Truth[1]

Marco Degano

Philosophical Logic 2025
10 November 2025

[1] Slides also draw on teaching material by Frank Veltman.

Readings

Optional:

- Tarski, Alfred (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research* 4(3), 341–376.
- Kripke, Saul (1975). Outline of a theory of truth. *The Journal of Philosophy* 72(19), 690–716.

Plan

1. Tarski on Truth and the Liar
2. Interlude: Ordinals & Transfinite Methods
3. Kripke's Theory of Truth

Outline

1. Tarski on Truth and the Liar

2. Interlude: Ordinals & Transfinite Methods

3. Kripke's Theory of Truth

Convention T and the Liar

- A language $\mathcal{L}$ is **semantically closed** when it contains:
  - names for *its own* sentences $'\varphi'$, and
  - a truth predicate T applying to such sentences.
- We take T to satisfy the T-schema:

$$\text{(T)} \qquad T('\varphi') \leftrightarrow \varphi$$

- The Liar sentence is $\lambda \equiv \neg T('\lambda')$.
- Then by (T) we reach a contradiction in classical logic:

$$T('\lambda') \leftrightarrow \lambda \leftrightarrow \neg T('\lambda')$$

Therefore (T) *cannot* hold unrestrictedly within a semantically closed language.

Self-reference is not necessary

A: A is false.

B: The next sentence is true.
C: The last sentence is false.

Tarski: informal picture

Alfred Tarski (1901–1983)

- The goal is to define *'is true'* for a *given formal language*.
- We aim to preserve classical logic.
- And avoid semantic paradoxes.
- **Strategy:** define truth for the **object-language** in a richer **metalanguage**.

Two levels

Object language $\mathcal{L}_0$:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \rightarrow \varphi)$$

Metalanguage $\mathcal{L}_1$ extends $\mathcal{L}_0$ with:

- **Naming:** if $\varphi \in \mathsf{Sent}(\mathcal{L}_0)$ then the name $'\varphi'$ is available in $\mathcal{L}_1$.
- **Truth predicate:** T_0 that applies *only to names of* $\mathcal{L}_0$*-sentences*. Thus $T_0('\varphi')$ is an $\mathcal{L}_1$-sentence.

No $\mathcal{L}_1$-sentence may occur as an argument of T_0.

Transparent truth across levels

The T-schema now lives in $\mathcal{L}_1$. For any $\varphi \in \mathsf{Sent}(\mathcal{L}_0)$,

$$T_0('\varphi') \ \leftrightarrow \ \varphi.$$

Blocking the Liar precisely (typed restriction)

Attempt $\lambda \equiv \neg T_0('\lambda')$.

- λ would be an $\mathcal{L}_1$-sentence (it mentions T_0).
- But T_0 takes only names of $\mathcal{L}_0$-sentences.
- Hence $T_0('\lambda')$ is **not well-formed**. The would-be Liar is a **type error**.

Hierarchy of languages

To speak about $\mathcal{L}_1$-sentences, ascend:

- $\mathcal{L}_2$ adds a truth predicate T_1 for names of $\mathcal{L}_1$-sentences.
- Transparency persists one level down:

$$T_1('T_0('p')') \leftrightarrow T_0('p') \leftrightarrow p.$$

Critical remarks: multiplicity of 'truths'

Multiplicity:

- A **hierarchy of languages** yields a family of truth predicates $T_0, T_1, T_2, \ldots$ each restricted to $\mathcal{L}_i$.
- Yet ordinary discourse uses a single word 'true'.

Reply:

- Each T_i expresses the *same concept* of truth, but restricted to a specific object-language.
- Natural languages are (potentially) *semantically closed*. Paradoxes show an unrestricted truth predicate is unsafe without typing.
- For formal work: keep truth *language-relative and stratified*. For ordinary talk: the untyped 'true' is a convenient gloss that misleads in paradoxical contexts.

[Sher 1999]

Kripke's challenge to Tarski: contingent paradoxicality

(A) *Alice:* "Most of Bob's assertions about Amsterdam are false."

(B) *Bob:* "Everything Alice says about Amsterdam is true."

Two contingent scenarios:

- *Scenario 1 (independent falsehoods):* Many of Bob's other assertions are false. Then (A) is true. Hence (B) is true. No paradox.
- *Scenario 2 (knife-edge count):* Bob has N other assertions, exactly $N/2$ true and $N/2$ false. Alice has exactly one assertion. One can verify that (A) and (B) are both paradoxical (true iff false).

- (A) evaluates Bob's assertions (incl. (B)) $\Rightarrow$ (A) must be *above* (B). (B) evaluates Alice's assertions (incl. (A)) $\Rightarrow$ (B) must be *above* (A). **No consistent level assignment**.
- Moreover, whether they're harmless or paradoxical depends on **contingent counts**. But Tarskian level-typing is fixed *syntactically*, not allowed to vary with the facts.

