Distinguishing one structure from another

Wilfrid Hodges
Queen Mary and Westfield College, London

Johannesburg, December 1999

1. The basic idea of model theory

A *formula* is a grammatical expression ϕ in some language (natural or formal) that can be used to make a true or false statement. But in fact ϕ may need interpreting in order to become either true or false.
So we have

Formula ϕ plus interpretation I	yields truth value

We say that ϕ is *true in* I,
or that I *satisfies* ϕ,
or that I is a *model of* ϕ,
or in symbols $I \models \phi$,
if under the interpretation I, ϕ is true.

FIRST SLOGAN: *A formula defines a class of interpretations, viz. those which make it true.*

So model theory is part of the theory of definition.

English example (based on Wason selection task experiment, 1966)

ϕ is the sentence

> If the letter on one side of the card is a vowel, the number on the other side is even.

A suitable card gives an interpretation of 'the letter on one side of the card', and of 'the number on the other side'.
The card is a model of ϕ if it satisfies the rule expressed in ϕ.

SECOND SLOGAN: *Model-theoretic truth is ordinary truth. Model-theoretic satisfaction is ordinary satisfaction.*

Typical examples of phrases needing interpretation:

1. **Singular definite descriptions**

'the card'

'the number on the back'

2. **Variables**

'X', 'G', 'g', 'g_1' etc. etc.

Model theory concentrates on phrases of these two kinds.

An interpretation for such phrases says *what they refer to*. (Such interpretations are also known as *assignments* or *valuations*.)

Sometimes an expression needs interpretation but gets a default interpretation from the *context* (or *point of reference*) where it is applied. For example

> 'the President of the Russian Federation'

currently refers to Boris Yeltsin; the default interpretation is the present one.

If we are studying a particular card, the default interpretation of

> 'the number on the back'

is the number on the back of that card. The card serves as a point of reference.

The interpretation of this phrase has to be a certain sort of thing, viz. a number. Expressions with this feature are called *sortal*.

Quantifiers

These involve interpretation at two levels. Take for example

> 'Everybody loves Madonna.'

1. This is true in a context if in that context every assignment to the sortal variable 'x_{person}' satisfies the formula

> 'x_{person} loves Madonna.'

2. The context must determine who or what counts as a 'person'; we say it has a *domain* or *universe* of persons.

Note: In 1 we split the interpretation into a context and an assignment. *This division is fundamental in model theory.*

The context by itself is called a *structure*.

In logic a formula that needs only a context to interpret it, and not a further assignment, is called a *sentence*. For example

> 'Everybody loves Madonna.'

is a sentence, but

> 'x_{person} loves Madonna.'

is only a formula, because it has a *free variable*.

A set of sentences is called a *theory*. A *model* of a theory is a model of all the sentences in the theory.

Mathematical example

A *partial ordering* is defined to be a structure which is a model of the following theory:

$$\forall x\ (x \leqslant x).$$

$$\forall x \forall y\ (x \leqslant y \wedge y \leqslant x \rightarrow x = y).$$

$$\forall x \forall y \forall z\ (x \leqslant y \wedge y \leqslant z \rightarrow x \leqslant z).$$

Here '$\forall x$' means 'for all x', '$\wedge$' means 'and', '$\rightarrow$' means 'if . . . then'. These are logical symbols with fixed meanings, except that a partial ordering must have a domain (to control the possible assignments to variables).

In this formal language an expression 'aRb' means: the pair (a, b) are in the relation R. The relation symbol '$\leqslant$' is short for:

'the relation called $\leqslant$'.

Each partial ordering A has a relation called $\leqslant$. We write this relation as '$\leqslant^A$'.

In model theory since 1950 the usual situation is that we have a class of structures and we classify them according to what sentences are true in them.

We need to make sure that the structures are suitable for the sentences, i.e. that they give interpretations for the right symbols.

A *signature* is a set of symbols needing a structure to interpret them. The symbols are called the *non-logical symbols* of the signature. A structure interpreting all and only the symbols in the signature σ is called a σ-*structure*.

A *language of signature* σ is one whose formulas need for interpretation only a structure interpreting the symbols in σ (and assignments to variables).

