Notre Dame Journal of Formal Logic
Volume 28, Number 2, April 1987

Invertible Definitions

TIMOTHY WILLIAMSON*

Introduction A concept of informational equivalence between relations is explicated to generalize some suggestions by Geach. It is shown that two relations are informationally equivalent if and only if each can be defined in terms of the other without the use of quantifiers. It is shown that there is a general method for listing the j-place relations informationally equivalent to an arbitrary given i-place relation if and only if $i \leq j$. The equivalence classes of the relation of informational equivalence are characterized as the invariants of the group of invertible quantifier-free definitions, for $i = j$.

Quantifier-free definition is contrasted with general first-order definition by means of an example of two first-order interdefinable relations which are not interdefined by any pair of mutually inverse first-order definitions.

1 Geach ([3], pp. xi–xii, 25–26, 33, 52) has suggested that one property or relation may be so closely connected to another that to have the concept of either is to have the concept of both. For example, to know what is *red* is to know what is *not red*, and to know what is *to the left of* what is to know what is *to the right of* what. If a relation is the same as its contradictory and its converse in this respect, it is presumably also, by transitivity, the same as the contradictory of its converse in the same respect. Since Geach claims to refute some theories of concept acquisition and possession on the basis of their neglect of this equivalence relation, a formal characterization of it would seem desirable, so that a general principle covering his remarks can be stated and tested.

Geach's suggestion is plausible if concepts are thought of as discriminative capacities. For vividness, imagine an ω-sequence of labeled objects to which we have no direct access, but about which we can gather information via a robot (the discriminative capacity). Suppose that we can use the robot to discover whether the object with a given label is red (when the instruction 'Red' is input, followed by the label, the robot moves along the line until it finds the object with

*I thank the anonymous referee for helpful comments on an earlier version.

Received March 1, 1985; revised November 29, 1985

that label, and on its return flashes a light if and only if that object was red): then *ipso facto* we can use the robot to discover whether the object with a given label is not red (similarly with 'red' and 'not red' reversed). 'Red' and 'not red' correspond to the same discriminative capacity. Again, suppose that for any labels a and b we can use the robot to discover whether the object labeled a is to the left of the object labeled b, by inputting the instruction 'Left', followed by a followed by b: then for any labels a and b we can use the robot to discover whether the object labeled a is to the right of the object labeled b, by inputting the instruction 'Left', followed by b followed by a (similarly with 'left' and 'right' reversed). 'Left' and 'right' correspond to the same discriminative capacity.

To generalize these examples, we need to define a relation—call it *informational equivalence*—which applies to pairs of converses and contradictories, such that a philosopher (whom I shall call "Geach") might plausibly claim that if R is informationally equivalent to S then to have the concept of R is to have the concept of S. Here are some *desiderata* for the definition of informational equivalence[1]:

(i) Geach's suggestion would become unnecessarily tendentious if we made first-order interdefinability entail informational equivalence. For consider the first-order interdefinable predicates Fx and $Fx \equiv (\exists yFy \;\&\; \exists y \neg Fy)$. In the example above, let the domain of quantification be the set of all objects in the ω-sequence, F be true of all and only the red ones and "Red" be the robot's only instruction. Suppose that, unknown to us, *all* the objects are red. Then, by means of the robot, we shall discover that $\exists yFy$ is true, but we shall never discover that $\exists y \neg Fy$ is false, so we shall never know the extension of $Fx \equiv (\exists yFy \;\&\; \exists y \neg Fy)$. Since we can use the robot to discover the extension of one property but not of the other in a finite way, it is not arbitrary to rule that they are not informationally equivalent. Moreover, even for a finite domain, we could know the extension of one over a proper subset of the domain without knowing the extension of the other. (For a generalization of (i), cf. (iii).)

(ii) However, Geach would lose one of his examples if we made the informational equivalence of two i-place relations entail that the extension of each could be recovered from the extension of the other over every set of i-tuples. For consider the singleton set $\{(x,y)\}$. If one is told only that (x,y) is *not* in the extension of *to the left of*, one cannot tell whether it is in the extension of *to the right of* (x may be on top of y). What we can make the informational equivalence of two i-place relations entail is that for every set of individuals, the extension of each relation could be recovered from the extension of the other over every set of i-tuples of members of that set. For example, if one is told the extension of *to the left of* over a set $\{(x,x), (x,y), (y,x), (y,y)\}$, one can deduce the extension of *to the right of* over that set.

(iii) Consider the predicates Fx and $Fx \equiv P$, where P is a closed formula whose truth value is contingent. Although in any given world $Fx \equiv P$ is extensionally equivalent either to $\neg Fx$ or to Fx itself, we should not treat Fx and $Fx \equiv P$ as informationally equivalent, since one cannot tell *how* to determine the extension of one from that of the other if one does not happen to know the truth value of P.

(iv) We should allow an i-place relation to be informationally equivalent to a j-place relation even when $i \neq j$. For example, suppose that Rxy is defined by Gx & Gy, so that Gx is logically equivalent to Rxx. Then, for any set, if we knew the extension of Gx over that set we could deduce the extension of Rxy over all ordered pairs of members of that set and *vice versa*.

We shall treat an i-place relation as a function from possible worlds to sets of i-tuples (a property is a one-place relation). Some relations in an intuitive sense (such as serial ordering) may not naturally correspond to a particular set of i-tuples in given circumstances; however, since possible worlds figure in the mathematics below only as indices, they can be regarded as including arbitrary stipulations for the extensions of such indeterminate relations (different stipulations will give different worlds). In any case, the results below can be interpreted in a more syntactic fashion, so that they are not as different as they look from the work of Beth and his successors on definability (see [4]). Think of the possible worlds as the models of a first-order theory T. Given an ordering of the variables of the language, each open formula in which at most the first i variables are free determines a function from models to sets of i-tuples in the obvious way. Two open formulas determine the same function if and only if they are satisfied by the same sequences in all models of T, hence if and only if they are provably equivalent in T; such relations can thus be thought of as the elements of a Lindenbaum algebra (cf. [2]). Since the definitions with which we shall be concerned are first-order, for each result below a corresponding result can be stated in syntactic terms.

Now we can define a sense in which R and G from *desideratum* (iv) necessarily determine each other's extensions as follows. For any n, suppose that the sequences $(x_0,\dots,x_{n-1})$ and $(y_0,\dots,y_{n-1})$ are alike with respect to R, in the sense that (x_i,x_j) is in the extension of R when and only when (y_i,y_j) is, for $i,j<n$; then $(x_0,\dots,x_{n-1})$ and $(y_0,\dots,y_{n-1})$ are alike with respect to G, in the sense that x_i is in the extension of G when and only when y_i is, for $i<n$. In virtue of this entailment we can say that R *determines* G. Similarly, the converse entailment holds, and in virtue of it we can say that G determines R. This definition can be generalized; if we then say that two relations are informationally equivalent if and only if each determines the other, it turns out that we meet all four *desiderata*. Note that, in order to meet *desideratum* (iii), we allow $x_0,\dots,x_{n-1}$ to be drawn from one world and $y_0,\dots,y_{n-1}$ from another: that is, the informational equivalence of two relations requires that the way in which one determines the other to be invariant over possible worlds, for the receiver of information may not know which possible world it is in. Before we give the general definition we require some technical apparatus.

