Limits of Representability

Professor Cian Dorr

22nd November 2022

New York University

Representability and semi-representability

Given a theory T in a signature Σ extending Str, and an n-ary relation R on strings:

Definition

R is **representable in** T iff there is a definitional extension T^+ of T with a new n-ary predicate F such that:

- Whenever $Rs_1 \dots s_n$, $T^+ \vDash F(\langle s_1 \rangle, \dots, \langle s_n \rangle)$
- Whenever it's not the case that $Rs_1 \dots s_n$, $T^+ \vDash \neg F(\langle s_1 \rangle, \dots, \langle s_n \rangle)$

Definition

R is **semi-representable in** T iff there is a definitional extension T^+ of T with a new n-ary predicate F such that:

- Whenever $Rs_1 \dots s_n$, $T^+ \vDash F(\langle s_1 \rangle, \dots, \langle s_n \rangle)$
- Whenever it's not the case that $Rs_1 \dots s_n$, $T^+ \nvDash F(\langle s_1 \rangle, \dots, \langle s_n \rangle)$

Capturability

Where T is a theory in a signature extending Str, and g is partial function from n-tuples of strings to strings:

Definition

g is **capturable in** T iff there is a definitional extension T^+ of T with a new n-ary function symbol f such that:

- Whenever $t = g(s_1, \ldots, s_n)$, $T^+ \vDash \langle t \rangle = f(\langle s_1 \rangle, \ldots, \langle s_n \rangle)$

Equivalently: if there is an $n + 1$-formula P with free variables $v_1, \ldots, v_{n+1}$ such that

(i) $T \vDash \forall v_1 \ldots \forall v_n \exists! v_{n+1} P$, and
(ii) Whenever $t = g(s_1, \ldots, s_n)$, $T \vDash P[\langle s_1 \rangle / v_1, \ldots, \langle s_n \rangle / v_n, \langle t \rangle / v_{n+1}]$.

Note: In the book, this is called ‘representability’ too; but this is confusing given that partial functions are relations.

A few simple observations about these concepts

(iv) if $T \subseteq T^+$, every relation representable in T is representable in T^+. However, some relations semi-representable in T may not be semi-representable in T^+.

Some observations about these concepts

(i) If T captures g and g', it captures $g' \circ g$.

- Definitionally extend T with function symbols f and f' such that $T^+ \vDash f(\langle s \rangle) = \langle g(s) \rangle$ and $T^+ \vDash f'(\langle s \rangle) = \langle g'(s) \rangle$ for all s. Then further definitionally extend with the definition

$$\forall x \forall y (y = f''(x) \leftrightarrow y = f'(f(x)))$$

(ii) If T captures a function f and (semi-)represents X, it (semi-)represents $\{y \mid fy \in X\}$ (the preimage of X under f—sometimes written $f^*(X)$).

- Definitionally extend T with a function symbol f such that $T^+ \vDash f(\langle s \rangle) = \langle g(s) \rangle$ and a predicate F such that $T^+ \vDash F(\langle s \rangle)$ whenever $s \in X$ and $T^+ \vDash \neg F(\langle s \rangle)$ ($T^+ \nvDash F(\langle s \rangle)$) otherwise. Then further definitionally extend with the definition

$$\forall x (Gx \leftrightarrow F(f(x)))$$

Tarski’s non-representability theorem

Consequences of Cantor's Theorem

When T is a theory in a signature extending Str, and P is a formula with free variables $v_1, \ldots, v_n$ (in alphabetical order), and $s_1, \ldots, s_n$ are strings, say that P **is T-provable of** $s_1, \ldots, s_n$ iff $T \vDash P[\langle s_1 \rangle / v_1, \ldots, \langle s_n \rangle / v_n]$.

Each of the countably many 1-formulae semi-represents at most one set of strings, and there are uncountably many sets of strings, so by Cantor's theorem, some sets of strings aren't semi-representable in T (and thus aren't representable in T if T is consistent).

And we can give an example! Consider any set Y that contains all 1-formula that are *not T-provable of themselves*, and no other 1-formulae. Y isn't semi-representable in T, since if 1-formula $NPOS(x)$ represented it, we would have both

- $NPOS(x) \in Y$ iff $T \nvDash NPOS(\langle NPOS(x) \rangle)$ (by the definition of Y).
- $T \vDash NPOS(\langle NPOS(x) \rangle)$ iff $NPOS(x) \in Y$ (since $NPOS(x)$ semi-represents Y).