Non-logical symbols normally have one of the following forms:

- *(individual) constants* a, b, c etc., standing for elements of the domain;

- *function symbols* F, G, H etc., of fixed arity, standing for functions over the domain;

- *relation symbols* P, Q, R etc., of fixed arity, standing for relations over the domain.

One can also have *propositional symbols* p, q, r etc., that stand for Truth or Falsehood depending on the structure.

Example: The signature of arithmetic has the following symbols.

- Individual constant symbol ‘0’.
- No propositional symbol.
- Relation symbol ‘$\leqslant$’ of arity 2.
- Function symbols ‘S’ of arity 1, and ‘$+$’, ‘.’, both of arity 2.

Hilary Putnam says in ‘Models and reality’:

‘Models are not lost noumenal waifs looking for someone to name them; they are constructions within our theory itself, and they have names from birth.’

He seems to mean that no structures exist except those for which we have names. This is Putnam’s private ontology; model theory makes no such restriction.

In particular we don’t require even that signatures can be named, or that their symbols can be physically written. The expression ‘the function named ‘F’ in the structure A’ need not be read literally either; it means the function F^A, which is a part of A.

2. The classical indistinguishability theorems

These are theorems about *first-order model theory*, i.e. where the formulas are built up from atomic formulas

$$R(x, y, a),$$

$$F(G(x), y) = G(z)$$

etc., using

- $\neg$ 'not' ;

- $\wedge$ 'and', $\vee$ 'or' ;

- $\forall$ 'for all elements',
  $\exists$ 'there is an element such that'.

Three useful notions

1. We say that two σ-structures A, B are *elementarily equivalent*, in symbols $A \equiv B$, if exactly the same first-order sentences are true in A as in B.

2. We say that two first-order theories T, U of signature σ are *(logically) equivalent* if exactly the same σ-structures are models of T as of U.

3. Two sentences ϕ and ψ are *equivalent modulo* the theory T if every model of T that is a model of ϕ is also a model of ψ and vice versa.

Recall that a structure A of signature σ has a domain dom(A).
Recall also that if c is an individual constant in σ then c^A must be in dom(A);
and if F is a function symbol in σ, say with arity n, then for any n elements $a_1, \ldots, a_n$ of dom(A),

$$F^A(a_1, \ldots, a_n)$$

is also an element of dom(A).
We express this by saying that dom(A) is *closed under* the constants and function symbols of σ.

We say that a σ-structure B is a *substructure* of A, and that A is an *extension* of B, if

$$\text{dom}(B) \subseteq \text{dom}(A)$$

and within dom(B), all of c^B, R^B, F^B etc. are just the restrictions of c^A, R^A, F^A etc.

Example.

In the signature with 0, $+$ and $-$, the structure $2\mathbb{Z}$ of *even* integers is a substructure of the structure $\mathbb{Z}$ of integers. In fact we can 'generate' $2\mathbb{Z}$ from 4 and 6, say, by closing off under 0, $+$ and $-$.

Generating is an important way of getting substructures.

By contrast $\mathbb{N}$, the structure of natural numbers $0, 1, 2, \ldots$, is *not* a substructure of $\mathbb{Z}$ in this signature,
since it is not closed under $-$.
But it is a substructure of $\mathbb{Z}$ in the smaller signature consisting of 0 and $+$.

A first-order sentence is said to be *universal* if it consists of a string of universal quantifiers $\forall x$ (or no quantifiers), followed by a formula with no quantifiers.
A theory is *universal* if all the sentences in it are universal.

For example the theory of partial orderings (previous lecture) is universal.

Łoś-Tarski Theorem. For any first-order theory T the following are equivalent:

(a) T is logically equivalent to a universal theory.

(b) If A is any model of T, then every substructure of A is also a model of T.

Corollary. If A is a model of a universal theory T and every first-order sentence is equivalent modulo T to a universal sentence, then every substructure of A is elementarily equivalent to A.

Augustus De Morgan (1846): ‘If language were copious enough, [existential] propositions would seldom occur: and the idioms of every tongue are probably influenced by its power of . . . converting [existentials] into the form of universals.’