$P(X)$ is the power set of X. Card (X) is the cardinality of X. $X\times Y$ is the Cartesian product of X and Y. Y^X is the set of functions from X to Y. i and j are finite ordinals, identified with natural numbers (thus $2=\{0,1\}$, etc.). The variables w, w', etc., range over a class whose elements are called the *possible worlds*. For each, w, D_w is a set whose elements are called the *individuals in w*. An i-place relation R is treated as a function taking each w to a subset R_w of D_w^i. e_i is the identity function on i and 1_i is the identity function on the class of all i-place relations. If f and g are functions, fg is g followed by f, $dom(f)$ is the domain of f and $ran(f)$ its range. The variable I can range over all sets, but in cases of interest is a finite ordinal.

If R and S are i- and j-place relations respectively, R *determines* S iff:

$$\forall I \forall w \forall w' \forall x \in D_w^I \forall y \in D_{w'}^I,$$
$$(\{f: f \in I^i \ \&\ xf \in R_w\} = \{f: f \in I^i \ \&\ yf \in R_{w'}\}$$
$$\rightarrow \{f: f \in I^j \ \&\ xf \in S_w\} = \{f: f \in I^j \ \&\ yf \in S_{w'}\}).$$

Determination is obviously reflexive and transitive. Since $\{f: f \in I^i \ \&\ xf \in D_w^i - R_w\} = I^i - \{f: f \in I^i \ \&\ xf \in R_w\}$, R determines $\neg R$. Since $\{f: f \in I^2 \ \&\ xf \in \breve{R}_w\} = \{g: g \in I^2 \ \&\ \exists h \in \{f: f \in I^2 \ \&\ xf \in R_w\}(g(0) = h(1) \ \&\ g(1) = h(0))\}$, R determines its converse $\breve{R}$. Thus the definition captures Geach's examples.

The proposal is to generalize Geach's original suggestion to all pairs of relations which are informationally equivalent in the sense of determining each other. We shall now show that one relation determines another if and only if the latter can be defined in terms of the former by use only of variables (that is, the permutation and identification of arguments) and truth functions. Hence the proposed generalization is in effect to all pairs of relations which are first-order interdefinable without the use of quantifiers. Although *desideratum* (i) is thereby met, Geach's suggestion may seem to be threatened when it is thus generalized. In order to know what it is for x to love y, does one really need to know what it is for x to love y if and only if z does not love itself? The condition hardly seems to be necessary if one thinks of the subject's ability to manifest understanding verbally: but Geach's original examples are already under threat at this level, since it seems that a child could master the predicate 'is red' without having mastered a sign for negation. Geach's claim is best defended for discriminative capacities, at which level the proposed generalization is not obviously wrong. We commend this issue to the reader, and turn to the proof that mutual determination and quantifier-free interdefinability are equivalent.

We now require a formal representation of definitions which use only truth functions and variables. Let $A \subseteq P(j^i)$. Define a function M_A taking subsets of D_w^i to subsets of D_w^j, for all w, by:

$$M_A(X) = \{x: x \in D_w^j \ \&\ \{f: f \in j^i \ \&\ xf \in X\} \in A\}.$$

Note that M_A is implicitly relativized to i and j; in practice this causes no confusion. A function M taking i-place relations to j-place relations is an *i, j-satisfaction function* iff for some $A \subseteq P(j^i)$, for all R and w $M(R)_w = M_A(R_w)$. If R is i-place, think of each $f \in j^i$ as taking variables from '$M_A(R)x_0 \ldots x_{j-1}$' to give $p(f) =$ '$Rx_{f(0)} \ldots x_{f(i-1)}$'; think of each $X \subseteq j^i$ as the conjunction $c(X)$ of $p(f)$ for each $f \in X$ and $\neg p(f)$ for each $f \in j^i - X$; think of each $A \subseteq P(j^i)$ as the disjunction of $c(X)$ for each $X \in A$. Thus A codes a definition of $M_A(R_w)$ which uses only truth functions and variables, in disjunctive normal form. Conversely, any definition which uses only truth functions and variables can be coded in this way. For example, $A = \{X: X \subseteq i^i \ \&\ e_i \notin X\}$ codes negation on i-place relations and if $a \in 2^2$ is given by $a(0) = 1$, $a(1) = 0$, $A = \{X: X \subseteq 2^2 \ \&\ a \in X\}$ codes conversion.

We now show that mutual determination and interdefinability by satisfaction functions are equivalent.

Proposition 1 *If R is i-place and S j-place, R determines S iff for some i, j-satisfaction function M, M(R) = S.*

Proof: Left to right: Define a subset of $P(j^i)$ by:

$$A = \{X: \exists w \exists x \in S_w(X = \{f: f \in j^i \ \&\ xf \in R_w\})\}.$$

If $x \in S_w$ then $\{f: f \in j^i \ \&\ xf \in R_w\} \in A$, so $x \in M_A(R_w)$. Thus $S_w \subseteq M_A(R_w)$.
Conversely, let $x \in M_A(R_w)$, so for some w' and $y \in S_{w'}$

$$\{f: f \in j^i \ \&\ xf \in R_w\} = \{f: f \in j^i \ \&\ yf \in R_{w'}\}.$$

Hence by definition of 'R determines S' for $I = j$,

$$\{f: f \in j^j \ \&\ xf \in S_w\} = \{f: f \in j^j \ \&\ yf \in S_{w'}\}.$$

Thus $xe_j \in S_w$ iff $ye_j \in S_{w'}$; but $ye_j(=y) \in S_{w'}$ so $x(=xe_j) \in S_w$ so $M_A(R_w) \subseteq S_w$. Thus $M_A(R_w) = S_w$.

Right to left: Suppose that $A \subseteq P(j^i)$, for all w $M(R)_w = M_A(R_w) = S_w$, for some w and w' $x \in D^I_w$, $y \in D^I_{w'}$ and

$$\{f: f \in I^i \ \&\ xf \in R_w\} = \{f: f \in I^i \ \&\ yf \in R_{w'}\}.$$

For $f \in I^j$ and $g \in j^i$, $fg \in I^i$, so $xfg \in R_w$ iff $yfg \in R_{w'}$. Hence for $f \in I^j$

$$\{g: g \in j^i \ \&\ xfg \in R_w\} = \{g: g \in j^i \ \&\ yfg \in R_{w'}\}$$

so

$$\{g: g \in j^i \ \&\ xfg \in R_w\} \in A \text{ iff } \{g: g \in j^i \ \&\ yfg \in R_{w'}\} \in A.$$

Since $xf \in D^j_w$, $yf \in D^j_{w'}$, $xf \in M_A(R_w)$ iff $yf \in M_A(R_{w'})$, so $xf \in S_w$ iff $yf \in S_{w'}$. Thus

$$\{f: f \in I^j \ \&\ xf \in S_w\} = \{f: f \in I^j \ \&\ yf \in S_{w'}\}.$$

Hence R determines S.

