The Decision Problem for First-Order Logic

Yuri Gurevich

Chapter 8. Existential Interpretation.

§ 1. Method.

Let T1 and T2 be interpreted theories. For the sake of simplicity we assume that T1 has no operation symbols and has a single, dyadic predicate symbol P . Let

$$\neg F(x) \sim G(x) \quad \text{and} \quad F(x) \mathbin{\&} F(y) \supset [C(x,y) \sim D(x,y)]$$

be theorems of the theory T2 , where the formulas F, G, C, and D are existential and have no free variables except for those shown.

For an arbitrary formula A of the theory T1 we indicate by A' the formula obtained from A by replacing P by C and limiting the quantifiers by means of F . For example, if

$A = \forall x \neg Pxx$, then

$A' = \forall x(F(x) \supset \neg C(x,x))$.

For the purposes of the following lemma formulas B_1 and B_2 of the theory T2 will be called F-equivalent, if they have exactly the same free variables, and the formula

$$\forall x_1 \ldots \forall x_m [F(x_1) \mathbin{\&} \ldots \mathbin{\&} F(x_m) \supset (B_1 \sim B_2)]$$

is a theorem of the theory T2 .

Lemma 1. There is an algorithm, which transforms an arbitrary prenex formula A of the theory T1 into a prenex formula A^* of the theory T2 , such that

(a) A^* is F-equivalent to the formula A' , and

(b) If the formula A is of the prefix class $\Pi(w)$, then formula A^* is of the prefix class $\Pi(w\exists\infty)$..

Proof. Let the sign $\equiv$ denote F-equivalence of formulas of the theory T2 . We show how to transform A to a suitable A^* by induction on the structure of A . Without loss of generality we may assume that the quantifier-free part of the formula A is constructed from atomic formulas and their negations with the help of the signs $\&$ and $\vee$.

Case 1. $A = Puv$. Then $A^* \rightleftharpoons C(u,v) = A'$.

Case 2. $A = \neg Puv$. Then $A^* \rightleftharpoons D(u,v) \equiv \neg C(u,v) = A'$.

Case 3. $A = A_1 \& A_2$. Then by the induction hypothesis $A_1^* \equiv \exists x_1 \dots x_m B_1$ and $A_2^* \equiv \exists y_1 \dots y_n B_2$, where B_1 and B_2 are quantifier-free. Therefore $A_1^* \rightleftharpoons \exists x_1 \dots x_m y_1 \dots y_n (B_1 \& B_2) \equiv A_1^* \& A_2^* \equiv A_1' \& A_2' = A'$.

Case 4. $A = A_1 \vee A_2$. Then by the induction hypothesis $A_1^* = \exists x_1 \dots x_m B_1$ and $A_2^* = \exists x_1 \dots x_n B_2$, where B_1 and B_2 are quantifier-free. Let $m \leq n$. Therefore

$A^* \rightleftharpoons \exists x_1 \dots x_n (B_1 \vee B_2) \equiv A_1^* \vee A_2^* \equiv A_1' \vee A_2' = A'$.

Case 5. $A = \exists x A_1$, where $A_1 = q_1 y_1 \dots q_k y_k A_2$ and A_2 is quantifier-free. By the induction hypothesis $A_1^* \equiv q_1 y_1 \ldots q_k y_k \exists y_{k+1} \ldots \exists y_m B$, where B is quantifier-free. Let $F(x) = \exists z_1 \ldots z_n F_0$, where F_0 is quantifier-free. Then

$$A^* \rightleftarrows \exists x q_1 y_1 \ldots q_k y_k \exists y_{k+1} \ldots \exists y_m \exists z_1 \ldots \exists z_n (F_0 \& B)$$
$$\equiv \exists x (F(x) \& A_1^*) \equiv \exists x (F x \& A_1') = A' .$$