Note that if T is consistent, it follows that Y is not representable in T.

More non-representable and non-semi-representable sets and relations

- Consider now the set of all 1-formulae that *are* T-provable of themselves. If T is consistent, it can't be representable in T, since if it were, its complement would be too, which we just ruled out. (However it could still be *semi*-representable.)
- The relation *P is T-provable of Q* also can't be representable in T if T is consistent (though it could be semi-representable). For if it were represented by a 2-formula $\mathrm{ProvOf}_T(x, y)$, the 1-formula $\neg\,\mathrm{ProvOf}(x, x)$ would represent the set of all 1-formulae that are T-provable of themselves.
- The relation *P is not T-provable of Q* can't even be semi-representable in T. For if it were semi-represented by $\mathrm{NotProvOf}_T(x, y)$, $\mathrm{NotProvOf}_T(x, x)$ would semi-represent the set of 1-formulae not T-provable of themselves.

Self-Application

Let an x-formula be a formula in which the only free variable is x.

Definition

For any signature Σ, Σ's *self-application* is the function that maps each x-formula P to the sentence $P[\langle P\rangle/x]$.

We are going to be interested in theories that *capture self-application*, which means they have a definitional expansion with a function symbol SelfApply such that for every x-formula P,

$$T \vDash \langle P[\langle P\rangle/x]\rangle = \mathrm{SelfApply}(\langle P\rangle)$$

Labelling, substitution, and self-application

Note that any theory that can capture the *labelling* function $\langle\cdot\rangle$ and the *substitution* function that takes a formula P, a variable v, and a term t and yields $P[t/v]$ can also capture self-application. For consider a definitional extension of T with an 1-ary function symbol Label and 3-ary function symbol Subst, and now introduce a further definition:

$$\forall y \forall z (z = \text{SelfApply}(y) \leftrightarrow z = \text{Subst}(\text{y}, \langle x \rangle, \text{Label}(y)))$$

For each x-formula P, we'll have $T \vDash \langle\langle P \rangle\rangle = \text{Label}(\langle P \rangle)$ and $T \vDash \langle P[\langle P \rangle / x] \rangle = \text{Subst}(\langle P \rangle, \langle x \rangle, \langle\langle P \rangle\rangle)$, hence $T \vDash \langle P[\langle P \rangle / x] \rangle = \text{SelfApply}(\langle P \rangle)$.

Later, we'll prove that Min *does* capture labelling and substitution, and hence captures self-application. (It follows that the same is true for all theories extending Min.)

Tarski's Non-Self-Representability Theorem

No consistent theory that captures self-application represents itself.

Proof: suppose for contradiction that T is consistent, function symbol SelfApply captures self-application in T, and predicate Prov_T represents T in T.

Now consider the 1-formula $\neg\,\mathrm{Prov}_T(\mathrm{SelfApply}(x))$.

Given our assumptions, it would have to represent a set containing all 1-formulae whose self-applications are not in T (i.e., which are not T-provable of themselves) and no other 1-formulae. But we've already shown that no such set is representable (given that T is consistent).

Non-semi-representability

Non-Semi-Representability Theorem

No theory T captures self-application and semi-represents any set of strings that doesn't contain any member of T and contains all sentences in T's signature that aren't in T.

Suppose X is such a set, semi-represented in (a definitional extension of) T by predicate F. Then the 1-formula $F(\mathrm{SelfApply}(x))$ would semi-represent a set containing all 1-formulae whose self-application is in X, i.e. which are not T-provable of themselves, and no other 1-formulae.

Non-semi-representability

Non-Self-Semi-Representability Theorem

No consistent theory that captures self-application and semi-represents itself is negation-complete.

Suppose T is negation complete and consistent, and has a function symbol SelfApply and predicate Prov_T that respectively capture self-application and semi-represent T in T. Then the 1-formula $\mathrm{Prov}_T(\text{"}\neg\text{"} \oplus \mathrm{SelfApply}(x))$ would semi-represent the set of all 1-formulae such that the negation of their self-application is a member of T. Since T is consistent, this set doesn't include any 1-formulae whose self-application is in T, and since (since T is negation-complete and closed under double negation elimination) it includes every 1-formula whose self-application isn't in T. But this is impossible.