Thus if

$$\forall x \exists y \forall z \exists w \; R(x, y, z, w),$$

is true in A, it can be rewritten as

$$\forall x \forall z \; R(x, F(x), z, G(x, z)),$$

where T contains, for example,

$$\forall x \forall y \forall z \forall w \; (R(x, y, z, w) \rightarrow$$

$$R(x, F(x), z, G(x, z))$$

which is a universal sentence.
The functions F^A, G^A (or the symbols F, G) are called *Skolem functions*.

By the *cardinality* of a structure A, $|A|$, we mean the number of elements of the domain of A.
By the *cardinality* of a signature σ, $|\sigma|$, we mean the number of symbols in σ, or $\aleph_0$, whichever the greater.
(The $\aleph_0$ is to allow for 'closing off' in the proof of the theorem below.)

Putting all these tricks together:

Downward Loewenheim-Skolem Theorem, Skolem 1920
Let A be a σ-structure of infinite cardinality κ, and let λ be a cardinal which is $< \kappa$ but $\geqslant |\sigma|$. Then A has a substructure which is elementarily equivalent to A and has cardinality λ.

Why just go downward?
The following is also true, though its proof is completely different.

Upward Loewenheim-Skolem Theorem, Mal'tsev 1938
Let A be a σ-structure of infinite cardinality κ, and let λ be a cardinal which is $> \kappa$ and $\geqslant |\sigma|$. Then A has an extension which is elementarily equivalent to A and has cardinality λ.

Putting these facts together, first-order logic is absolutely hopeless at distinguishing between infinite cardinals.

The Upward Loewenheim-Skolem Theorem rests on the following important result.
We say that a first-order theory T is *consistent* if it has at least one model.

Compactness Theorem (Goedel, Mal'tsev, Henkin)
Let T be a first-order theory. If every finite subset of T is consistent, then T is consistent.

Typical application. Consider the structure $\mathbb{R}$ of real numbers, in any reasonable signature.
Let T be the set of all first-order sentences true in $\mathbb{R}$, together with the sentences

$$c > 0, c < 1, c < 0.1, c < 0.01, \ldots$$

This theory has a model A, by the compactness theorem. The element c^A is 'infinitesimal', i.e. greater than 0 but less than any positive fraction.

The Compactness Theorem tells us that first-order logic is not just bad at distinguishing infinite cardinals from one another; it can't even distinguish finite from infinite.
(Apply the Compactness Theorem to the set of sentences expressing:

> There is at least one element.
> There are at least two elements.
> There are at least three elements.
> etc.)

Also the theorems of this section tell us nothing about finite structures. To get useful information about distinguishing between finite structures, we must look elsewhere.

3. Distinguishing elements within a structure

We go back to the first lecture and recall that interpretations can be split into two parts: a *structure* that provides a context, and an *assignment* to free variables. So far we have ignored assignments (though we would have had to mention them in some proofs). Now we do the opposite: we fix a structure and look at assignments.

Throughout this lecture, σ is a fixed signature.
For technical reasons we assume σ has at least one individual constant.
A is a fixed σ-structure.

We can list the variables of the language as

$$x_1, x_2, x_3, \ldots$$

An *assignment* (or *tuple*) *of length* n *in* A is a list of n elements of the domain of A:

$$\bar{a} = (a_1, a_2, \ldots, a_n).$$

It assigns a_1 to x_1, a_2 to x_2 and so on.
If ϕ is a formula whose free variables are all among $x_1, \ldots, x_n$, then ϕ separates the assignments of length n into those that satisfy ϕ and those that don't.

We write $\phi(A^n)$ for the set of assignments of length n that satisfy ϕ in A.
$\phi(A^n)$ is an n-ary relation on dom(A);
we say it is a *first-order definable relation on* A.

The big question: In any given structure A, what are the relations of the form $\phi(A^n)$?

Example: A is $\mathbb{R}$, the field of real numbers; $n = 2$.
If ϕ is the formula

$$x_1^2 + x_2^2 = 1$$

then $\phi(\mathbb{R}^2)$ is the unit circle.

Fact. The class of first-order definable n-ary relations on A

- contains the empty set;
- contains the set $\text{dom}(A)^n$ of all tuples of length n in A;
- is closed under intersection $\cap$;
- is closed under union $\cup$;
- is closed under complement $(\text{dom}(A)^n \setminus X)$.

The last three clauses say that this set is *closed under boolean combinations*.

