Definability and Min-representability

Professor Cian Dorr

29th November 2022

New York University

Key facts about Min

1. Min can capture the labelling function.
2. Min can capture the substitution function.
3. For any axiom-set Ax, if Min can represent Ax, then Min can represent the relation *A is a proof of P from Ax*.

Key facts about Min

1. Min can capture the labelling function.
2. Min can capture the substitution function.
3. For any axiom-set Ax, if Min can represent Ax, then Min can represent the relation *A is a proof of P from Ax*.

And more boringly, we can add:

0. Min can represent any finite set of srtrings.

A generalization behind these facts

The previous facts can all be subsumed under the following generalizations:

4. Min can represent every *decidable* relation among strings.
5. Min can capture every *computable* function on strings.

A generalization behind these facts

The previous facts can all be subsumed under the following generalizations:

4. Min can represent every *decidable* relation among strings.
5. Min can capture every *computable* function on strings.

Intuitively: function f is *computable* if we can set up an ideal computer such that when it's run with strings $s_1, \ldots, s_n$ as input, it will stop and print out $f(s_1, \ldots, s_n)$ as output, and relation R among strings is *decidable* if we can set up an ideal computer such that when it's run with strings $s_1, \ldots, s_n$ as input, it'll stop and print out `yes` if $Rs_1 \ldots s_n$ and `no` otherwise.

We'll give precise versions of these informal definitions next week. But at our current informal level, it should seem obvious enough that the labelling and substitution functions are computable, and that the relation *A is a proof of P from Ax* is decidable whenever Ax is.

How we will prove these facts (preview)

We will prove a theorem to the following effect (I'll explain 'sufficiently simple' later):

Representability Theorem (rough statement)

Every relation that is definable in $\mathbb{S}$ *by a sufficiently simple formula* is represented in Min, and every function that is definable in $\mathbb{S}$ *by a sufficiently simple formula* is capturable in Min.

How we will prove these facts (preview)

We will prove a theorem to the following effect (I'll explain 'sufficiently simple' later):

Representability Theorem (rough statement)

Every relation that is definable in $\mathbb{S}$ *by a sufficiently simple formula* is represented in Min, and every function that is definable in $\mathbb{S}$ *by a sufficiently simple formula* is capturable in Min.

Given this, we can establish the facts on the previous slide by showing that:

1,2. The labelling and substitution functions have a sufficiently simple definitions in $\mathbb{S}$.

3. Whenever axiom-set Ax that has a sufficiently simple definition in $\mathbb{S}$, the relation *s is a proof of P from Ax* has a sufficiently simple definition in $\mathbb{S}$.

4. Every finite set has a sufficiently simple definition in $\mathbb{S}$.

5,6. Every decidable set and every computable function has a sufficiently simple definition in $\mathbb{S}$.

Semantical Gödel's Theorem

But first let's discuss what we can do with the facts, starting with a more detailed explanation of the following theorem (sketched last week):

Semantic Gödel's Theorem

No theory T has the following four properties:

1. String-accuracy: every Str-sentence in T is true in $\mathbb{S}$.
2. Sufficient strength: T extends Min.
3. Sufficient simplicity: T has a Min-representable (e.g., finite) axiomatization.
4. Negation-completeness: if $P \notin T$ and P is a sentence of T's signature, $\neg P \in T$.

Semantical Gödel's Theorem

But first let's discuss what we can do with the facts, starting with a more detailed explanation of the following theorem (sketched last week):

Semantic Gödel's Theorem

No theory T has the following four properties:

1. String-accuracy: every Str-sentence in T is true in $\mathbb{S}$.
2. Sufficient strength: T extends Min.
3. Sufficient simplicity: T has a Min-representable (e.g., finite) axiomatization.
4. Negation-completeness: if $P \notin T$ and P is a sentence of T's signature, $\neg P \in T$.

We prove this by showing that if T has properties 2 and 3, there is a Str-sentence G_T—the *Gödel sentence* of T—such if $T \vDash G_T$, T is inconsistent, and if $T \nvDash G_T$, G_T is true in $\mathbb{S}$. So if T has property 1, neither G_T nor $\neg G_T$ is in T, so T lacks property 4.

Proving Semantical Gödel's Theorem

So, suppose T extends Min and has a Min-representable axiomatization Ax. Then there's a formula $\mathrm{Proof}_{Ax}(x, y)$ that represents the relation *A is a proof of P from Ax* in Min (and hence also in T). Define $\mathrm{Prov}_T(x)$ as $\exists y\, \mathrm{Proof}_{Ax}(y, x)$.

