Possible Worlds: A Logical Model of Coherence between Identity and Veracity

Omar Pavel Ancka Quispe

Facultad de Ingeniería Eléctrica y Electrónica,
Escuela de Ingeniería Electrónica, Universidad Nacional del Callao, Callao, Perú

(Dated: 18 de Mayo de 2021)

This article presents a general logical-semantic model that allows the evaluation of the global consistency between ontological identities and statements in finite worlds of individuals. Although the model is inspired by the classic problems of truth-teller and liar —such as those found in the so-called "islands of knights and knaves"—, its formal structure allows for a broader application in contexts where it is necessary to analyze the coherence between what an agent *is* and what it *says*. The model is based on assigning an ontological identity to each individual (e.g., honest or liar, reliable or faulty sensor) together with a statement that expresses a proposition. Two semantic operators are established: one representing the agent's identity and another evaluating the truth value of its statement in a possible world. From these assignments, a finite tabular semantics is defined that allows the construction of all possible worlds and the application of a consistency rule that filters those worlds which present coherence between identity and statement. This approach offers a formal and explicit alternative to classical inferential methods, allowing applications in agent logic, diagnosis, model theory and epistemological analysis.

1. Introduction

The problems of truth-teller and liar have been widely used both in the teaching of logic and in the exploration of epistemological principles about truth, reference, and consistency. In these scenarios, one considers a set of individuals who may be truthful (who always tell the truth) or liars (who always lie), and whose statements involve references to themselves or to others.

Although these problems can be formalized within the framework of propositional or predicate logic, they generally require the assumption of ad hoc premises or axioms to be solved. In contrast, this paper proposes a new semantic model that does not require prior assumptions about the veracity of statements or the use of deductive inference rules, but rather exhaustively explores the space of possible configurations.

The model introduces two operators: one that determines the ontological identity of an individual, and another that interprets the truth content of his statement. By a systematic tabulation over all possible combinations of identities in the Universe $\mathcal{M}$, and an evaluation rule based on the consistency between what an agent *is* and what it *says*, worlds with inconsistent configurations are eliminated and those satisfying where the semantic matching rule are preserved. This methodology, which we call *Tabular Semantic Model of Coherence between Identity and Veracity*, provides a rigorous, intuitive and extensible tool for the analysis of such logical problems.

2. Formal Definitions of the Model

In this section, we formally introduce the logical-semantic model proposed for analyzing problems involving truthful and deceitful agents. We begin with a a finite universe of individuals, each of whom belongs to one of two ontological classes: truthful (honest) or liars. Using formal operators, we study the coherence between the identity of each agent and the truth value of their statements, evaluated within all possible worlds.

2.1 Universe and ontological classes

Let $U = \{X_1, X_2, \ldots, X_n\}$ be the finite set of individuals involved in the problem. Each individual X_i is assigned an ontological identity, for example: $\sigma(X_i) \in \{\mathrm{H}, \mathrm{M}\}$, given by:

$$\sigma : U \to \{\mathrm{H}, \mathrm{M}\}$$

where σ is the identity operator. H represents an Honest individual (always tells the truth). M represents a Liar individual (always lies). Furthermore, after the assignment

$$\sigma(X_i) \quad \Rightarrow \quad \sigma \in \mathcal{M}$$

represents a complete assignment of identities to all individuals in U, which we will call **possible world**. Then, we define the set of all possible identity assignments (possible worlds) as **the Universe** $\mathcal{M}$, given by:

$$\mathcal{M} := \{\sigma \mid \sigma : U \to \{\mathrm{H}, \mathrm{M}\}\}$$

2.2 Truth value according to identity and evaluation

We introduce two functions that define the truth values involved in the model:

Definition 1: ***Truth value according to its Identity***.
Let $\sigma \in \mathcal{M}$ be a possible world, and let $X_i \in U$ be. We define the truth value associated with the identity of X_i as:

$$v(\sigma(X_i)) = \begin{cases} \mathrm{V} & \text{if } \sigma(X_i) = \mathrm{H} \\ \mathrm{F} & \text{if } \sigma(X_i) = \mathrm{M} \end{cases}$$

Definition 2: ***Truth value of the Statement issued by $\boldsymbol{X_i}$.***
Let φ_{X_i} be the specific statement issued by individual X_i, and let $\sigma \in \mathcal{M}$ be a possible world. We define the truth value of the statement evaluated at σ as:

$$v(\varphi_{X_i}, \sigma) \in \{\mathrm{V}, \mathrm{F}\}$$

This value is obtained by interpreting φ_{X_i} under the identity assignment given by σ.