In some cases we can describe a simple set of first-order formulas, called *basic formulas*, and prove that every formula defines the same relation in A as some boolean combination of basic formulas.
Then the first-order definable relations are exactly the boolean combinations of the relations defined by basic formulas.

Arguments along these lines are known as *the method of quantifier elimination*. They are usually not easy, though some relatively recent adaptations of back-and-forth games make them easier.

Fortunately the method of quantifier elimination works (and gives important information) for many interesting mathematical structures.

The term algebra $\mathbb{T}$

(Martin Davis is responsible for logic programmers calling $\mathbb{T}$ the *Herbrand universe*, but he has since regretted this name which is historically inaccurate.)

The set of *terms* of the signature σ is defined inductively by:

- Every individual constant is a term.
- Every variable is a term.
- If F is a function symbol of arity n, and $t_1, \ldots, t_n$ are terms, then
$$F(t_1, \ldots, t_n)$$
is a term.

A *closed term* is a term with no variables in it.

We make the set C of closed terms of σ into a σ-structure $\mathbb{T}$ as follows.

The domain is C.

For each individual constant c, $c^{\mathbb{T}}$ is c.

For each n-ary function symbol F, if $t_1, \ldots, t_n$ are closed terms, then

$$F^A(t_1, \ldots, t_n) = F(t_1, \ldots, t_n).$$

For each relation symbol R, $R^{\mathbb{T}}$ is empty.

Universal algebraists know $\mathbb{T}$ as the free σ-algebra on no generators.

Given a closed term

$$F(G(a), G(F(c, d)))$$

we shall call $G(a)$ the (1)-th subterm
and $G(F(c, d))$ the (2)-th subterm.
Likewise we call a the (1,1)-th-subterm,
$F(c, d)$ the (2,1)-th subterm,
d the (2,1,2)-th subterm, etc.
We call the whole term the ()-th subterm.

Quantifier elimination theorem. As the basic formulas we can take all formulas of the forms

> The (...)-th subterm of x_i is equal to the (...)-th subterm of x_j.
>
> The (...)-th subterm of x_i begins with an F (or a c, etc.).

Hence none of the following can be expressed by first-order formulas:

1. x appears in y as a subterm.

2. The symbol 'F' appears in the term x.

3. x and y have a subterm in common.

4. Games for comparing structures

The first-order languages in this lecture have only finitely many symbols in their signatures, and *none of these symbols are function symbols*.

A typical example is the language of *graphs*. It has just one 2-ary relation symbol E;

$$E(x, y)$$

means that there is an edge between node x and node y.

In a graph G we say that a list of nodes

$$(a_1, \ldots, a_n)$$

is a *path from a_1 to a_n* if there are an edge from a_1 to a_2, an edge from a_2 to a_3, . . . , an edge from a_{n-1} to a_n.
The *length* of this path is $n - 1$.
The list (a) counts as a path from a to a of length 0.

We say that nodes a, b in a graph are *connected* if there is a path from a to b; their *distance* is the length of a shortest such path.
We say that the graph itself is *connected* if every pair of nodes of the graph is connected.

Our main task in this lecture is to show that there is no first-order sentence distinguishing the connected graphs from the unconnected ones.

We want to compare two σ-structures A and B.
We shall do it by setting up a game $EF_\omega(A, B)$ (EF after Ehrenfeucht and Fraïssé).

The game will use the notion of an atomic formula.

For any signature σ, the *atomic formulas* of σ are the formulas of these two kinds:

- $t_1 = t_2$, where t_1 and t_2 are terms.

- $R(t_1, \ldots, t_n)$, where R is a relation symbol of arity n and $t_1, \ldots, t_n$ are terms.

So in the signature for graphs, the only atomic formulas are

$$x_i = x_j, \qquad E(x_i, x_j).$$

There are two players, ∀belard (often known as Spoiler) and ∃loise (often known as Duplicator).
Note that Spoiler is male and Duplicator is female.

Intuitively, ∀belard is trying to show that A and B are different by finding a feature of one structure that can't be matched in the other. ∃loise is trying to show that A and B are the same, by finding a match in the other structure for each feature that ∀belard points to.

The game $EF_\omega(A, B)$ takes place in steps, starting at the 1st.
In each step (say the n-th), first ∀belard chooses an element from either dom(A) or dom(B), then ∃loise chooses one from the other of these sets.
We write a_n for the element chosen from A and b_n for the element chosen from B.

