Notre Dame Journal of Formal Logic
Volume 39, Number 2, Spring 1998

Note on Supervenience and Definability

LLOYD HUMBERSTONE

Abstract The idea of a property's being supervenient on a class of properties is familiar from much philosophical literature. We give this idea a linguistic turn by converting it into the idea of a predicate symbol's being supervenient on a set of predicate symbols relative to a (first order) theory. What this means is that according to the theory, any individuals differing in respect to whether the given predicate applies to them also differ in respect to the application of at least one of the predicates in the set. The latter relationship we show turns out to coincide with something antecedently familiar from work on definability: with what is called the piecewise (or modelwise) definability, in the theory in question, of the given predicate in terms of those in the set.

1 The idea of the supervenience of a property P on a set S of properties has become increasingly familiar in the philosophical literature of the past twenty-five years.[1] At its simplest, this relation holds between P and S when it is impossible for there to be two individuals alike in respect to each property in S which are not alike in respect to P. (Objects "are alike" or "agree" in respect to a property when both have the property or both lack the property.) In possible worlds terms, we can take this as saying that two individuals in any arbitrarily selected world which agree, in that world, on all properties in S, also agree on P in that world. This is what is sometimes called *weak* or *intraworld* supervenience, and it contrasts, in particular, with *strong* or *interworld* supervenience, by which is meant that any two individuals in an arbitrarily selected pair of worlds which agree, in the respective worlds, on the properties in S, also agree on P.[2] The purpose of the present note is to consider a version of supervenience applied to monadic predicates in the language of a first-order theory and to relate it to a well-known concept from definability theory. The reasoning will be entirely elementary.

2 Philosophical discussions of supervenience often involve the question of whether this or that notion of supervenience amounts to "reducibility" in some sense. Our interest will be in how what we shall be calling supervenience relative to (or *in*) a theory

Received November 5, 1997; revised October 21, 1998

is related to various theory-relative notions of definability. We therefore review the latter concepts here. (A full discussion with references to the original sources may be found in the useful survey in Rantala [7]; the reason we define explicit definability as a special case of piecewise definability is that we shall need the latter, more general, concept below.) An n-ary predicate symbol F in the language of a theory T is *piecewise definable* in T if for some formulas $\varphi_1, \ldots, \varphi_m$ in that language, none of which contains occurrences of F or has variables other than $x_1, x_2, \ldots, x_n$ free in it, we have

$$\vdash_T \forall x_1, \ldots, \forall x_n (Fx_1, \ldots, x_n \longleftrightarrow \varphi_1) \vee \cdots \vee \forall x_1, \ldots, \forall x_n (Fx_1, \ldots, x_n \longleftrightarrow \varphi_m).$$

We add "in terms of such-and-such items of nonlogical vocabulary" (presumed not to include F) when the only nonlogical expressions to appear in $\varphi_1, \ldots, \varphi_m$ are drawn from the listed items. For present purposes, the items in question will always be predicate symbols. F is *explicitly definable* in T ("in terms of a given set of predicates") if something of the above form is provable in T where $m = 1$ (and only predicates in the given set occur in φ_1). Finally, F is *implicitly definable* in T in terms of predicates $G_1, \ldots, G_k$ when any two models of T with the same domain and the same extensions for $G_1, \ldots, G_2$ also assign the same extension to F. ("Model of T" means, of course: structure for the language of T in which all sentences of T are true.)