Kripke: informal picture

Saul Kripke (1940-2022)

- Start with a language where the truth predicate is **initially uninterpreted.**
- **Extend** its interpretation by declaring more sentences true in stages.
- The operator is **monotone**: once a sentence is true, it remains true at later stages.
- A **fixed point** is a stage where further extensions yield no change.
- The Liar comes out **ungrounded** at the least fixed point: neither true nor false.

Outline

1. Tarski on Truth and the Liar

2. Interlude: Ordinals & Transfinite Methods

3. Kripke's Theory of Truth

Why ordinals here?

- We will iterate a **revaluation operator** for the interpretation of the truth predicate through many stages until a **fixed point**.
- Finite iteration can miss fixed points.
- We need a notion of **stage** that goes *beyond all finite numbers* and keeps going in a well-behaved way.
- Ordinals give canonical **well-ordered** stages: $0, 1, 2, \ldots, \omega, \omega+1, \ldots$

Well-orders

Definition (Well-order)

A strict order $<$ on a set X is a **well-order** iff:

- (Linear) $\forall x \neq y \in X \left(x < y \vee y < x\right)$, transitive, irreflexive.
- (No infinite descent) Every nonempty $A \subseteq X$ has a $<$-least element.

$(\mathbb{N}, <)$ is well-ordered.

$(\mathbb{Z}, <)$ is not well-ordered.

Ordinals

- A **well-order** is like a perfect queue: every subqueue has a *first* person.
- An **ordinal** is the *position* in such a queue. We package positions as sets so that 'is earlier than' becomes just $\in$.
- Given any well-ordered $\langle X, < \rangle$, **rename** each $x \in X$ by the set of elements before it: $x \;\mapsto\; \{y \in X : y < x\}$
- After everyone is renamed, the collection of new names is an *ordinal* α

Ordinal Construction

$$
\begin{aligned}
0 &:= \varnothing, \\
1 &:= \{0\} = \{\varnothing\} \\
2 &:= \{0, 1\} = \{\varnothing, \{0\}\} \\
3 &:= \{0, 1, 2\} = \ \ldots \\
\vdots \quad & \quad \vdots \\
n &:= \{0, 1, \ldots, n-1\} \\
n{+}1 &:= n \cup \{n\} \qquad \text{(successor).}
\end{aligned}
$$

Facts:

- $m < n \iff m \in n$
- each n is *transitive*: $\forall x \in n\,(x \subseteq n)$
- $\langle n, \in \rangle$ is a well-order

From the finite ordinals we form the first limit ordinal $\omega := \{0, 1, 2, \ldots\}$, which corresponds to the well-order $(\mathbb{N}, <)$.

Successor and limit ordinals

Every ordinal is exactly one of:

- 0
- a **successor** $\alpha{+}1 := \alpha \cup \{\alpha\}$ (has a greatest element α)
- a **limit** λ (nonzero ordinal with *no* greatest element)

$$\lambda \text{ is limit} \quad \Longleftrightarrow \quad \lambda \;=\; \sup\{\gamma : \gamma < \lambda\} \;=\; \bigcup_{\gamma<\lambda} \gamma$$

Transfinite induction

Given a property $P(\alpha)$ for any ordinal α. To prove $\forall\alpha\ P(\alpha)$ it suffices to show:

1. **Base:** $P(0)$
2. **Successor:** for every ordinal α, if $P(\alpha)$ holds, then $P(\alpha+1)$ holds as well.
3. **Limit:** for every limit ordinal λ, if $P(\beta)$ holds for every $\beta < \lambda$, then $P(\lambda)$ holds.

Building sequences

Let S be a set and $F : \mathcal{P}(S) \to \mathcal{P}(S)$

We construct the following sequence (X_α) recursively:

$$X_0 := X \subseteq S, \qquad X_{\alpha+1} := F(X_\alpha), \qquad X_\lambda := \bigcup_{\beta<\lambda} X_\beta \ \ (\lambda \text{ limit}).$$

This abstract pattern is exactly what we will use for the revaluation sequence later on.

Worked example

Let $S = \mathbb{N}$ and $F(X) = X \cup \{\, n{+}1 : n \in X \,\}$ (add all successors).

$$X_0 = \{0\}, \qquad X_1 = F(X_0) = \{0, 1\}, \qquad X_2 = \{0, 1, 2\}, \ \ldots$$

$$X_\omega = \bigcup_{n<\omega} X_n = \mathbb{N}$$

By definition,

$$F(X_\omega) = X_\omega \ \cup \ \{\, n{+}1 : n \in X_\omega \,\} = \mathbb{N} \ \cup \ \{1, 2, 3, \ldots\} = \mathbb{N} = X_\omega.$$

Thus $F(X_\omega) = X_\omega$, so $\rho = \omega$ is the 'stabilization' stage.

(Importantly, this was possible because this F was **monotone**: $X \subseteq Y \Rightarrow F(X) \subseteq F(Y)$. A property relevant also for our revaluation operator.)