Proposition 1 could also be extended to the case where S is defined in terms of several relations $R, R', \ldots$. That is, if $R, R', \ldots$ and S are i-, i'-, $\ldots$ and j-place relations respectively, $\{R, R', \ldots\}$ *collectively determines* S iff:

$$\begin{aligned}&\forall I \forall w \forall w' \forall x \in D^I_w \forall y \in D^I_{w'}\\ &\quad((\{f: f \in I^i \ \&\ xf \in R_w\} = \{f: f \in I^i \ \&\ yf \in R_{w'}\} \ \&\\ &\quad\{f: f \in I^{i'} \ \&\ xf \in R'_w\} = \{f: f \in I^{i'} \ \&\ yf \in R'_{w'}\} \ \&\ldots)\\ &\quad\rightarrow \{f: f \in I^j \ \&\ xf \in S_w\} = \{f: f \in I^j \ \&\ yf \in S_{w'}\}).\end{aligned}$$

The corresponding enlargement of the definition of a satisfaction function would be, for $A \subseteq P(j^i) \times P(j^{i'}) \times \ldots$, to define M_A taking arguments $Z \times Z' \times \ldots$, where $Z \subseteq D^i_w$, $Z' \subseteq D^{i'}_w, \ldots$, to subsets of D^j_w, for all w, by:

$$\begin{aligned}M_A(Z \times Z' \times \ldots) = \{x: x \in D^j_w \ \&\ &\{f: f \in j^i \ \&\ xf \in Z\} \times\\ &\{f: f \in j^{i'} \ \&\ xf \in Z'\} \times \ldots \in A\}.\end{aligned}$$

These definitions would also allow certain privileged relations to be used as fixed parameters. For instance, if one treated identity as a logical constant one might be interested in the equivalence relation which holds between R and S iff each is definable in terms of the other using only truth functions, variables, and identity: in other words, when $\{R, =\}$ collectively determines S and $\{S, =\}$ collectively determines R.

Proposition 2 *If R is i-place and S is j-place, R determines S and S determines R iff for some i,j-satisfaction function M and some j,i-satisfaction function N, $M(R) = S$ and $N(S) = R$.*

Proof: Immediate from Proposition 1.

Since satisfaction functions make the generalization of Geach's idea more perspicuous, we list some of their elementary properties. We first show the coding of i,j-satisfaction functions by subsets of $P(j^i)$ to be unique except for small domains.

Proposition 3 *If $A, B \subseteq P(j^i)$, $M_A = M_B$ and for some w either $j \leq card(D_w)$ or $i = 0$ or both $i < card(D_w)$ and $i = 1$, then $A = B$.*

Proof: Let $I \subseteq j^i$. For $j \leq card(D_w)$, choose $x \in D^j_w$ to be $1-1$; for $i = 0$ choose any $x \in D^j_w$; for $1 < card(D_w)$ and $i = 1$, take $a, b \in D_w$ ($a \neq b$) and define $x \in D^j_w$ by: for $k \in j$, if the function $g \in j^i$ such that $g(0) = k$ is in I, let $x(k) = a$, otherwise let $x(k) = b$. In all cases, let $R_w = \{y: y \in D^i_w \;\&\; \exists g \in I(y = xg)\}$. Hence for $f \in j^i$, $xf \in R_w$ iff for some $g \in I$ $xf = xg$. Thus if $f \in I$, $xf \in R_w$. Conversely, let $xf \in R_w$: if x is $1-1$, $xf = xg$ ($g \in I$) entails $f = g$, so $f \in I$; if $i = 0$, $g \in I$ entails $f = g$, so $f \in I$; for $1 < card(D_w)$ and $i = 1$, $xf = xg$ ($g \in I$) entails that $f \in I$ iff $g \in I$ by the choice of x, so $f \in I$. Thus $\{f: f \in j^i \;\&\; xf \in R_w\} = I$. But $M_A(R_w) = M_B(R_w)$, so $\{f: f \in j^i \;\&\; xf \in R_w\} \in A$ iff $\{f: f \in j^i \;\&\; xf \in R_w\} \in B$, so $I \in A$ iff $I \in B$. Since I was arbitrary, $A = B$.

Proposition 4 *If for all w $card(D_w) < j$ and $1 < i$, there exist $A, B \subseteq P(j^i)$ such that $A \neq B$ but $M_A = M_B$.*

Proof: Let $A = \{\ \}$, $B = \{\{f: f \in j^i \;\&\; \forall m, n \in i(f(m) = f(n))\}\}$. Exercise: $M_A = M_B$.

We now list some closure properties of satisfaction functions.

Proposition 5 1_i is an i,i-satisfaction function.

Proof: Let $A = \{X: X \subseteq i^i \;\&\; e_i \in X\}$. Hence for $R_w \subseteq D^i_w$, $x \in M_A(R_w)$ iff $\{f: f \in i^i \;\&\; xf \in R_w\} \in A$ iff $e_i \in \{f: f \in i^i \;\&\; xf \in R_w\}$ iff $xe_i \in R_w$ iff $x \in R_w$. Thus $M_A(R_w) = 1_i(R)_w$.

Proposition 6 *If M is an i,j-satisfaction function and N is a j,k-satisfaction function, then NM is an i,k-satisfaction function.*

Proof: Let $M(R)_w = M_A(R_w)$ and $N(R)_w = M_B(R_w)$, for all R and w, where $A \subseteq P(j^i)$ and $B \subseteq P(k^j)$. Define

$$C = \{X: X \subseteq k^i \;\&\; \{f: f \in k^j \;\&\; \{g: g \in j^i \;\&\; fg \in X\} \in A\} \in B\}.$$

Hence for $R_w \subseteq D^i_w$ and $x \in D^k_w$,

$x \in M_C(R_w)$ iff
$\{h: h \in k^i \;\&\; xh \in R_w\} \in C$ iff

$\{f: f \in k^j \;\&\; \{g: g \in j^i \;\&\; fg \in \{h: h \in k^i \;\&\; xh \in R_w\}\} \in A\} \in B$ iff
$\{f: f \in k^j \;\&\; \{g: g \in j^i \;\&\; xfg \in R_w\} \in A\} \in B$ iff
$\{f: f \in k^j \;\&\; xf \in M_A(R_w)\} \in B$ iff
$x \in M_B(M_A(R_w))$.

Hence $M_C = M_B M_A$.

Proposition 7 *If M is an i,i-satisfaction function and $MN = 1_i$ or $MN = 1_i$ then N is an i,i-satisfaction function.*

Proof: By Propositions 5 and 6, the iterations of $M(M^0, M^1, M^2, \ldots$, i.e. 1_i, M, $MM, \ldots)$ can be indexed by elements of $P(P(i^i))$, which is a finite set. Hence for some m, $n(1 \leq n)$, $M^{m+n} = M^m$, so $N^{m+1}M^{m+n} = N^{m+1}M^m$ or $M^{m+n}N^{m+1} = M^m N^{m+1}$, so $M^{n-1} = N$, so by Propositions 5 and 6 N is an i,i-satisfaction function.

