Carnap's Problem, Definability, and Compositionality

Pedro del Valle-Inclan

Scuola Normale Superiore

Outline

What is Carnap's Problem?

- Some **'non-normal' interpretations** of the connectives and quantifiers are **consistent with** the classical consequence relation. (Carnap 1943)

Carnap's Problem (CPL)

Fix a language $\mathcal{L}$ with $\wedge, \neg,$.

- **Interpretation:** valuation v from $\mathcal{L}$-formulas to $\{1, 0\}$.
- **Normal:** v respects the usual truth-tables.
- **Consistent with:** Truth preserving. In other words, if $\Gamma \vdash_{\mathcal{L}} A$, then $v(\Gamma) = 1$ implies $v(A) = 1$.

Here $\vdash_{\mathcal{L}}$ is understood 'syntactically', as a relation between sets of sentences and sentences.

Examples

- Valuation v_T such that $v_T(A) = 1$ for all A. Trivially truth-preserving.
- Valuation v_{cl} such that $v_{cl}(A) = 1$ iff A is a classical tautology.

Bonnay and Westerståhl (2016)

The space of possible interpretations is **a priori restricted by universal semantic principles** [...]. *As speakers, we know that our language is going to have some true and some false sentences, that it will be compositional, and that its logical constituents will be topic-neutral.* (Bonnay and Westerståhl 2016)

Bonnay and Westerståhl (2016)

The space of possible interpretations is **a priori restricted by universal semantic principles** [...]. *As speakers, we know that our language is going to have* **some true and some false sentences***, that it will be* **compositional***, and that its logical constituents will be* **topic-neutral***.* (Bonnay and Westerståhl 2016)

Restrictions: non-triviality

- **Non-triviality:** A valuation v is trivial if $v(\varphi) = 1$ for all sentences φ.
- Requiring non-triviality just amounts to banning v_T.

Restrictions: compositionality

- **Compositionality:** Let μ be an assignment of semantic values to the expressions of some language. Then μ is compositional if:

  (PC): For every syntactic rule $\mathcal{O}$ there is a function on semantic values $\mathrm{F}_{\mathcal{O}}$ such that given a complex expression $\mathcal{O}(e_1, ...e_n)$, we have $\mu(\mathcal{O}(e_1, ...e_n)) = F_{\mathcal{O}}(\mu(e_1), ...\mu(e_n))$.

- In propositional logic **(PC)** just says that v must interpret each connective by a truth-function.

Solution for PL

Theorem 1

For any two-valued valuation v, if v is compositional, non-trivial and consistent with $\vdash_{\mathcal{L}}$, then v is normal for all connectives.

What is a normal interpretation? $(\wedge, \neg, \forall)$

(i) $\mathcal{M}, \sigma \models P(x_1, ...x_n)$ iff $(\sigma(x_1), ..., \sigma(x_n)) \in I(P)$.

(ii) $\mathcal{M}, \sigma \models \varphi \wedge \psi$ iff $\mathcal{M}, \sigma \models \varphi$ and $\mathcal{M}, \sigma \models \varphi$.

(iii) $\mathcal{M}, \sigma \models \neg\varphi$ iff $\mathcal{M}, \sigma \not\models \varphi$.

(iv)$\mathcal{M}, \sigma \models \forall x \varphi$ iff for all $a \in D$ $\mathcal{M}, \sigma[a/x] \models \varphi$.

What is a normal interpretation? $(\wedge, \neg, \forall)$

(i) $\mathcal{M}, \sigma \models P(x_1, ...x_n)$ iff $(\sigma(x_1), ..., \sigma(x_n)) \in I(P)$.

(ii) $\mathcal{M}, \sigma \models \varphi \wedge \psi$ iff $\mathcal{M}, \sigma \models \varphi$ and $\mathcal{M}, \sigma \models \varphi$.

(iii) $\mathcal{M}, \sigma \models \neg\varphi$ iff $\mathcal{M}, \sigma \not\models \varphi$.

(iv) $\mathcal{M}, \sigma \models \forall x \varphi$ iff for all $a \in D$ $\mathcal{M}, \sigma[a/x] \models \varphi$.