Outline

Language

Let $\mathcal{L}$ be a first-order language with

- a one-place predicate T (read: $T(a)$ means 'a is true'), and
- for every sentence φ, an individual term $'\varphi'$ naming φ.

We write $S_{\mathcal{L}}$ for the set of all $\mathcal{L}$-sentences.

Models

A model is a triple $M = \langle D, I, \mathcal{T} \rangle$ where

- $S_{\mathcal{L}} \subseteq D$.
- I assigns:
  - an element $I(a) \in D$ to each individual constant a.
  - the sentence φ to the name $'\varphi'$.
  - to each n-ary predicate $P \neq T$ a function $I(P) : D^n \to \{0, 1\}$.
- $\mathcal{T} \subseteq S_{\mathcal{L}} \times \{0, 1\}$ is a (possibly partial) interpretation of T.

Satisfaction relations

We give the satisfaction clauses in a bilateral representation, which is Strong Kleene in disguise.

For $P \neq T$:

$$M \vDash P(a_1, \ldots, a_n) \iff I(P)(\langle I(a_1), \ldots, I(a_n)\rangle) = 1$$

$$M \dashv P(a_1, \ldots, a_n) \iff I(P)(\langle I(a_1), \ldots, I(a_n)\rangle) = 0$$

Boolean connectives:

$$M \vDash \neg\varphi \iff M \dashv \varphi$$

$$M \dashv \neg\varphi \iff M \vDash \varphi$$

$$M \vDash \varphi \wedge \psi \iff M \vDash \varphi \text{ and } M \vDash \psi$$

$$M \dashv \varphi \wedge \psi \iff M \dashv \varphi \text{ or } M \dashv \psi$$

$$M \vDash \varphi \vee \psi \iff M \vDash \varphi \text{ or } M \vDash \psi$$

$$M \dashv \varphi \vee \psi \iff M \dashv \varphi \text{ and } M \dashv \psi$$

$$M \vDash \varphi \to \psi \iff M \dashv \varphi \text{ or } M \vDash \psi$$

$$M \dashv \varphi \to \psi \iff M \vDash \varphi \text{ and } M \dashv \psi$$

Satisfaction relations

Quantifiers:

$$M \vDash \exists x\, \varphi \iff M \vDash \varphi[a/x] \text{ for some individual constant } a$$
$$M \mathrel{\text{⫤}} \exists x\, \varphi \iff M \mathrel{\text{⫤}} \varphi[a/x] \text{ for all individual constant } a$$
$$M \vDash \forall x\, \varphi \iff M \vDash \varphi[a/x] \text{ for all individual constant } a$$
$$M \mathrel{\text{⫤}} \forall x\, \varphi \iff M \mathrel{\text{⫤}} \varphi[a/x] \text{ for some individual constant } a$$

Truth predicate:

$$M \vDash T(a) \iff \langle I(a), 1\rangle \in \mathcal{T}$$
$$M \mathrel{\text{⫤}} T(a) \iff \langle I(a), 0\rangle \in \mathcal{T}$$

Total vs Partial

Definition (Consistency / No-gluts)

For each $\varphi \in S_{\mathcal{L}}$,

$$\text{not both } \langle \varphi, 1 \rangle \in \mathcal{T} \text{ and } \langle \varphi, 0 \rangle \in \mathcal{T}$$

At most one label per sentence, but gaps are allowed.

Definition (Totality / No-gaps)

For each $\varphi \in S_{\mathcal{L}}$,

$$\langle \varphi, 1 \rangle \in \mathcal{T} \text{ or } \langle \varphi, 0 \rangle \in \mathcal{T}$$

Every sentence gets a label. *Consistency* $+$ *Totality* $\Rightarrow$ exactly one label per sentence.

Transparent truth?

Can we have full transparency

$$M \vDash T('\varphi') \iff M \vDash \varphi$$

for *all* φ while allowing self-reference and a total $\mathcal{T}$?

No: in the presence of a Liar name l with $I(l) = \neg T(l)$, transparency forces T to be both true and false of $\neg T(l)$.

Example: the Liar

Let l be a constant with $I(l) = \neg T(l)$. Then

$$M \vDash T('\neg T(l)') \iff \langle I('\neg T(l)'), 1\rangle \in \mathcal{T} \iff \langle \neg T(l), 1\rangle \in \mathcal{T}$$

But:

$$M \vDash \neg T(l) \iff M \mathrel{=\!\!|} T(l) \iff \langle I(l), 0\rangle \in \mathcal{T} \iff \langle \neg T(l), 0\rangle \in \mathcal{T}$$

Thus full transparency with total $\mathcal{T}$ is not possible.

We assume Consistency but not Totality, so unlabelled (undefined) sentences can occur.