The way implicit definability was just characterized renders it recognizably a supervenience relation: "no difference here (on interpretation of F) without a difference there (on interpretation of $G_1, \ldots, G_k$)." Accordingly, Beth's Theorem, which states that whenever F is implicitly definable in T in terms of $G_1, \ldots, G_k$, F is explicitly definable in T in terms of $G_1, \ldots, G_k$, may be seen as claiming that some sort of supervenience implies some sort of reducibility.[3] Note, however, that the objects among whom agreement in certain respects implies agreement in another are certain first-order structures, namely, models of the theory T. (It is these which, if they interpret each of $G_1, \ldots, G_k$ alike, must also treat F alike.) Accordingly, this is the wrong sort of supervenience to connect with the supervenience notions (weak, strong, . . .) we began with. Those concerned agreement among the objects having or lacking the properties—which we can think of here simply as sets—represented by such predicates, rather than among models assigning various sets as the predicates' extensions. Let us, then, consider a theory-relative notion of supervenience in which the objects among which agreement counts are the individuals in the domains of the theory's models, and what constitutes agreement is agreement as to whether or not these predicates are true of the individuals concerned. This would be the most direct analogue of the notion of supervenience of properties on sets of properties introduced in our initial paragraph, except that now properties give way to predicates, and the whole thing is done relative to a first-order theory. The most straightforward adaptation of those initial ideas will be to monadic predicates, since it is these that are true or false of the individuals in the domain (of a model of a theory). It may be of interest to consider what would become of the developments to follow if the definition to be given below were stated more generally for n-adic predicates ($n \geq 1$), but we shall not consider this here. (Certainly, some of the points made apropos of Corollaries 3.3 and 3.4 at the end of our discussion would not hold in the more general setting.)

Suppose, then, that T is a first-order theory and among the primitive monadic predicates of the language of T are F, $G_1, \ldots, G_k$. (We reserve these letters to stand for distinct monadic predicates until further notice.) Then we say that F *is supervenient* on $G_1, \ldots, G_k$ in T just in case

$$\vdash_T \ \forall x \forall y (((G_1 x \longleftrightarrow G_1 y) \wedge \cdots \wedge (G_k x \longleftrightarrow G_k y)) \to (Fx \longleftrightarrow Fy)).$$

(Note that the '$\longleftrightarrow$' in the consequent can be replaced by '$\to$' without loss of logical strength.) We adapt the above talk of agreement and likeness to the present setting and say that two objects in the domain of a first-order structure agree on (or are alike with respect to) a predicate if both lie in the extension of that predicate in the structure or if both lie outside that extension. Then the present notion is akin to weak, rather than strong, supervenience in the sense of our opening paragraph because the relation just defined holds when in any model of T (compare "in an arbitrarily selected world"), agreement on each of $G_1, \ldots, G_k$ implies agreement on F.

Clearly, if F is explicitly definable in T in terms of $G_1, \ldots, G_k$, then F is supervenient on $G_1, \ldots, G_k$ in T. For any individuals agreeing on the latter predicates in a model of T will agree on the open formula constructed from them and serving as the φ in a definition $\forall x(Fx \longleftrightarrow \varphi(x))$,[4] and hence, since we are only considering structures which are models of T, will agree on F. We can see that the converse does not hold—supervenience does not imply explicit definability (relative to an arbitrarily given theory)—because the argument just provided for the implication from explicit definability to supervenience works just as well to show that piecewise definability implies supervenience. This is because of the way models of T are considered one at a time. (The point here is that, where $Th(M)$—*the theory of M*—is the set of closed first-order formulas in the language of a (first-order) structure M which are true in M, then piecewise definability in T amounts to explicit definability in $Th(M)$ for every model M of T. Indeed, piecewise definability is sometimes called modelwise definability.) Thus we have *piecewise definability* $\Longrightarrow$ *supervenience*, so we could not also have *supervenience* $\Longrightarrow$ *explicit definability*; otherwise we should have—something that is certainly not the case—*piecewise definability* $\Longrightarrow$ *explicit definability*. Readers for whom it is clear that piecewise definability is strictly weaker than explicit definability should skip the following paragraph.

(A referee for this journal has suggested that an example along the following lines be included for the benefit of readers who do not find it obvious that piecewise definability does not imply explicit definability. Let T be the theory comprising the consequences of $\forall x(Fx \longleftrightarrow G_1 x) \vee \forall x(Fx \longleftrightarrow G_2 x)$, in the language whose nonlogical vocabulary consists of the predicate symbols figuring in this axiom. By the choice of axiom, F is evidently piecewise definable in terms of G_1 and G_2 in T. Let M be a structure with the three-element domain $\{a, b, c\}$ with $\{a, b\}$ as the extension of G_1 and $\{a, c\}$ as the extension of G_2. We can expand M to a model of T in each of two distinct ways: (i) by taking $\{a, b\}$ as the extension of F, and (ii) by taking $\{a, c\}$ as the extension of F. Since the extension of F is thus not fixed by those of G_1 and G_2 in this model of T, T does not implicitly define F in terms of G_1 and G_2, and so F is not explicitly definable in T in terms of them, its piecewise definability notwithstanding.)