Case 6. $A = \forall x A_1$, where $A_1 = q_1 y_1 \ldots q_k y_k A_2$, and A_2 is quantifier-free. By the induction hypothesis $A_1^* \equiv q_1 y_1 \ldots q_k y_k \exists y_{k+1} \ldots \exists y_m B$, where B is quantifier-free. Let $G(x) = \exists z_1 \ldots z_n G_0$, where G_0 is quantifier-free. Then

$$A^* \rightleftarrows \forall x q_1 y_1 \ldots q_k y_k \exists y_{k+1} \ldots \exists y_m \exists z_1 \ldots \exists z_n (G_0 \vee B)$$
$$\equiv \forall x (G(x) \vee A_1^*) \equiv \forall x (F(x) \supset A_1') = A' .$$

QED Lemma 1.

Lemma 2. There is an algorithm which transforms an arbitrary closed prenex formula A of the theory T1 into a closed prenex formula A^+ of the theory T2, such that

(a) $A^+ \equiv A' \& \exists x F(x)$ in the theory T2,

(b) If the formula A is of the prefix class $\Pi(w)$, then the formula A^+ is of the prefix class $\Pi(w\exists^\infty)$.

Proof. Let $F(x_0) = \exists x_1 \ldots \exists x_m F_0$ and $A^* = q_1 y_1 \ldots q_n y_n B$, with F_0 and B quantifier-free. Then the desired formula

$$A^+ \rightleftarrows q_1 y_1 \ldots q_n y_n \exists x_0 \ldots \exists x_m (B \& F_0)$$
$$\equiv A^* \& \exists x_0 F(x_0) .$$

QED Lemma 2.

In the interpreted theory T there are neither predicate nor operation variables. Thus $\Phi(w,all,all) = \Phi(w,0,0)$. We call this class simply $\Phi(w)$.

Corollary. If A is a formula of the class $\Phi(w)$ of the theory $T1$, then A^+ is a formula of the class $\Phi(w\exists^\infty)$ of the theory $T2$. In particular, if A is of the class $\Phi(\forall^r\exists^\infty)$ of the theory $T1$, then A^+ is of the class $\Phi(\forall^r\exists^\infty)$ of the theory $T2$.

Lemma 3. If the formula A is logically inconsistent, so is A^+ .

Proof. Obvious.

For brevity we call an interpreted theory T r-inseparable, if the set of numbers of logically inconsistent formulas of the class $\Phi(\forall^r\exists^\infty)$ of the theory T and the set of numbers of formulas of this class in T having finite models are effectively inseparable.

Lemma 4. If the theory $T1$ is r-inseparable and for each finitely satisfiable formula A in $T1$ of the class $\Phi(\forall^r\exists^\infty)$ of the theory $T1$, the formula A^+ is finitely satisfiable in $T2$, then the theory $T2$ is also r-inseparable.

Proof. See theorem 2, § 2 of chapter 1.

On the strength of lemma 2, we may replace A^{+} by $A' \ \& \ \exists x\, F(x)$ in the statement of lemma 4.

In the sequel we assume that the theory T1 is p-inseparable. And let the theory T2 be obtained from some theory T3 by adding a finite number of axioms $\beta_1, \ldots, \beta_n$ of the class $\Phi(\forall^{q}\exists^{\infty})$, where $p \leq q$. We explain "adding a finite number of axioms". A formula α of the theory T3 is a theorem of the theory T2 if and only if the formula $\beta_1 \ \& \ldots \& \ \beta_n \supset \alpha$ is a theorem of T3. Obviously:

<u>Lemma 5.</u> $\forall x_1 \ldots x_r \exists y_1 \ldots y_m A \ \& \ \forall x_1 \ldots x_r \exists z_1 \ldots z_n B \equiv$
$\forall x_1 \ldots x_r \exists y_1 \ldots y_m z_1 \ldots z_n (A \ \& \ B)$.

<u>Theorem 1.</u> If for each finitely satisfiable formula A in T1 of the class $\Phi(\forall^{p}\exists^{\infty})$ the formula $A' \ \& \ \exists x\, F(x)$ is finitely satisfiable in T2, then the theory T3 is q-inseparable.