2.3 Semantic coherence criterion

We now define the fundamental rule of consistency to accept a possible world σ as consistent and belonging to the solution set of the problem in question:

Semantic consistency rule: A possible world $\sigma \in \mathcal{M}$ is **coherent** with the statement X_i if and only if:

$$v(\sigma(X_i)) = v(\varphi_{X_i}, \sigma)$$

That is, the truth value corresponding to the identity of X_i coincides with the truth value of its statement evaluated within the same possible world.

2.3.1 Definition of valid world

Since a possible world σ may be coherent for some individuals but not for others, we define:

Valid world: A possible world $\sigma \in \mathcal{M}$ is a **valid world** for the given problem if it is consistent for all individuals who have made explicit statements.

2.4 Model operators

Let's explicitly summarize the semantic operators used:

- $\sigma : U \to \{\mathrm{H}, \mathrm{M}\}$: Identity assignment (possible world). After the assignment, $\sigma(X_i)$ is the individual identity for a particular individual.
- $v(\sigma(X_i))$: Truth value assigned according to identity.
- $v(\varphi_{X_i}, \sigma)$: Truth value of the statement in the model.

Furthermore, the model can be extended to any number of identities. For example, let $U = \{X_1, X_2, \ldots, X_n\}$ be the finite set of n individuals. Let $k = 4$ be the number of identities, then the identity operator σ is defined as:

$$\sigma : U \to \{\mathrm{H}, \mathrm{M}, \mathrm{S}, \mathrm{R}\}$$

This operator represents a possible world, that is, an assignment that tells each individual $X_i \in U$ whether they are H, M, S, or R, and the universe $\mathcal{M}$ will have k^n possible worlds.

$$\mathcal{M} = \{\, \sigma_j \,\}_{j=1}^{k^n}$$

2.5 General procedure

The general logical procedure for solving a problem is summarized in the following steps:

1. Enumerate all possible worlds $\sigma \in \mathcal{M}$, where each world assigns to each individual in the set U an identity (e.g., honest or liar).
2. For each statement φ_{X_i} uttered by some individual $X_i \in U$, evaluate its truth value $v(\varphi_{X_i}, \sigma)$ in every possible world σ. That is, ask yourself:

   "Would this statement be true if the world were σ?"

3. Calculate in each world σ the value associated with the identity of the individual X_i, that is, $v(\sigma(X_i))$, where:

$$v(\sigma(X_i)) = \begin{cases} \mathrm{V}, & \text{if } \sigma(X_i) = \mathrm{H} \\ \mathrm{F}, & \text{if } \sigma(X_i) = \mathrm{M} \end{cases}$$

4. Check whether the consistency rule is met for each individual who made a statement:

$$v(\varphi_{X_i}, \sigma) = v(\sigma(X_i))\,,$$

   This ensures that X_i's statement matches its identity in the world σ.

5. Keep only those worlds $\sigma \in \mathcal{M}$ in which all the speakers' statements are consistent with their identity. These worlds are the **valid worlds** of the problem. This consistency check can be done either **simultaneously** for all statements (evaluating in parallel), or **step-by-step**, evaluating one statement at a time and discarding inconsistent worlds at each step. In the sequential approach, one starts with all possible worlds and progressively filters according to the consistency of each individual statement.

3. Examples of Application of the Model

We present here two examples to illustrate how the set of possible worlds $\mathcal{M}$ is defined from the universe U of individuals, and how each world is represented as a total assignment of identities.

3.1 Example of use 1: Honest and lying agents

Consider a three-person universe: $U = \{A, B, C\}$, where each individual can be either **honest** (H) or **liar** (M). An honest person always tells the truth; a liar always lies.

The assignment of identities is represented by a function:

$$\sigma : U \to \{\mathrm{H}, \mathrm{M}\}$$

and each distinct assignment constitutes a **possible world**. Since there are two options for each of the three individuals, there are $2^3 = 8$ possible worlds, and thus, the set of possible worlds $\mathcal{M}$ is constructed, which is presented in the following table.