Revaluation operator

Given $M = \langle D, I, \mathcal{T} \rangle$, define $J(M) = \langle D, I, J(\mathcal{T}) \rangle$ by

$$\langle \varphi, 1 \rangle \in J(\mathcal{T}) \iff M \vDash \varphi$$

$$\langle \varphi, 0 \rangle \in J(\mathcal{T}) \iff M \Dashv \varphi$$

We call $J(M)$ the *revaluation* of M, and $J(\mathcal{T})$ the *revaluation* of $\mathcal{T}$.

A valuation $\mathcal{T}$ is *coherent* iff $\mathcal{T} \subseteq J(\mathcal{T})$ (soundness w.r.t. the induced semantics).

Warm-up example

Let $M = \langle D, I, \mathcal{T} \rangle$ with $\mathcal{T} = \varnothing$. For a predicate $P \neq T$ with $I(P)(I(a)) = 1$,

Initial Model:

- $M \vDash P(a)$
- $M \nvDash T('P(a)')$
- …

First revaluation:

- $J(M) \vDash P(a)$
- $J(M) \vDash T('P(a)')$
- $J(M) \nvDash T('T('P(a)')')$
- …

Second revaluation:

- $J(J(M)) \vDash P(a)$
- $J(J(M)) \vDash T('P(a)')$
- $J(J(M)) \vDash T('T('P(a)')')$
- …

Example: the Liar

'This sentence is false.'

Assume a constant l with $I(l) = \neg T(l)$ and abbreviate $\lambda := \neg T(l)$.

Let $\mathcal{T} := \{\langle \lambda, 0 \rangle\}$ (i.e., we set the Liar to be false). To be coherent we need $\langle \lambda, 0 \rangle \in J(\mathcal{T})$:

$$\begin{aligned} \langle \lambda, 0 \rangle \in J(\mathcal{T}) &\iff M \Dashv \lambda \\ &\iff M \vDash T(l) \\ &\iff \langle I(l), 1 \rangle \in \mathcal{T} \\ &\iff \langle \lambda, 1 \rangle \in \mathcal{T} \end{aligned}$$

But $\langle \lambda, 1 \rangle \notin \mathcal{T}$, so $\langle \lambda, 0 \rangle \notin J(\mathcal{T})$ and $\mathcal{T} \not\subseteq J(\mathcal{T})$. Hence $\mathcal{T}$ **is incoherent**.

$\mathcal{T}' := \{\langle \lambda, 1 \rangle\}$ is incoherent, too.

($\mathcal{T} = \{\langle \neg T(l), 0 \rangle, \langle \neg T(l), 1 \rangle\}$ would be coherent, but inconsistent.)

The Truth-teller

'This sentence is true.'

Assume a constant t with $I(t) = T(t)$ and write $\tau := T(t)$.

The valuations

$$\mathcal{T}_+ := \{\langle \tau, 1 \rangle\} \qquad \text{and} \qquad \mathcal{T}_- := \{\langle \tau, 0 \rangle\}$$

are both *coherent*

$$\begin{aligned} \langle \tau, 1 \rangle \in J(\mathcal{T}_+) &\iff M \vDash \tau \\ &\iff M \vDash T(t) \\ &\iff \langle I(t), 1 \rangle \in \mathcal{T}_+ \\ &\iff \langle T(t), 1 \rangle \in \mathcal{T}_+ \end{aligned}$$

Stability of Revaluation

A useful lemma is the monotonicity of the revaluation J.

Lemma (Monotonicity of J)

Fix D and I. Let $M = \langle D, I, \mathcal{T} \rangle$ and $M' = \langle D, I, \mathcal{T}' \rangle$ with $\mathcal{T} \subseteq \mathcal{T}'$. Then the revaluation operator J is monotone:

$$J(\mathcal{T}) \subseteq J(\mathcal{T}')$$

(The proof is by induction, and you will prove this in your assignment. Be careful on what counts as the atomic case)

Revaluation sequence

Let $M = \langle D, I, \mathcal{T} \rangle$ be a model.

Define the sequence of valuations $(\mathcal{T}_\sigma)$ and models $M_\sigma = \langle D, I, \mathcal{T}_\sigma \rangle$ recursively as follows:

- $\mathcal{T}_0 \coloneqq \mathcal{T}$
- $\mathcal{T}_{\sigma+1} \coloneqq J(\mathcal{T}_\sigma)$ for every ordinal σ
- $\mathcal{T}_\lambda \coloneqq \bigcup_{\tau<\lambda} \mathcal{T}_\tau$ for every limit ordinal λ

Visually: $M_0,\ M_1,\ \ldots,\ M_\omega,\ M_{\omega+1},\ \ldots$

Monotonicity

Theorem (Monotonicity of the revaluation sequence)