Though supervenience does not coincide with explicit definability, the possibility remains open of a coincidence with the weaker property of piecewise definability.

Let us introduce this possibility (which Proposition 3.2 (below) shows is indeed realized) by looking at a special—and especially manageable—case, in which $k = 1$. Suppose, then, that F is supervenient on G in some theory T.

$$\vdash_T \forall x \forall y((Gx \longleftrightarrow Gy) \to (Fx \longleftrightarrow Fy)). \tag{1}$$

Thus we have (2) and (3).

$$\vdash_T \forall x \forall y((Gx \wedge Gy) \to (Fx \to Fy)). \tag{2}$$

$$\vdash_T \forall x \forall y((\neg Gx \wedge \neg Gy) \to (Fx \to Fy)). \tag{3}$$

We can manipulate (2) and (3) so that atomic subformulas in the same variable are grouped together, getting (4) and (5), respectively.

$$\vdash_T \forall x \forall y((Gx \wedge Fx) \to (Gy \to Fy)). \tag{4}$$

$$\vdash_T \forall x \forall y((\neg Gx \wedge Fx) \to (\neg Gy \to Fy)). \tag{5}$$

Having separated the variables, we can now massage (4) and (5) into the disjunctive forms (6) and (7).

$$\vdash_T \forall x(Gx \to \neg Fx) \vee \forall x(Gx \to Fx). \tag{6}$$

$$\vdash_T \forall x(Fx \to Gx) \vee \forall x(\neg Fx \to Gx). \tag{7}$$

Calling the first and second disjunct of the formula in (6), (6i), and (6ii), respectively, and similarly in the case of (7), we note that

(6i), (7i)	$\vdash \forall x(\neg Fx)$
(6i), (7ii)	$\vdash \forall x(Fx \longleftrightarrow \neg Gx)$
(6ii), (7i)	$\vdash \forall x(Fx \longleftrightarrow Gx)$
(6ii), (7ii)	$\vdash \forall x(Fx)$

where $\vdash$ is (classical) first-order logical consequence. Thus from (6), (7), and therefore from (1), it follows that

$$\vdash_T \forall x(Fx \longleftrightarrow Gx) \vee \forall x(Fx \longleftrightarrow \neg Gx) \vee \forall x(Fx) \vee \forall x(\neg Fx). \tag{8}$$

It is easy to see that each of the four disjuncts here, and hence the disjunction, similarly implies the formula in (1), which is therefore equivalent to that in (8). The latter is not quite in the form required for piecewise definability, but we can adjust the last two disjuncts so that the letter of the condition is satisfied; we abbreviate $Gx \vee \neg Gx$ to $\top x$ and its negation to $\bot x$.

$$\begin{aligned} \vdash_T \forall x(Fx \longleftrightarrow Gx) \vee \forall x(Fx \longleftrightarrow \neg Gx) \vee \\ \forall x(Fx \longleftrightarrow \top x) \vee \forall x(Fx \longleftrightarrow \bot x). \end{aligned} \tag{9}$$

Since (9) follows from (1), for the special case of supervenience of F on a single predicate G, we have shown that such supervenience implies piecewise definability. Noting that the four disjuncts involve, on the right, each of the four nonequivalent Boolean compounds that can be constructed from Gx, we would expect something similar in the general case, in which the hypothesis is that F is supervenient in T on $G_1, \ldots, G_k$. For the proof in the general case, however, it would be cumbersome in the extreme to retrace the analogue of the passage from (1) to (9), and we use a different style of argument—one for which no particular originality is claimed: see note 2.