Universe $\mathcal{M}$	$\sigma(A)$	$\sigma(B)$	$\sigma(C)$
σ_1	H	H	H
σ_2	H	H	M
σ_3	H	M	H
σ_4	H	M	M
σ_5	M	H	H
σ_6	M	H	M
σ_7	M	M	H
σ_8	M	M	M

then, explicitly

$$\mathcal{M} = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8\} = \{\sigma_j\},$$
$$\mathcal{M} = \{\mathrm{HHH}, \mathrm{HHM}, \mathrm{HMH}, \mathrm{HMM}, \mathrm{MHH}, \mathrm{MHM}, \mathrm{MMH}, \mathrm{MMM}\}$$

Proposals issued

Suppose individuals make the following propositions:

- A says: “B is a liar”
- B says: “C is honest”
- C says: “A is a liar”

Each proposition φ_i is assigned a truth value according to the world σ, evaluating whether what it says is true or not. This is represented as:

$$v(\varphi_i, \sigma) \in \{\mathrm{V}, \mathrm{F}\}$$

Evaluation in a specific world

Let's look at the full procedure for the world σ_4, where:

$$\sigma_4(A) = \text{H}, \quad \sigma_4(B) = \text{M}, \quad \sigma_4(C) = \text{M}$$

- A says: "B is a liar"
  - In σ_4, B is really a liar: the proposition is true $\Rightarrow v(\varphi_A, \sigma_4) = \text{V}$
  - A is honest $\Rightarrow v(\sigma(A)) = \text{V}$
  - They match $\Rightarrow$ consistently
- B says: "C is honest"
  - In σ_4, C is a liar: the proposition is false $\Rightarrow v(\varphi_B, \sigma_4) = \text{F}$
  - B is a liar $\Rightarrow v(\sigma(B)) = \text{F}$
  - They match $\Rightarrow$ consistently
- C says: "A is a liar"
  - In σ_4, A is honest: the proposition is false $\Rightarrow v(\varphi_C, \sigma_4) = \text{F}$
  - C is a liar $\Rightarrow v(\sigma(C)) = \text{F}$
  - They match $\Rightarrow$ consistently

Since all statements are consistent with each individual's identity, the world σ_4 is a **coherent world**.

Filtering inconsistent worlds

To solve the problem, all worlds $\sigma \in \mathcal{M}$ are evaluated. Those that contain any inconsistency are discarded, that is, for which:

$$v(\varphi_i, \sigma) \neq v(\sigma(X_i)) \quad \text{for some } i$$

The remaining worlds are the **valid worlds**. In many classical problems, only one survives, thus determining the identity of each individual.

Use Example 2: Reliable or faulty sensors

Suppose a set of digital sensors is given by $U = \{S_1, S_2\}$. Each sensor can have one of two possible functional identities:

- C: Reliable sensor *(it tells the truth: its output reflects reality)*.
- D: Faulty sensor *(it lies: its output is always opposite to reality)*.

Possible worlds

Each sensor can be either C or D, therefore there are $2^2 = 4$ possible worlds:

$\mathcal{M}$	$\sigma(S_1)$	$\sigma(S_2)$
σ_1	D	D
σ_2	D	C
σ_3	C	D
σ_4	C	C

Assertions and sensor output

Suppose both sensors are trying to measure the same binary quantity (e.g., whether an object is present or not), and that:

- The real value of the world is **1** (present object).
- The observed outputs are:
  - S_1 emits: 0
  - S_2 emits: 1

So, for each world σ_j, we evaluate whether the sensor output matches (in the reliable case) or contradicts (in the faulty case) reality. We denote the truth value of the sensor's statement as:

$$v(\varphi_{S_i}, \sigma_j) = \begin{cases} 1 & \text{if the sensor output matches its type (C/D) in } \sigma_j, \\ 0 & \text{otherwise.} \end{cases}$$

For example:

- If $\sigma_j(S_1) = \text{D}$, then your output is expected to be $\neg 1 = 0$, and since you output 0, then $v(\varphi_{S_1}, \sigma_j) = 1$.
- If $\sigma_j(S_2) = \text{C}$, output 1 is expected, and since it gave 1, then $v(\varphi_{S_2}, \sigma_j) = 1$.

