

Validity and Entailment

Entailment (also called logical consequence or logical implication) guarantees the truth of a conclusion, relative to a set of true premises; as such it provides a semantic basis for judging whether proposed inference rules are sound

Examples of truth in a model

Consider a model $\mathcal{M} = \{\mathcal{D}, \mathcal{I}\}$ for a first-order language that contains individual constants A, B, unary predicates P and Q, and variables x, y. Suppose that $\mathcal{D} = \{a, b\}$ (where we're using a, b are metalinguistic names for two distinct entities), and

$$\begin{aligned} A^{\mathcal{I}} &= a \\ B^{\mathcal{I}} &= b \\ P^{\mathcal{I}} &= \{a, b\} \quad (\text{we could also write } \{\langle a \rangle, \langle b \rangle\}) \\ Q^{\mathcal{I}} &= \{a\} \end{aligned}$$

Example 1

Is \mathcal{M} a model of $(\forall x P(x))$? (I.e., $\models_{\mathcal{M}} (\forall x P(x))$?)

To answer this we must check if $\models_{\mathcal{M}} (\forall x P(x)) [\mathcal{U}]$ for all variable assignments \mathcal{U} (mapping x and y to elements of $\{a, b\}$). We'll write

$$\mathcal{U}_{x:\delta}$$

to denote the variable assignment that is just like \mathcal{U} except that it assigns x the value δ . (Here δ is a metalinguistic *variable* that ranges over the individuals we're referring to by a and b . We could have used Latin letter d , as we do elsewhere, but here this might lead to confusion, suggesting that d stands for some individual *other* than a or b .) Consider any particular \mathcal{U} . (Note: there are only 4 possible choices here, since there are two possible values for each of x, y.)

$$\models_{\mathcal{M}} (\forall x P(x)) [\mathcal{U}] \text{ iff for all } \delta \in \mathcal{D} (= \{a, b\}), \models_{\mathcal{M}} P(x) [\mathcal{U}_{x:\delta}].$$

Checking the first possibility, $\delta = a$:

$$\models_{\mathcal{M}} P(x) [\mathcal{U}_{x:a}] \text{ iff } a \in \{a, b\}, \text{ which is true.}$$

$$\models_{\mathcal{M}} P(x) [\mathcal{U}_{x:b}] \text{ is similarly true since } b \in \{a, b\}.$$

Note that the choice of \mathcal{U} didn't matter, so this is true for all \mathcal{U} , so that we have shown $\models_{\mathcal{M}} (\forall x P(x))$.

Example 2

Now consider whether \mathcal{M} is a model of $(\exists x Q(x))$. In this case, we require that for all variable assignments \mathcal{U} , there must be *some* $\delta \in \{a, b\}$ such that

$$\models_{\mathcal{M}} Q(x) [\mathcal{U}_{x:\delta}].$$

But this is clearly so for the choice $\delta = a$, given that $Q^{\mathcal{I}} = \{a\}$.

N.B.: Obviously we also have $(\exists x P(x))$ true in \mathcal{M} . In fact, since we always assume a nonempty domain \mathcal{D} , the truth of $(\forall x P(x))$ always assures the truth of $(\exists x P(x))$, regardless of the interpretation \mathcal{I} .

Example 3

Finally, consider whether $(\forall x (Q(x) \Rightarrow P(x)))$ is true in \mathcal{M} . We require that for all variable assignments \mathcal{U} ,

$$\models_{\mathcal{M}} (Q(x) \Rightarrow P(x)) [\mathcal{U}_{x:\delta}] \text{ for all } \delta \in \{a, b\}.$$

For the choice $\delta = a$, we thus require

$$\not\models_{\mathcal{M}} Q(x) [\mathcal{U}_{x:a}] \text{ or } \models_{\mathcal{M}} P(x) [\mathcal{U}_{x:a}]$$

The latter condition is true, since $a \in P^{\mathcal{I}} = \{a, b\}$.

For the choice $\delta = b$ we similarly require

$$\not\models_{\mathcal{M}} Q(x) [\mathcal{U}_{x:b}] \text{ or } \models_{\mathcal{M}} P(x) [\mathcal{U}_{x:b}]$$

and the former condition is true since $b \notin \{a\}$.

Thus $(\forall x (Q(x) \Rightarrow P(x)))$ is true in \mathcal{M} .

Note that the formula would be true in \mathcal{M} even if *nothing* had property Q, i.e., $Q^{\mathcal{I}} = \{\}$. For instance if we think of the formula as saying “All Martians are green”, then this will be judged true in any model in which there are no Martians. In English, saying “All Q are P” carries a strong implicature that there *are* some Q’s. Still, we can get away with saying (without genuine contradiction),

All Martians are green, for the simple reason that there are no Martians.

Note also that while “All Q are P” allows for the possibility that there are no Q’s, the situation is different for “Some Q is P” (e.g., “Some Martians are green”). In fact, the logical form here is

$$(\exists x (Q(x) \wedge P(x))),$$

i.e., we have a conjunction, not a conditional. So here some Q must exist, and in addition it must have property P.