3 Formulas of the form $\pm G_1x \wedge \pm G_2x \wedge \cdots \wedge \pm G_kx$, where $\pm G_ix$ is either the formula G_ix or else the formula $\neg G_ix$, will be called *elementary G-conjunctions*; we reserve 'ψ' as a variable ranging over these conjunctions. By ψ^M we mean the set of elements satisfying such a formula in the structure M. (See note 4.) Similarly F^M (alias $(Fx)^M$) is the extension assigned to F by M. We call ψ *F-favorable* just in case $\psi^M \cap F^M \neq \varnothing$. (We should strictly say 'F-favorable relative to M' but the relevant structure will be clear from the context.)

Lemma 3.1 *If M is a model for a theory in which F is supervenient on $G_1, \ldots, G_k$ then $F^M = \cup\{\psi^M | \psi$ is an F-favorable elementary G-conjunction$\}$.*

Proof:

Case 1: $\subseteq$ For each element a of the domain of any structure M for a language with the predicates mentioned in the statement of Lemma 3.1, we have $a \in \psi^M$ for some (indeed for exactly one) elementary G-conjunction ψ (corresponding to a's pattern of membership in the sets $\pm G_i^M$). So if $a \in F^M$, the ψ^M to which a belongs satisfies the F-favorability condition in the union term, showing that F^M is included in that union.

Case 2: $\supseteq$ Suppose a belongs to the union, and hence to some ψ^M with ψF-favorable. The latter means that for some $b \in F^M$, we have $b \in \psi^M$. Now M is supposed to be a model for a theory in which F is supervenient on $G_1, \ldots, G_k$, and since we have $a, b \in \psi^M$, a and b agree on membership in each G_i^M, so they must agree in respect to F^M. But $b \in F^M$. Therefore $a \in F^M$. □

We are now in a position to state the main observation of this note.

Proposition 3.2 *Let T be a first-order theory with monadic predicates $F, G_1, \ldots, G_n$ in its language. Then F is supervenient on $G_1, \ldots, G_n$ in T if and only if F is piecewise definable in terms of $G_1, \ldots, G_n$ in T.*

Proof:

Case 1: *If* Clear.

Case 2: *Only if* Suppose that T is as described in Proposition 3.2, with F supervenient on $G_1, \ldots, G_n$ in T. For each model M of T, we show that there is a formula $\forall x(Fx \longleftrightarrow \varphi(x))$ verified by M, with φ a disjunction of elementary G-conjunctions. Since there are only finitely many pairwise nonequivalent such disjunctions φ, the disjunction of all such $\forall x(Fx \longleftrightarrow \varphi(x))$ is a formula which will be provable in T. To get each of these *inner disjunctions* φ from M, we appeal to Lemma 3.1. Let $\psi_1, \ldots, \psi_r$ be all of the F-favorable elementary G-conjunctions (relative to M). Then according to Lemma 3.1, $F^M = \psi_1^M \cup \cdots \cup \psi_r^M$, so we can take $\varphi(x)$ as

$$\psi_1(x) \vee \cdots \vee \psi_r(x).$$

Suppose there are altogether s nonequivalent such disjunctions obtainable from the models of T. Then we have, for suitable values of the indices:

$$\vdash_T \forall x(Fx \longleftrightarrow (\psi_1^1 \vee \cdots \vee \psi_{r_1}^1)) \vee \cdots \vee \forall x(Fx \longleftrightarrow (\psi_1^S \vee \cdots \vee \psi_{r_s}^S))$$

Since the disjunctions of the ψ^i_j contain no predicates other than $G_1, \ldots, G_n$ and no free variables other than x, this shows that F is piecewise definable in terms of these predicates in T. □

In fact—continuing from the last sentence of the above proof—it shows rather more, since in our disjunctions of the ψ^i_j , there are not only no free variables other than x, but there are no variables at all other than x. It is worth having some terminology for such quantifier-free *definientia*, so we introduce the following specialization of the notion of piecewise definability. For its statement, 'F' and 'G_1', 'G_2', . . . revert to their use to stand in for predicates of arbitrary arity. We say that an n-ary predicate symbol F in the language of a theory T is piecewise *Boolean-definable* in T (in terms of predicates $G_1, \ldots, G_k$) if for some formulas $\varphi_1, \ldots, \varphi_m$, each of these being a Boolean combination of atomic formulas (whose nonlogical vocabulary is drawn from the list $G_1, \ldots, G_k$), we have:

$$\vdash_T \forall x_1, \ldots, \forall x_n(Fx_1, \ldots, x_n \longleftrightarrow \varphi_1) \vee \cdots \vee \forall x_1, \ldots, \forall x_n(Fx_1, \ldots, x_n \longleftrightarrow \varphi_m)$$

