamma_{ag}^{O}(t_1) = \Gamma_{ag}^{O}(t_2) = \cdots \Gamma_{ag}^{O}(t_w),\ w > w_0 > 0$, then $\Gamma_{ag}(t_w) \supset \Gamma_{ag}^{O}(t_1)$.

D_4: The agent sequentially studies the selected path on the graph Γ, and also deviates to the graph vertices that are closest in the sense of the graph metric l_Γ to the given path.

One can attribute different attractiveness for exploration by an agent $ag \in Ag$ to vertices and edges Γ. The information $\Gamma_1 \subset \Gamma$ may be less useful than it seems, but the agent is inclined to study it because it is similar to the very useful information $\Gamma_2 \subset \Gamma$ in the sense of the graph dissimilarity metric ρ, while in the sense of the distance on the graph l_Γ, the subgraphs Γ_1 and Γ_2 are located far from each other. Figuratively speaking, the agent is looking for a horned hare and spends a lot of time catching the hare, which seemed to him to be horned, because in the dark the agent mistook its ears for horns.

4.2 *Multi-Agent Interaction and Strategic Games*

One can consider cooperation and competition of agents, egoism, altruism, or even parasitism, commensalism, and similar forms of interaction.

Model Example 4 (information warfare):

States always have a struggle for some resource, for dominance in the world, etc. Spies of two different states express the antagonistic interests of these states, even if they are looking for different information. For example, a spy of country X is looking for information about the weapons of country Y, and a spy of country Y is looking for information about the weapons of country X. Although they are looking for different things, the acquisition of information about country X by country Y will harm country X and vice versa. Those. It is important for each country and its spies that the spies of other countries know as little as possible about this country and, conversely, that their own spies know as much as possible about other countries.

Two spies process information and at the same time try to slow each other down. Each spy spends T seconds processing each block of information B. Spies sequentially process blocks $A_1^j, A_2^j, \dots, A_t^j$, $j = 1$ or $j = 2$, respectively. At the same time, not all blocks of information are needed by every spy to obtain critical information; some are useless and simply waste valuable time.

Spy j can, to save time, discard the block A_t^j at the time moment t, but this decision will lead to an erroneous conclusion, with probability $P(t)$, and it will be necessary to process the erroneously discarded block again.

A spy or counterintelligence agent usually does not know that his counterparty is an enemy spy or counterintelligence agent, since intelligence and counterintelligence officers do their best to hide this fact. Counterintelligence agents do their best to misinform spies by posing as ordinary citizens and passing completely irrelevant information to the spies. The spy who receives the information may suspect his agent of something. In general, security personnel are very suspicious of others. Typically, a security officer works in several modes and can choose either to collect information about the enemy for himself or to misinform his opponents.

There is a repeated game between two persons ag_1 and ag_2, each of whom has the following alternatives:

- $D_1(m_1)$: Study the graph Γ in the volume of the fragment Γ_1 independently.
- $D_2(m_2)$: Misinform another agent about Γ by making a false statement Γ_2.
- $D_3(m_3)$: Misinform another agent by issuing a true statement Γ_3 about Γ, far from the interests of the agents.
- $D_4(m_4)$: Misinform another agent by making a meaningless statement Γ_4 about Γ.
- $D_5(m_5)$: Inform another agent by producing a true statement Γ_5 about Γ that is interesting to all agents.
- $D_6(m_6)$: Check information Γ_6 received from outside.

Here $m_i = \mu(\Gamma_i)$, $i = \overline{1,6}$.

Agents have a degree of trust $0 \le b_j \le 1$ in each other, which decreases upon detection of misinformation $D_2 - D_4$ and increases upon receipt of reliable

information D_5. Agent ag_j will apply D_5 with probability b_j. The agent also has a tendency to misinform $0 \le \overline{b}_j \le 1$, which increases when he uses alternatives $D_2 - D_4$ with subsequent acceptance and verification of misinformation by another agent D_5 and falls when another agent refuses to use D_5. Agent ag_j will apply D_i, $i = 2,3,4$ with probability proportional to $\overline{b}_j$.

