Herbrand's Theorem for a Modal Logic

Melvin Fitting
fitting@alpha.lehman.cuny.edu
Dept. Mathematics and Computer Science
Lehman College (CUNY), Bronx, NY 10468
Depts. Computer Science, Philosophy, Mathematics
Graduate Center (CUNY), 33 West 42nd Street, NYC, NY 10036

August 16, 1996

1 Introduction

Herbrand's theorem is a central fact about classical logic, [9, 10]. It provides a constructive method for associating, with each first-order formula X, a sequence of formulas X_1, X_2, X_3, ..., so that X has a first-order proof if and only if some X_i is a tautology. Herbrand's theorem serves as a constructive alternative to Gödel's completeness theorem. It provides the theoretical basis for automated theorem proving, reducing a first-order problem to a search through an infinite sequence of propositional problems, [12]. It provides machinery for theoretical investigations,[2]. But it does not travel well. Unlike Gentzen's cut elimination theorem, or Gödel's completeness theorem, analogs of Herbrand's result essentially do not exist for non-classical logics.

In this paper we sketch how a Herbrand theorem can be obtained for the modal logic **K**, after making a natural addition to the customary first-order modal machinery. A similar result can be shown for several other modal logics, though it is an interesting problem to determine the range of modal logics for which this is possible.

Actually, in this paper we can present only a summary of the basic ideas, and provide motivation for the approach we take. A full proof is too long to meet space limitations here. A longer version of this paper, with full proofs, can be found in [8].

2 Why extra machinery is needed

Presentations of the classical Herbrand theorem often begin by putting formulas in prenex form. It is well-known that this is not possible for any standard modal logic. Fortunately, this is no obstacle. Herbrand expansions for classical formulas can be defined without making use of prenex form, thus avoiding the issue altogether.

The real problems begin with the next step: Skolemization. In order to present the modal difficulties let us use *validity functional form*, in which quantifiers that are essentially universal are eliminated, introducing new function symbols in a way that preserves validity. The simplest example is $(\forall x)Px$, where P is a one-place relation symbol—its (classical) validity functional form is just Ps, where s is a new constant symbol. Now, consider the modal formula $\Diamond(\forall x)Px$; what should its (modal) validity functional form be? A reasonable guess is $\Diamond Ps$, where again s is a new constant symbol. If $\Diamond(\forall x)Px$ is false at possible world p of a Kripke model $\mathcal{M}$, then at each world accessible from p, $(\forall x)Px$ must be false and so, at each world accessible from p, Px must be false of some object that exists at that world. If q_1 and q_2 are two worlds accessible from p, it could happen that while Px is false of some object at each of them, the object is not the same at the two worlds—Px might be false only of object a at q_1, and false only of b at q_2, where $a \neq b$. But then, if $\Diamond Ps$ is to be the Skolemized version of $\Diamond(\forall x)Px$, and we want it to be false at p, we are forced to have the constant symbol s designate a at q_1 and b at q_2.

Terms that can designate different things at different possible worlds are called *non-rigid* (by philosophers), or *flexible* (by computer scientists). But, allowing non-rigidity introduces a new set of problems. This time consider $\Box Ps$, and assume it is true at world p of a Kripke model, in which worlds q_1 and q_2 are accessible from p. What, exactly, should this mean? One possible meaning to give to $\Box Ps$ being true at p is: in every world accessible from p, the formula Ps is true, taking this to mean that the P property holds in q_1 of the object that s designates at q_1, and the P property holds in q_2 of the object that s designates at q_2. But, another possible meaning to give to $\Box Ps$ being true at p is: the object that s designates at p has the $\Box P$ property, and thus *that* object has the P property in both q_1 and q_2. These two readings of $\Box Ps$ can be very different, since what s designates at p need not be the same as what it designates at q_1 or q_2. In short, if non-rigidity is allowed, the act of designation and the act of passing to an alternative world need not commute.