Fix D and I. Let $M = \langle D, I, \mathcal{T} \rangle$ be a model with *coherent* $\mathcal{T}$ (i.e. $\mathcal{T} \subseteq J(\mathcal{T})$), where J is the revaluation operator computed relative to D, I. Define the revaluation sequence recursively:

$$\mathcal{T}_0 \coloneqq \mathcal{T} \qquad \mathcal{T}_{\sigma+1} \coloneqq J(\mathcal{T}_\sigma) \qquad \mathcal{T}_\lambda \coloneqq \bigcup_{\tau<\lambda} \mathcal{T}_\tau \ \ (\lambda \text{ limit})$$

Then for all ordinals σ, τ with $\sigma < \tau$,

$$\mathcal{T}_\sigma \subseteq \mathcal{T}_\tau$$

Stepwise growth

A standard result for increasing sequences:

Lemma (Stepwise growth)

For every ordinal α, we have $\mathcal{T}_\alpha \subseteq \mathcal{T}_{\alpha+1}$

Proof (transfinite induction on α):

Base $(\alpha = 0)$: $\mathcal{T}_0 \subseteq J(\mathcal{T}_0) = \mathcal{T}_1$.

Successor: If $\mathcal{T}_\beta \subseteq \mathcal{T}_{\beta+1}$, then by monotonicity $\mathcal{T}_{\beta+1} = J(\mathcal{T}_\beta) \subseteq J(\mathcal{T}_{\beta+1}) = \mathcal{T}_{\beta+2}$.

Limit: Let λ be limit and take $x \in \mathcal{T}_\lambda = \bigcup_{\beta<\lambda} \mathcal{T}_\beta$. Then $x \in \mathcal{T}_\beta$ for some $\beta < \lambda$, hence by the inductive hypothesis $x \in \mathcal{T}_{\beta+1} = J(\mathcal{T}_\beta) \subseteq J(\mathcal{T}_\lambda) = \mathcal{T}_{\lambda+1}$. Thus $\mathcal{T}_\lambda \subseteq \mathcal{T}_{\lambda+1}$.

Monotonicity

If $\sigma < \tau$ then $\mathcal{T}_\sigma \subseteq \mathcal{T}_\tau$

Proof (transfinite induction on τ, with σ fixed).

- *Base* $(\tau = 0)$: vacuous, since there is no $\sigma < 0$.
- *Successor* $(\tau = \kappa + 1)$: then $\sigma \leq \kappa$. By IH, $\mathcal{T}_\sigma \subseteq \mathcal{T}_\kappa$. By Stepwise Growth, $\mathcal{T}_\kappa \subseteq \mathcal{T}_{\kappa+1} = \mathcal{T}_\tau$. Hence $\mathcal{T}_\sigma \subseteq \mathcal{T}_\tau$.
- *Limit* $(\tau = \lambda)$: $\mathcal{T}_\lambda = \bigcup_{\rho<\lambda} \mathcal{T}_\rho$, so for any $\sigma < \lambda$, $\mathcal{T}_\sigma \subseteq \mathcal{T}_\lambda$.

Fixed points

Theorem (Fixed point and stabilization)

Fix D and I. Let $M = \langle D, I, \mathcal{T} \rangle$ be a model with *coherent* $\mathcal{T}$ (i.e. $\mathcal{T} \subseteq J(\mathcal{T})$), and define the revaluation sequence by:

$$\mathcal{T}_0 := \mathcal{T} \qquad \mathcal{T}_{\sigma+1} := J(\mathcal{T}_\sigma) \qquad \mathcal{T}_\lambda := \bigcup_{\tau<\lambda} \mathcal{T}_\tau \;\; (\lambda \text{ limit})$$

Then there exists a unique ordinal ρ such that:

$$\text{(i) } \forall\, \sigma < \tau \leq \rho : \; \mathcal{T}_\sigma \subsetneq \mathcal{T}_\tau \qquad \text{(ii) } \forall\, \sigma \geq \rho : \; \mathcal{T}_\sigma = \mathcal{T}_\rho$$

In particular, $J(\mathcal{T}_\rho) = \mathcal{T}_\rho$, so $\mathcal{T}_\rho$ is the **fixed point** of J generated by $\mathcal{T}$. We write $\mathcal{T}^* := \mathcal{T}_\rho$ for this point.

The sequence is increasing

Fix D and I. Write $(\mathcal{T}_\alpha)_{\alpha\in\mathrm{Ord}}$ for the transfinite sequence $\alpha \mapsto \mathcal{T}_\alpha$ of valuations.

Let $S := S_{\mathcal{L}} \times \{0,1\}$, where $S_{\mathcal{L}}$ is the set of all $\mathcal{L}$-sentences.

Each stage $\mathcal{T}_\alpha$ is a set of labeled sentences, hence

$$\mathcal{T}_\alpha \subseteq \{\langle\varphi, t\rangle : \varphi \in S_{\mathcal{L}},\ t \in \{0,1\}\} = S \quad \text{for all } \alpha.$$

By Monotonicity, for all ordinals $\sigma < \tau$,

$$\mathcal{T}_\sigma \subseteq \mathcal{T}_\tau$$

Thus $(\mathcal{T}_\alpha)_{\alpha\in\mathrm{Ord}}$ forms a (weakly) increasing chain of subsets of S (i.e., $\sigma < \tau \Rightarrow \mathcal{T}_\sigma \subseteq \mathcal{T}_\tau$, and $\mathcal{T}_\alpha \subseteq S$ for all α).

