Chapter 7
Herbrand Method

§7.1 Introduction

In Propositional Logic, it is possible to determine satisfiability and logical entailment by looking at the set of all interpretations for the logical constants of that language, i.e. its truth table. Although the truth table for a propositional language can be large, it is always finite; and so there is a simple procedure for checking whether a sentence is satisfiable or valid or whether a sentence logically entails another.

In Relational Logic, we also have the notion of an interpretation; and this raises the question of whether there is an approach to checking satisfiability and logical entailment for Relational Logic that is analogous to the truth table method for Propositional Logic.

Consider a set of sentences in Relational Logic with two object constants *a* and *b* and one unary relation constant *r*. The table below shows all of the interpretations using a universe of discourse consisting of just two objects, o and •. In this case, the set of all possible interpretations is finite.

∀	a	b	r
{o,•}	o	o	{}
{o,•}	o	o	{o}
{o,•}	o	o	{•}
{o,•}	o	o	{o,•}
{o,•}	o	•	{}
{o,•}	o	•	{o}
{o,•}	o	•	{•}
{o,•}	o	•	{o,•}
{o,•}	•	o	{}
{o,•}	•	o	{o}
{o,•}	•	o	{•}
{o,•}	•	o	{o,•}
{o,•}	•	•	{}
{o,•}	•	•	{o}
{o,•}	•	•	{•}
{o,•}	•	•	{o,•}

Unfortunately, using just this table to check for satisfiability and logical entailment is not adequate, since it covers just one possible universe of discourse. There are many other possibilities. The problem with Relational Logic is that, in general, the

set of possible interpretations is necessarily infinite. Moreover, there is not even a systematic way of enumerating all of the possibilities.

The good news is that all is not lost. At least in some cases. The trick is to use a set of special interpretations, called *Herbrand interpretations*, to check satisfiability and logical entailment. It can be shown that in some cases checking these models alone is sufficient. Furthermore, in some cases, there are only finitely many Herbrand models; and so the process of checking for these conditions is algorithmic.

Sadly, this trick does not always work. Sometimes, Herbrand interpretations alone do not suffice. And, sometimes, even though Herbrand interpretations suffice, they are infinitely large; and so checking satisfaction takes forever.

In order to understand when the trick works and when it does not, we look at the method on increasingly complex subsets of Relational Logic. First, we examine Basic Logic, i.e. Relational Logic without variables, quantifiers, or function constants. We then extend this to Universal Logic by allowing variables. After looking at Universal Logic, we switch to Existential Logic (Universal Logic with quantifiers) but still avoid functions. Then, we look at Functional Logic but avoid quantifiers.

§7.2 Basic Logic

Basic logic is Relational Logic without variables, quantifiers, or function constants. All ground sentences are included and all logical sentences that can be formed from ground sentences.

The *Herbrand universe* for a set of sentences in Basic Logic is the set of all object constants used in the sentences. For example, for sentences with just two object constants a and b, the Herbrand universe is the set $\{a,b\}$.

A *Herbrand interpretation* for Basic Logic is an interpretation in which (1) the universe of discourse is the Herbrand universe for the language and (2) each object constant maps into itself.

For example, the interpretation shown below is a Herbrand interpretation for a set of sentences with objects constants a and b and binary relation constant r.

$$\forall^i = \{a,b\}$$
$$a^i = a$$
$$b^i = b$$
$$r^i = \{\langle a,b\rangle, \langle b,b\rangle\}$$

Note that the object constants are interpreted as themselves, as required by the definition of Herbrand interpretation. The interpretation of the relation constant r is arbitrary; any other binary relation on the Herbrand universe would be acceptable.

One interesting thing to note about Herbrand interpretations for basic logic is that there are only finitely many of them. Given a finite set of constants, the universe of discourse is finite. The interpretation of the object constants is fixed; and there are only finitely many relations that can be formed from a finite universe of discourse.

The other interesting thing about Herbrand interpretations for basic logic is that they alone are sufficient for checking satisfiability and logical entailment, because of the Basic Herbrand Theorem.

Basic Herbrand Theorem: A set of sentences in Basic Logic has a model if and only if it has a Herbrand model.