Analogously, we say that F is *explicitly Boolean-definable* in T (in terms of $G_1, \ldots, G_k$) when a formula of the above form is provable in T, with $m = 1$.[5] Although we give the definition here in general terms, for present applications we are concerned only with the special case in which $F, G_1, \ldots, G_k$ are all monadic.

We are now in a position to say what more the proof of the proposition above yields than is stated in the proposition itself, namely, that whenever F is supervenient on $G_1, \ldots, G_k$ in the theory T, the predicate F is piecewise Boolean-definable in terms of $G_1, \ldots, G_k$ in T. Hence we extract the following information, deleting all reference to supervenience.

Corollary 3.3 *Where T is a first-order theory in a language with monadic predicates $F, G_1, \ldots, G_k$ in its language, F is piecewise definable in terms of $G_1, \ldots, G_n$ in T if and only if F is piecewise Boolean-definable in terms of $G_1, \ldots, G_n$ in T.*

Since explicit definability is a special case of piecewise definability, we have the following corollary.

Corollary 3.4 *With T as before: if F is explicitly definable in terms of $G_1, \ldots, G_k$ in T, then F is piecewise Boolean-definable in terms of $G_1, \ldots, G_k$ in T.*

We devote the remainder of our discussion to illustrating the phenomenon involved here and relating it to some of the foregoing remarks.

4 We left the proof of the *If* direction of Proposition 3.2 above at the word "Clear," having remarked earlier that our argument that explicit definability implied supervenience in fact showed, more generally, that piecewise definability implied supervenience. The argument went like this: suppose that F is explicitly definable in T in terms of $G_1, \ldots, G_k$, say because $\vdash_T \forall x(Fx \longleftrightarrow \varphi(x))$, where φ is built from $G_1, \ldots, G_k$ and contains no free variables other than x. Then F is supervenient on $G_1, \ldots, G_k$ in T. For any individuals agreeing on the latter predicates in a model of T will have to agree on $\varphi(x)$, and hence on F. That was the argument. But in particular cases we can see that the form of φ may enable us to eliminate some of $G_1, \ldots, G_k$

in the conclusion about what F is supervenient on in T, thereby getting a stronger conclusion. (Of course, if F is supervenient on a set of predicates, it is automatically supervenient on any superset thereof.) For example, $\varphi(x)$ is $G_1x \wedge \exists y(G_2y)$ and $\vdash_T \forall x(Fx \longleftrightarrow (G_1x \wedge \exists y(G_2y)))$. Then F is explicitly definable in terms of G_1, G_2 (in T), and we cannot eliminate the reference to G_2 in this definability statement. But F is supervenient on G_1 all by itself (in T), since a and b agree on $(G_1x \wedge \exists y(G_2y))^M$ if they agree on G_1^M : if $\exists y(G_2y)$ is true in M then a and b belong to $(G_1x \wedge \exists y(G_2y))^M$ if and only if they belong to G_1^M— on which we are supposing them to agree—and if $\exists y(G_2y)$ is false in M then a and b automatically agree on $(G_1x \wedge \exists y(G_2y))^M$, since this is $\varnothing$ in that case.[6] The internal structure of the formula $\exists y(G_2y)$ in the example just considered was irrelevant to the point, and so we represent it here by the schematic letter 'χ' (mnemonic: *chi* for *closed*). If the $\varphi(x)$ which provides the *definiens* in an explicit definition (of F) does not itself secure Boolean-definability, because $\varphi(x)$ is of the form $\varphi_1(x) \wedge \chi$, where x occurs free in $\varphi_1(x)$ but not in χ, then we can move to piecewise Boolean-definability by noting that (11) follows from (10).