By the Diagonal Lemma, there is a sentence G_T such that

$$\mathrm{Min} \vDash G_T \leftrightarrow \neg \mathrm{Prov}_T(\langle G_T \rangle)$$

Whenever $P \in T$, there is a proof A of P from Ax, so $\mathrm{Min} \vDash \mathrm{Proof}_{Ax}(\langle A \rangle, \langle P \rangle)$, hence $\mathrm{Min} \vDash \mathrm{Prov}_T(\langle P \rangle)$, hence $T \vDash \mathrm{Prov}_T(\langle P \rangle)$. It follows that if $T \vDash G_T$, both $T \vDash \mathrm{Prov}_T(\langle G_T \rangle)$ and $T \vDash \neg \mathrm{Prov}_T(\langle G_T \rangle)$, so T is inconsistent.

It remains to show that if $T \nvDash G_T$, G_T is true in $\mathbb{S}$. Suppose $T \nvDash G_T$. Then $\mathrm{Min} \vDash \neg \mathrm{Proof}_{Ax}(\langle A \rangle, \langle G_T \rangle)$ for every string A, so $\mathrm{Proof}_{Ax}(\langle A \rangle, \langle G_T \rangle)$ is false in $\mathbb{S}$ for every string A, hence $\neg \mathrm{Prov}_T(\langle G_T \rangle)$ is true in $\mathbb{S}$, hence G_T is true in $\mathbb{S}$.

Note

I could also have derived the theorem as an easy corollary of the "Non-Semi-Representability" theorem from last week (the proof of which is essentially the same).

If Ax axiomatises T, then $\mathrm{Prov}_T(x)$ semi-represents T in Min: for any P, $P \in T$ iff $\mathrm{Min} \vDash \mathrm{Prov}_T(\langle P \rangle)$. Moreover, T is string-theoretically accurate and extends Min, it also semi-represents T in T, since if $P \in T$, $\mathrm{Prov}_T(\langle P \rangle)$ is in T since it's in Min, and if $P \notin T$, $\mathrm{Prov}_T(\langle P \rangle)$ isn't in T since it's false in $\mathbb{S}$. But we already proved that no consistent, negation-complete theory that extends Min semi-represents itself.

Strengthening the theorem

Gödel himself proved a theorem a bit stronger than Semantical Gödel's Theorem, replacing property 1 (string-theoretic accuracy) with a weaker property called ω-consistency. I won't bother telling you what this is, since a few years later Rosser improved the theorem further showing that property 1 can be replaced with plain old consistency, yielding what we have been calling

Gödel's Theorem (version 1)

No theory T has the following four properties:

1. Consistency.
2. Sufficient strength: T extends Min.
3. Sufficient simplicity: T has a Min-representable (e.g., finite) axiomatization.
4. Negation-completeness: if $P \notin T$ and P is a sentence of T's signature, $\neg P \in T$.

Rosser's trick

The semantical version of the theorem leaves it open in principle that we could have a consistent, negation complete, sufficiently strong, sufficiently simple T that isn't string-theoretically accurate. In that case we'd have $T \vDash \neg G_T$, hence $T \vdash \mathrm{Prov}_T(\langle G_T \rangle)$ and $T \vdash \mathrm{Prov}_T(\langle \neg G_T \rangle)$: although T isn't inconsistent, T wrongly thinks that it is.

Rosser's trick

The semantical version of the theorem leaves it open in principle that we could have a consistent, negation complete, sufficiently strong, sufficiently simple T that isn't string-theoretically accurate. In that case we'd have $T \vDash \neg G_T$, hence $T \vdash \mathrm{Prov}_T(\langle G_T \rangle)$ and $T \vdash \mathrm{Prov}_T(\langle \neg G_T \rangle)$: although T isn't inconsistent, T wrongly thinks that it is.

Rosser's trick was to replace the formula $\mathrm{Prov}_T(x)$ with a stronger formula for which this case can't arise.

Rosser's trick

First we define $\mathrm{RProof}_{Ax}(y, x)$ as

$$\mathrm{Proof}_{Ax}(y, x) \wedge \forall y' \forall x' (y' \leq y \wedge \mathrm{Proof}_{Ax}(y', x') \rightarrow \neg \, \mathrm{neg}(x', x))$$

where neg is defined as follows:

$$\forall x \forall y (\mathrm{neg}(x, y) \leftrightarrow y = "\neg" \oplus x \vee x = "\neg" \oplus y)$$

In words: a Rosser-proof of a sentence from Ax is a proof of it from Ax such that there is no shorter-or-equal proof of its negation from Ax.

Note that if Ax is consistent, A is a proof of P from Ax iff it is a Rosser-proof of P from Ax.

Rosser's trick

At this point we need to appeal to a new fact about Min:

7. If $\text{Proof}_{Ax}(y, x)$ represents the relation *A is a proof of P from Ax* in Min, then $\text{RProof}_{Ax}(y, x)$ represents the relation *A is a Rosser-proof of P from Ax* in Min.