For an model i, we can construct a corresponding Herbrand model h as follows. The Herbrand interpretation for each object constant is itself. The Herbrand interpretation h for each relation constant ρ is the set of all tuples of object constants $\tau_1,\ldots,\tau_n$ such that interpretation i satisfies the sentence $\rho(\tau_1,\ldots,\tau_n)$. From this construction, it is easy to see that h satisfies the same ground atomic sentences as i, and their agreement on logical sentences is assured by the semantics of the logical operators.

As an example of this construction, consider a set of sentences with object constants a and b and a single binary relation r. Now, consider the interpretation i shown below. There are just two elements in the universe of discourse. The constant a maps to o; b maps to •; and r is an arbitrary binary relation in terms of o and •.

$$\forall^i = \{\text{o},\bullet\}$$
$$a^i = \text{o}$$
$$b^i = \bullet$$
$$r^i = \{\langle \text{o},\bullet \rangle, \langle \bullet,\bullet \rangle\}$$

The corresponding Herbrand interpretation is shown below. The constants a and b maps to themselves. In order to get the interpretation of r, we consider each relational sentence that can be formed from a, b, and r. For each, we check whether or not it is satisfied by i. If so, we include the corresponding pair of constants in our interpretation. If not, we do not include them. In this case, this leads to the interpretation shown.

$$\forall^i = \{a,b\}$$
$$a^i = a$$
$$b^i = b$$
$$r^i = \{\langle a,b \rangle, \langle b,b \rangle\}$$

As a more interesting example, consider the following interpretation for the same language. In this case, both object constants refer to the same object, viz. •; and one of the elements in the universe of discourse, viz. o, has no corresponding object constant.

$$\forall^i = \{o, \bullet\}$$
$$a^i = \bullet$$
$$b^i = \bullet$$
$$r^i = \{\langle o, \bullet \rangle, \langle \bullet, \bullet \rangle\}$$

The corresponding Herbrand interpretation is shown below. The constants *a* and *b* map to themselves. In order to get the interpretation of *r*, we consider each relational sentence involving *r* and the constants *a* and *b*. For each, we check whether it is satisfied by *i*. If so, we include the corresponding pair of constants in our interpretation. If not, we do not include them. In this case, we get all pairs.

$$\forall^i = \{a, b\}$$
$$a^i = a$$
$$b^i = b$$
$$r^i = \{\langle a, a \rangle, \langle a, b \rangle, \langle b, a \rangle, \langle b, b \rangle\}$$

In the first example, the "structure" of the interpretation and its corresponding Herbrand interpretation are very similar. In this second example, the structure is very different. In the second interpretation, both object constants refer to the same object whereas in the corresponding Herbrand interpretation, they refer to different objects. In the Herbrand interpretation, there is an object constant that refers to every element in the universe of discourse, whereas in the ordinary interpretation, there is an object o with no corresponding object constant. These differences might lead one to think that the two interpretations do not satisfy the same sentences. The Herbrand theorem assures us that, despite these differences, they satisfy exactly the same sentences, provided that we restrict ourselves to Basic Logic. The reader is encouraged to try out the interpretations on various sentences to confirm that this is indeed correct.

The importance of the Herbrand theorem is that it gives us a way of checking logical entailment in a finite amount of time. This is called the Herbrand Method. Starting with a set of premises and a possible conclusion, we loop over Herbrand interpretations, crossing out each interpretation that does *not* satisfy the premises. Then, we check whether each remaining interpretation satisfies the conclusion. If so, the premises logically entail the conclusion. Otherwise, logical entailment does not hold.