The diagonal lemma

The Diagonal Lemma

A different and useful route to the same result goes via the *diagonal lemma*, also known as the *fixed point lemma*. Define:

Definition

Sentence G is a **fixed point** of x-formula $H(x)$ in T iff $T \vDash G \leftrightarrow H(\langle G \rangle)$.

- It's traditional (though potentially misleading) to think of a sentence G that is the fixed point of $H(x)$ in T as "saying of itself that it is H"—as if it were the sentence 'I am H' or 'This very sentence is H'.

Generalizing an idea that we in effect already employed in the Unrepresentability Theorem, we can prove

The Diagonal Lemma

If T captures self-application, then every 1-formula has a fixed point in T.

Proof of the Diagonal Lemma

Proof: Definitionally extend T with a function symbol SelfApply that captures self-application, and for any 1-formula $H(x)$, let G_H be the sentence

$$H(\text{SelfApply}(\langle H(\text{SelfApply}(x))\rangle))$$

i.e., the self-application of the 1-formula $H(\text{SelfApply}(x))$. Call this the **diagonalization** of $H(x)$.

Since SelfApply captures self-application in T,

$$T \vDash \text{SelfApply}(\langle H(\text{SelfApply}(x))\rangle) = \langle G \rangle$$

Since T is closed under logical consequence, this implies

$$T \vDash H(\text{SelfApply}(\langle H(\text{SelfApply}(x))\rangle)) \leftrightarrow H(\langle G_H \rangle)$$

i.e.

$$T \vDash G_H \leftrightarrow H(\langle G_H \rangle)$$

From the Diagonal Lemma to the Unrepresentability Theorem

Suppose for contradiction that T captures self-application (with a function symbol SelfApply) and also represented T itself (with a predicate Prov_T). By the Diagonal Lemma, the formula $\neg Prov_T(x)$ has a fixed point in T. That is: a sentence G_T such that

$$T \vDash G_T \leftrightarrow \neg \text{Prov}_T \langle G_T \rangle$$

Since T is a theory, it follows that

$$T \vDash G_T \text{ iff } T \vDash \neg \text{Prov}_T(\langle G_T \rangle)$$

Suppose $T \vDash G_T$. Then $T \vDash \text{Prov}_T(\langle G_T \rangle)$ (since Prov_T represents T in T), and also $T \vDash \neg \text{Prov}_T(\langle G_T \rangle)$ (by the above biconditional), so T is inconsistent, contradicting our assumption.

Suppose on the other hand that $T \nvDash G_T$. Then, $T \vDash \neg \text{Prov}_T \langle G_T \rangle$ (since Prov_T represents T in T), so $T \vDash G_T$ by the above biconditional: contradiction.

What Min represents

The next order of business is to show that Min can represent the labelling and substitution functions (and hence also self-application).

We will also show that Min represents the set of *proofs* (in our proof system $\vdash$). And as a consequence of this: for any set Ax that's representable in Min, Min represents the relation *p is a proof whose final sequent has Q on the right and only members of Ax* on the left.

A version of Gödel's theorem

Suppose that a theory T that extends Min is axiomatised by a set Ax that's representable in Min (and hence also in T). (For example, Ax could be any finite set). Let $\mathrm{Proof}_{Ax}(x, y)$ be a predicate that represents the relation *p is a proof of Q from Ax*. Then the predicate Prov_T defined by $\exists x\, \mathrm{Proof}_{Ax}(x, y)$ *defines* T, and also *semi-represents* T in Min.

This means that we have $T \vDash \mathrm{Prov}_T(\langle P \rangle)$ for all $P \in T$. So in particular, if $T \vDash G_T$ (where G_T is the diagonalization of $\neg\,\mathrm{Prov}_T(x)$), then T is inconsistent. So G_T is not in T, so there's no proof of it from Ax. It follows from this that it is true in $\mathbb{S}$. So we have a variant of Gödel's theorem: no consistent finitely (or more generally, Min-representably) axiomatizable theory that extends Min contains every sentence true in $\mathbb{S}$.