<u>Proof.</u> We assume that the theory T1 is p-inseparable. On the strength of Lemma 4 and the conditions of the theorem (?) theory T2 is also p-inseparable. Let α be a formula of the class $\Phi(\forall^{p}\exists^{\infty})$ of the theory T2. On the strength of Lemma 5 the formula $\alpha \ \& \ \beta_1 \ \& \ldots \& \ \beta_n$ is logically equivalent to some formula α' of the class $\Phi(\forall^{q}\exists^{\infty})$. If α is logically inconsistent, so is α'. If α is finitely satisfiable in T2, then α' is finitely satisfiable in T3. It remains to apply Theorem 2, § 2,

Chapter 1.

QED Theorem 1.

Let N be a model for theory T2 . Assuming that in it $Pab \rightleftarrows C(a,b)$, we obtain a model N' for the language of the theory T1 . We denote by N^* that submodel (if it is nonempty) of the model N' , such that $|N^*| = \{a \in |N| : N \models F(a)\}$.

Theorem 2. If for each finite model M of the theory T1 we can find a finite model N of the theory T2 , such that N^* is isomorphic to M , then the theory T3 is q-inseparable.

Indeed, with the conditions of theorem 2, it is obvious that the conditions of theorem 1 are satisfied.

Theorem 2 is employed below for several corollaries.

The stated method can be applied in several directions. Instead of finite satisfiability one might be interested in ordinary satisfiability. Instead of inseparability one might speak simply of unsolvability. Instead of $\Phi(\forall^r \exists^\infty)$ one might be interested in other classes $\Phi(w \exists^\infty)$. In some cases one can get by without G . But more of t' s in another place.

§ 2. The theory of two dyadic predicates.

We here call a model $M = (|M|, P)$, where P is a dyadic predicate, a graph. The theory of graphs TG has a corresponding meaning. Put differently, it could be called the theory of a single dyadic predicate. This is an interpreted theory with a single predicate symbol P and without operation symbols. Since the class $\Phi(A\ E\ A\ E \ldots E\ ,\ 2^1)$ of the theory LP is a conservative reduction class,

Lemma 1. The class $\Phi(A\ E\ A\ E \ldots E)$ of the theory TG is a conservative reduction class.

We denote by T2P the theory of two dyadic predicates. The language of this theory is obtained from the language of the theory TG by adding a new dyadic predicate symbol Q . Every model of the language of the theory T2P is a model of the same theory T2P .

Theorem 1. The theory T2P is 3-inseparable.

Proof. Let

$$\alpha = \forall x \exists u \forall y \exists z_1 \ldots z_n\ A$$

be a formula of the class $\Phi(A\ E\ A\ E \ldots E)$ of the theory TG . Let

$$\beta = (\forall x \exists v\ Qxv)\ \&\ \forall xuy\ \exists z_1 \ldots z_n\ (Qxu \supset A)\ ,$$

$$\gamma = \forall xuy\ \exists vz_1 \ldots z_n\ [Qxv\ \&\ Qxu \supset A)]\ .$$

Obviously, formulas α and β are satisfiable in exactly the same universes, and the formulas β and γ are equivalent in T2P . It remains to refer to Theorem 2, § 2, chapter 1.

§ 3. Theory of graphs.

In this § we employ the following abbreviations:

$Pxy = 0$	for	$\neg Pxy \;\&\; \neg Pyx$
$Pxy = 1$	for	$Pxy \;\&\; \neg Pyx$
$Pxy = 2$	for	$\neg Pxy \;\&\; Pyx$
$Pxy = 3$	for	$Pxy \;\&\; Pyx$
$Pxy = Puv$	for	$(Pxy \sim Puv) \;\&\; (Pyx \sim Pvu)$

We call a graph M economical , if for each pair a,b of distinct elements of M there is in M an element c , such that $Pac \neq Pbc$.