Suppose, for example, we were asked to determine whether $(p(a) \Rightarrow q(b))$ and $(p(a) \lor q(b))$ logically entail $q(b)$. We start by writing out the set of all Herbrand interpretations and crossing out those that do not satisfy the premises. The first table below on the left is the complete set of Herbrand interpretations for this language. The second table results from crossing out those interpretations that do not satisfy $(p(a) \Rightarrow q(b))$. The third table results from crossing out the interpretations that do not satisfy $(p(a) \lor q(b))$. The fourth table is the result of crossing out all rows that do not satisfy both premises.

p	q	p	q	p	q	p	q
{}	{}	{}	{}	×	×	×	×
{}	{a}	{}	{a}	×	×	×	×
{}	{b}	{}	{b}	{}	{b}	{}	{b}
{}	{a,b}	{}	{a,b}	{}	{a,b}	{}	{a,b}
{a}	{}	×	×	{a}	{}	×	×
{a}	{a}	×	×	{a}	{a}	×	×
{a}	{b}	{a}	{b}	{a}	{b}	{a}	{b}
{a}	{a,b}	{a}	{a,b}	{a}	{a,b}	{a}	{a,b}
{b}	{}	{b}	{}	×	×	×	×
{b}	{a}	{b}	{a}	×	×	×	×
{b}	{b}	{b}	{b}	{b}	{b}	{b}	{b}
{b}	{a,b}	{b}	{a,b}	{b}	{a,b}	{b}	{a,b}
{a,b}	{}	×	×	{a,b}	{}	×	×
{a,b}	{a}	×	×	{a,b}	{a}	×	×
{a,b}	{b}	{a,b}	{b}	{a,b}	{b}	{a,b}	{b}
{a,b}	{a,b}	{a,b}	{a,b}	{a,b}	{a,b}	{a,b}	{a,b}

Now, we check that each remaining interpretation satisfies the desired conclusion. In this case, the check is successful. Consequently, we can conclude that the premises in this case logically entail the conclusion.

Although this method is simple, it is not necessarily very efficient. The Herbrand model for a set of sentences in Basic Logic can be very large. If there are n object constants, then the Herbrand universe is of size n. This, in itself, is not so bad. What hurts is the number of interpretations of the relation constants. With n constants, one can form n^k k-tuples, leading to $2^{\wedge}n^k$ possible k-ary relations. With m k-ary relation constants, this leads to $(2^{\wedge}n^k)^{\wedge}m$ possible Herbrand interpretations.

To make this concrete, consider a language with ten object constants and three binary relation constants. In this case, the Herbrand universe has size 10. There are 100 2-tuples, leading to 2^{100} binary relations, leading to 2^{300} possible Herbrand interpretations. For problems of this size, the Herbrand method is impractical.

7.3 Universal Logic

Universal Logic differs from Basic Logic in that we are permitted to use variables. Function constants are still forbidden. Explicit quantifiers are excluded as well. (Free variables have implicit universal quantification, and there are no embedded universal quantifiers and no existential quantifiers.)

As with Basic Logic, the *Herbrand universe* for Universal Logic is the set of all object constants. If there are no object constants, then we include a single, arbitrary object constant, say a.

As with Basic Logic, we can form the Herbrand interpretation corresponding to any given interpretation. The process is exactly the same. Every object constant is mapped into itself. For every relation constant ρ, we take the set of tuples of object constants $\tau_1,\ldots,\tau_n$ for which $\rho(\tau_1,\ldots,\tau_n)$ is true under the given interpretation.

The apparent bad news is that this sometimes produces a Herbrand interpretation that does not agree with the interpretation on which it is based. Consider the interpretation shown below.

$$\forall^i = \{o,\bullet,*\}$$
$$a^i = o$$
$$b^i = \bullet$$
$$r^i = \{\langle o,o\rangle, \langle\bullet,\bullet\rangle, \langle o,*\rangle\}$$

The following interpretation is the corresponding Herbrand interpretation.

$$\forall^i = \{a,b\}$$
$$a^i = a$$
$$b^i = b$$
$$r^i = \{\langle a,a\rangle, \langle b,b\rangle\}$$

Now, consider the sentence $r(x,x)$, which asserts that r is reflexive. This is clearly true in the Herbrand interpretation, since every object constant is r-related to itself. However, it is not true in the preceding interpretation.

In the case of Basic Logic, the Herbrand interpretation corresponding to an ordinary interpretation satisfies exactly the same sentences as the ordinary interpretation. While this relationship does not hold for Universal Logic, a weaker version does hold.

The *Skolemization* of a sentence ψ is the sentence ψ' obtained by replacing all variables in ψ with new constants. The importance of Skolemization stems from the fact that a sentence in Universal Logic is satisfiable if and only if its Skolemization is satisfiable.