2 When a concept has been defined, one wants some idea of its extension. Now the extension of the concept of informational equivalence may include, for any given i and j, infinitely many pairs (R, S), where R and S are i- and j-place relations respectively (for instance, infinitely many pairs of converses). However, by Proposition 1 and our definition, for each such pair there is an i,j-satisfaction function M for which $M(R) = S$, and the number of i,j-satisfaction functions, for any particular values of i and j, is finite (each function has a distinct index in the finite set $P(j^i)$). Thus it is natural to seek a finite representation of the infinite extension of the concept of informational equivalence in terms of satisfaction functions. More precisely, suppose that there are i,j-satisfaction functions M, M', M'', ... such that, for every i-place relation R, $M(R)$, $M'(R)$, $M''(R)$, ... is a complete list of the j-place relations informationally equivalent to R; then the finite list M, M', M'', ... is a perspicuous and general answer to the question: to which j-place relations is an i-place relation informationally equivalent? It will be shown that such a list exists whenever $i \leq j$.

We can sharpen the issue by means of an observation. Suppose that the i,j-satisfaction function M appears on such a list. Then $M(R)$ is always informationally equivalent to R, so for every i-place relation R there is a j,i-satisfaction function N such that $N(M(R)) = R$. We can argue that N is independent of the choice of R, as follows. Since satisfaction functions are definable in terms of variables and truth functions, for any j,i-satisfaction function N there is a first-order sentence $A(M,N)$ in prenex normal form with only universal quantifiers and whose only atomic predicate is i-place such that for any i-place relation R the extensions of R and $N(M(R))$ coincide in a possible world w if and only if (D_w, R_w) is a model of $A(M,N)$. Suppose, for a contradiction, that for no j,i-satisfaction function N is $A(M,N)$ universally valid: then it is an easy exercise to show, from the form of such sentences, that there is a model (X, Y), whose domain X is countable, such that for every j,i-satisfaction function N $A(M,N)$ is false in (X, Y). Hence if there is a relation R and a world w such that $(X, Y) = (D_w, R_w)$, there is no j,i-satisfaction function N such that $N(M(R)) = R$, contrary to our assumption. But such an R exists for any world w for which D_w is countably infinite. Thus we are warranted in assuming that $M(R)$ is informationally equivalent to R for every i-place relation R only on condition that $A(M,N)$ is universally valid for some j,i-satisfaction function N, in which case $N(M(R)) = R$ for every i-place relation R. Such an N is a *left inverse* of M; in what follows, a left inverse will always be required to be a satisfaction function. Thus we seek i,j-satisfaction functions M, M', M'', . . . , each with a left inverse, such that, for every i-place relation R, $M(R)$, $M'(R)$, $M''(R)$, . . . is a complete list of the j-place relations informationally equivalent to R.

Trivially, if the i,j-satisfaction function M does have a left inverse, every i-place relation R is informationally equivalent to $M(R)$. Thus we can assume the list M, M', M'', . . . to include all i,j-satisfaction functions with left inverses. So our question becomes: has every j-place relation informationally equivalent to an i-place relation R the form $M(R)$ for some i,j-satisfaction function M with a left inverse?

Of course, if the j-place relation S is informationally equivalent to the i-place relation R, there is an i,j-satisfaction function M and a j,i-satisfaction function N such that $M(R) = S$ and $N(S) = R$. This, however, does not guarantee that N, or any other j,i-satisfaction function, is a left inverse of M (there may be another i-place relation T such that $N(M(T)) \neq T$). The question is whether there must be some i,j-satisfaction function M', perhaps different from M, where $M'(R) = S$ and M' does have a left inverse.

The problem does not arise for Geach's original examples, since the satisfaction functions which he uses already have left inverses, being self-inverse: every relation is the contradictory of its contradictory and every two-place relation is the converse of its converse. The following is an example in which the problem does arise. Let R and S be the two-place relations such that x has R to y if and only if x is red and y is red, while x has S to y if and only if x is red or y is red. Let M and N be the 2,2-satisfaction functions which take each two-place relation Txy to the relations $Txx \vee Tyy$ and $Txx \,\&\, Tyy$ respectively. Then $M(R) = S$ and $N(S) = R$, so R and S are informationally equivalent. However, M does not have a left inverse, for if T is any nontrivial reflexive relation, $M(T)$ will be the tautologous relation, and there will be no 2,2-satisfaction function N' such that $N'(M(T)) = T$. In this case, we can show that there is a 2,2-satisfaction function M' with a left inverse such that $M'(R) = S$, for we can define M' as taking each Txy to $Txy \equiv (Txx \equiv Tyy)$ (it is easy to show that $M'(R) = M(R)$ and that M' is self-inverse). Proposition 10 is simply the generalization of this example to any i-place relation R and j-place relation S, for $i \leq j$.

The extension of Proposition 10 to the case where $i > j$ is false. Consider again any nontrivial reflexive two-place relation: the only one-place relations which can be defined in terms of it using only variables and truth functions will be the tautologous and the self-contradictory ones, from which the original two-place relation will not be recoverable. Thus there is no 2,1-satisfaction function M such that, for every two-place relation R, $M(R)$ is informationally equivalent to R. In contrast, *desideratum* (iv) above gave an example of a two-place relation which was informationally equivalent to a one-place relation. Hence there are no 2,1-satisfaction functions M, M', M'', . . . such that, for every two-place relation R, $M(R)$, $M'(R)$, $M''(R)$, . . . is a complete list of the one-place relations to which R is informationally equivalent, although (by Proposition 10) there are 1,2-satisfaction functions M, M', M'', ... such that, for every one-place relation R, $M(R)$, $M'(R)$, $M''(R)$, ... is a complete list of the two-place relations to which R is informationally equivalent. Similar examples can be given whenever $i \neq j$. For $i < j$, there is a general answer to the question "To which j-place relations is an i-place relation informationally equivalent?", but no general answer to the question "To which i-place relations is a j-place relation informationally equivalent?". As Russell said of relations in a different context, 'You cannot reduce them downward, but you can reduce them upward' ([5], p. 58).

It should be noted that these lists of informational equivalents may contain repetitions. That is, we may have $M(R) = M'(R)$ even when M and M' are distinct satisfaction functions with left inverses. For example, the identity and the operation of taking converses are distinct self-inverse 2,2-satisfaction functions which give the same result when applied to a symmetric relation. *Any* list of functions M, M', M'', ... will have this feature if $M(R)$, $M'(R)$, $M''(R)$, ... are all and only the j-place relations informationally equivalent to R for every i-place relation R, since of two i-place relations one can be informationally equivalent to more j-place relations than is the other.