Applying the alternative $D_i(m)$ takes time $T_i(m) \propto m$. The payoff function is formalised as follows. Agent ag_i seeks to maximise the size $Volume_+$ of the studied fragment Γ (which increases by m when $D_1(m)$ is applied, can increase by m when $D_6(m)$ is applied, if another agent applied $D_5(m)$ before, remains unchanged in all other cases), to minimise its study time T (which increases by $T_i(m)$ when applied $D_i(m)$, $i = \overline{1,6}$) and minimise the size of the interesting fragment Γ studied by other agents $Volume_-$ (which increases by m when they use $D_1(m)$, as well as $D_6(m)$, if this agent has previously used $D_5(m)$, remains unchanged in all other cases).

Within the framework of the described simplified model, it can be assumed that information is provided only by a complete analysis of a fragment and that fragments discarded after analysis do not bring additional information. Using standard game theory notations D_i (alternative chosen by the agent) and D_{-i} (alternatives chosen by other agents), $D = (D_i, D_{-i})$, we obtain the expression for the payoff function:

$$u_i(D) = Volume_+(D_i) - T(D_i) - Volume_-(D_{-i}).$$

The strategy for the game as a whole for any agent will be a sequence of alternatives $D_{i_1}, D_{i_2}, \dots, D_{i_k}, \dots$, chosen by the agent in each round $1,2, \dots, k, \dots$, and the payoff function for such a strategy must be constructed in accordance with the specific applied problem and the required degree of detail of the process.

4.3 *Illustrative Applications*

In *Model Example 1*, it is enough for *Il Dottore* to take a measure like time to the most recent common ancestor as the ontological "distance" between people. It is obvious that relationships dating back more than a thousand years ago are unlikely to affect the distribution of funds in the present, and the share of funds due to the self-proclaimed "brother" as well as the value of his claims is inversely proportional to the time elapsed.

In *Model Example 2*, the relationship between the *Judge* and *Il Dottore* is described by a nondeterministic (or rather probabilistic) finite state machine $\mathcal{A}$. However, the global relations of the *Judge*, *Il Dottore* and *Pantalone* are described by a deterministic finite state machine $\mathcal{A}'$, the sub-automaton of which is $\mathcal{A}$, and indetermination is not in constructive feature of $\mathcal{A}'$ but in lack of *Il Dottore*'s knowledge.

In *Model Example 3*, *Il Dottore* thinks that the set of events is $A = \{white, black\}$, and the probability of getting a white ball is $P\{white\} = 1/|A|$.

However, in reality, if we assume that at time $t = 2$ *Brighella* approaches the box, the probability is defined as

$$P_t\{white\} = \begin{cases} 1/2, t = 1, \\ 0, t = 2, \end{cases}$$

and the process loses its probabilistic character. Thus, in fact, the set of events is

$$A = \{(white, 1), (black, 1), (white, 2), (black, 2)\}$$

and it is more about a random process for *Il Dottore*.

So, our framework formalises how different agents can interpret the same statement. Consider a meaningless statement of the second type, S (as defined in Section 3.1). Two agents, ag_1 and ag_2, may use their own expansion operators, W_{ag_1} and W_{ag_2}, to convert S into a meaningful (true or false) statement. Each agent ag_i treats the resulting statement $\mathrm{W}_{ag_i}S$ as a possible truth about the world. An external observer ag_O, with access to the complete ontology graph Γ, can then assess these expansions against reality. For example, the observer might find that $\mathrm{W}_{ag_1}S$ remains meaningless (i. e., it cannot be grounded even by ag_O), or that a further expansion $\mathrm{W}_O\mathrm{W}_{ag_1}S$ is demonstrably false, while $\mathrm{W}_{ag_2}S$ is verified as true. This formal model explains how miscommunication, deception, and differential understanding arise from the same vague input.

4.4 *Bridging to Hybrid AI and Neuro-Symbolic Systems*

The formal framework developed in this article provides a mathematical foundation for key challenges in modern artificial intelligence, particularly in the design of *hybrid (neuro-symbolic)* systems and the validation of generative models.

The core constructs of our theory map directly onto the architecture of a hybrid AI agent:

- The ontology graph Γ represents a symbolic knowledge base (KB), encapsulating structured, interpretable facts and rules.
- An agent's reflection graph Γ_{ag} models the internal, often subsymbolic, representation of a statistical learner (e.g., a neural network or a Large Language Model). This representation is typically incomplete, approximate, and derived from data.
- The process of cognition described in Section 4.1 (the optimisation problems) formalises the iterative interaction between statistical and symbolic components. The neural component proposes modifications to Γ_{ag} (generating candidates, making "meaningless statements of the first type"), while the symbolic component evaluates, corrects, and grounds these proposals using the measures $diss(\Gamma_{ag}, \Gamma)$ and $correct(\cdot)$ against the KB Γ.

