Part VIII

FORMAL SCIENCES

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192990 Published online by Cambridge University Press

On the Logic of Interrogative Inquiry

Jaakko Hintikka and Stephen Harris

Florida State University

In earlier publications Jaakko Hintikka has introduced the *interrogative model of inquiry* and studied some of its applications.[1] At its simplest, the interrogative model takes the form of a game between a player known as *the Inquirer* and a source of information we call *Nature*. The inquirer is trying to derive a conclusion **C** from a given set of premises **T** by standard deductive means augmented by additional information gained from Nature. (We can think of **C** as a set of formulas related disjunctively to each other.) The inquirer may obtain this additional information by means of questions put to Nature. The answers to these questions, when available, are then added to the premise set **T**. When a question is answerable but no true answer is available, Nature may respond with a false answer.

The deductive steps employed are subject to certain limitations. For instance, if the derivation is being carried out in a Beth-style *tableau* system, we require that the deductive steps obey the subformula principle.[2] Also, we disallow in general the arbitrary addition of tautological disjunctions to the left-hand column of the *tableau*. (The addition of contradictory conjunctions to the right column is allowed if one half of the contradiction occurs in the *subtableau*).[3] In addition, the movement of a formula by a negation rule from the right column to the left one is prohibited. (The converse movement from the left column to the right one is admissible.)

Likewise, limitations are placed upon the conditions under which questions may be put to Nature. One is to allow a question only if the presupposition of the question is present in the left-hand column of the *subtableau*. Additionally, one may further limit the set of possible questions by allowing only those questions whose answers are of a certain quantificational complexity.

An interrogative game proceeds relative to a particular model **M** of the given language, which is assumed to be a first-order language. When **C** is derivable from **T** in **M**, we say that **C** is a *model consequence* of **T** in **M**, and express this by

(1) $\mathbf{M:T \vdash C}$

Clearly, if no questions are allowed, **M** becomes irrelevant and the relation of model consequence reduces to the usual relation of deductive consequence.

PSA 1988, Volume 1, pp. 233-240

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192990 Published online by Cambridge University Press

Conversely, with certain qualifications it can be shown that a sentence is true in **M** if it can be established by means of an unrestricted interrogative procedure without any initial premise **T**. Thus in a sense both the concept of deductive consequence and the concept of truth are special cases of the relation (1).

The interest that the interrogative model has is largely due to the fact that it enables us to study strategies of scientific inquiry and perhaps even strategies of discovery in the form of strategies of question selection. Moreover, it turns out that the principles of interrogative strategy selection are closely related to the principles governing the choice of deductive strategies. This lends a new significance to the old phrases "*logic* of science" and "*logic* of scientific discovery".

Because of this partial strategic parallelism between empirical (interrogative) inquiry and deductive reasoning, it is of crucial importance to study the precise extent to which metatheoretical results obtained in logical theory have counterparts in interrogative inquiry. The purpose of this paper is to launch such a study. It will be shown that some of the most important results concerning the deductive consequence relation can be extended to the relation of model consequence. One such result is Craig's interpolation theorem.[4] This result has as an important consequence Beth's theorem relating implicit and explicit definability.[5] We shall extend Craig's result to the interrogative model, and then proceed to define the interrogative counterpart to definability. This notion turns out to be in effect what is commonly called identifiability.[6] We shall then prove a result analogous to Beth's relating implicit and explicit identifiability.

Interpolation Theorem: Assume that **M:T⊢C**, and that (a) **T** is consistent; (b) not **⊢C**; and (c) **C** does not contain individual constants. (It may contain dummy names, i.e., free individual variables.) Then there is a formula **I** such that: (i) each nonlogical constant of **I** occurs in both **T** and **C**, except for a finite number individual constants $\mathbf{b_1, b_2, \ldots, b_n}$ which name members of the domain **Do(M)** of **M**. These constants will have been introduced by Nature's answers to wh-questions in the course of the derivation of **C** from **T** in **M**. (ii) **M:T⊢I**. (iii) **I⊢C**.

