Sep 11, 2025

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

DOI
https://dx.doi.org/10.17504/protocols.io.n92ld6wzxg5b/v1

Omar Pavel Ancka Quispe[1], omarancka[2]

[1]Facultad de Ingeniería Eléctrica y Electrónica, Escuela de Ingeniería Electrónica, Universidad Nacional del Callao, Callao, Perú;
[2]Universidad Nacional del Callao

Omar Ancka Quispe

omarancka

OPEN ACCESS

DOI: https://dx.doi.org/10.17504/protocols.io.n92ld6wzxg5b/v1

External link: https://philarchive.org/rec/ANCMPU

Protocol Citation: Omar Pavel Ancka Quispe, omarancka 2025. Possible Worlds: A Logical Model of Coherence between Identity and Veracity. **protocols.io** https://dx.doi.org/10.17504/protocols.io.n92ld6wzxg5b/v1

Protocol status: In development
We are still developing and optimizing this protocol

Created: September 11, 2025

Last Modified: September 11, 2025

Protocol Integer ID: 227059

Keywords: Logical semantics, Model theory, Semantic evaluation, Ontological identity, Veracity and coherence, Agent logic, Truth-tellers and liars, Possible worlds semantics, logical model of coherence, agent logic, applications in agent logic, ontological identity, finite tabular semantic, ontological identity to each individual, semantic operator, epistemological analysis, statements in finite world, tabular semantic, semantic model, coherence, logical model, coherence between identity, truth value, global consistency, evaluation of the global consistency, classic problems of truth, construction of all possible world, agent, formal structure, statement, finite world, consistency rule, veracity, possible world, explicit alternative to classical inferential method, context

Abstract

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.

Attachments

PDF

Possible Worlds - A ...

403KB

Guidelines

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

2.1 Universe and ontological classes

- Let U = {X1, X2, ..., Xn} be the finite set of individuals involved in the problem. Each individual Xi is assigned an ontological identity, for example: σ(Xi) ∈ {H, M}, given by: σ : U → {H, M}
- H represents an Honest individual (always tells the truth). M represents a Liar individual (always lies).
- σ(Xi) ⇒ σ ∈ M represents a complete assignment of identities to all individuals in U, which we call possible world.
- Define the set of all possible identity assignments (possible worlds) as the Universe \mathcal{M}, given by: \mathcal{M} := {σ | σ : U → {H, M}}

2.2 Truth value according to identity and evaluation

- Definition 1: Truth value according to its Identity. For σ ∈ M and Xi ∈ U, the truth value associated with the identity of Xi is: v(σ(Xi)) = { V if σ(Xi) = H; F if σ(Xi) = M }

- Definition 2: Truth value of the Statement issued by Xi. Let φ*Xi be the specific statement issued by individual Xi, and let σ ∈ M be a possible world. The truth value of the statement evaluated at σ is: v(φXi, σ)* ∈ {V, F}.

2.3 Semantic coherence criterion

- Semantic consistency rule: A possible world σ ∈ M is coherent with the statement Xi if and only if: v(σ(Xi)) = v(φ_Xi, σ)
- That is, the truth value corresponding to the identity of Xi coincides with the truth value of its statement evaluated within the same possible world.

2.3.1 Definition of valid world

- Valid world: A possible world σ ∈ 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

- σ : U → {H, M}: Identity assignment (possible world). After the assignment, σ(Xi) is the individual identity for a particular individual.
- v(σ(Xi)): Truth value assigned according to identity.
- v(φ_Xi, σ): Truth value of the statement in the model.
- The model can be extended to k identities, e.g., σ : U → {H, M, S, R}, producing k^n possible worlds and universe \mathcal{M} = {σ*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 σ ∈ \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*Xi uttered by some individual Xi ∈ U, evaluate its truth value v(φXi,* σ) 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 Xi, that is, v(σ(Xi)), where: v(σ(Xi)) = { V, if σ(Xi) = H; F, if σ(Xi) = M }
4. Check whether the consistency rule is met for each individual who made a statement: v(φ_Xi, σ) = v(σ(Xi)), This ensures that Xi's statement matches its identity in the world σ.
5. Keep only those worlds σ ∈ \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

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). The assignment of identities is represented by a function:

σ : U → {H, 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 M is constructed, which is presented in the following table. (Universe M table with σ1..σ8 assignments H/M for A, B, C.)

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 φi is assigned a truth value according to the world σ, evaluating whether what it says is true or not. This is represented as: v(φi, σ) ∈ {V, F}.

Evaluation in a specific world

Let’s look at the full procedure for the world σ4, where:

σ4(A) = H, σ4(B) = M, σ4(C) = M

- A says: “B is a liar”
  - In σ4, B is really a liar: the proposition is true ⇒ v(φ_A, σ4) = V
  - A is honest ⇒ v(σ(A)) = V
  - They match ⇒ consistently

- B says: “C is honest”
  - In σ4, C is a liar: the proposition is false ⇒ v(φ_B, σ4) = F
  - B is a liar ⇒ v(σ(B)) = F
  - They match ⇒ consistently

- C says: “A is a liar”
  - In σ4, A is honest: the proposition is false ⇒ v(φ_C, σ4) = F
  - C is a liar ⇒ v(σ(C)) = F
  - They match ⇒ consistently

Since all statements are consistent with each individual’s identity, the world σ4 is a coherent world.