Generalised Quantifiers

Quantifiers denote sets of subsets of the domain. For example, $\forall$ denotes $\{D\}$. Instead of:

(iv)$\mathcal{M}, \sigma \models \forall x \varphi$ iff for all $a \in D$ $\mathcal{M}, \sigma[a/x] \models \varphi$.

We say:

(iv)′ $\mathcal{M}, \sigma \models \forall x \varphi$ iff $\{a \in D | \mathcal{M}, \sigma[a/x] \models \varphi\} \in \{D\}$.

What is an interpretation? $(\wedge, \neg, \forall)$

- *Since we assume compositionality* [...] *and non-triviality, we need not worry about the interpretation of connectives, which has to be standard by* [Theorem 1].

What is an interpretation? $(\wedge, \neg, \forall)$

- *Since we assume compositionality* [...] *and non-triviality, we need not worry about the interpretation of connectives, which has to be standard by* [Theorem 1].

- *Hence, our interpretations can be taken to be pairs of the form $(\mathcal{M}, Q)$ where $\mathcal{M}$ is a standard $\mathcal{L}$-structure* [...], *and Q is a set of subsets of the domain, interpreting* $\forall$.

(Bonnay and Westerståhl 2016, p. 729.)

What is an interpretation? $(\wedge, \neg, \forall)$

An interpretation for $\mathcal{L}$ is **weak model** $\mathcal{M}, Q$ where $\mathcal{M} = (D, I)$ is an $\mathcal{L}$-structure and $Q \subseteq \mathcal{P}(D)$ interprets $\forall$.

$\mathcal{M}, Q, \sigma \models P(x_1, ...x_n)$ iff $(\sigma(x_1), ..., \sigma(x_n)) \in I(P)$.

$\mathcal{M}, Q, \sigma \models \varphi \wedge \psi$ iff $\mathcal{M}, Q, \sigma \models \varphi$ and $\mathcal{M}, Q, \sigma \models \varphi$.

$\mathcal{M}, Q, \sigma \models \neg\varphi$ iff $\mathcal{M}, Q, \sigma \not\models \varphi$.

$\mathcal{M}, Q, \sigma \models \forall x \varphi$ iff $\{a \in D | \mathcal{M}, Q, \sigma[a/x] \models \varphi\} \in \mathbf{Q}$.

Normal, consistent with $\vdash_{\mathcal{L}}$

- $\mathcal{M}, Q$ is normal if $Q = \{D\}$.
- $\mathcal{M}, Q$ is consistent with $\vdash_{\mathcal{L}}$ if it is truth preserving:

  If $\Gamma \vdash_{\mathcal{L}} \varphi$ and $\mathcal{M}, Q \models \Gamma$, then $\mathcal{M}, Q \models \varphi$

Bonnay and Westerståhl (2016)

The space of possible interpretations is **a priori restricted by universal semantic principles** [...]. *As speakers, we know that our language is going to have* **some true and some false sentences**, *that it will be* **compositional**, *and that its logical constituents will be* **topic-neutral**. (Bonnay and Westerståhl 2016)

Invariance under permutation

- A permutation π of the domain D of a structure $\mathcal{M}$ is a bijection of D onto itself.
- Permutations are lifted point-wise to sets: if π is a permutation of D and $X \subseteq D$, let $\pi(X) = \{\pi(a) \in D \mid a \in X\}$.
- $Q \subseteq \mathcal{P}(D)$ is invariant under permutations if $Q = \{\pi(X) \mid X \in Q\}$ for any permutation π.
- Examples: $\{D\}$, E_D, E_n for any n...

Summary

- Interpretation is a **weak model** $\mathcal{M}, Q$, where $\mathcal{M}$ is an $\mathcal{L}$-structure, $Q \subseteq \mathcal{P}(D)$ interprets $\forall$.
- Normal interpretation of connectives is fixed. Non-normality can only come from Q.
- Invariance under permutations is supposed to get rid of it.