Proposition 9 (of which Proposition 10 is a corollary) shows that, in the important special case where $i = j$, the left inverse of Proposition 10 can be taken to be a two-sided inverse. That is, any two informationally equivalent i-place relations are mapped to each other by invertible i, i-satisfaction functions. Since (by Propositions 5, 6, and 7) the invertible i, i-satisfaction functions form a group, we can think of what informationally equivalent i-place relations have in common with each other as the invariants of this group. Proposition 9 also in effect shows that an apparently more cautious generalization of Geach's examples is in fact equivalent to our own, in the case $i = j$. For since the operations of taking the contradictory or the converse of a relation are invertible, one might have restricted his suggestion to pairs of relations which can be interdefined by quantifier-free definitions that invert each other everywhere. This criterion picks out the same pairs as our own (quantifier-free interdefinability).

When $i < j$, however, the left inverse of Proposition 10 cannot be taken to be two-sided. We saw above that a one-place relation R can be informationally equivalent to a two-place relation S even though no 2,1-satisfaction function has a left inverse. Now if M were a 1,2-satisfaction function with a two-sided inverse N such that $M(R) = S$, then N would be a 2,1-satisfaction function with a two-sided, and therefore left, inverse M, which is impossible.

The main work toward proving Propositions 9 and 10 is for a lemma, Proposition 8, about subsets of $P(i^i)$, the indices of i, i-satisfaction function (i is assumed to be fixed). For this we require some auxiliary notions. Invertible i, i-satisfaction functions will turn out to correspond to total *apt* functions, in a sense explained below.

We abbreviate e_i to e. The variables f, g, and h range over i^i and X, Y, Z, etc., range over $P(i^i)$. For any sets I and J, for $a \in I^i$ and $J \subseteq I^i$, $a \,\%\, J$ is $\{f\colon af \in J\}$. Thus $a \,\%\, J \subseteq i^i$, $af \,\%\, J = f \,\%\, a \,\%\, J$, $a \,\%\, (J \cup K) = (a \,\%\, J) \cup (a \,\%\, K)$, $a \,\%\, (J - K) = (a \,\%\, J) - (a \,\%\, K)$, etc., and if $J \subseteq i^i$, $e \,\%\, J = J$. An *apt* function is a partial or total 1-1 function $F\colon P(i^i) \to P(i^i)$ such that if $X \in dom(F)$, then for any f, $f \,\%\, X \in dom(F)$ and $F(f \,\%\, X) = f \,\%\, F(X)$. F/J is the restriction of F to the domain J.

Proposition 8 *If F is apt, there is an apt F'' such that $dom(F'') = P(i^i)$ and $F''/dom(F) = F$.*

Proof: (a) We show the existence of an apt F' such that $dom(F) \subseteq dom(F')$, $dom(F) \neq dom(F')$, and $F'/dom(F) = F$. Since $P(i^i)$ is finite, iteration of this procedure gives Proposition 8. We first define an order £ on $P(i^i)$ by

$$X \pounds Y \text{ iff } card(\{Z: \exists f(Z = f \% X)\}) \leq card(\{Z: \exists f(Z = f \% Y)\}).$$

£ is obviously reflexive, transitive, and connected. We can also define an equivalence relation ££ on $P(i^i)$ by: $X \pounds\pounds Y$ iff $X \pounds Y$ and $Y \pounds X$.

(b) If $Y \in dom(F)$, $Y \pounds\pounds F(Y)$. *Proof:* For $Y \in dom(F)$, since F is apt, $f \% Y = g \% Y$ iff $F(f \% Y) = F(g \% Y)$ iff $f \% F(Y) = g \% F(Y)$. Thus the function G given by $G(f \% Y) = f \% F(Y)$ is a 1-1 correspondence between $\{Z: \exists f(Z = f \% Y)\}$ and $\{Z: \exists f(Z = f \% F(Y))\}$.

(c) $P(i^i) - dom(F)$ can be assumed to be nonempty (otherwise we are done); hence it contains a £-minimal element U: for $Y \notin dom(F)$, $U \pounds Y$.

(d) If $Y \notin ran(F)$, $U \pounds Y$. *Proof:* Let E be the ££ -equivalence class of Y. Since F is 1-1, by (b) it induces a 1-1 correspondence between $E \cap dom(F)$ and $E \cap ran(F)$. Since E is finite, $card(E - dom(F)) = card(E - ran(F))$; but $Y \in E - ran(F)$, so $E - dom(F)$ is also nonempty. Choose $Z \in E - dom(F)$. By (c), $U \pounds Z$; since $Z \in E$, $Z \pounds Y$. By transitivity, $U \pounds Y$.

(e) We make the following definitions:

$$\begin{aligned}
W &= \{f: f \% U \in dom(F)\}\\
W' &= W \cap \{f: e \in F(f \% U)\}\\
A &= \{Y: \forall f \forall g(f \% U = g \% U \rightarrow f \% Y = g \% Y)\}\\
B &= \{Y: Y \cap W = U \cap W\}\\
B' &= \{Y: Y \cap W = W'\}\\
G(Y) &= (Y - W) \cup W'.
\end{aligned}$$

(f) G maps $A \cap B$ 1-1 into $A \cap B'$. *Proof:* G is 1-1 on $A \cap B$, for if Y, $Z \in A \cap B$ and $(Y - W) \cup W' = (Z - W) \cup W'$, $Y - W = Z - W$ (since $W' \subseteq W$) and $Y \cap W = U \cap W = Z \cap W$ (since $Y, Z \in B$); thus $Y = Z$. Let $Y \in A \cap B$. $G(Y) \in B'$, for since $W' \subseteq W$, $W \cap ((Y - W) \cup W') = W'$. To show $G(Y) \in A$, let $f \% U = g \% U$. Since $Y \in A$, $f \% Y = g \% Y$. Now for any h, $h \in f \% W$ iff $fh \in W$ iff $fh \% U \in dom(F)$ iff $h \% f \% U \in dom(F)$ iff $h \% g \% U \in dom(F)$ iff $h \in g \% W$; thus $f \% W = g \% W$. Similarly, $f \% W' = g \% W'$. Thus $f \% G(Y) = ((f \% Y) - f \% W) \cup f \% W' = ((g \% Y) - g \% W) \cup g \% W' = g \% G(Y)$.

(g) For $Y \in dom(F)$, $Y \in A$ iff $F(Y) \in A$. *Proof:* For $Y \in dom(F)$ and any f, g, $f \% Y$, $g \% Y \in dom(F)$. Hence $f \% Y = g \% Y$ iff $F(f \% Y) = F(g \% Y)$ iff $f \% F(Y) = g \% F(Y)$.