Filtering inconsistent worlds

To solve the problem, all worlds σ ∈ M are evaluated. Those that contain any inconsistency are discarded, that is, for which:

v(φ*i, σ) ≠ v(σ(Xi))* 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 = {S1, S2}. 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:

M | σ(S1) | σ(S2)
σ1 | D | D
σ2 | D | C
σ3 | C | D
σ4 | 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:
  - S1 emits: 0
  - S2 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(φ*{Si}, σj) = {1 if the sensor output matches its type (C/D) in σj*, 0 otherwise.}

For example:
- If σ_j(S1) = D, then your output is expected to be ¬1 = 0, and since you output 0, then v(φ_{S1}, σ_j) = 1.
- If σ_j(S2) = C, output 1 is expected, and since it gave 1, then v(φ_{S2}, σ_j) = 1.

Let us evaluate in all worlds:

M | Inputs | Output | To Compare
| σ(S1) | σ(S2) | v(φ{*S1*}, *σj*) | v(φ{*S2*}, *σj*)
σ1 | D | D | 1 | 0
σ2 | D | C | 1 | 1
σ3 | C | D | 0 | 0
σ4 | 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(φ_{S_i}, σ) = v(σ(S_i))

Since we used 1 to mean “identity-consistent statement”, we apply:
- σ2 is the only world where both statements are consistent.

Conclusion

The valid world is:

σ2 = {S1 = D, S2 = C}

This means that:
- S1 is faulty (it outputs 0, as opposed to the actual 1).
- S2 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.

M | σ_j(A) | σ_j(B) | σ_j(C)
σ1 | H | H | H
σ2 | H | H | M
σ3 | H | M | H
σ4 | H | M | M
σ5 | M | H | H
σ6 | M | H | M
σ7 | M | M | H
σ8 | 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:

φA := "I am *σ*verbal(A)*", where σverbal(A)* ∈ {H, M}

This statement is interpreted as a proposition about the identity of A that must be evaluated under each possible world σ ∈ M_A. Since in our system honest agents always tell the truth and liars always lie, we have the criteria:

- v(φ*A, σm*): truth value of the specific proposition issued by A, evaluated in the possible world *σm of M*A.

v(φ_A, σ) = {
 V if σ(A) = H and φ_A := "I am H"
 F if σ(A) = H and φ_A := "I am M"
 F if σ(A) = M and φ_A := "I am H"
 V if σ(A) = M and φ_A := "I am M"
}

(before_start section retained from earlier pages)

Before start

Section 2: Formal Definitions of the Model

2.1 Universe and ontological classes
- Let U = {X1, X2, ..., Xn} be the finite set of individuals involved in the problem. Each individual Xi is assigned an ontological identity, for example: σ(Xi) ∈ {H, M}, given by: σ : U → {H, M}
- H represents an Honest individual (always tells the truth). M represents a Liar individual (always lies).
- σ(Xi) ⇒ σ ∈ M represents a complete assignment of identities to all individuals in U, which we call possible world.
- Define the set of all possible identity assignments (possible worlds) as the Universe \mathcal{M}, given by: \mathcal{M} := {σ | σ : U → {H, M}}

2.2 Truth value according to identity and evaluation
- Definition 1: Truth value according to its Identity. For σ ∈ M and Xi ∈ U, the truth value associated with the identity of Xi is: v(σ(Xi)) = { V if σ(Xi) = H; F if σ(Xi) = M }
- Definition 2: Truth value of the Statement issued by Xi. Let φ*Xi be the specific statement issued by individual Xi, and let σ ∈ M be a possible world. The truth value of the statement evaluated at σ is: v(φXi, σ)* ∈ {V, F}.

2.3 Semantic coherence criterion
- Semantic consistency rule: A possible world σ ∈ M is coherent with the statement Xi if and only if: v(σ(Xi)) = v(φ_Xi, σ)
- That is, the truth value corresponding to the identity of Xi coincides with the truth value of its statement evaluated within the same possible world.

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

1 We then write down the results of v(φ_B, σ_j) and evaluate the consistency of B across worlds σ1..σ8, comparing v(φ_B, σ_j) (which is False in all worlds) with v(σ_j(B)) to determine Consistency B. The worlds consistent with A and B are: {σ3, σ4, σ7, σ8}.

2 Step 4: Evaluating C's proposition — C said: "B is a liar", that is φ_C := (σ_C(B) = M). Evaluate φ_C in the remaining worlds:

3

	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	×

Fig. 1: Evaluation table for v(φ_C, σ) across the remaining worlds {σ3, σ4, σ7, σ8} (placeholder figure from the document).

4 but it is observed that the coherent worlds σ3 and σ7 are the same for B and C. So we finally get for A, B and C: σ_{3,7} : A = {H, M}, B = M, C = H

5 Result — The only possible world compatible with all the statements is: A = {H, M}, B = M, 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.

Protocol references

[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.