Some equality must occur

Each stage is a valuation set $\mathcal{T}_\alpha \subseteq S$, where $S := S_{\mathcal{L}} \times \{0, 1\}$. Hence there are *at most* $|\mathcal{P}(S)|$ distinct stages.

Let $\kappa := |\mathcal{P}(S)|^+$ (the successor of $|\mathcal{P}(S)|$), so $|\mathcal{P}(S)| < \kappa$. Consider the function $f : \{\alpha \in \mathrm{Ord} : \alpha < \kappa\} \mapsto \mathcal{P}(S), \qquad f(\alpha) := \mathcal{T}_\alpha$.

If the family $(\mathcal{T}_\alpha)_{\alpha<\kappa}$ were pairwise distinct, then f would be injective, hence $\kappa \leq |\mathcal{P}(S)|$, which is impossible.

Thus there exist $\alpha < \beta < \kappa$ with $\mathcal{T}_\alpha = \mathcal{T}_\beta$. Since the sequence is increasing, for every γ with $\alpha \leq \gamma \leq \beta$,

$$\mathcal{T}_\alpha \subseteq \mathcal{T}_\gamma \subseteq \mathcal{T}_\beta = \mathcal{T}_\alpha,$$

so $\mathcal{T}_\gamma = \mathcal{T}_\alpha$. In particular $\mathcal{T}_{\alpha+1} = \mathcal{T}_\alpha$.

Set $E := \{\delta : \mathcal{T}_\delta = \mathcal{T}_{\delta+1}\}$. Then $E \neq \varnothing$. Let $\rho := \min E$ (exists by well-ordering). Then:

- If $\sigma < \tau \leq \rho$, we cannot have $\mathcal{T}_\sigma = \mathcal{T}_\tau$. By monotonicity, $\mathcal{T}_\sigma \subsetneq \mathcal{T}_\tau$.
- $\mathcal{T}_\rho = \mathcal{T}_{\rho+1} = J(\mathcal{T}_\rho)$, so $\mathcal{T}_\rho$ is a fixed point.

Stability from ρ on

We prove by transfinite induction on $\sigma \geq \rho$ that $\mathcal{T}_\sigma = \mathcal{T}_\rho$.

- **Base** $\sigma = \rho$: $\mathcal{T}_\rho = \mathcal{T}_\rho$.
- **Successor** $\sigma = \gamma + 1 \geq \rho + 1$: By IH, $\mathcal{T}_\gamma = \mathcal{T}_\rho$, and since $\mathcal{T}_\rho$ is a fixed point ($\mathcal{T}_\rho = J(\mathcal{T}_\rho)$),

$$\mathcal{T}_\sigma = J(\mathcal{T}_\gamma) = J(\mathcal{T}_\rho) = \mathcal{T}_\rho$$

- **Limit** $\lambda > \rho$: By IH, $\mathcal{T}_\beta = \mathcal{T}_\rho$ for all $\rho \leq \beta < \lambda$, and by monotonicity $\mathcal{T}_\beta \subseteq \mathcal{T}_\rho$ for $\beta < \rho$. Thus

$$\mathcal{T}_\lambda = \bigcup_{\beta<\lambda} \mathcal{T}_\beta = \Big(\bigcup_{\beta<\rho} \mathcal{T}_\beta \Big) \cup \Big(\bigcup_{\rho\leq\beta<\lambda} \mathcal{T}_\beta \Big) \subseteq \mathcal{T}_\rho \cup \mathcal{T}_\rho = \mathcal{T}_\rho$$

  while $\mathcal{T}_\rho \subseteq \mathcal{T}_\lambda$ since $\rho < \lambda$. Hence $\mathcal{T}_\lambda = \mathcal{T}_\rho$.

Therefore $\mathcal{T}_\sigma = \mathcal{T}_\rho$ for every $\sigma \geq \rho$.

Minimal fixed point

Corollary

Fix D and I, and let J be the revaluation operator (relative to D, I). Let $\mathcal{T}, \mathcal{T}'$ be coherent with $\mathcal{T} \subseteq \mathcal{T}'$. Let $\mathcal{T}^*$ and $\mathcal{T}'^*$ be the fixed points generated by $\mathcal{T}$ and $\mathcal{T}'$, respectively. Then

$$\mathcal{T}^* \subseteq \mathcal{T}'^*$$

In particular, $\varnothing^*$ is the least fixed point.