Bonnay and Westerståhl's strategy

- (BW1) Let $\mathcal{L}$ be a language, $\mathcal{M}, Q$ a weak model for it. Then $\mathcal{M}, Q$ is consistent with $\vdash_{\mathcal{L}}$ iff Q is a **principal filter** of the domain.

- (BW2) The only principal filter on D that's invariant under permutations is $Q = \{D\}$.

Bonnay and Westerståhl's strategy

- (BW1) Let $\mathcal{L}$ be a language, $\mathcal{M}, Q$ a weak model for it. Then $\mathcal{M}, Q$ is consistent with $\vdash_{\mathcal{L}}$ iff Q is a **principal filter** of the domain.
- (BW2) The only principal filter on D that's invariant under permutations is $Q = \{D\}$.

Why does (BW1) fail?

$\mathcal{M}, \{D\}, \sigma \models \forall x \varphi$ iff $\{a \in D | \mathcal{M}, \{D\}, \sigma[a/x] \models \varphi\} \in \{D\}$.

Strictly speaking we don't need the denotation of $\forall$ to contain *only* D. We can clutter it with extra subsets that the language cannot describe.

Roughly, we clutter it with subsets that are not blue for any φ and σ.

Why does (BW1) fail?

Given a weak model $\mathcal{M}, Q$:

- The **extension of a formula** φ relative to σ and x is: $||\varphi||^{\mathcal{M},Q}_{\sigma,x} := \{a \in D \mid (\mathcal{M}, Q), \sigma[a/x] \models \varphi\}$.
- $X \subseteq D$ is **definable** if it's the extension of some formula.
- The definable subsets of $\mathcal{M}, Q$ are denoted $\mathrm{Def}(\mathcal{M}, Q)$.

Why does (BW1) fail?

Lemma 2

Let $\mathcal{M}, Q$ and $\mathcal{M}, Q'$ be weak models such that $Q' = Q \cup B$ for some $B \subseteq \mathcal{P}(D)$ such that $B \cap \mathit{Def}(\mathcal{M}, Q) = \emptyset$. Then $\mathcal{M}, Q, \sigma \models \varphi$ iff $\mathcal{M}, Q', \sigma \models \varphi$ for any φ and σ.

How does (BW1) fail?

Corollary 3

There is a first-order language $\mathcal{L}_1$ *with a non-normal weak model* $\mathcal{N}, Q$ *consistent with the classical consequence relation* $\vdash_{\mathcal{L}_1}$.

How does (BW1) fail?

Corollary 3

There is a first-order language $\mathcal{L}_1$ with a non-normal weak model $\mathcal{N}, Q$ consistent with the classical consequence relation $\vdash_{\mathcal{L}_1}$.

Proof.

Let $\mathcal{L}_1$ have a single unary predicate symbol P, set $|D| \geq 1$, $I(P) = D$. The only definable subsets are $D, \emptyset$, so the normal model $\mathcal{N}, \{D\}$ and the non-normal model $\mathcal{N}, E_D$ are as described in Lemma 2. Note that E_D is invariant under permutations. □

Summary (definability)

- (BW1) fails because we can exploit the undefinability of subsets of the domain.
- There are non-normal interpretations **even if we assume that the interpretation of all connectives is normal.**

Compositionality

- We're assuming the normal interpretation of connectives is fixed by **compositionality and non-triviality**.
- The Tarskian clauses don't assign semantic values to formulas. They define a ternary relation between structures, assignments and formulas.

Compositionality

Let $\mathcal{M}$ be a normal model. Instead of $\mathcal{M}, \sigma \models \varphi$ we write $v_\sigma(\varphi) = 1$.

$v_\sigma(\varphi \wedge \psi) = 1$ iff $v_\sigma(\varphi) = 1$ and $v_\sigma(\psi) = 1$.

$v_\sigma(\neg\varphi) = 1$ iff $v_\sigma(\varphi) = 0$

$v_\sigma(\forall x \varphi) = 1$ iff for all σ' st. $\sigma' \sim_x \sigma$, we have $v_{\sigma'}(\varphi) = 1$

Compositionality

Since we assume compositionality [...] *and non-triviality, we need not worry about the interpretation of connectives, which has to be standard by* [Theorem 1]. (p. 729)

Problem

- **The NORMAL valuations v_σ are not compositional.**

- The value of $v_\sigma(\forall x \varphi)$ is not a function of the value of $v_\sigma(\varphi)$.