Thus, our framework offers a language to describe the neuro-symbolic integration and can be used as a versatile formal toolkit for engineering the next generation of hybrid AI systems [30, 31].

Possible practical applications of the proposed theory are automated collection and analysis of the truth of information, including within the framework of information warfare, and validation of the output of Large Language Models (LLM) and similar neural network mechanisms to discard hallucinations (see, for example, [32, 33]). Validation is carried out by discarding a conclusion that is assessed as unlikely (in the sense of logical probability) within the framework of a given ontology. A primary application is the validation of generative model outputs, such as those from LLMs. An LLM's natural language generation can be parsed into a proposed ontology subgraph Γ_{LLM}.

- The graph dissimilarity measure $diss(\Gamma_{LLM}, \Gamma_{KB})$ quantifies the structural and semantic deviation of the generation from a trusted knowledge base Γ_{KB}.
- The correctness measure $correct(P_{LLM})$, where P_{LLM} is the logical form of the generation, evaluates its integrity and cost of grounding.

Outputs with a high $diss$ value or a low $correct$ score can be flagged as potential *hallucinations* [32, 33, 34] and subjected to review, rejection, or iterative refinement. This provides an alternative to purely statistical confidence scores.

The meta-loss function is an objective function used in meta-learning (learning to learn) to evaluate and improve the performance of an entire learning algorithm, including the standard task-specific loss function, rather than just the model's parameters. The global optimisation problem formulated in Section 4.1.2 (maximise awareness/accuracy, minimise time) can be interpreted as a meta-loss function for training hybrid systems. It balances the exploration of new symbolic structures (expanding $Volume_+$) against the cost of verification ($T(D_{ag})$). This formalises the trade-off between the *expressivity* of a neural component and the *verifiability* of a symbolic one, a central concern in scalable neuro-symbolic AI [31].

The game-theoretic model of Section 4.2 (Information Warfare) roughly models scenarios critical to AI safety and robustness. The strategies $D_2 - D_4$ (misinformation) model an adversary corrupting a training dataset or a communication channel. The trust parameter b_j and the payoff function $u_i(D)$ describe the dynamics of robustness and detection.

In summary, the theory advanced herein is not only a model of human-like cognition under uncertainty but also a formal toolkit for engineering of hybrid AI systems. It directly addresses the challenges of integration, validation, and robust multi-agent interaction that define the forefront of the field.

5. Conclusion

The article developed a theory of levels of cognition located one above the other, which makes it possible to formally describe certain artefacts of cognition and thinking, referring to the language of levels lower than those at which the mentioned artefacts arose. Using the developed theory, it is possible to search for a solution to a particular formally described problem in a certain neighbourhood

of this problem, described using the graph dissimilarity metric. This system can be used, for example, when building a route for a robot. In a stable environment, the robot's route consists of simple elements - segments, arcs, etc. and control is carried out by issuing commands to move according to previously known primitives. When uncertainty and sudden changes in the environment arise, the robot analyses data from its sensors with a level of detail dependent on the uncertainty of the environment. As a result, it becomes possible to save robot resources in a stable environment and avoid accidents in a dynamic environment.

Another possible practical application of the proposed theory is the automated collection and analysis of the truth of information, including within the framework of information warfare, and validation of the output of Large Language Models and similar neural network mechanisms in order to discard hallucinations (see, for example, [32], [33]). Validation is carried out by discarding a conclusion that is assessed as unlikely (in the sense of logical probability) within the framework of a given ontology. So, the theory provides formal criteria for the holistic evaluation of generative models [35] and the principled design of neuro-symbolic architectures [30].

Declarations

- Funding. The authors did not receive support from any organisation for the submitted work.
- Conflict of interest. The authors have no relevant financial or non-financial interests to disclose.
- Data availability. Data sharing is not applicable to this article as no new data were created or analysed in this study.