Let us evaluate in all worlds:

	Inputs		Output	To Compare
$\mathcal{M}$	$\sigma(S_1)$	$\sigma(S_2)$	$v(\varphi_{S_1}, \sigma_j)$	$v(\varphi_{S_2}, \sigma_j)$
σ_1	D	D	1	0
σ_2	D	C	1	1
σ_3	C	D	0	0
σ_4	C	C	0	1

Consistency evaluation

We apply the consistency rule to keep only worlds where both outputs are consistent with their sensor type:

$$v(\varphi_{S_i}, \sigma) = v(\sigma(S_i))$$

Since we used 1 to mean "identity-consistent statement", we apply:

- σ_2 is the **only** world where both statements are consistent.

Conclusion

The **valid world** is:

$$\sigma_2 = \{S_1 = \text{D},\ S_2 = \text{C}\}$$

This means that:

- S_1 is faulty (it outputs 0, as opposed to the actual 1).
- S_2 is reliable (it issued 1, which matches reality).

This example shows how the model can be applied to sensor diagnostics by filtering possible worlds based on the consistency between functional identity (C or D) and observable behavior.

4. Step-by-step practical example: Three inhabitants of an island of truth-tellers and liars

You are on an island where the inhabitants can be either **honest (H)** or **liars (M)**. You meet three people: A, B, and C. Each makes a statement:

- A says, "I am..." but you can't hear it.
- B says: "A said he's a liar."
- C says: "B is lying; he is a liar."

We want to determine who is H and who is M using the logical model of consistency between identity and truthfulness.

Step 1: Table of possible worlds

Each person can be honest (H) or a liar (M), so there are $2^3 = 8$ possible worlds. We call each world σ_j.

$\mathcal{M}$	$\sigma_j(A)$	$\sigma_j(B)$	$\sigma_j(C)$
σ_1	H	H	H
σ_2	H	H	M
σ_3	H	M	H
σ_4	H	M	M
σ_5	M	H	H
σ_6	M	H	M
σ_7	M	M	H
σ_8	M	M	M

Step 2: Evaluating A's proposition

From what is heard from A, and from the statement of B, it can be established that the statement of A must be modeled as a verbal self-assignment of identity, in the form:

$$\varphi_A := \text{“I am } \sigma_{\text{verbal}}(A)\text{”}, \quad \text{where } \sigma_{\text{verbal}}(A) \in \{H, M\}$$

This statement is interpreted as a proposition about the identity of A that must be evaluated under each possible world $\sigma \in \mathcal{M}_A$. Since in our system honest agents always tell the truth and liars always lie, we have the criteria:

- $v(\varphi_A, \sigma_m)$: truth value of the specific proposition issued by A, evaluated in the possible world σ_m of $\mathcal{M}_A$.

$$v(\varphi_A, \sigma) = \begin{cases} \text{V} & \text{if } \sigma(A) = \text{H and } \varphi_A := \text{“I am } H'' \\ \text{F} & \text{if } \sigma(A) = \text{H and } \varphi_A := \text{“I am } M'' \\ \text{F} & \text{if } \sigma(A) = \text{M and } \varphi_A := \text{“I am } H'' \\ \text{V} & \text{if } \sigma(A) = \text{M and } \varphi_A := \text{“I am } M'' \end{cases}$$

- $v(\sigma(A))$: This is the truth value that A would always say if its identity were the one given in $\mathcal{M}$.

$$v(\sigma(A)) = \begin{cases} \text{V} & \text{if } \sigma(A) = \text{H} \\ \text{F} & \text{if } \sigma(A) = \text{M} \end{cases}$$

- Coherence: If $v(\sigma(A)) = v(\varphi_A, \sigma_j)$, then the possible world σ_j is coherent.

	Inputs		Output	To Compare	
$\mathcal{M}_A$	$\sigma(A)$	$\sigma_{\text{verbal}}(A)$	$v(\varphi_A, \sigma)$	$v(\sigma(A))$	Consistency A
	H	H	V	V	✓
	H	M	F	V	×
	M	H	F	F	✓
	M	M	V	F	×

Consequently, by applying the coherence rule in all possible worlds, it follows that in all coherent worlds of A it is:

$$\sigma_{\text{verbal}}(A) = \text{H}\,.$$

Now we evaluate $v(\varphi_A, \sigma_j)$ with respect to the entries of $\mathcal{M}$, and then the consistency of A taking as inputs: $v(\varphi_A, \sigma_j)$ and $v(\sigma_j(A))$, the table becomes:

	Inputs		Output	To Compare	
$\mathcal{M}$	$\sigma_j(A)$	$\sigma_{\text{verbal}}(A)$	$v(\varphi_A, \sigma_j)$	$v(\sigma_j(A))$	Consistency A
σ_1	H	H	V	V	✓
σ_2	H	H	V	V	✓
σ_3	H	H	V	V	✓
σ_4	H	H	V	V	✓
σ_5	M	H	F	F	✓
σ_6	M	H	F	F	✓
σ_7	M	H	F	F	✓
σ_8	M	H	F	F	✓

All worlds are consistent with the identity of A, so $A = \{\text{H}, \text{M}\}$. This information allows us to record A's **verbal self-assignment** as an observational constant within the Universe $\mathcal{M}$:

$$\sigma_{\text{verbal}}(A) := \text{H}$$

This verbal identity does not represent what A *is*, but what A *claims to be*, and should be used as input for evaluating the truth value of B's statements.

Step 3: Evaluating B's proposition

Person B says *"A said he is a liar"*. But we already deduced that A said: "I am H". Therefore, the proposition B is false in all worlds:

$$\varphi_B := \text{“}A \text{ said it's M”} \quad \Rightarrow \quad v(\varphi_B, \sigma_j) = \text{False}$$

We then write down the results of $v(\varphi_B, \sigma_j)$ and evaluate the consistency of B:

	Inputs		Output	To Compare	
$\mathcal{M}$	$\sigma_{\text{verbal}}(A)$	$\sigma_j(B)$	$v(\varphi_B, \sigma_j)$	$v(\sigma_j(B))$	Consistency B
σ_1	H	H	F	V	×
σ_2	H	H	F	V	×
σ_3	H	M	F	F	✓
σ_4	H	M	F	F	✓
σ_5	H	H	F	V	×
σ_6	H	H	F	V	×
σ_7	H	M	F	F	✓
σ_8	H	M	F	F	✓

then, the worlds consistent with A and B, are: $\{\sigma_3, \sigma_4, \sigma_7, \sigma_8\}$

Step 4: Evaluating C's proposition

C said: *"B is a liar"*, that is:

$$\varphi_C := (\sigma_C(B) = M)$$

We evaluate this in the remaining worlds:

	Inputs		Output	To Compare	
$\mathcal{M}$	$\sigma_j(B)$	$\sigma_j(C)$	$v(\varphi_C, \sigma_j)$	$v(\sigma_j(C))$	Consistency C
σ_3	M	H	V	V	✓
σ_4	M	M	V	F	×
σ_7	M	H	V	V	✓
σ_8	M	M	V	F	×

but it is observed that the coherent worlds σ_3 and σ_7 are the same for B and C. So we finally get for A, B and C:

$$\sigma_{3,7}: \quad A = \{H, M\}, \quad B = M, \quad C = H$$

Result

The only possible world compatible with all the statements is:

$$\boxed{A = \{H, M\}, \quad B = M, \quad C = H}$$

The model has filtered out the inconsistent worlds, leaving only one that satisfies the coherence between the identity of individuals and the value of their statements.

5. Conclusion

The proposed model provides a formal semantic framework that does not depend on prior axioms or inference rules. It can be used to analyze the consistency between an agent's identity and what it declares or represents. In the case of sensors, their connection topology can be logically transformed into the proposition to be evaluated. Its tabular approach allows for computational processing and, when applied iteratively, can improve the model's reliability in detecting consistent worlds, for example, for random liars. This method is generalizable to other problems and enables new applications in logic, artificial intelligence, and epistemology. One potential limitation is its scalability to very large universes $\mathcal{M}$, which may require optimization strategies.

[1] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning About Knowledge*. The MIT Press, Cambridge, Massachusetts; London, England, First MIT Press paperback edition, 2003. Originally published in 1995.

[2] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. First edition. Springer Science+Business Media B.V., Dordrecht, 2007. XI, 296 pages. DOI: 10.1007/978-1-4020-5839-4.

[3] Graham Priest. *An Introduction to Non-Classical Logic: From If to Is*, second edition. Cambridge University Press, Cambridge, 2008. ISBN 978-0-521-85433-7 (hardback), ISBN 978-0-521-67026-5 (paperback). Originally published in 2001. Available at `www.cambridge.org/9780521854337`.