The upshot of this is that a version of the Herbrand method can still be used in Universal Logic. It is essentially the same as the Herbrand Method for Basic Logic. The only difference is that any variables that occur in the desired conclusion must first be replaced by new object constants. If a sentence can be proved with arbitrary new object constants substituted for variables, then it must be true for all objects in the universe of discourse.

Checking whether φ logically entails ψ is a two-step process. We first create the Skolemization ψ' by replacing all variables in ψ with new constants. We then loop over Herbrand interpretations for the (possibly enlarged) language in the usual fashion. If every Herbrand interpretation that satisfies the premises also satisfies the Skolemized conclusion, then the premises logically entail the conclusion.

Suppose, for example, we were asked to determine whether $(p(x) \Rightarrow q(x))$. and $(q(x) \Rightarrow r(x))$ logically entail $(p(x) \Rightarrow r(x))$. Remember that, in Relational Logic, free

variables are interpreted as universally quantified variables; hence the implication here must be true for all objects in the universe of discourse.

We start by Skolemizing the desired conclusion. We introduce a new object constant a and then create the Skolemized conclusion $(p(a) \Rightarrow r(a))$. Next, we write out the set of all Herbrand interpretations. We then crossout those that do not satisfy the premises. The first table below on the left is the complete set of Herbrand interpretations for this language. The second table results from crossing out those interpretations that do not satisfy $(p(x) \Rightarrow q(x))$. The third table results from crossing out the interpretations that do not satisfy $(q(x) \Rightarrow r(x))$. The fourth table is the result of crossing out all rows that do not satisfy both premises.

p	q	r		p	q	r		p	q	r		p	q	r
{}	{}	{}		{}	{}	{}		{}	{}	{}		{}	{}	{}
{}	{}	{a}		{}	{}	{a}		{}	{}	{a}		{}	{}	{a}
{}	{a}	{}		{}	{a}	{}		×	×	×		×	×	×
{}	{a}	{a}		{}	{a}	{a}		{}	{a}	{a}		{}	{a}	{a}
{a}	{}	{}		×	×	×		{a}	{}	{}		×	×	×
{a}	{}	{a}		×	×	×		{a}	{}	{a}		×	×	×
{a}	{a}	{}		{a}	{a}	{}		×	×	×		×	×	×
{a}	{a}	{a}		{a}	{a}	{a}		{a}	{a}	{a}		{a}	{a}	{a}

Now, we verify that the remaining interpretations satisfy the desired conclusion. Consequently, we can conclude that the premises in this case logically entail the conclusion.

§7.4 Existential Logic

Existential Logic is the same as universal logic except that it includes quantified sentences. It is called existential logic because the problems that arise are due to the presence of existential quantifiers. Universal quantifiers can also cause problems but only when they are equivalent to existential quantifiers, i.e. when used in a "negative context" such as the target of a negation or the antecedent of an implication or reduction.

The bad news is that no version of the Herbrand Theorem holds for Existential Logic. If a set of sentences in an Existential Logic has a model, then it does not necessarily have a Herbrand model.

To see that this is true, consider the following example. Suppose we have a language with two object constants a and b and a single unary relation constant r. Now, consider the following sentences in existential logic.

$$\neg r(a)$$
$$\neg r(b)$$
$$\exists x.r(x)$$

The interpretation shown below is a model of these sentences. There are three objects in the universe of discourse. The object constants *a* and *b* map to two of these objects, and the relation constant *r* maps to a relation that is true of the third object only.

$$\forall^i = \{o, \bullet, *\}$$
$$a^i = o$$
$$b^i = \bullet$$
$$p^i = \{*\}$$

The corresponding Herbrand interpretation in this case is shown below. We have a universe of discourse with just two objects.

$$\forall^i = \{a, b\}$$
$$a^i = a$$
$$b^i = b$$
$$r^i = \{\}$$

Unhappily, this interpretation is not a model of the sentences above. The third sentence requires that there be an object that satisfies *r*, and this interpretation does not contain any such object.