$v_\sigma(\forall x \varphi) = 1$ iff for all σ' st. $\sigma' \sim_x \sigma$ we have $v_{\sigma'}(\varphi) = 1$

Summary (compositionality)

- If the only semantic values are 0 and 1, then compositionality rules out **normal** interpretations. If there are more semantic values, Theorem 1 does not apply.

General idea

General idea

- Using second-order variables won't do. But if we look at the way we prove things, something similar to it can be justified.

General idea

- Using second-order variables won't do. But if we look at the way we prove things, something similar to it can be justified.
- So far, an interpretation is consistent with the classical consequence relation if it makes valid arguments truth-preserving.

General idea

- Using second-order variables won't do. But if we look at the way we prove things, something similar to it can be justified.
- So far, an interpretation is consistent with the classical consequence relation if it makes valid arguments truth-preserving.
- **This is too weak!**

General idea

- $P(c) \wedge Q(c) \vdash Q(c)$ is valid. We draw inferences like this without knowing the extension of P, Q or the reference of c.

General idea

- $P(c) \wedge Q(c) \vdash Q(c)$ is valid. We draw inferences like this without knowing the extension of P, Q or the reference of c.
- $\forall x P(x) \vdash P(c)$ is valid. We draw inferences like this without knowing the extension of P or the reference of c.

General idea

- $P(c) \wedge Q(c) \vdash Q(c)$ is valid. We draw inferences like this without knowing the extension of P, Q or the reference of c.
- $\forall x P(x) \vdash P(c)$ is valid. We draw inferences like this without knowing the extension of P or the reference of c.
- We take valid inferences to be truth-preserving **regardless of the interpretation of non logical vocabulary.**

General idea

'Consistency with respect to $\vdash$' should be truth-preservation **regardless of the interpretation of non-logical vocabulary.**

What is an interpretation?

- We're going to re-interpret non-logical vocabulary, so instead of starting with an $\mathcal{L}$-structure $\mathcal{M} = (D, I)$, we start with a domain D.
- An interpretation is a triple $\mathcal{T} = (F_\wedge, F_\neg, Q)$ where the $F_\diamond$ are truth-functions, $Q \subseteq \mathcal{P}(D)$ is the extension of $\forall$.
- Given a function I interpreting non-logical vocabulary, the truth-value of sentences is given in the obvious way:

What is an interpretation?

$$\mathcal{T}_\sigma(Px_1, ...x_n) = 1 \text{ iff } (\sigma(x_1), ..., \sigma(x_n)) \in I(P).$$

$$\mathcal{T}_\sigma(\varphi \wedge \psi) = F_\wedge[\mathcal{T}_\sigma(\varphi), \mathcal{T}_\sigma(\psi)].$$

$$\mathcal{T}_\sigma(\neg\varphi) = F_\neg[\mathcal{T}_\sigma(\varphi)].$$

$$\mathcal{T}_\sigma(\forall x\varphi) = 1 \text{ iff } \{a \in D |\, \mathcal{T}_{\sigma[a/x]}(\varphi) = 1\} \in Q.$$

Consistent^{+} with $\vdash_{\mathcal{L}}$

An interpretation $\mathcal{T} = (F_{\wedge}, F_{\neg}, Q)$ is consistent^{+} with $\vdash_{\mathcal{L}}$ if it is **truth-preserving for any** I interpreting non-logical vocabulary:

$$\Gamma \vdash_{\mathcal{L}} \varphi \text{ and } \mathcal{T}_{\sigma}(\Gamma) = 1 \text{ imply } \mathcal{T}_{\sigma}(\varphi) = 1.$$

Solution

Theorem 4

Let $\mathcal{T} = (F_\wedge, F_\neg, Q)$ be an interpretation for $\mathcal{L}$. If $\mathcal{T}$ is consistent$^+$ with $\vdash_\mathcal{L}$, then it is normal.