The reason this is true is that where $s_1, \ldots, s_n$ are all the strings that are no longer than s, Min proves

$$\forall x(x \leq \langle s \rangle \leftrightarrow x = \langle s_1 \rangle \vee \cdots \vee x = \langle s_n \rangle)$$

so when each string $\leq A$ is not a proof of a sentence that negates P, Min proves

$$\forall y'(y' \leq \langle A \rangle \to \forall x'(\text{Proof}_{Ax}(y', x') \to \neg\, \text{neg}(x', \langle P \rangle))$$

Rosser's trick

So, suppose T is consistent, extends Min, and is axiomatised by a Min-representable set Ax. Define $\text{Rosser}_T(x)$ as $\exists y(\text{RProof}_{Ax}(y, x))$. By the Diagonal lemma there is a sentence R_T (T's *Rosser sentence*) such that

$$(*) \qquad \text{Min} \vDash R_T \leftrightarrow \neg\,\text{Rosser}_T(\langle R_T \rangle)$$

When $P \in T$, there's a Rosser-proof A of P from Ax, so $\text{Min} \vDash \text{RProof}_{Ax}(\langle A \rangle, \langle P \rangle)$, hence $\text{Min} \vDash \text{Rosser}_T(\langle P \rangle)$ and $T \vDash \text{Rosser}_T(\langle P \rangle)$. Thus if $T \vDash R_T$, both $T \vDash \text{Prov}_T(\langle R_T \rangle)$ and $T \vDash \neg\,\text{Prov}_T(\langle R_T \rangle)$, in which case T is inconsistent. Thus $T \nvDash R_T$.

Since T is negation complete, $T \vDash \neg R_T$, so there's a Rosser-proof of $\neg R_T$ from Ax, so $T \vDash \text{Rosser}_T(\langle \neg R_T \rangle)$. Also $T \vDash \text{Rosser}_\text{T}(\langle R_T \rangle)$ by (*). But this can't happen when T extends Min, because Min includes the axiom $\forall x \forall y(x \leq y \vee y \leq x)$ which implies that whenever we have a proof of P and one of $\neg P$ from Ax, at least one of them isn't a Rosser-proof.

The Representability Theorem

Bounded formulae

Now it's time to get precise about what we were calling 'sufficiently simple formulae'.

There are actually two relevant notions of simplicity.

Definition

The set of **bounded formulae** of Str is the smallest set of formulae such that:

- Every atomic formula is bounded
- When P and Q are bounded, $\neg P$, $P \to Q$, $P \vee Q$, and $P \wedge Q$ are bounded.
- When P is bounded and t is any term, $\forall v(v \leq t \to P)$ and $\exists v(v \leq t \wedge P)$ are bounded.

Bounded formulae

Now it's time to get precise about what we were calling 'sufficiently simple formulae'. There are actually two relevant notions of simplicity.

Definition

The set of **bounded formulae** of Str is the smallest set of formulae such that:

- Every atomic formula is bounded
- When P and Q are bounded, $\neg P$, $P \to Q$, $P \vee Q$, and $P \wedge Q$ are bounded.
- When P is bounded and t is any term, $\forall v(v \leq t \to P)$ and $\exists v(v \leq t \wedge P)$ are bounded.

Definition

P is a Σ_1 **formula** of Str iff either P is a bounded formula of Str, or P is $\exists v Q$ for some bounded formula Q of Str.

Min, the minimal theory of strings

$M1$	$\forall x \forall y(\neg(c \oplus x = \text{""}))$ (for $c \neq \text{""}$)
$M2$	$\forall x \forall y(c \oplus x = c \oplus y) \to x = y$
$M3$	$\forall x \neg(c_1 \oplus x = c_2 \oplus x)$
$M4$	$\forall x(\text{""} \oplus x = x)$
$M5$	$\forall x \forall y((c \oplus x) \oplus y = c \oplus (x \oplus y))$
$M6$	$c = c \oplus \text{""}$
$M7$	$\forall x(\text{""} \leq x)$
$M8$	$\forall x(x \leq \text{""} \leftrightarrow x = \text{""})$
$M9$	$\forall x \forall y((c_1 \oplus x \leq c_2 \oplus y) \leftrightarrow x \leq y)$
$M10$	$\forall x \forall y(x \leq y \vee y \leq x)$
$M11$	$\forall x(x = \text{""} \vee \exists y(x = c_1 \oplus y \vee \cdots \vee x = c_n \oplus y))$

Representation theorem

(i) If a set is defined in $\mathbb{S}$ by a bounded formula, it is representable in Min.

(ii) If a set is defined in $\mathbb{S}$ by a Σ_1 formula, it is semi-representable in Min.

(iii) If a partial function is defined in $\mathbb{S}$ by a bounded or Σ_1 term, it is capturable in Min.