References

[1] N. Guarino, D. Oberle, and S. Staab, *What Is an Ontology?* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–17 (2009). [Online]. Available: https://doi.org/10.1007/978-3-540-92673-3_0

[2] R. Quintero, M. Torres-Ruiz, R. Menchaca-Mendez, M. A. Moreno-Armendariz, G. Guzman, and M. Moreno-Ibarra, "Dis-c: conceptual distance in ontologies, a graph-based approach," *Knowl. Inf. Syst.*, vol. 59, no. 1, p. 33–65 (Apr. 2019). [Online]. Available: https://doi.org/10.1007/s10115-018-1200-3

[3] J. Euzenat, C. Allocca, J. David, M. d'Aquin, C. Le Duc, and O. Sv´ab-Zamazal, "Ontology distances for contextualisation," *INRIA, Contract*, (Aug. 2009). [Online]. Available: https://inria.hal.science/hal-00793450

[4] S. Harispe, S. Ranwez, J. Montmain, and S. Janagi, Semantic Similarity from Natural Language and Ontology Analysis. Cham, Switzerland: Springer (2022).

[5] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, Eds., The Description Logic Handbook: Theory, Implementation and Applications. Cambridge, UK: Cambridge University Press (2003).

[6] B. Courcelle, "On the expression of graph properties in some fragments of monadic second-order logic," in Descriptive Complexity and Finite Models. Proceedings of a DIMACS Workshop. Princeton University: American Mathematical Society, pp. 33–62 (1997). [Online]. Available: https://api.semanticscholar.org/CorpusID: 7858432

[7] O. Pikhurko and O. Verbitsky, "Logical complexity of graphs : a survey," in Model theoretic methods in finite combinatorics : AMS-ASL joint special session (January 5-8, 2009), Washington, DC, ser. Contemporary mathematics, M. Grohe and J. A. Makowsky, Eds. Providence, R.I.: American Mathematical Society, vol. Vol. 558, pp. 129–180 (2011). [Online]. Available: http://www.ams.org/bookstore?fn=20&arg1=whatsnew&ikey=CONM-558

[8] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, ser. Theory and Decision Library D. Dordrecht, Netherlands: Kluwer Academic Publishers, vol. 9 (1991). [Online]. Available: https://link.springer.com/book/10. 1007/978-94-011-3534-4

[9] F. P. Ramsey, The Foundations of Mathematics and Other Logical Essays, R. B. Braithwaite, Ed. New York,: Harcourt, Brace and company (1931).

[10] A. H´ajek, "Interpretations of probability," in The Stanford Encyclopedia of Philosophy, winter 2023 ed., E. N. Zalta and U. Nodelman, Eds. Stanford, California: Metaphysics Research Lab, Stanford University (2023). [Online]. Available: https://plato.stanford.edu/archives/win2023/entries/probability-interpret/

[11] R. Carnap, Logical Foundations of Probability. Chicago]: Chicago University of Chicago Press (1950).

[12] A. Fern, "Weighted logic," College of Engineering, Oregon State University, Technical report (2010). [Online]. Available: https://web.engr.oregonstate.edu/~afern/classes/cs532/notes/weighted-logic.pdf

[13] D. Dubois, L. Godo, and H. Prade, "Weighted logics for artificial intelligence –an introductory discussion," *International Journal of Approximate Reasoning,* vol. 55, no. 9, pp. 1819–1829 (2014). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888613X14001364

[14] R. T. Cox, Algebra of Probable Inference. Baltimore, MD: Johns Hopkins University Press (2001).

[15] A. P. Dempster, "Technical report no. 20. a generalization of bayesian inference," Department of statistics Harvard University a generalization of Bayesian inference, Contract Nonr 1866(37) NR-042-097 (November 15, 1967). [Online]. Available: https://apps.dtic.mil/sti/tr/pdf/AD0664659.pdf

[16] A. Chkhartishvili, "Reflexive games: Transformation of awareness structure," *Automation and Remote Control*, vol. 71, p. 1208–1216 (2010). [Online]. Available: https://doi.org/10.1134/S0005117910060214

[17] A. Tarski, "The semantic conception of truth and the foundations of semantics," *Philosophy and Phenomenological Research*, vol. 4, no. 3, pp. 341–375 (1944). [Online]. Available: https://www.jstor.org/stable/2102968

[18] M. Hiriyanna, Outlines of Indian Philosophy. Delhi: Motilal Banarsidass, 2000, first published 1932. Chapter 8 covers Vai´ses.ika atomism in detail.