(h) For $Y \in dom(F)$, $Y \in B$ iff $F(Y) \in B'$. *Proof:* Note that if $f \in W$, $fg \in W$ for any g, because $f \% U \in dom(F)$, so $fg \% U = g \% f \% U \in dom(F)$. Left to right: Let $Y \in B$, so $Y \cap W = U \cap W$. Hence for any $f \in W$ and any g, $fg \in Y$ iff $fg \in U$, so $g \in f \% Y$ iff $g \in f \% U$, so $f \% Y = f \% U$. Hence $f \in F(Y)$ iff $fe \in F(Y)$ iff $e \in f \% F(Y)$ iff $e \in F(f \% Y)$ iff $e \in F(f \% U)$ iff $f \in W'$. Thus $W \cap F(Y) = W \cap W' = W'$, so $F(Y) \in B'$. Right to left: Let $F(Y) \in B'$, so $W \cap F(Y) = W'$. Hence for any $f \in W$ and any g, $g \in F(f \% Y)$ iff $g \in f \% F(Y)$ iff $fg \in F(Y)$ iff $fg \in W'$ iff $e \in F(fg \% U)$ iff $e \in F(g \% f \% U)$ iff $e \in g \% F(f \% U)$ iff $g \in F(f \% U)$; thus $F(f \% Y) = F(f \% U)$, so $f \% Y = f \% U$. Thus $f \in Y$ iff $e \in f \% Y$ iff $e \in f \% U$ iff $f \in U$, so $Y \cap W = U \cap W$.

(i) $(A \cap B')$-$ran(F)$ is nonempty. *Proof:* By (f), $card(A \cap B) \leq card(A \cap B')$. By (g) and (h), $card(A \cap B \cap dom(F)) = card(A \cap B' \cap ran(F))$. Trivially, $U \in (A \cap B) - dom(F)$. Since these are finite sets, $1 \leq card((A \cap B) - dom(F)) = card(A \cap B) - card(A \cap B \cap dom(F)) \leq card(A \cap B') - card(A \cap B' \cap ran(F)) = card((A \cap B') - ran(F))$.

(j) By (i) we can choose $V \in (A \cap B') - ran(F)$. We now define F' by:

$F'(Z) = F(Z)$ for $Z \in dom(F)$
$F'(f \% U) = f \% V$ for $f \% U \notin dom(F)$
F' is undefined in all other cases.

The second clause is legitimate because $V \in A$. Note that $U(=e \% U) \in dom(F') - dom(F)$. Thus $dom(F) \subseteq dom(F')$ and $dom(F) \neq dom(F')$. Trivially, $F'/dom(F) = F$. Thus it remains only to show that F' is apt.

(k) For any f, $F'(f \% U) = f \% V$. *Proof:* We assume $f \% U \in dom(F)$, else we are done. Then for any g, $g \in f \% V$ iff $fg \in V$ iff $fg \in W'$ (because $V \in B'$ and $f \in W$, so $fg \in W$) iff $e \in F(fg \% U)$ iff $e \in g \% F(f \% U)$ iff $g \in F(f \% U)$ iff $g \in F'(f \% U)$.

(l) F' is 1-1. *Proof:* Let $F'(X) = F'(Y)$. There are four cases.

(I) $X, Y \in dom(F)$. Then $F(X) = F'(Y) = F'(Y) = F(X)$, so $X = Y$.

(II) $X \in dom(F)$, $Y \notin dom(F)$. Thus for some h, $Y = h \% U$. Hence for any f, $f \% Y = f \% h \% U = hf \% U$. Thus $\{Z\colon \exists f(Z = f \% Y)\} \subseteq \{Z\colon \exists f(Z = f \% U)\}$. But $Y \notin dom(F)$, so by (c) U £ Y, i.e., $card(\{Z\colon \exists f(Z = f \% U)\}) \leq card(\{Z\colon \exists f(Z = f \% Y)\})$. Since the sets are finite, $\{Z\colon \exists f(Z = f \% Y)\} = \{Z\colon \exists f(Z = f \% U)\}$. Thus for some f, $e \% U = f \% Y = hf \% U$. Since $V \in A$, $V = e \% V = hf \% V$. Since $X \in dom(F)$, $f \% X \in dom(F)$, so $F(f \% X) = f \% F(X) = f \% F'(X) = f \% F'(Y) = f \% h \% V = hf \% V = V$. This contradicts $V \notin ran(F)$, so the case cannot arise.

(III) As (II), with X and Y reversed.

(IV) $X, Y \notin dom(F)$. Hence for some g, h, $X = g \% U$ and $Y = h \% U$. By (k), F' maps $\{Z\colon \exists f(Z = f \% U)\}$ onto $\{Z\colon \exists f(Z = f \% V)\}$. But $V \notin ran(F)$, so by (d) $card(\{Z\colon \exists f(Z = f \% U)\}) \leq card(\{Z\colon \exists f(Z = f \% V)\})$. Since these sets are finite, F' must be 1-1 on $\{Z\colon \exists f(Z = f \% U)\}$. Since $F'(g \% U) = F'(h \% U)$, $g \% U = h \% U$.

(m) F' is apt. *Proof:* It remains only to show that if $X \in dom(F')$ then for any f, $f \% X \in dom(F')$ and $F'(f \% X) = f \% F'(X)$. If $X \in dom(F)$, this follows from the corresponding property of F. Otherwise, for some g, $X = g \% U$. Hence $f \% X = f \% g \% U = gf \% U \in dom(F')$ and $F'(f \% X) = gf \% V = f \% g \% V = f \% F'(X)$.

Proposition 9 *Suppose that R,S are i-place and M,N are i,i-satisfaction functions such that $M(R) = S$, $N(S) = R$. Then there are i,i-satisfaction functions M', N' such that $M'N' = N'M' = 1_i$ and $M'(R) = S$, $N'(S) = R$.*

Proof: We translate the problem about satisfaction functions into a finite combinatorial one about their indices, which is solved by Proposition 8, and then translate the answer back again.

(a) The following defines an apt function F:
$F(x \% R_w) = x \% S_w$ for any w and $x \in D^i_w$
$F(X)$ is undefined if X is not $x \% R_w$ for any w and $x \in D^i_w$.

Proof: By Proposition 2, for $x \in D^i_w$ and $y \in D^i_{w'}$, $x \% R_w = y \% R_{w'}$ iff $x \% S_w = y \% S_{w'}$, so F is well-defined on its domain and 1-1. If $X \in dom(F)$, for some w and $x \in D^i_w$ $X = x \% R_w$, so for any f, $f \% X = f \% x \% R_w = xf \% R_w \in dom(F)$, and $F(f \% X) = F(xf \% R_w) = xf \% S_w = f \% x \% S_w = f \% F(X)$. Thus F is apt.

(b) By Proposition 8, there is an apt function F'' such that $dom(F'') = P(i^i)$ and $F''/dom(F) = F$. Since F'' is 1-1 and its domain is finite, it has a two-sided inverse G''. Moreover, G'' is apt too, since it is 1-1, total and for any X and f, $F''(f \% G''(X)) = f \% F''(G''(X)) = f \% X = F''(G''(f \% X))$, so $f \% G''(X) = G''(f \% X)$.

(c) We define subsets of $P(i^i)$ by:

$$A = \{X \colon e \in F''(X)\}$$
$$B = \{X \colon e \in G''(X)\}.$$

For any w, $x \in D^i_w$ and $I \subseteq D^i_w$, $x \in M_A(I)$ iff $x \% I \in A$ iff $e \in F''(x \% I)$ and $x \in M_B(I)$ iff $x \% I \in B$ iff $e \in G''(x \% I)$, by definition. We shall have $M'(T)_w = M_A(T_w)$ and $N'(T)_w = M_B(T_w)$, for all w and T.