If non-rigidity is allowed, syntax like $\Box Ps$ becomes ambiguous. This is sometimes sorted out by attaching metalanguage qualifiers: s has *narrow scope* or *broad scope*. For our purposes, both are needed. Validity functional form Skolemization of $\Box(\forall x)Px$ should yield $\Box Ps$ where s has narrow scope, but Skolemization of $(\forall x)\Box Px$ should yield $\Box Ps$ where s has broad scope. We may also need both at once, as in Skolemizing $(\forall x)\Box(\forall y)Rxy$. And $\Box(\forall x)\Box Px$ shows that a broad/narrow scope distinction is not sufficient to cover all the cases we are interested in.

Continuing with problems, in the classical Herbrand theorem, after Skolemizing, the remaining essentially existential quantifiers are replaced with disjunctions of instances. But, these instances introduce broad/narrow scope problems of their own, and the difficulties outlined above simply compound.

3 Disambiguating scope

In [13, 14] a formal scoping of terms was introduced into modal syntax and semantics by Stalnaker and Thomason. This device was further applied and elaborated by the present author, [3, 4, 5, 6, 7]. We present it here under the name *predicate abstraction*—it provides the solution to the problems of the previous section.

The "usual" syntax of first-order modal logic is taken to be that of classical first-order logic, with terms built up from variables, constant and function symbols, but allowing $\Box$ and $\Diamond$ to appear in formulas. We modify this usual syntax in two ways. First, an *atomic formula* is now an expression of the form $Px_1 \ldots x_n$, where the x_i are *variables*. More complex terms are not allowed to appear at the atomic level. Second, and most important, we add one more formation rule to the usual list.

- If φ is a formula, x is a variable, and t is a term, $\langle \lambda x.\varphi \rangle(t)$ is a formula, and its free variable occurrences are those of φ, except for x, together with all variable occurrences in t.

Essentially, think of φ as a formula, and from it a predicate can be abstracted, a predicate denoted $\langle \lambda x.\varphi \rangle$. It is such predicate abstracts that are applied to terms.

A predicate abstraction mechanism does not turn up in classical logic because all the classical connectives and quantifiers are transparent to it. On the other hand, $\langle \lambda x.\Box\phi \rangle(t)$ and $\Box\langle \lambda x.\phi \rangle(t)$ can have very different meanings semantically. Also, though it does not play a role here, the effects of predicate abstraction can show up even at the classical level if non-designating terms are allowed—something Russell observed in his well-known treatment of definite descriptions ([17], reprinted in [16]).

We use the following conventions and terminology. We take as primitive $\neg$, $\supset$, $\forall$, and $\Box$; all other logical operations are defined. A *frame* is a structure $\langle \mathcal{G}, \mathcal{R}, \mathcal{D} \rangle$ where $\mathcal{G}$ is a non-empty set of possible worlds, $\mathcal{R}$ is a binary relation of accessibility on $\mathcal{G}$, and $\mathcal{D}$ is a *domain function* from members of $\mathcal{G}$ to non-empty sets, meeting the *monotonicity condition*, $p\mathcal{R}q \Longrightarrow \mathcal{D}(p) \subseteq \mathcal{D}(q)$. An *interpretation* in a frame is a mapping $\mathcal{I}$ that assigns:

1. to each constant symbol c and each $p \in \mathcal{G}$ some member $\mathcal{I}(p, c)$ of $\mathcal{D}(p)$;
2. to each n-ary function symbol f and each $p \in \mathcal{G}$ some n-ary function $\mathcal{I}(p, f)$ on $\mathcal{D}(p)$;
3. to each n-ary relation symbol R and each $p \in \mathcal{G}$ some n-ary relation $\mathcal{I}(p, R)$ on $\mathcal{D}(p)$.

Note that constant and function symbols are explicitly allowed to vary their designation from world to world. A structure $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{D}, \mathcal{I} \rangle$ is a *model* if $\langle \mathcal{G}, \mathcal{R}, \mathcal{D} \rangle$ is a frame and $\mathcal{I}$ is an interpretation in it.

If $\mathcal{M} = \langle \mathcal{G}, \mathcal{R}, \mathcal{D}, \mathcal{I} \rangle$ is a model, its *domain* is $\cup\{\mathcal{D}(p) \mid p \in \mathcal{G}\}$. An *assignment* in a model is a mapping from variables to the domain of the model.