After infinitely many steps, ∀belard wins if there is some atomic formula ϕ which is satisfied by exactly one of $(a_1, a_2, \ldots)$ and $(b_1, b_2, \ldots)$; otherwise ∃loise wins.

Note also the game $EF_n(A, B)$ where n is a finite number. This is the same as above, except that the players stop as soon as they have chosen a_n, b_n.
The smaller n, the easier it is for ∃loise to win.

EXAMPLE FOR GAME

We say that A is *n-equivalent to B*, in symbols

$$A \sim_n B,$$

if ∃loise has a winning strategy for the game $EF_n(A, B)$.
If she has a winning strategy for the infinite game $EF_\omega(A, B)$, we say that A and B are *back-and-forth equivalent*,

$$A \sim B.$$

Note:

$$A \sim B \;\Rightarrow\; \ldots \;\Rightarrow\; A \sim_n B \;\Rightarrow\; A \sim_{n-1} B \;\Rightarrow \ldots$$

because the games get easier for ∃loise to win as they get shorter.

FACT. All the relations $\sim$, $\sim_n$ are equivalence relations.

The main theorem about these games needs the notion of the *quantifier rank* $qr(\phi)$ of a formula ϕ. This is defined inductively:

- If ϕ is atomic then $qr(\phi) = 0$.

- $qr(\neg\phi) = qr(\phi)$.

- $qr(\phi \wedge \psi) = qr(\phi \vee \psi) = \max(qr(\phi), qr(\psi))$.

- $qr(\forall x\phi) = qr(\exists x\phi) = qr(\phi) + 1$.

Important fact: Every first-order formula has a finite quantifier rank.

Fraïssé's Theorem. Let A, B be σ-structures. Then for each n the following are equivalent:

(i) $A \sim_n B$.

(ii) If ϕ is any sentence of quantifier rank at most n, then A is a model of ϕ if and only if B is a model of ϕ.

Corollary. Let **K** be a class of σ-structures. Suppose that for every finite n there are a structure A_n in **K** and a structure B_n not in **K**, such that $A_n \sim_n B_n$. Then there is no first-order sentence whose models are the structures in **K**.

Let k, m be positive integers with $k \geqslant 3$.
We write $[k, m]$ for the graph consisting of m copies of a cycle of length k.

Then $[k, m]$ is connected if and only if $m = 1$.

Suppose A, B are graphs of the form $[k, m]$ (possibly different k, m).
Suppose $\bar{a} = (a_1, \ldots, a_p)$ are nodes of A and $\bar{b} = (b_1, \ldots, b_p)$ are nodes of B.
We say $\bar{a}$ and $\bar{b}$ are *h-matched* if for all i, j and every $h' \leqslant h$,

$$\text{distance}(a_i, a_j) = h' \Leftrightarrow \text{distance}(b_i, b_j) = h'.$$

One can show: If $A = [k, m]$ and $B = [k', m']$ and $\bar{a}, \bar{b}$ are as above and are $(2^{n-1} - 1)$-matched, then $(\bar{a}, \bar{b})$ is a winning position for ∃loise in $EF_{p+n}(A, B)$. Then:

Fact. If k, k' are both $\geqslant 2^n$ then $[k, m] \sim_n [k', m']$.

Let n be a positive integer. Then by the Fact above,

$$[2^n, 1] \sim_n [2^n, 2].$$

Since one of these graphs is connected and the other is not, this shows that *no first-order sentence distinguishes between connected graphs and unconnected graphs*.

Hence *there is no first-order formula* $\phi(x, y)$ *that expresses in graphs that* x, y *are connected*.
Otherwise the sentence

$$\forall x \forall y \ \phi(x, y)$$

would have distinguished between connected and unconnected graphs.

5. Ways of cheating

During the 1980s several authors proposed using first-order logic as either a programming language or a specification language.
At first glance this is absurd.
We saw in Lecture Two that first-order logic can't express that a structure is finite.
We saw in Lecture Three that first-order logic can't express that a graph is connected.