Kinds of sentences

fixed points behaviour		**example**
true in all, false in no	**grounded** true	
true in some, false in no		(Exercise)[2]
false in all, true in no	**grounded** false	
false in some, true in no		(Assignment)
true in no, false in no	**paradoxical**	the liar
true in some, false in some	**biconsistent**	the truth teller

Read *some = some but not all*

[2] $\varphi = T(t) \vee \neg T(t)$ in models in which $I(t) = T(t)$.

Some reminders

- Before proving the statements, some reminders and facts.
- Throughout the examples we fix D and specify I. The initial valuation $\mathcal{T}$ varies.
- For any initial $\mathcal{T}$, iterating the revaluation operator J by the transfinite sequence yields a fixed point $\mathcal{T}^*$ with $J(\mathcal{T}^*) = \mathcal{T}^*$, where transparency holds. We write $M^* := \langle D, I, \mathcal{T}^* \rangle$.
- Evaluation at a fixed point:

  φ is **true** at $\mathcal{T}^*$ $\iff$ $\langle \varphi, 1 \rangle \in \mathcal{T}^*$ $[M^* \models \varphi$ by transparency$]$
  φ is **false** at $\mathcal{T}^*$ $\iff$ $\langle \varphi, 0 \rangle \in \mathcal{T}^*$ $[M^* \models\!\!\!| \, \varphi$ by transparency$]$
  otherwise φ is **unlabelled/undefined** at $\mathcal{T}^*/M^*$

- All initial $\mathcal{T}$ need to be coherent and consistent.
- Consistency is preserved by the J-iteration and at fixed points (you will prove this in your assignment).

Grounded (T-free) sentences are decided uniformly

Lemma (Interpretation-independence for T-free)

Fix D and I. If φ is T-free and $\mathcal{T}^*, \mathcal{T}'^*$ are fixed points, then

$$\langle \varphi, 1\rangle \in \mathcal{T}^* \iff \langle \varphi, 1\rangle \in \mathcal{T}'^*, \quad \langle \varphi, 0\rangle \in \mathcal{T}^* \iff \langle \varphi, 0\rangle \in \mathcal{T}'^*$$

Proof by induction on φ

It thus follows that for any T-free φ, exactly one holds:

- φ is true in **all** fixed points and false in **no** fixed point (grounded true);
- φ is false in **all** fixed points and true in **no** fixed point (grounded false).

Paradoxical: the Liar is true in no FP and false in no FP

Let $\lambda := \neg T(l)$ with $I(l) = \lambda$.

For every $\mathcal{T}^*$, we have:

$$\langle \lambda, 1 \rangle \notin \mathcal{T}^* \quad \text{and} \quad \langle \lambda, 0 \rangle \notin \mathcal{T}^*$$

Proof:

- If $\langle \lambda, 1 \rangle \in \mathcal{T}^*$, then $M^* \vDash \lambda$, so $M^* \Dash T(l)$. Hence $\langle I(l), 0 \rangle = \langle \lambda, 0 \rangle \in \mathcal{T}^*$, contradicting consistency.
- If $\langle \lambda, 0 \rangle \in \mathcal{T}^*$, then $M^* \Dash \lambda$, so $M^* \vDash T(l)$. Hence $\langle I(l), 1 \rangle = \langle \lambda, 1 \rangle \in \mathcal{T}^*$, again contradicting consistency.

Hence, λ is **undefined** at any fixed point.

Biconsistent: the Truth-teller varies across fixed points

Let $\tau := T(t)$ with $I(t) = \tau$.

$\mathcal{T}_+ := \{\langle\tau, 1\rangle\}$ and $\mathcal{T}_- := \{\langle\tau, 0\rangle\}$

Proof: We have established before that these initial valuations are coherent. In $\mathcal{T}_+$, $\langle\tau, 1\rangle$ forces $M_+ \vDash \tau$, hence $\langle\tau, 1\rangle \in J(\mathcal{T}_+)$. Similarly for $\mathcal{T}_-$. By the fixed-point theorem, the J-iterations yield fixed points

$$\mathcal{T}_+^* \supseteq \{\langle\tau, 1\rangle\}$$

$$\mathcal{T}_-^* \supseteq \{\langle\tau, 0\rangle\}$$

Hence τ is **true** in $\mathcal{T}_+^*$ and **false** in $\mathcal{T}_-^*$.

The Truth-teller at the least fixed point

Let $\tau := T(t)$ with $I(t) = \tau$.

Consider the $\varnothing$-initial sequence given by $\mathcal{T}_0 = \varnothing$, $\mathcal{T}_{\alpha+1} = J(\mathcal{T}_\alpha)$, and $\mathcal{T}_\lambda = \bigcup_{\beta<\lambda} \mathcal{T}_\beta$.

For all ordinals α, neither $\langle \tau, 1 \rangle$ nor $\langle \tau, 0 \rangle$ is in $\mathcal{T}_\alpha$. *(prove this by transfinite induction on α)*

In particular, if ρ is the first stage with $\mathcal{T}_\rho = \mathcal{T}_{\rho+1}$ (the stabilization index), then τ is **unlabelled/undefined** in $\mathcal{T}_\rho = \varnothing^*$.