__Theorem.__ The theory of graphs TG is 3-inseparable.

__Proof__, carried out with the aid of an existential interpretation of the theory T2P of the previous § in the Theory TG . Let

$$\beta_1 \rightleftarrows \forall xyz\, (Pxx \;\&\; Pyy \supset Pxz = Pyz) .$$

In economical models of the formula β_1 there is not more than one self-tied element, i.e., not more than one a such that Paa holds.

For each $\mu = 0,1,2,3$ we make the abbreviation $fx = \mu$ for $\exists u(\neg Pxx \;\&\; Puu \;\&\; Pxu = \mu)$. An element a of a graph M such that $fa = \mu$ we call an element of the type μ . If in a model of the formula

β_1 there is a selftied element, then each element of this model which is not selftied has a well-defined type. Hereafter, let

$\beta_2 \rightleftharpoons \forall x \exists u_1 u_2 u_3$ $(fx = 0 \supset \bigwedge_{1\leq\mu\leq 3} (fu_\mu = \mu \,\&\, Pxu_\mu))$,

$\beta_3 \rightleftharpoons \forall xyz \bigwedge_{1\leq\mu\leq 3} (fx = 0 \,\&\, fy = fz = \mu \,\&\, Pxy \,\&\, Pxz \supset \neg Pyz)$,

$\beta_4 \rightleftharpoons \forall xyz \bigwedge_{1\leq\mu\leq 3} (fx = fy = \mu \,\&\, \neg Pxy \supset Pxz = Pyz)$.

In accordance with § 1 we denote by T2 the theory which is yielded within the theory TG by axioms β_1-β_4 . In economical models of the theory T2 for each element a of type 0 and for each $\mu = 1,2,3$ there is a unique element b of type μ such that Pab . To this one, various elements of type μ are tied. Finally, let

$F(x) \rightleftharpoons (fx = 0)$,

$G(x) \rightleftharpoons Pxx \vee fx = 1 \vee fx = 2 \vee fx = 3$,

$C_p(x,y) \rightleftharpoons \exists uv \; (fu = 1 \,\&\, Pxu \,\&\, fv = 2 \,\&\, Pyv \,\&\, Puv)$,

$D_p(x,y) \rightleftharpoons \exists uv \; (fu = 1 \,\&\, Pxu \,\&\, fv = 2 \,\&\, Pyv \,\&\, \neg Puv)$,

$C_Q(x,y) \rightleftharpoons \exists uv \; (fu = 1 \,\&\, Pxu \,\&\, fv = 3 \,\&\, Pyv \,\&\, Puv)$,

$D_Q(x,y) \rightleftharpoons \exists uv \; (fu = 1 \,\&\, Pxu \,\&\, fv = 3 \,\&\, Pyv \,\&\, \neg Puv)$.

The last five formulas do not appear to be existential, but they have become such after substituting with preceeding formulas. Clearly, in the theory T2 , $\neg F(x) \equiv G(x)$, and $Fx.Fy \supset (\neg C_p\, xy \sim D_p\, xy)$ and $Fx.Fy \supset (\neg C_Q\, xy \sim D_Q\, xy)$ hold.

It remains to verify the conditions of Theorem 2, § 1. Let M be a finite model of the theory T2P . We construct a graph N .

$|N| \rightleftharpoons |M| \times \{0,1,2,3\}$,

$P[(a,0),\ (b,\mu)] \rightleftharpoons (a = b \ \&\ \mu \neq 0)$,

$P[(a,\mu),\ (b,\mu)] \rightleftharpoons (a \neq b)$,

$P[(a,1),\ (b,2)] \rightleftharpoons Pab$,

$P[(a,1),\ (b,3)] \rightleftharpoons Qab$.

Otherwise, P is arbitrary in N .

We construct an expansion N of the graph N : $|N_1| \rightleftharpoons |N| \cup \{e\}$, where e is some new element. Pee holds, and for each $(a,\mu) \in |N|$, $P[(a,\mu),\ e] = \mu$ holds.