$$\vdash_T \forall x(Fx \longleftrightarrow (\varphi_1(x) \wedge \chi)). \tag{10}$$

$$\vdash_T \forall x(Fx \longleftrightarrow \varphi_1(x)) \vee \forall x(\neg Fx). \tag{11}$$

As before (transition from (8) to (9) above) to make it clear that we are dealing with piecewise (Boolean-)definability, we can rewrite (11) as (12).

$$\vdash_T \forall x(Fx \longleftrightarrow \varphi_1(x)) \vee \forall x(Fx \longleftrightarrow \bot x). \tag{12}$$

The first disjunct of the formula in (11) (or (12)) is the 'χ-true' case, and the second, the "χ-false" case. The reader is invited to explore the analogues of (11)/(12) for the cases in which the '$\wedge$' of (10) is replaced by '$\vee$' or by '$\longleftrightarrow$'. (By judiciously negating $\varphi_1(x)$ or χ, one sees that no other binary Boolean connectives need be considered.) We have lost information, it should be noted, in passing from (10) to (12), because the conditions under which $F^M = \varphi_1^M$ and those under which $F^M = \varnothing$ are no longer specified. (What Corollary 3.3 says are equivalent are piecewise Boolean-definability and piecewise definability: it is not claimed that a given "piecewise definition" and the promised "piecewise Boolean definition" are themselves equivalent.)[7] We can convert (12) into something recovering this lost information easily enough.

$$\vdash_T (\chi \rightarrow \forall x(Fx \longleftrightarrow \varphi_1(x))) \wedge (\neg\chi \rightarrow \forall x(Fx \longleftrightarrow \bot x)). \tag{13}$$

The formula in (13) is a conjunction of what are sometimes called conditional definitions, though of a rather special form here, since the antecedent does not contain the variable 'x' free. (To qualify strictly as conditional definitions we should write, for example, for the first conjunct:

$$\forall x(\chi \rightarrow (Fx \longleftrightarrow \varphi_1(x)),$$

but of course there is no difference in content between this and the first conjunct in (13) as written).

In general, a representation of $\varphi(x)$ as $\varphi_1(x)\ \#\ \chi$, where # is some binary Boolean mode of combination (as is #′ below), is not guaranteed to be available because the open and closed subformulas in a Boolean compound are too intermingled.

For example, if $\varphi(x)$ is $\chi_1 \wedge (\varphi_1(x) \vee \chi_2)$, there is no representation of $\varphi(x)$ in the form $\varphi_1(x) \# (\chi_1 \#' \chi_2)$.[8] So for the general case, we should allow $\varphi(x)$ to have the form $\mathbf{f}(\chi_1, \ldots, \chi_m, \varphi_1(x), \ldots, \varphi_n(x))$, where $\mathbf{f}$ indicates some Boolean combination (some polynomial in $\wedge$ and $\neg$, say).

Note that since only monadic predicate letters are involved, we can assume $\varphi_1(x), \ldots, \varphi_n(x)$ are themselves quantifier-free. Let the $(m+n)$-ary truth-function corresponding in the obvious sense to $\mathbf{f}$ be f. We can treat the general case adequately and avoid a mire of subscripts and ellipses if we consider putting $m = 3$, $n = 2$. Then $\varphi(x)$ has the form indicated on the right of (14).

$$\forall x(Fx \longleftrightarrow \mathbf{f}(\chi_1, \chi_2, \chi_3, \varphi_1(x), \varphi_2(x))) \tag{14}$$

We denote the truth-values *True*, *False* by 1, 0, respectively, and use 'u,' 'v,' as variables ranging over $\{1, 0\}$. From the 5-ary truth-function f associated with $\mathbf{f}$, we define the binary truth-functions f^{111}, f^{110}, f^{101}, and so on, by holding fixed the first three arguments as indicated by the superscripts, that is, by putting:

$$\begin{aligned} f^{111}(u, v) &= f(1, 1, 1, u, v), \\ f^{110}(u, v) &= f(1, 1, 0, u, v), \\ f^{101}(u, v) &= f(1, 0, 1, u, v), \text{ etc.} \end{aligned}$$