[19] T. Stcherbatsky, Buddhist Logic. Delhi: Motilal Banarsidass, vol. 1 (2019).

[20] T. Stcherbatsky, Buddhist Logic. Delhi: Motilal Banarsidass, vol. 2 (2019).

[21] M. Heidegger, Being and Time, trans. J. Macquarrie and E. Robinson. New Delhi, India: Motilal Banarsidass (2020). (Indian reprint of the 1962 English translation; includes the original pagination.)

[22] S. Umaa Mageswari, C. Mala, A. Santhanavijayan, and G. Deepak, "A noncollaborative approach for modeling ontologies for a generic IoT lab architecture," *Journal of Information and Optimization Sciences,* vol. 41, no. 2, pp. 395–402 (2020).

[23] L. V. Shcherba, "The threefold aspect of linguistic phenomena and experiment in linguistics," *Historiographia Linguistica,* vol. 47, no. 2-3, pp. 245–272 (2021).

[24] A. Bansal, M. Balakrishnan, and V. Sorge, "Evaluating cognitive complexity of algebraic equations," *The Journal on Technology and Persons with Disabilities,* vol. 9, pp. 170–200 (2021).

[25] F. Serratosa, "Graph edit distance: Restrictions to be a metric," *Pattern Recognition,* vol. 90, pp. 250–256 (2019). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320319300639

[26] M. Gromov, "Groups of polynomial growth and expanding maps," Publications Mathématiques de l'Institut des Hautes Scientifiques, vol. 53, no. 1, pp. 53–78 (Dec. 1981).

[27] A. Schumann and A. V. Kuznetsov, "Foundations of mathematics under neuroscience conditions of lateral inhibition and lateral activation," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 33, no. 3, pp. 237–256 (2018). [Online]. Available: https://doi.org/10.1080/17445760.2018.1439490

[28] W. L. Hamilton, Graph Representation Learning, ser. Synthesis Lectures on Artificial Intelligence and Machine Learning. Cham, Switzerland: Springer (2020).

[29] T. S. Kuhn, The Structure of Scientific Revolutions, O. Neurath, Ed. Chicago: University of Chicago Press (1962).

[30] A. S. d'Avila Garcez, M. Gori, L. C. Lamb, L. Serafini, M. Spranger, and S. N. Tran, "Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning," *Journal of Applied Logics*, vol. 6, no. 4, pp. 611–632 (2019). [Online]. Available: https://www.collegepublications.co.uk/downloads/ifcolog00024.pdf

[31] T. R. Besold, A. d'Avila Garcez, S. Bader, H. Bowman, P. Domingos, P. Hitzler, K.-U. K¨uhnberger, L. C. Lamb, D. Lowd, P. M. V. Lima, L. de Penning, G. Pinkas, H. Poon, and G. Zaverucha, "Neural-symbolic learning and reasoning: A survey and interpretation," *Neuro-Symbolic Artificial Intelligence: The State of the Art*, vol. 342, pp. 1–51 (2022).

[32] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Bang, D. Chen, W. Dai, A. Madotto, and P. Fung, "Survey of hallucination in natural language generation," *ACM Computing Surveys*, vol. 55, pp. 1 – 38 (2022). [Online]. Available: https://api.semanticscholar.org/CorpusID:246652372

[33] K. Šekrst, "Unjustified untrue "beliefs": Ai hallucinations and justification logics," in Logic, Knowledge, and Tradition. Essays in Hon-

or of Srecko Kovac, K. ´Swiętorzecka, F. Grgić, and A. Brozek, Eds. Leiden: Brill (2025). [Online]. Available: https://philarchive.org/archive/EKRUUQ

[34] E. M. Bender and A. Koller, "Climbing towards NLU: On meaning, form, and understanding in the age of data," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, pp. 5185–5198 (jul 2020). [Online]. Available: https://aclanthology.org/2020.acl-main.463

[35] P. Liang, R. Bommasani, T. Lee, "Holistic evaluation of language models," *Annals of the New York Academy of Sciences,* vol. 1525, no. 1, pp. 140–146 (2023), special issue on AI and Society.

Received August, 2025