Clearly the model N_1^* (see § 1) is isomorphic to M .

QED Theorem.

In an obvious way one obtains from the theorem,

Corollary. The class $\Phi(A^3 E \ldots E)$ of the theory LP is a conservative reduction class.

§ 4. Theory of unoriented graphs.

By TNG we denote the theory obtained from the theory of graphs TG by adding the axiom $\forall xy\ (Pxy \sim Pyx)$.

<u>Theorem</u>. The theory TNG is 3-inseparable.

<u>Proof</u>, with the aid of an existential interpretation of the theory of graphs TG in the theory TNG . Let

$\beta_1 \rightleftarrows \forall xyz\, [Pxx.\ Pyy.\ \ Pxy \supset (Pxz \sim Pyz)]$,

$\beta_2 \rightleftarrows \forall xyz\, [Pxx.\ Pyy.\ Pzz \supset Pxy \vee Pyz \vee Pzx]$.

We make the abbreviations:

$fx = 0$	for	$\exists uv\, (Puu.\ Pvv.\ \neg Puv.\ \neg Pxu.\ \neg Pxv)$,
$fx = 1$	for	$\exists uv\, (Puu.\ Pvv.\ \neg Puv.\ \neg Pxu.\ \ Pxv)$,
$fx = 2$	for	$\exists uv\, (Puu.\ Pvv.\ \neg Puv.\ Pxu.\ Pxv)$.

Moreover, we let

$\beta_3 \rightleftarrows \forall x \exists u_1 u_2\, [fx = 0 \supset \bigwedge_{\mu=1}^{2} (fu_\mu = \mu\ \&\ Pxu_\mu)]$

$\beta_4 \rightleftarrows \forall xyz \bigwedge_{\mu=1}^{2} [fx = 0.\ fy = fz = \mu.\ Pxy.\ Pxz \supset \neg Pyz]$,

$\beta_5 \rightleftarrows \forall xyz \bigwedge_{\mu=1}^{2} [fx = fy = \mu\ .\ \neg Pxy \supset (Pxz \sim \neg Pyz)]$

Let the theory T2 be given within the theory TNG by the axioms β_1-β_5 . We explain these axioms. Let M be an economical (cf. § 3) model for theory T2 . On the strength of β_1 , the various selftied elements of M are not tied among themselves. On the strength of β_2 , there are no more than two different selftied elements of M . Moreover we assume that in M there actually are distinct selftied elements. We call an element a of the graph M such that $fa = \mu$ an element of type μ .

Clearly, each element of M which is not selftied has a well-defined type. On the strength of axioms β_3-β_5 each element of type 0 is tied, for each $\mu = 1,2$, to exactly one element of type μ, and various elements of type μ are tied among themselves.

Finally, let

$F(x) \rightleftharpoons (fx = 0)$,

$G(x) \rightleftharpoons Pxx \vee fx = 1 \vee fx = 2$,

$C(x,y) \rightleftharpoons \exists uv\ [fu = 1.\ Pxy.\ fv = 2.\ Pyv.\ Puv]$

$D(x,y) \rightleftharpoons \exists uv\ [fu = 1.\ Pxu.\ fv = 2.\ Pyv.\ \neg Puv]$.

The last three formulas become existential after substitution by the preceding form. Clearly, in the theory T2 $\neg F(x) \sim G(x)$, $Fx.Fy \rightarrow (\neg Cxy \sim Dxy)$. The conditions of theorem 2 § 1 are also clear. Therefore the theory TNG is 3-inseparable.

QED Theorem.

§ 5. Other theories.

The theory of unoriented graphs is naturally called the theory of a symmetric predicate. According to the cited article of the author [1965], the theory of a reflexive predicate is 3-inseparable, the theory of a symmetric and reflexive predicate is 6-inseparable, the theory of metabelian groups is 6-inseparable, and the theories of a partial order and of a lattice are r-inseparable for some r . Since then the author has obtained further results in this area and in time hopes to publish them.