True in some, false in no: $\chi := \tau \vee \neg\tau$ (Exercise)

Let $\chi := \tau \vee \neg\tau$ with τ being the truth-teller sentence.

For every $\mathcal{T}^*$, $\langle \chi, 0 \rangle \notin \mathcal{T}^*$. Moreover, $\langle \chi, 1 \rangle \in \mathcal{T}^*_+$ and $\langle \chi, 1 \rangle \in \mathcal{T}^*_-$, while χ is unlabelled at the least fixed point $\varnothing^*$.

Proof:

- $\varphi \vee \psi$ is false iff both disjuncts are false. If $\langle \tau, 0 \rangle \in \mathcal{T}^*$, then $M \models \neg\tau$, so the right disjunct is true. Dually if $\langle \tau, 1 \rangle \in \mathcal{T}^*$. Thus χ is **never false**.
- It is true at $\mathcal{T}^*_+$ (left disjunct true) and at $\mathcal{T}^*_-$ (right disjunct true). Thus χ is **true in some** fixed point.
- At $\varnothing^*$, both disjuncts are unlabelled (see previous slide), hence so is χ. Thus, χ is true in some, **but not all** fixed points.

Revenge: the 'not true' Liar

'This sentence is not true ([neither true nor false] or false)'

Add a unary operator $\sim$ with clauses:

$$M \vDash {\sim}\varphi \iff M \nvDash \varphi \qquad M \Dashv {\sim}\varphi \iff M \vDash \varphi$$

Let $l^{\sim}$ name ${\sim}T(l^{\sim})$ and abbreviate $\chi := {\sim}T(l^{\sim})$ (so $I(l^{\sim}) = \chi$).

Assume **full transparency** for T in the expanded language.[3] Note that ${\sim}\varphi$ is never gappy.

- If $M^* \vDash \chi$, then by the clause for $\sim$, $M^* \nvDash T(l^{\sim})$. But by transparency and $I(l^{\sim}) = \chi$, $M^* \vDash T(l^{\sim}) \iff M^* \vDash \chi$, contradiction.
- If $M^* \Dashv \chi$, then by the clause for $\sim$, $M^* \vDash T(l^{\sim})$. By transparency $M^* \vDash \chi$, so χ is both true and false, contradicting consistency.

No consistent fully transparent model can exist: χ is a **revenge** liar.

[3] So we are supposing that we have a model M^* with transparency, even though $\sim$ is actually not monotone, so the fixed-point theorem would not work if we add $\sim$ to the language.

Exercise 1: A sentence of exact stage ω

Work in a language extended with numerals, and add a *fixed* family of binary predicates $\mathrm{Tr}(n, a)$ intended as 'a codes a sentence that is true at stage n of the *bottom* Kripke chain'.

Formally, fix D, I and define the bottom chain $\mathcal{T}_0 = \varnothing$, $\mathcal{T}_{k+1} = J(\mathcal{T}_k)$, $\mathcal{T}_\lambda = \bigcup_{\beta<\lambda} \mathcal{T}_\beta$. Interpret $\mathrm{Tr}(n, x)$ so that

$$\langle n,' \varphi' \rangle \in I(\mathrm{Tr}) \quad \Longleftrightarrow \quad \langle \varphi, 1 \rangle \in \mathcal{T}_n$$

Define the sentence

$$H \ \equiv \ \forall n \, \neg \, \mathrm{Tr}(n,' H')$$

(a) Show H is *not* labelled at any finite stage n.

(b) Show H becomes decided at stage ω of the bottom chain and remains stable thereafter.

(c) Determine whether H is true or false at stage ω.

Exercise 2: Non-ω-continuity of J

A monotone F is ω-continuous if $F(\bigcup_n X_n) = \bigcup_n F(X_n)$ for every increasing ω-chain (X_n).

(a) Prove J is *not* ω-continuous by constructing an increasing chain $\mathcal{T}_0 \subseteq \mathcal{T}_1 \subseteq \cdots$ with union U such that

$$J(U) \neq \bigcup_n J(\mathcal{T}_n)$$

(b) Explain how your counterexample connects to Exercise 1.

Another exercise

Extend the language by a binary operator $\bullet$ with semantic clauses:

$$M \vDash \phi \bullet \psi \iff M \dashv \phi$$
$$\text{or } M \vDash \psi \text{ or } \big(M \nvDash \phi \;\wedge\; M \nvDash \psi \;\wedge\; M \not\dashv \phi \;\wedge\; M \not\dashv \psi\big)$$
$$M \dashv \phi \bullet \psi \iff M \vDash \phi \text{ and } M \dashv \psi.$$

Show that, in the expanded language, one can generate a revenge paradox with $\bullet$ assuming full transparency for T.