Solution

Theorem 4

Let $\mathcal{T} = (F_\wedge, F_\neg, Q)$ be an interpretation for $\mathcal{L}$. If $\mathcal{T}$ is consistent$^+$ with $\vdash_\mathcal{L}$, then it is normal.

Proof.

Negation: Suppose e.g. that $F_\neg(0) = 0$. Set $I(P) = \emptyset$ and let σ be arbitrary. Since $I(P) = \emptyset$ we must have $v_\sigma^\mathcal{T}[Px] = 0$. Then $v_\sigma^\mathcal{T}[(Px \wedge \neg Px)] = 0$, since one conjunct is false. But $F_\neg(0) = 0$, so $v_\sigma^\mathcal{T}[\neg(Px \wedge \neg Px)] = 0$. Thus, $\mathcal{T}$ invalidates the classical tautology $\vdash \neg(Px \wedge \neg Px)$. Remaining cases are similar. □

Summary

- Bonnay and Westerståhl assume compositionality, non-triviality, invariance under permutations, and predicate variables.
- We take their notion of interpretation and use consistency$^+$ instead of predicate variables.

Comparing solutions

- Appeals to open-endedness assume that inference rules remain valid under any extension of the language. This proposal only assumes that inferences remain valid under (well-behaved!) reinterpretations of the vocabulary we already have.

- We can read classical semantics both from the classical consequence relation and from **standard rules** for CL.

Thanks!

What if we change the semantics?

- Two-valued semantics is not compositional. But what if we apply Bonnay and Westerståhl's strategy to a different semantics that is?

Normal Compositional Semantics

- The semantic value $[\![\varphi]\!]^{\mathcal{M}}$ of φ in $\mathcal{M}$ is the **set of assignments** σ **such that** $\mathcal{M}, \sigma \models \varphi$.

- Connectives and quantifiers are **operations on semantic values**.

Normal Compositional Semantics

$$[\![P(x_1, ...x_n)]\!]^{\mathcal{M}} = \{\sigma \in A^{\mathcal{M}} \mid (\sigma(x_1), ..., \sigma(x_n)) \in I(P)\}$$

$$[\![\varphi \wedge \psi]\!]^{\mathcal{M}} = [\![\varphi]\!]^{\mathcal{M}} \cap [\![\psi]\!]^{\mathcal{M}}$$

$$[\![\neg\varphi]\!]^{\mathcal{M}} = A^{\mathcal{M}} - [\![\varphi]\!]^{\mathcal{M}}$$

$$[\![\forall x \varphi]\!]^{\mathcal{M}} = \{\sigma \in A^{\mathcal{M}} \mid \sigma' \in [\![\varphi]\!]^{\mathcal{M}} \text{ for all } \sigma' \text{ st. } \sigma \sim_x \sigma'\}$$

Normal Compositional Semantics

$$[\![P(x_1, ...x_n)]\!]^{\mathcal{M}} = \{\sigma \in A^{\mathcal{M}} \mid (\sigma(x_1), ..., \sigma(x_n)) \in I(P)\}$$

$$[\![\varphi \wedge \psi]\!]^{\mathcal{M}} = [\![\varphi]\!]^{\mathcal{M}} \cap [\![\psi]\!]^{\mathcal{M}}$$

$$[\![\neg\varphi]\!]^{\mathcal{M}} = A^{\mathcal{M}} - [\![\varphi]\!]^{\mathcal{M}}$$

$$[\![\forall x \varphi]\!]^{\mathcal{M}} = \{\sigma \in A^{\mathcal{M}} \mid \sigma' \in [\![\varphi]\!]^{\mathcal{M}} \text{ for all } \sigma' \text{ st. } \sigma \sim_x \sigma'\}$$

What is an interpretation?