Where $\mathbf{f}^{111}$, $\mathbf{f}^{110}$, $\mathbf{f}^{101}$, and so on, are Boolean compoundings corresponding to these truth-functions, we then have, in the style of (13), a representation of the content of (14) as the conjunction of

$$\begin{array}{lcl} (\chi_1 \wedge \chi_2 \wedge \chi_3) & \rightarrow & \forall x(Fx \longleftrightarrow \mathbf{f}^{111}(\varphi_1(x), \varphi_2(x))), \\ (\chi_1 \wedge \chi_2 \wedge \neg\chi_3) & \rightarrow & \forall x(Fx \longleftrightarrow \mathbf{f}^{110}(\varphi_1(x), \varphi_2(x))), \\ (\chi_1 \wedge \neg\chi_2 \wedge \chi_3) & \rightarrow & \forall x(Fx \longleftrightarrow \mathbf{f}^{101}(\varphi_1(x), \varphi_2(x))), \text{ etc.} \end{array}$$

Throwing away the information, given in the antecedents of these (eight) conditionals, about the conditions saying exactly how the extension of F is determined by those of $\varphi_1(x)$ and $\varphi_2(x)$, we obtain a piecewise Boolean definition of F by disjoining their consequents.

We have concentrated on illustrating Corollary 3.4 showing how explicit definability, with possible internal quantificational complexity, implies piecewise definability without such complexity. But similar "Booleanizing" of piecewise definability proceeds along the same lines, disjunct by disjunct. For example, suppose we have

$$\vdash_T \forall x(Fx \longleftrightarrow (\chi_1 \wedge \varphi_1(x))) \vee \forall x(Fx \longleftrightarrow (\chi_2 \vee \varphi_1(x))). \tag{15}$$

Then, in any model of T, the truth of the first disjunct implies (16), and the second implies (17).

$$\forall x(Fx \longleftrightarrow \varphi_1(x)) \vee \forall x(Fx \longleftrightarrow \bot x). \tag{16}$$

$$\forall x(Fx \longleftrightarrow \top x) \vee \forall x(Fx \longleftrightarrow \varphi_2(x)). \tag{17}$$

The disjunction of (16) with (17) is provable in T, establishing piecewise Boolean-definability of F on the basis of (15), as is promised by Corollary 3.3.

NOTES

1. We date the emergence of supervenience as a topic in its own from the appearance of Kim [5] and Hellman and Thompson [2], though the term is not used in the latter paper. For a discussion of [2] in the general framework of the theory of supervenience, see Haugeland [1], Teller [8]. [5] is now mainly of historical interest, its main claim having since been retracted by Kim (see [6], p. 65, especially note 18).

2. Strong and weak supervenience are distinguished in various publications of Kim; see especially the papers reproduced as chapters 4, 5, 7, and 9 of [6]. The reader should be warned that Haugeland (in [1] and elsewhere) uses "weak supervenience" for another concept, which in Kim's terminology (in the papers just cited) is "global supervenience." This is a matter of whole worlds agreeing in certain respects (the supervenient respects) given agreement in others (the subvenient respects). Another point to note is that Kim's favored characterizations of supervenience (weak or strong) involve a greater complexity than those sketched here, with a detour down to the subvenient properties and then back up to the supervenient: thus the class of properties S' is said to be supervenient on the class of properties S when any individual with a property $P' \in S'$ has some (intuitively, "maximally specific") property $P \in S$ such that any other individual possessing P possesses P'. The equivalence of this characterization with the traditional characterization (agreement on every property in S implies agreement on every property in S') is then shown by Kim at, for example, p. 58 f. of [6]. The discussion there concentrates on weak supervenience, and the proof of Lemma 3.1 below involves essentially the same argument. (The "elementary G-conjunctions" we employ there correspond to maximally specific subvenient properties.) However, it is hoped that the somewhat different conceptual orientation of the present discussion will be of interest. The same (pattern of) argument may also be found—this time apropos of the relation between necessary equivalence and supervenience—in several works of Frank Jackson; see, for example, the first (new) paragraph on p. 85 of [4]. Finally, I have been informed by my colleague Richard Holton that similar reasoning is to be found in an unpublished paper (dated 1989) by Bas van Fraassen. For some further methodological remarks on the traditional characterization of (notions of) supervenience, see §1 of Humberstone [3].

