Remark 34.0.1 Sometimes information is explicit: Euclid proved the square root of 2 is irrational. Often times we find an implicit characterization instead: Euclid proved the positive root of the equation $x^2 - 2 = 0$ is irrational. Puzzles often involve turning implicit characterizations into explicit ones. Beth's Definability Theorem, due to the Dutch logician Evert Beth, essentially says that such puzzles can always be solved in classical logic. This is a fundamental result that says that classical logic has a kind of completeness result where definability is concerned. The proof we give here comes from Craig's 1957 paper.

Definition 34.0.2 (Explicit Definition) Let $\mathcal{L}$ be a first-order language and R an n-place predicate symbol not in $\mathcal{L}$. We say that R is *explicitly definable* in $\mathcal{L}$ with respect to a set S of sentences in $\mathcal{L} \cup \{R\}$ if there is a formula $\Phi(x_1, \ldots, x_n)$ with free variables among $x_1, \ldots, x_n$ (and, of course, no occurrence of R) such that

$$S \models (\forall x_1) \ldots (\forall x_n)\big[R(x_1, \ldots, x_n) \leftrightarrow \Phi(x_1, \ldots, x_n)\big].$$

Definition 34.0.3 (Implicit Definition) Again let R be an n-place predicate symbol. R is *implicitly definable* with respect to a set S of sentences, provided S determines R uniquely, in the following sense. Let R^* be an n-place predicate different from R and let S^* be like S except every occurrence of R has been replaced by R^*. The S determines R uniquely if

$$S \cup S^* \models (\forall x_1) \ldots (\forall x_n)\big[R(x_1, \ldots, x_n) \leftrightarrow R^*(x_1, \ldots, x_n)\big].$$

Example 34.0.4 Let S be the set consisting of the following three sentences:

$$(\forall x)(R(x) \rightarrow A(x))$$
$$(\forall x)(R(x) \rightarrow B(x))$$
$$(\forall x)((A(x) \wedge B(x)) \rightarrow R(x))$$

It is easy to check that R is uniquely determined in any structure which provides an interpretation of the predicates A and B, so that R is implicitly definable from S. In face, R has the explicit definition

$$(\forall x)\big[R(x) \leftrightarrow (A(x) \wedge B(x))\big].$$

Remark 34.0.5 It is easy to see that if R has an explicit definition with respect to a set S of sentences, then it has an implicit definition with respect to the same set S of sentences. Suppose

$$S \models (\forall x_1) \ldots (\forall x_n)\big[R(x_1, \ldots, x_n) \leftrightarrow \Phi(x_1, \ldots, x_n)\big].$$

Let S^* be the result of replacing R by R^* (where R^* is new to S and Φ, so that it is also true that

$$S^* \models (\forall x_1) \ldots (\forall x_n)\big[R^*(x_1, \ldots, x_n) \leftrightarrow \Phi(x_1, \ldots, x_n)\big].$$

Hence

$$S \cup S^* \models (\forall x_1) \ldots (\forall x_n)\big[R(x_1, \ldots, x_n) \leftrightarrow R^*(x_1, \ldots, x_n)\big],$$

because

$$(\forall x_1) \ldots (\forall x_n)\big[R(x_1, \ldots, x_n) \leftrightarrow \Phi(x_1, \ldots, x_n)\big] \wedge (\forall x_1) \ldots (\forall x_n)\big[R^*(x_1, \ldots, x_n) \leftrightarrow \Phi(x_1, \ldots, x_n)\big]$$
$$\rightarrow (\forall x_1) \ldots (\forall x_n)\big[R(x_1, \ldots, x_n) \leftrightarrow R^*(x_1, \ldots, x_n)\big].$$

The converse is also true, this is the result of Beth's.

Theorem 34.0.6 (Beth's Definability Theorem.) If R is implicitly defined by a set S of sentences, the R has an explicit definition with respect to S as well.

Proof. Nothing in the proof depends on the arity of the predicate R, so I will reduce clutter by assuming that R is a 1-place predicate. Suppose R is implicitly defined by S. Let R^* be a new predicate symbol and S^* the result of replacing R by R^* in S. Then

$$S \cup S^* \models (\forall x)\ \big[R(x) \leftrightarrow R^*(x)\big].$$

Although $S \cup S^*$ may be infinite sets, it follows by the compactness theorem 30.3.6, that there are finite subsets S_0 of S and S_0^* of S^* so that

$$S_0 \cup S_0^* \models (\forall x)\ \big[R(x) \leftrightarrow R^*(x)\big].$$

Let X be the conjunction of the sentences in S_0 and X^* be a conjunction of the sentences in S_0^*, so that the following is valid by the Deduction theorem (see Remark 24.1.8(2))

$$(X \wedge X^*) \rightarrow (\forall x)\ \big[R(x) \leftrightarrow R^*(x)\big].$$

Let a be a parameter new to $S \cup S^*$, so the following is also valid (by Theorem 23.2.6)

$$(X \wedge X^*) \rightarrow\ \big[R(a) \leftrightarrow R^*(a)\big].$$

By propositional logic, this latter sentence is equivalent to

$$\big(X \wedge R(a)\big) \rightarrow \big(X^* \rightarrow R^*(a)\big).$$

By the Craig Interpolation Theorem there is an interpolant for this conditional $\Phi(a)$ which may contain the parameter a, so that each of the following are valid

$$\big(X \wedge R(a)\big) \rightarrow \Phi(a)$$
$$\Phi(a) \rightarrow \big(X^* \rightarrow R^*(a)\big)$$

Now, the key fact is that neither predicate R nor R^* can occur in $\Phi(a)$, since $\Phi(a)$ must share its nonlogical symbols with the antecedent (which does not contain R^*) and the consequent (which does not contain R).

Using propositional logic we can re-arrange the subformulae

$$X \rightarrow \big(R(a) \rightarrow \Phi(a)\big)$$
$$X^* \rightarrow \big(\Phi(a) \rightarrow R^*(a)\big)$$

In the second sentence replace all occurrences of R^* by R, this will not change the validity of the second sentence (since R does not occur in it). There is no change to $\Phi(a)$ in this replacement, so the following are valid

$$X \rightarrow \big(R(a) \rightarrow \Phi(a)\big)$$
$$X \rightarrow \big(\Phi(a) \rightarrow R(a)\big)$$

It follows again by propositional logic that

$$X \rightarrow \big(R(a) \leftrightarrow \Phi(a)\big)$$

and so using the definition of X together with the Deduction theorem again

$$S_0 \models \big(R(a) \leftrightarrow \Phi(a)\big).$$

Since the parameter a does not occur in S, we can quantify it out

$$S_0 \models (\forall x)\big(R(x) \leftrightarrow \Phi(x)\big).$$

Finally since $S_0 \subseteq S$,

$$S \models (\forall x)\big(R(x) \leftrightarrow \Phi(x)\big).$$

Thus R is explicitly definable from S by means of $\Phi(x)$. □