(d) For $x \in D^i_w$ and $I \subseteq D^i_w$, $x \% M_A(I) = F''(x \% I)$ and $x \% M_B(I) = G''(x \% I)$. *Proof:* For any f, $f \in x \% M_A(I)$ iff $xf \in M_A(I)$ iff $e \in F''(xf \% I)$ iff $e \in F''(f \% x \% I)$ iff $e \in f \% F''(x \% I)$ iff $f \in F''(x \% I)$. The other case is parallel.

(e) For $I \subseteq D^i_w$, $M_A M_B(I) = M_B M_A(I) = I$. *Proof:* For $X \in D^i_w$ $x \in M_A(M_B(I))$ iff $e \in F''(x \% M_B(I))$ iff (by (d)) $e \in F''(G''(x \% I))$ iff $e \in x \% I$ iff $x \in I$. Similarly, $x \in M_B(M_A(I))$ iff $x \in I$.

(f) $M_A(R_w) = S_w$, $M_B(S_w) = R_w$. *Proof:* For $x \in D^i_w$, $F(x \% R_w) = F''(x \% R_w) = x \% S_w$ so $x \in M_A(R_w)$ iff $e \in F''(x \% R_w)$ iff $e \in F(x \% R_w)$ iff $e \in x \% S_w$ iff $x \in S_w$. $M_B(S_w) = R_w$ follows by (e).

Proposition 10 *Suppose that $1 \leq i \leq j$, R is i-place and S j-place, M is an i,j-satisfaction function, N is a j,i-satisfaction function, $M(R) = S$ and $N(S) = R$. Then there is an i,j-satisfaction function M' and a j,i-satisfaction function N' such that $N'M' = 1_i$, $M'(R) = S$ and $N'(S) = R$.*

Proof: Define $g \in j^i$ by $g(n) = n$ and $h \in i^j$ by $h(n) = n$ for $n \in i$ and $h(n) = 0$ otherwise. Thus $hg = e$. Now define $C \subseteq P(j^i)$ and $C' \subseteq P(i^j)$ by:

$$C = \{X \colon X \subseteq j^i \;\&\; g \in X\}$$
$$C' = \{X \colon X \subseteq i^j \;\&\; h \in X\}.$$

Now for $I \subseteq D^i_w$ and $x \in D^i_w$, $x \in M_{C'}(M_C(I))$ iff $\{f \colon f \in i^j \;\&\; xf \in M_C(I)\} \in C'$ iff $h \in \{f \colon f \in i^j \;\&\; xf \in M_C(I)\}$ iff $xh \in M_C(I)$ iff $\{f \colon f \in j^i \;\&\; xhf \in I\} \in C$ iff $g \in \{f \colon f \in j^i \;\&\; xhf \in I\}$ iff $xhg \in I$ iff $x \in I$. Thus if P is the j,i-

satisfaction function such that $P(T)_w = M_{C'}(T_w)$ everywhere and Q is the i,j-satisfaction function such that $Q(T)_w = M_C(T_w)$ everywhere then $PQ = 1_i$.

By Proposition 6, there are j,j-satisfaction functions P', Q' such that $P' = MP$ and $Q' = QN$. Then $P'(Q(R)) = MPQ(R) = M(R) = S$ and $Q'(S) = QN(S) = Q(R)$. Hence by Proposition 9 there are j,j-satisfaction functions P'', Q'' such that $P''Q'' = Q''P'' = 1_j$, $P''(Q(R)) = S$ and $Q''(S) = Q(R)$. By Proposition 6, there is an i,j-satisfaction function M' and a j,i-satisfaction function N' such that $M' = P''Q$ and $N' = PQ''$. Then $N'M' = PQ''P''Q = PQ = 1_i$; $M'(R) = S$ and $N'(S) = PQ''(S) = PQ(R) = R$.

The proofs of Propositions 8–10 can be extended to the case, already mentioned, where auxiliary relations are allowed as parameters in the definitions; details are omitted. The results also allow one to compute the maximum number of relations with which a relation may be interdefinable in terms of satisfaction functions, by consideration of the corresponding total apt functions. For example, a binary relation may be interdefinable with up to 192 binary relations in this way.

3 The proof of Proposition 9 depended on the codability of i,j-satisfaction functions by the finite set $P(P(j^i))$, for a 1-1 partial function from a finite set to itself can be extended to a 1-1 total function, which automatically has a two-sided inverse. First-order definitions, which may use quantifiers, are not finitely codable, and thus allow a 1-1 total function which, not being onto, lacks a two-sided inverse. We substantiate these remarks by showing the analogue of Proposition 9 for first-order definitions to fail for a suitable choice of possible worlds. We shall exhibit first-order interdefinable relations which are not interchanged by any pair of everywhere mutually inverse first-order definitions.

Let the sets of pairs of natural numbers $X \subseteq \omega^2$ be correlated 1-1 with the possible worlds $w(X)$, with $D_{w(X)} = \omega$. We make the following definitions:

$$\begin{aligned}
x \overset{R}{=} y&: \forall z((Rxz \equiv Ryz) \;\&\; (Rzx \equiv Rzy))\\
P(R)&: \forall x \exists y \forall z(Rxy \;\&\; \neg Ryx \;\&\; ((Rxz \;\&\; \neg Rzx) \to y \overset{R}{=} z) \;\&\\
&\quad ((Rzy \;\&\, \neg Ryz) \to x \overset{R}{=} z))\\
Q(R)&: \forall x \exists y(Rxx \to (Ryx \;\&\; \neg Rxy))\\
M(R)xy&: (x \overset{R}{=} y \;\&\; P(R) \;\&\; \exists z(Rzz \;\&\; Rzx \;\&\; \neg Rxz)) \vee\\
&\quad (\neg(x \overset{R}{=} y \;\&\; P(R)) \;\&\; Rxy)\\
N(R)xy&: (x \overset{R}{=} y \;\&\; P(R) \;\&\; \exists z(Rzz \;\&\; Rxz \;\&\; \neg Rzx)) \vee\\
&\quad (\neg(x \overset{R}{=} y \;\&\; P(R)) \;\&\; Rxy).
\end{aligned}$$