We write $\mathcal{M}, p \Vdash \varphi\,[s]$ to indicate that formula φ is true at world p of model $\mathcal{M}$ under assignment s. In order to define this formally we first introduce two pieces of notation. First, for an assignment s, by $s\left[\begin{smallmatrix}x\\a\end{smallmatrix}\right]$ we mean the assignment that is like s on all variables except x, and that assigns a to x. Second, for an assignment s and an interpretation $\mathcal{I}$, we define a mapping $(s \star \mathcal{I})$ on worlds and terms, as follows:

1. If x is a free variable, $(s \star \mathcal{I})(p, x) = s(x)$.
2. If c is a constant symbol, $(s \star \mathcal{I})(p, c) = \mathcal{I}(p, c)$.
3. If f is an n-place function symbol,
$$(s \star \mathcal{I})(p, ft_1, \ldots, t_n) = \mathcal{I}(p, f)((s \star \mathcal{I})(p, t_1), \ldots, (s \star \mathcal{I})(p, t_n))$$

Now, here is the definition of truth at a world, most of which is standard.

1. For an n-ary relation symbol R, $\mathcal{M}, p \Vdash Rx_1, \ldots, x_n\,[s]$ iff $\langle s(x_1), \ldots, s(x_n) \rangle \in \mathcal{I}(p, R)$.
2. $\mathcal{M}, p \Vdash \neg\varphi\,[s]$ iff $\mathcal{M}, p \nVdash \varphi\,[s]$.
3. $\mathcal{M}, p \Vdash \varphi \supset \psi\,[s]$ iff $\mathcal{M}, p \Vdash \varphi\,[s]$ implies $\mathcal{M}, p \Vdash \psi\,[s]$.
4. $\mathcal{M}, p \Vdash \Box\varphi\,[s]$ iff $\mathcal{M}, q \Vdash \varphi\,[s]$ for all $q \in \mathcal{G}$ such that $p\mathcal{R}q$.
5. $\mathcal{M}, p \Vdash (\forall x)\varphi\,[s]$ iff $\mathcal{M}, p \Vdash \varphi\,[s\left[\begin{smallmatrix}x\\a\end{smallmatrix}\right]]$ for all $a \in \mathcal{D}(p)$.
6. $\mathcal{M}, p \Vdash \langle \lambda x.\varphi \rangle(t)\,[s]$ iff $\mathcal{M}, p \Vdash \varphi\,[s\left[\begin{smallmatrix}x\\a\end{smallmatrix}\right]]$ where $a = (s \star \mathcal{I})(p, t)$.

Of course, the last item is the key new one. Loosely, for $\langle \lambda x.\varphi \rangle(t)$ to be true at a world, φ should be true there provided we take the value of x to be whatever the term t designates at p.

4 Skolemization

The problems presented in section 2 now go away. A straightforward model-theoretic argument can be used to show the following, [6].

Proposition 4.1 *Suppose φ is a closed formula, $(\forall x)\psi$ is a positively occurring subformula of φ, $(\forall y_1)$, ..., $(\forall y_k)$ are all the quantifiers in φ within whose scope $(\forall x)\psi$ occurs, and each of $(\forall y_i)$ occurs negatively in φ. Let φ^* be the result of* *replacing* $(\forall x)\psi$ *in* φ *with* $\langle\lambda x.\psi\rangle(fy_1,\ldots,y_k)$*, where* f *is a function symbol not occuring in* φ*. Then* φ *is valid (true at all worlds of all models) if and only if* φ^* *is valid.*

By repeated applications of this Proposition, all essentially universal quantifiers can be eliminated from a modal formula φ, producing an equi-valid formula. We call the result of doing so the *validity functional form* of φ. Thus Skolemization carries over to the modal setting in a simple way, provided predicate abstraction is used.