Example from Prolog

Sterling and Shapiro, *The Art of Prolog*, give the following Prolog program to define connectedness in graphs:

```
connected(Node,Node).
connected(Node1,Node2) ←
    edge(Node1,Link),
    connected(Link,Node2).
```

This is a first-order theory T:

$$\forall x\ C(x, x).$$

$$\forall x \forall y \forall z\ (E(x, y) \wedge C(y, z) \to C(x, z)).$$

$$\forall x \ C(x, x).$$

$$\forall x \forall y \forall z \ (E(x, y) \wedge C(y, z) \rightarrow C(x, z)).$$

Note 1: This theory T is not in the language of graphs, because it has the extra symbol C.

Note 2: If we read '$E(x, y)$' as 'there is an edge from x to y', and '$C(x, y)$' as 'there is a path from x to y', then both sentences are true in any graph.

Note 3: Both sentences are still true in any graph if we read E as before but take '$C(x, y)$' to be true for all nodes x and y (regardless of whether they are connected).

So T is hopeless as a straightforward model-theoretic definition of connectedness. So why is it called a *definition* by Prolog people?

Clue: T is a *definite Horn clause theory*. This means that each sentence in T has one of the forms

$$\forall x_1 \ldots \forall x_n \; \psi,$$

$$\forall x_1 \ldots \forall x_n \; (\phi_1 \wedge \ldots \phi_k \to \psi)$$

where $\phi_1, \ldots, \phi_k, \psi$ are atomic formulas.

Suppose G is any graph.
Give each node of G a name, by adding new individual constants to the signature if necessary.
There is a smallest set U of atomic sentences that contains:

(i) All the atomic sentences true in G.

(ii) All the sentences $C(a, a)$ (a an individual constant).

(iii) The sentence $C(a, c)$ whenever U contains sentences $E(a, b)$ and $C(b, c)$.

We read the second and third clauses off from the Horn theory T.

Then U is a complete description of a structure H which is a model of T.

After clause (i) we never added new individual constants or equalities ‘$a = b$’ or sentences beginning with ‘E’.
So H is an exact copy of G, except that it also has a new binary relation C^H.

This relation holds between a and b if and only if a and b are connected.

There are several ways of saying this.
One (the universal algebraist’s) is that the only model of T that we are concerned with is the ‘free model over G’.
Another (the logician’s) is that T is being used not as a first-order theory but as *part of an inductive definition*.

Algebraic example

Suppose A is any infinite structure.
Let $G(A)$ be the group of all permutations of dom(A).
What can we say about A, using *first-order sentences about* $G(A)$?

Let $Transp(x)$ say:

> $x^2 = 1$ and $x \neq 1$ and for all y there is a subgroup of order at most 6 containing both x and $y^{-1}xy$.

Then $Transp(x)$ expresses that x is a transposition.

Let $Overlap(x, y)$ be

$$Transp(x) \wedge Transp(y) \wedge xy \neq yx.$$

Then $Overlap(x, y)$ expresses that x, y are transpositions $(a, b), (a, c)$ with $b \neq c$; we call a the *pivot* of x, y.

One can write a formula $Equiv(x, y, z, w)$ which expresses that $Overlap(x, y)$, $Overlap(z, w)$ and x, y have the same pivot as z, w.

Now to express 'w moves every element of dom(A)' we say:

> Whenever $Overlap(x, y)$, it is false that $Equiv(x, y, wx, wy)$.

To express 'w is cyclic' we say:

> Every x that commutes with w either moves every element of dom(A) moved by w, or moves no element of dom(A) moved by w.

So we can write a first-order sentence ϕ in the language of groups, such that for every infinite structure A, $G(A)$ satisfies ϕ if and only if A is countable.
Namely: 'There is a cyclic permutation that moves every element.'

Question: Why doesn't this contradict the Upward Loewenheim-Skolem Theorem?

Answer: Because ϕ is not about A but about a structure built up from A using set theory.

One can use first-order sentences to make other strong statements about A, by talking instead about other structures built up from A:
for example the universe of sets with elements of dom(A) as urelements.

References

For Lectures 2–4,

Wilfrid Hodges, A shorter model theory, Cambridge University Press 1997.

For back-and-forth games on finite structures there is a fuller account in

Heinz-Dieter Ebbinghaus and Jörg Flum, Finite model theory, Springer 1999 (2nd edition).