Min includes all true quantifier-free sentences

Here's the first thing we'll prove on the way to showing this.

Quantifier-Free Omniscience If A is a *quantifier-free* sentence of the language of strings, A is true in $\mathbb{S}$ iff Min $\vDash A$

Note that this immediately immediately implies:

- If a relation R is definable in $\mathbb{S}$ by a quantifier-free formula P, P represents R in Min.

This doesn't yet give us what we need, since the terms that define the functions we're interested in, e.g. substitution and labelling, are not quantifier-free. But it's a start.

First batch of facts to prove

For any closed terms t_1, t_2 in the language of strings:

Fact 1 If $t_1 = t_2$ is true in $\mathbb{S}$, $\mathrm{Min} \models t_1 = t_2$.

Fact 2 If $t_1 = t_2$ is false in $\mathbb{S}$, $\mathrm{Min} \models t_1 \neq t_2$.

Fact 3 If $t_1 \leq t_2$ is true in $\mathbb{S}$, $\mathrm{Min} \models t_1 \leq t_2$.

Fact 4 If $t_1 \leq t_2$ is false in $\mathbb{S}$, $\mathrm{Min} \models \neg t_1 \leq t_2$.

These straightforwardly entail Quantifier-Free Omniscience, by induction on the construction of formulae.

The proof of Facts 1–4 turns on:

Label Lemma For any closed term t in the language of strings, and any string s, if $t = \langle s \rangle$ is true in $\mathbb{S}$, it's a theorem of Min.

Problem Set 11 walks you through the proof of this, and the proof of facts 1–4 from it.

The next thing we'll show is

Bounded Omniscience If P is a bounded sentence of the language of strings, P is true in $\mathbb{S}$ iff $\mathsf{Min} \vDash P$.

Given that the negation of a bounded formula is bounded, this immediately implies the first part of the Representation Theorem:

- If R is definable in $\mathbb{S}$ by a bounded formula P, P represents Y in Min.

Proving Bounded Omniscience

Listing Lemma When s is any string, and $s_1, \ldots, s_n$ are *all* the strings which are no longer than s, Min $\vDash \forall x(x \leq \langle s \rangle \leftrightarrow (x = \langle s_1 \rangle \vee \cdots \vee x = \langle s_n \rangle))$.

Proving Bounded Omniscience

Listing Lemma When s is any string, and $s_1, \ldots, s_n$ are *all* the strings which are no longer than s, Min $\vDash \forall x(x \leq \langle s \rangle \leftrightarrow (x = \langle s_1 \rangle \vee \cdots \vee x = \langle s_n \rangle))$.

Proof: by induction on the length of s, using M10 together with earlier results.

Given the Listing Lemma, we can show the following

Equivalence Lemma For every bounded formula P and assignment g, there is a quantifier-free sentence P' such that Min $\vDash P[v \mapsto \langle gv \rangle] \leftrightarrow P'$.

Proof: by induction on the complexity of P.

Bounded Omniscience follows immediately from this together with Quantifier-Free Omniscience.

Positive Σ_1 Omniscience

As an immediate consequence of Bounded Omniscience, we have

Σ_1 Omniscience If A is a Σ_1 sentence of the language of strings, A is true in $\mathbb{S}$ iff Min $\vDash A$.

Proof: The right-to-left direction follows from the truth of Min in $\mathbb{S}$. For the left-to-right direction, let P be $\exists v Q$ where Q is bounded. If P is true in $\mathbb{S}$, then there must be some string s such that Q is true in $\mathbb{S}$ on the assignment $[v \mapsto s]$; then the sentence $Q[v \mapsto \langle s \rangle]$ is true in $\mathbb{S}$, so by Bounded Omniscience it is a theorem of Min, so by Existential Generalization $\exists v Q$ is also a theorem of Min.

This implies the second part of the Representation Theorem: if a Σ_1 formula defines a set in $\mathbb{S}$, it semi-represents it in Min. But it need not *represent* the set, since the negation of a Σ_1 formula is no Σ_1, and generally not even Σ_1-equivalent.

Σ_1-equaivalence

Definition

A Σ_1-**equivalent** formula P is one such that for some Σ_1-formula Q, Min $\vDash P \leftrightarrow Q$

Definition

A Σ_1-**equivalent** formula P is one such that for some Σ_1-formula Q, Min $\models P \leftrightarrow Q$

Note: the disjunction of two Σ_1-equivalent formulae is Σ_1-equivalent, and so is their conjunction.

Σ_1-equaivalence

Definition

A Σ_1-**equivalent** formula P is one such that for some Σ_1-formula Q, $\mathrm{Min} \models P \leftrightarrow Q$

Note: the disjunction of two Σ_1-equivalent formulae is Σ_1-equivalent, and so is their conjunction.

Also, if P is Σ_1-equivalent, so is $\exists v P$.