3. Thus, Hellman and Thompson pay considerable attention in [2] to the bearing of Beth's Theorem on their "determinationism" (a model-theoretic analogue of the concept of global supervenience alluded to in note 2 above). For further discussion of this point, see Tennant [9] and for background on Beth's Theorem, [7].

4. Here we have written '$\varphi(x)$' as a reminder that no variable other than x may appear free in φ. And we make a natural extension of the talk of agreement on a monadic predicate letter to agreement on such an open formula: the intention is that for the two individuals concerned, the formula is true (in the structure in question) when one of them is assigned to the free variable if and only if it is true when the other is assigned to the free variable. Below, we denote the set of such φ-satisfiers in M by φ^M.

5. This concept goes under the name 'quantifier-free definability' in Williamson's fascinating study [10] of invertible definitions.

6. Of course, it may be that a closed subformula of $\varphi(x)$ involves no predicate letter other than figures in open subformulas of $\varphi(x)$, in which case there will not be the kind of reduction, illustrated by this case, in the set of predicates on which F is supervenient. For example, Fx may be defined as $G_1x \wedge \exists y(\neg G_1 y)$—perhaps the simplest example of a predicate which can be satisfied by some things but not by everything (in a given domain).

7. By “piecewise (Boolean) definition” is here meant simply the disjunction (as in the second paragraph of this note) whose provability attests to the piecewise (Boolean-) definability of F.

8. This is a point about truth-functional logic, in the context of which let us say that q is *extractible* from a formula $\varphi(q, p_1, \ldots, p_l)$ built up from the displayed sentence letters (all presumed distinct) if there is some Boolean connective # for which φ is equivalent to $q \,\#\, \psi(p_1, \ldots, p_l)$, ψ being any formula built from $p_1, \ldots, p_l$. Then q (corresponding to the ‘$\varphi(x)$’ of our discussion) is not extractible from $p_1 \wedge (q \vee p_2)$, as one sees by a truth-table examination.

REFERENCES

[1] Haugeland, J., “Ontological supervenience,” *Southern Journal of Philosophy*, vol. 22 (1985), Supplement, Spindel Conference, edited by T. Horgan, pp. 1–12. 4, 4

[2] Hellman, G., and F. W. Thompson, “Physicalism: Ontology, determination, and reduction,” *Journal of Philosophy*, vol. 72 (1975), pp. 551–64. 4, 4, 4

[3] Humberstone, I. L. “Some structural and logical aspects of supervenience,” *Logique et Analyse*, vol. 35 (1992), pp. 101–37. Zbl 0832.03007 4

[4] Jackson, F., and P. Pettit, “Moral functionalism, supervenience and reductionism,” *Philosophical Quarterly*, vol. 45 (1996), pp. 82–86. 4

[5] Kim, J., “Supervenience and nomological incommensurables,” *American Philosophical Quarterly*, vol. 15 (1978), pp. 149–56. 4, 4

[6] Kim, J., *Supervenience and Mind: Selected Philosophical Essays*, Cambridge University Press, Cambridge, 1993. 4, 4, 4

[7] Rantala, V., “Definitions and definability,” pp. 135–59 in *Definitions and Definability: Philosophical Perspectives*, edited by J. H. Fetzer, D. Shatz, and G. Schlesinger, Kluwer, Dordrecht, 1991. 2, 4

[8] Teller, P., “A poor man’s guide to supervenience and determination,” *Southern Journal of Philosophy*, vol. 22 (1985), Supplement, Spindel Conference, edited by T. Horgan, pp. 137–67. 4

[9] Tennant, N., “Beth’s theorem and reductionism,” *Pacific Philosophical Quarterly*, vol. 66 (1985), pp. 342–54. 4

[10] Williamson, T. “Invertible definitions,” *Notre Dame Journal of Formal Logic*, vol. 28 (1987), pp. 244–58. Zbl 0637.03004 MR 88e:03008 4

Department of Philosophy
Monash University
Wellington Road
Clayton Victoria 3168
AUSTRALIA
email: Lloyd.Humberstone@arts.monash.edu.au