We first prove that any relation R has the same extension as $N(M(R))$ in each possible world. Consider any world. If $P(R)$ is false, R and $M(R)$ coincide, so $P(M(R))$ is also false, so $N(M(R))$ coincides with $M(R)$ and thence with R. Thus we can assume $P(R)$ to be true. We next show that $\overset{R}{=}$ and $\overset{M(R)}{=}$ coincide. $x \overset{R}{=} y$ clearly entails $x \overset{M(R)}{=} y$, so assume that $x \overset{M(R)}{=} y$. By $P(R)$, choose y' so that $\forall z(Ryy' \;\&\; \neg Ry'y \;\&\; ((Rzy' \;\&\; \neg Ry'z) \to y \overset{R}{=} z))$. Since Ryy' without $Ry'y$, $y \overset{R}{=} y'$ fails, so $M(R)yy'$ and $M(R)y'y$ are equivalent to Ryy' and $Ry'y$ respectively, so $M(R)yy'$ without $M(R)y'y$. Since $x \overset{M(R)}{=} y$, $M(R)xy'$

and $M(R)y'x$ are equivalent to $M(R)yy'$ and $M(R)y'y$ respectively, so $M(R)xy'$ without $M(R)y'x$. Thus $x \overset{R}{=} y'$ fails, so $M(R)xy'$ and $M(R)y'x$ are equivalent to Rxy' and $Ry'x$ respectively, so Rxy' without $Ry'x$. By the choice of y', $x \overset{R}{=} y$. Thus $\overset{R}{=}$ and $\overset{M(R)}{=}$ coincide, so the former can replace the latter in $P(M(R))$. Hence $P(R)$ and $P(M(R))$ can differ in truth value only if Rxy & $\neg Ryx$ and $M(R)xy$ & $\neg M(R)yx$ are not equivalent for all x, y. But if $x \overset{R}{=} y$, Rxy & $\neg Ryx$ and $M(R)xy$ & $\neg M(R)yx$ both fail, while if $x \overset{R}{=} y$ fails, $M(R)xy$ and $M(R)yx$ are equivalent to Rxy and Ryx respectively. Thus $M(R)xy$ & $\neg M(R)yx$ is equivalent to Rxy & $\neg Ryx$, so $P(R)$ and $P(M(R))$ are equivalent, so $P(M(R))$ is true. Hence $N(M(R))$ is equivalent to $(x \overset{R}{=} y \,\&\, \exists z(M(R)zz \,\&\, Rxz \,\&\, \neg Rzx)) \vee (\neg x \overset{R}{=} y \,\&\, M(R)xy)$. Thus if $x \overset{R}{=} y$ fails, $N(M(R)xy$ iff $M(R)xy$ iff Rxy. If $x \overset{R}{=} y$, by $P(R)$ x' can be chosen so that $\forall z(Rxx' \,\&\, \neg Rx'x \,\&\, ((Rxz \,\&\, \neg Rzx) \rightarrow x' \overset{R}{=} z) \,\&\, ((Rzx' \,\&\, \neg Rx'z) \rightarrow x \overset{R}{=} z))$. Hence $\exists z(M(R)zz \,\&\, Rxz \,\&\, \neg Rzx)$, and so $N(M(R))xy$, is equivalent to $M(R)x'x'$, which is equivalent to $\exists z(Rzz \,\&\, Rzx' \,\&\, \neg Rx'z)$, which by the choice of x' is equivalent to Rxx and so to Rxy. Thus R and $N(M(R))$ coincide.

Now define a relation R by $R_{w(X)} = X$. Since $N(M(R)) = R$, R and $M(R)$ are first-order interdefinable, so if the analogue of Proposition 9 for first-order definability held, there should be first-order definable operations M' and N' such that $M'(R) = M(R)$ and for all binary relations T, $N'(M'(T)) = M'(N'(T)) = T$. Moreover, if for each world w we put $w' = w(T_w)$, then $R_{w'} = T_w$, so by the first-order definability of M and M', $M'(T)_w = M'(R)_{w'} = M(R)_{w'} = M(T)_w$. Thus for any binary relation T, $M'(T) = M(T)$.

Let $Y = \{(m, n)\colon (m, n) \in \omega^2 \,\&\, (m = n \vee m + 1 = n)\}$. We work in $w(Y)$, in which Y is the extension of R. One can easily check that $\overset{R}{=}$ coincides with identity, $P(R)$ is true and $N(R)$ coincides with R, so that $P(N(R))$ is also true. Since for any x, $x \overset{N(R)}{=} x$, if we assume that $M(N(R))xx$ then for some z, $N(R)zx$ without $N(R)xz$. In that case, $x \overset{N(R)}{=} z$ fails, so $M(N(R))xz$ and $M(N(R))zx$ are equivalent to $N(R)xz$ and $N(R)zx$ respectively, so $M(N(R))zx$ without $M(N(R))xz$. Thus $Q(M(N(R)))$ is true. But $N(R) = N(M'(N'(R))) = N(M(N'(R))) = N'(R)$, so $M(N(R)) = M(N'(R)) = M'(N'(R)) = R$. Hence $Q(R)$ should be true too, but it is not, for $R00$ even though for no y do we have $Ry0$ without $R0y$. Thus no such operations as M' and N' can exist: R and $M(R)$ are not interchanged by any pair of mutually inverse first-order definitions.

In proving this negative result, we have treated possible worlds purely as mathematical objects. Its interest thus to some extent depends on whether or not they represent "real possibilities". The crucial assumption of the proof is just that some relation has as its possible extensions precisely the subsets of the Cartesian product of a countably infinite set with itself. But if we define R to hold between x and y if and only if x and y are natural numbers such that there is a bag which contains exactly x red marbles and exactly y green ones, R seems to meet just that constraint. Hence there is no reason to suppose that its interest is undermined in that way.

We can also modify the above example so that the possible worlds are simply the models of a first-order theory. Consider a first-order language whose only atomic predicate R is two-place, and the theory in it generated by the standard logical axioms. In effect, we showed above that $\forall x \forall y(Rxy \equiv N(M(R))xy)$ is true in all countable models; hence (by the Löwenheim-Skolem theorems) it is true in all models and therefore (by the Completeness theorem) provable. As before, we can use a model in which the extension of R is Y to show that R and $M(R)$ are not interchanged (up to provable equivalence) by any pair of provable mutually inverse first-order definitions. Thus the result of Part III can be given a purely syntactic form.

Problem: Are there first-order interdefinable relations R and S such that $M(R) = S$ for no first-order definition M with a one-sided inverse?

NOTE

1. If relations are treated as genuine universals, rather than—as in this paper— set-theoretic particulars, it can be argued that the concepts of *to the right of* and *to the left of* are concepts of the *same* relation (cf. [6] and [1], pp. 42 and 94). Similarly, Armstrong rejects negative universals and treats at least some pairs of contradictory predicates as differently related to the *same* universal ([1], pp. 23–29). Hence, informational equivalence might also be proposed as a criterion of identity for genuine universals.

REFERENCES

[1] Armstrong, D. M., *Universals and Scientific Realism, Volume II: A Theory of Universals*, Cambridge University Press, Cambridge, 1978.

[2] Bell, J. L. and A. B. Slomson, *Models and Ultraproducts*, North Holland, Amsterdam, 1969.

[3] Geach, P. T., *Mental Acts*, 2nd ed., Routledge, London, 1971.

[4] Hintikka, J. and V. Rantala, "Systematizing definability theory," pp. 40–62 in *Proceedings of the Third Scandinavian Logic Symposium*, ed. S. Langer, North Holland, Amsterdam, 1975.

[5] Pears, D. F., ed., *Russell's Logical Atomism*, Fontana, London, 1972.

[6] Williamson, T., "Converse relations," *The Philosophical Review*, vol. 94 (1985), pp. 249–262.

Department of Philosophy
University of Dublin
Trinity College
Dublin 2 Ireland