5 Herbrand expansions

Classically, after Skolemization, the next step is to replace the remaining negatively occurring universal quantifiers with conjunctions of instances. Thus, for example, if a and b are closed terms, the formula $\neg(\forall x)Px$ can be converted into $\neg(Pa \wedge Pb)$. In the present modal setting, predicate abstraction complicates this. Consider the formula $\neg\Box(\forall x)P(x)$, and again assume a and b are closed terms. Replacing the quantifier by a conjunction, using predicate abstraction, can lead to any of the following:

$$\neg\Box\langle\lambda x.\langle\lambda y.Px \wedge Py\rangle(b)\rangle(a)$$
$$\neg\langle\lambda x.\Box\langle\lambda y.Px \wedge Py\rangle(b)\rangle(a)$$
$$\neg\langle\lambda x.\langle\lambda y.\Box(Px \wedge Py)\rangle(b)\rangle(a)$$

or even things like

$$\neg\langle\lambda x.\langle\lambda y.\Box\langle\lambda z.Px \wedge Pz\rangle(y)\rangle(b)\rangle(a)$$

The specification of what is a conjunction of instances becomes non-trivial. We use a simple sequent calculus for this purpose.

Definition 5.1 X' is a *modal Herbrand transform* of the formula X if $X \to X'$ is derivable in the following calculus.

Literal For A atomic, $A \to A$ and $\neg A \to \neg A$.

Propositional

$$\frac{X \to X'}{\neg\neg X \to \neg\neg X'}\ Neg$$

$$\frac{\neg X \to \neg X' \quad Y \to Y'}{X \supset Y \to X' \supset Y'}\ +Imp \qquad \frac{X \to X' \quad \neg Y \to \neg Y'}{\neg(X \supset Y) \to \neg(X' \supset Y')}\ -Imp$$

Modal

$$\frac{X \to X'}{\Box X \to \Box X'}\ +Nec \qquad \frac{\neg X \to \neg X'}{\neg\Box X \to \neg\Box X'}\ -Nec$$

Abstraction

$$\frac{X \to X'}{\langle\lambda x.X\rangle(t) \to \langle\lambda x.X'\rangle(t)}\ +Lambda$$

$$\frac{\neg X \to \neg X'}{\neg\langle\lambda x.X\rangle(t) \to \neg\langle\lambda x.X'\rangle(t)}\ -Lambda$$

Quantification For new variables $x_1, \ldots, x_n$,

$$\frac{\neg\varphi(x) \to \neg\varphi_1(x) \quad \ldots \quad \neg\varphi(x) \to \neg\varphi_n(x)}{\neg(\forall x)\varphi(x) \to \neg[\varphi_1(x_1) \wedge \ldots \wedge \varphi_n(x_n)]}\ -Quant$$

Binding For x not free in X,

$$\frac{X \to X'}{X \to \langle\lambda x.X'\rangle(t)}\ +Bind \qquad \frac{\neg X \to \neg X'}{\neg X \to \neg\langle\lambda x.X'\rangle(t)}\ -Bind$$

Definition 5.2 We say Y is a *modal Herbrand expansion* of X provided there is a formula X^* that is a validity functional form of X, Y is a modal Herbrand transform of X^*, and Y is closed.

Example 5.3 Consider modal formula $\Box(\forall x)\neg(\forall y)Rxy \supset (\forall x)\Box\neg(\forall y)Rxy$. For it, $\Box(\forall x)\neg\langle\lambda y.Rxy\rangle(fx) \supset \langle\lambda x.\Box\neg(\forall y)Rxy\rangle(c)$ is a validity functional form, and a closed modal Herbrand transform of it is:

$$\langle\lambda z.\Box\langle\lambda x.\neg\langle\lambda y.Rxy\rangle(fx)\rangle(z)\rangle(c) \supset \langle\lambda x.\Box\neg\langle\lambda y.Rxy\rangle(fx)\rangle(c).$$

Consequently, this is a modal Herbrand expansion of

$$\Box(\forall x)\neg(\forall y)Rxy \supset (\forall x)\Box\neg(\forall y)Rxy.$$

6 Results

First, a result that is the direct analog of the classical Herbrand theorem.

Theorem 6.1 *A closed modal formula φ is valid (in all modal models) if and only if some modal Herbrand expansion of φ is valid.*

Consider Example 5.3 again. As a matter of fact, we began with a valid formula, $\Box(\forall x)\neg(\forall y)Rxy \supset (\forall x)\Box\neg(\forall y)Rxy$, and the modal Herbrand expansion we produced for it is likewise valid.