Validity (Logical Truth) and (Un)satisfiability

While truth in general is relative to a model, some formulas are true in *all* models. Examples are

$$P(C) \vee \neg P(C), (\forall x (x = x)), P(C) \Leftrightarrow \neg \neg P(C).$$

In fact we can generalize to whole classes of such formulas, e.g., writing ϕ for an arbitrary formula and ν for an arbitrary variable,

$$\phi \vee \neg \phi, (\forall \nu (\nu = \nu)), \phi \Leftrightarrow \neg \neg \phi$$

are examples of formulas (more accurately, formula *schemas*) that are true in all models. Such formulas are called *valid*, or *logically true*, and we write

$$\models \phi$$

to say that $\models_{\mathcal{M}} \phi$ for all models \mathcal{M} .

Example

Consider $(\phi \vee \neg \phi)$. To show that $\models (\phi \vee \neg \phi)$, we need to show that for any model \mathcal{M} , $\models_{\mathcal{M}} (\phi \vee \neg \phi)$. This in turn requires that we show that for any model \mathcal{M} and any variable assignment \mathcal{U} ,

$$\models_{\mathcal{M}} (\phi \vee \neg \phi) [\mathcal{U}].$$

But by the satisfaction conditions for “ \vee ”, this is true iff

$$\models_{\mathcal{M}} \phi [\mathcal{U}] \text{ or } \models_{\mathcal{M}} \neg \phi [\mathcal{U}].$$

By the satisfaction conditions for “ \neg ”, this is true iff

$$\models_{\mathcal{M}} \phi [\mathcal{U}] \text{ or } \not\models_{\mathcal{M}} \phi [\mathcal{U}],$$

which is obviously true.

Logicians are very interested in valid formulas, since these are true in virtue of logic alone, without regard for the domain of discourse and the interpretations of the constants of a first-order language. In that sense the valid formulas characterize the logic. Moreover, we would like to have logical axioms (or axiom schemas) and rules of inference so that every valid formula is either an axiom (instance of an axiom schema) or is deducible as a *theorem* of the logic.

At the other extreme we have *logically false*, or *unsatisfiable* formulas, i.e., ones with no models. Note that we cannot simply write $\not\models \phi$ to say that ϕ is unsatisfiable, if the “slash” just means negation. After all, a formula can perfectly well *not* be true in *all* models, yet be true in *some* models. Such formulas are called *contingent* formulas; examples are

$$P(A), \text{ Loves}(\text{Bob}, \text{Alice}), \neg \text{Loves}(\text{Bob}, \text{Alice}), (\forall x (P(x) \Rightarrow Q(x))),$$

etc. So, to express that a formula is unsatisfiable (logically false), we either use no special notation (just the word “unsatisfiable”), or we say that the *negation* of the formula is valid:

$$\models \neg \phi.$$

Formulas that have some models – this covers the valid as well as the contingent formulas – are called *satisfiable*. Admittedly the terminology of “satisfiable” and “unsatisfiable” formulas is a little misleading, given that the notion of “satisfaction” as defined earlier involved a variable assignment, as well as a model. However, it is easy to see that a formula has a model (is satisfiable) iff there is a model *and* a variable assignment such that the model and variable assignment satisfy the formula (in the previously defined sense).

Entailment (logical consequence)

From a KRR (knowledge representation and reasoning) perspective, we are more interested in what is true in those models that are compatible with the facts in the knowledge base (the *nonlogical axioms*, or *world knowledge*), rather than *all* models, without regard for the world.

For this, we define a notion closely related to validity, namely *entailment* (or *logical consequence*). First, we generalize the notion of \mathcal{M} being a model of a formula ϕ (ϕ being true in \mathcal{M}) to apply to a *set* of formulas Δ :

$$\models_{\mathcal{M}} \Delta \text{ (}\mathcal{M}\text{ is a model of } \Delta\text{) iff for all } \phi \in \Delta, \models_{\mathcal{M}} \phi.$$

What we are interested in, given some true statements Δ about a world, is *what additional* true statements we can make about that world. In such a case, we say that the latter statements are *logical consequences of* (or *entailed by*) the given statements Δ . Formally we write

$$\Delta \models \phi \text{ (}\Delta\text{ entails } \phi, \phi\text{ is a logical consequence of } \Delta\text{)}$$

iff

$$\text{for all models } \mathcal{M} \text{ such that } \models_{\mathcal{M}} \Delta, \models_{\mathcal{M}} \phi.$$

In other words, whenever all formulas in Δ are true in \mathcal{M} , so is ϕ .

We can extend the notion satisfiability or unsatisfiability to a set of sentences Δ in an obvious way:

$$\Delta \text{ is satisfiable iff there is a model } \mathcal{M} \text{ such that } \models_{\mathcal{M}} \Delta;$$

otherwise Δ is unsatisfiable. Note that for unsatisfiable Δ , we have

$$\Delta \models \phi$$

for *any* formula ϕ , since it is certainly true that all models of Δ – of which there are none – are models of ϕ ! So from unsatisfiable premises, “everything follows”.