Given $\mathcal{M} = (D, I)$:

- **Interpretation:** function from formulas to $\mathcal{P}(A^{\mathcal{M}})$.
- **Compositional:** $\mathcal{T} = (f_0, F_\wedge, F_\neg, F_\forall)$, where the $F_\diamond$ are operations on $\mathcal{P}(A^{\mathcal{M}})$.
- **Normal:** the $F_\diamond$ are the intended operations.
- **Consistent:** whenever $\Gamma \vdash_{\mathcal{L}} \varphi$, we have that $\bigcap_{\gamma \in \Gamma} [\![\gamma]\!]^{\mathcal{T}} \subseteq [\![\varphi]\!]^{\mathcal{T}}$.

Permutation invariance (McGee 1996)

Let π is a permutation of D, X a set of variable assignments.

- π is lifted to sets of assignments pointwise: $\pi^*(X) = \{\pi \circ \sigma \in A^{\mathcal{M}} | \sigma \in X\}$

- An n-ary operation $\mathcal{F}_\diamond$ on $A^{\mathcal{M}}$ is **invariant under permutations** if $\pi^*(\mathcal{F}_\diamond(X_1, ..., X_n)) = \mathcal{F}_\diamond(\pi^*(X_1), ..., \pi^*(X_n))$ for all permutations π.

Problem

- **More semantic values** means **more (non-normal) interpretations.**
- There are non-normal valuations **even if every subset of the domain is definable**.
- Why? Given a sufficiently large domain, some sets of assignments can *never* be the value of *any* formula on *any* normal interpretation.

More semantic values than formulas

- In a normal structure, $\mathcal{M}, \sigma \models \varphi$ depends on the value σ assigns to the **(finitely many)** free variables $\vec{x}$ of φ.

- If $\sigma \in [\![\varphi]\!]^{\mathcal{M}}$, there are finitely many variables $\vec{x}$ st. if σ' differs from σ only in $\vec{x}$, then $\sigma' \notin [\![\varphi]\!]^{\mathcal{M}}$

- $Y \subseteq A^{\mathcal{M}}$ is **dependent** if, for some **finite** $\vec{x}$, there are σ and a σ' that differ at most in the values of $\vec{x}$ but such that $\sigma \in Y$ and $\sigma' \notin Y$. Otherwise Y is **independent**.

Compositionality

Observation 5

The value of $[\![\varphi]\!]^{\mathcal{M}}$ in a normal interpretation based on $\mathcal{M}$ is always $A^{\mathcal{M}}, \emptyset$, or a **dependent** set of assignments, for arbitrary φ and $\mathcal{M}$.

Compositionality

Observation 5

The value of $[\![\varphi]\!]^{\mathcal{M}}$ in a normal interpretation based on $\mathcal{M}$ is always $A^{\mathcal{M}}, \emptyset$, or a **dependent** set of assignments, for arbitrary φ and $\mathcal{M}$.

Observation 6

Given a structure $\mathcal{M}$ for $\mathcal{L}$, some **independent** sets of assignments are **invariant** under permutations.

Compositionality

Observation 5

The value of $[\![\varphi]\!]^{\mathcal{M}}$ in a normal interpretation based on $\mathcal{M}$ is always $A^{\mathcal{M}}, \emptyset$, or a **dependent** set of assignments, for arbitrary φ and $\mathcal{M}$.

Observation 6

Given a structure $\mathcal{M}$ for $\mathcal{L}$, some **independent** sets of assignments are **invariant** under permutations.

Example: the set C_∞ of assignments that give the same value to infinitely many variables.

Compositionality

Lemma 7

Let $\mathcal{L}$ be any first-order language. Then there is a compositional, non-trivial, non-normal interpretation $\mathcal{T}$ that is consistent with $\vdash_{\mathcal{L}}$.

Sketch: Let $|D| \geq \omega$. Define an interpretation that behaves normally for all sets of assignments *except* for C_∞, where they are identity. The resulting operations are invariant, because C_∞ is invariant. Also, the non-normality is not 'felt', because C_∞ is never the semantic value of a formula, so the interpretation is consistent.

Summary (compositionality)

- (FO) compositional interpretations require *a lot* of semantic values, and more semantic values means more problems.

- **Compositionality does not solve Carnap's Problem, it only makes it worse.**