The classical Herbrand theorem reduces a first-order validity problem to a sequence of propositional problems, and the theorem above does not quite do this. It is true that a modal Herbrand expansion is quantifier-free, but validity for such a formula is not entirely a propositional issue. The difficulty is that modal Herbrand expansions still involve predicate abstractions, and their semantics still requires the first-order machinery of non-empty domains to characterize. Nonetheless, part of the point of a Herbrand reduction to a sequence of propositional problems is that we then have a sequence of decidable problems. This aspect carries over to the modal setting.

Theorem 6.2 *Validity for quantier free closed modal formulas—in particular, for modal Herbrand expansions—is decidable.*

Proofs of both of these theorems makes essential use of tableau methods. We have a modal tableau system that is sound and complete, even when predicate abstraction and non-rigid designators are present. A valid modal Herbrand expansion for a formula φ can be extracted from a tableau proof of φ. There is not space enough here to present this work—we refer to [8] for details.

7 Conclusions

Versions of Herbrand's theorem for modal logic have appeared before [11, 1], though the particular approaches were quite different. Predicate abstraction is, we believe, not only the key to a natural treatment, but is, in a sense, the missing piece of machinery that first-order modal logic needs. It was used in [4] to give a Herbrand-like theorem, but details of the expansion were very different.

Predicate abstraction is basic to first-order modal logic. By using it, a satisfactory treatment of equality in a modal setting can be given, definite descriptions can be dealt with properly, and traditional problems like the morning star/evening star puzzle become straightforward issues. That it makes Skolemization and a Herbrand theorem possible is more evidence for its essential nature. We hope that, as time goes on, it will become a familiar part of the toolkit of a modal logician.

References

[1] M. Cialdea and L. Cerro. A modal Herbrand property. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 32:523–530, 1986.

[2] B. Dreben and W. D. Goldfarb. *The Decision Problam. Solvable cases of quantificational formulas.* Advanced Book Program. Addison-Wesley, Reading, MA, 1979.

[3] M. C. Fitting. An epsilon-calculus system for first-order S4. In W. Hodges, editor, *Conference in Mathematical Logic, London '70*, pages 103–110, 1972. *Springer Lecture Notes in Mathematics, No. 255.*

[4] M. C. Fitting. A modal logic analog of Smullyan's fundamental theorem. *Zeitschrift für mathematische Logik und Gründlagen der Mathematik*, 19:1–16, 1973.

[5] M. C. Fitting. A modal logic epsilon-calculus. *Notre Dame Journal of Formal Logic*, 16:1–16, 1975.

[6] M. C. Fitting. Modal logic should say more than it does. In J.-L. Lassez and G. Plotkin, editors, *Computational Logic, Essays in Honor of Alan Robinson*, pages 113–135. MIT Press, Cambridge, MA, 1991.

[7] M. C. Fitting. Basic modal logic. In D. M. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 1, pages 368–448. Oxford University Press, 1993.

[8] M. C. Fitting. A modal Herbrand theorem. To appear in *Fundamenta Informaticae*, 1996.

[9] J. Herbrand. Investigations in proof theory. 1930. English translation in [10] and in [15].

[10] J. Herbrand. *Logical Writings*. Harvard University Press, Cambridge, MA, 1971. Translation of *Écrits logiques*, Jean Van Heijenoort, Ed., Presses Universitaires de France, Paris.

[11] K. Konolige. *A deduction model of belief*. Morgan Kaufman, Los Altos, CA, 1986.

[12] J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM*, 12:23–41, 1965.

[13] R. Stalnaker and R. Thomason. Abstraction in first-order modal logic. *Theoria*, 34:203–207, 1968.

[14] R. Thomason and R. Stalnaker. Modality and reference. *Nous*, 2:359–372, 1968.

[15] J. van Heijenoort. *From Frege to Gödel*. Harvard University Press, Cambridge, MA, 1967.

[16] A. N. Whitehead and B. Russell. Incomplete symbols: Descriptions. In *[15]*, pages 217–223. 1910.

[17] A. N. Whitehead and B. Russell. *Principia Mathematica*. Cambridge University Press, Cambridge, 2nd edition, 1927. First edition, 1910.