What's more, it is easy to see that there can be no model built on the Herbrand universe. Because of the first two sentences, the relation corresponding to *r* must not be true of these constants, and there is no other object in the universe of discourse to satisfy the third sentence.

The upshot of this is that the Herbrand method does not work for Existential Logic. It can be used to show that a set of sentences is satisfiable. However, the failure to satisfy all Herbrand interpretations does not mean that there are no interpretations that satisfy the sentences, as this example demonstrates.

§7.5 Functional Logic

Functional Logic is the same as Universal Logic except that it includes functional terms. We are still not permitted to use quantifiers, though as with Universal Logic free variables are implicitly universally quantified.

For Universal Logic, the Herbrand universe is the set of all object constants. For Functional Logic, we must also include complex terms that can be formed by applying function constants to ground terms. If there are no object constants, then we include an arbitrary object constant, say *a*.

As an example, consider a set of sentences with two object constants *a* and *b*. In this case, the Herbrand universe consists of just these two constants.

$$\{a, b\}$$

8

Now, let us add two unary function constants, say *f* and *g*. In this case, we must add additional terms, as shown below. Note that the Herbrand universe is infinite in this case, as it is for all functional languages.

$$\{a, b, f(a), f(b), g(a), g(b), f(f(a)), f(f(b)), f(g(a)), f(g(b)), ...\}$$

A *Herbrand interpretation* for functional logic is an interpretation in which (1) the universe of discourse is the Herbrand Universe for the language, (2) each object constant maps into itself, and (3) each function constant denotes a function that maps ground terms to themselves.

For example, the interpretation shown below is a Herbrand interpretation with objects constants *a* and *b* and unary function constants *f* and *g*. Each of the object constants maps into itself, and each function constant is interpreted as a function that maps terms into a new term comprised of that function constant applied to those terms.

$$\forall^i = \{a, b, f(a), f(b), g(a), g(b), ...\}$$
$$a^i = a$$
$$b^i = b$$
$$f^i = \{\langle a \rightarrow f(a) \rangle, \langle b \rightarrow f(b) \rangle, \langle f(a) \rightarrow f(f(a)) \rangle, ...\}$$
$$g^i = \{\langle a \rightarrow g(a) \rangle, \langle b \rightarrow g(b) \rangle, \langle f(a) \rightarrow g(f(a)) \rangle, ...\}$$
$$r^i = \{\langle a, b \rangle, \langle b, b \rangle, \langle a, f(a) \rangle, \langle b, f(b) \rangle\}$$

The interpretation shown below is also a Herbrand interpretation for this language. The object constants and function constants are interpreted as before, as required by the definition of Herbrand interpretation. The only thing different, the only thing that can be different is the interpretation of the relation constant *r*.

$$\forall^i = \{a, b, f(a), f(b), g(a), g(b), ...\}$$
$$a^i = a$$
$$b^i = b$$
$$f^i = \{\langle a \rightarrow f(a) \rangle, \langle b \rightarrow f(b) \rangle, \langle f(a) \rightarrow f(f(a)) \rangle, ...\}$$
$$g^i = \{\langle a \rightarrow g(a) \rangle, \langle b \rightarrow g(b) \rangle, \langle f(a) \rightarrow g(f(a)) \rangle, ...\}$$
$$r^i = \{\langle a, b \rangle, \langle b, b \rangle, \langle f(a), a \rangle, \langle f(b), b \rangle\}$$

We define a *Herbrand variable assignment* as a variable assignment in which the universe is the Herbrand Universe for the language. The only difference is that the universe of discourse now contains more objects.

$$x^i = a$$
$$y^i = a$$
$$z^i = f(g(b))$$

The good news about Functional Logic is that the Herbrand Method works, provided we use the algorithm used for Universal Logic. Many people are surprised by this result. It is possible to have a finite model for a functional language; the universe of discourse of the corresponding Herbrand interpretation is infinitely large; yet the two interpretations satisfy the same sentences.

Unfortunately, in the case of Functional Logic, the result is not as useful as it is for Universal Logic. The size of the Herbrand universe for Functional Logic is infinite. Consequently, checking the Herbrand interpretations for a language to determine satisfiability and logical entailment is not feasible.