Comment: We are not making any assumptions as to what kinds of answers Nature will give, as long as their totality remains the same throughout the argument.

Proof: By induction on the length **l** of a derivation of **M:T⊢C**. A basis for induction is obtained from case **l=0**.

Let **l=0**. Then the *tableau* corresponding to **M:T⊢C** is closed without the application of any rules. By the *tableau* closure conditions, there is a formula **F** such that (1) **F** occurs both in **T** and in **C**; or (2) **F** and **-F** occur in **T**; or (3) **F** and **-F** occur in **C**. But (2) is ruled out by (a) and (3) by (b). In the case (1), **F** serves as the interpolation formula.

In general, we assume the result for **l=n**; and prove it for **l=n+1**. We shall here consider only a few sample cases. It will be clear from these how to extend the proof to the remaining cases.

Sample Case 1: The first step in the derivation of **M:T⊢C** is the rule for a disjunction **(F ∨ G)** in the left column. Two *subtableaux* result, each of length **n** or less, like the original except that by **F** and **G**, respectively, has been added to their left column. By hypothesis, each *subtableau* has an interpolation formula. Let us call them **I'** and **I"**, respectively. Then **I = (I' ∨ I")** will serve as an interpolation formula for the derivation of length **l=n+1**, as one can easily show.

https://doi.org/10.1086/psaprocbienmeetp.1988.1.192990 Published online by Cambridge University Press

Sample Case 2: The first step in the derivation $M{:}T\vdash C$ is an universal instantiation of a formula $\forall xS[x]$ in the left column with respect to an individual symbol appearing only in the right column. Then this symbol must be a dummy name **z**, for **C** did not initially contain individual constants. For this reason, any individual constant occurring in the right column also occurs in the left one. Let the universal formula be $\forall xS[x]$ and its instantiation $S[z]$. But by the inductive hypothesis, there is a derivation of length **n** of $M{:}T\cup\{S[z]\}\vdash C$ and hence an interpolation formula $J[z]$ for this derivation. Then it can be shown that $\forall xJ[x]$ serves as an interpolation formula for the longer derivation. Obviously it contains the right symbols, and clearly, $\forall xJ[x]\vdash C$, for $\forall xJ[x]\vdash J[z]$ and $J[z]\vdash C$. We can also show that $M{:}T\vdash\forall xJ[x]$ as follows: Start a *tableau* with **T** in the left column and $\forall xJ[x]$ in the right one. It can be built further as follows:

T	∀xJ[x]
∃x S[x]	J[z]
(member of T)	(instantiation)
S[z]	--
(instantiation)	--
--	--
--	
--	

By the inductive hypothesis we can close this *tableau*, which shows that $M{:}\ T\vdash\forall\ xJ[x]$ and completes the proof of this subcase.

Sample Case 3: The first step of the derivation is an application of existential instantiation to a formula in the right column with respect to an individual symbol occurring only in the left column. Let the existential formula be $\exists xS[x]$ and the symbol a constant **b**. Then there is a derivation of length **n** of $M{:}T\vdash C\cup\ \{S[b]\}$, and by the inductive hypothesis there is an interpolation formula $J[b]$ for this derivation. Then $\exists xJ[x]$ will serve as an interpolation formula for the derivation of length **n+1**. Obviously, it has the right symbols, and since by the inductive hypothesis $M{:}T\vdash J[b]$, we now have $M{:}T\vdash\exists xJ[x]$. Also, $\exists xJ[x]\vdash C$, for by assumption $J[b]\vdash C$ where **b** does not occur in **C**.

The case in which the instantiation is with respect to a dummy name **z** is dealt with similarly.

The remaining deductive cases proceed similarly. We will now address the two interrogative cases.

Sample Case 4: The first step of the derivation is a propositional question whose presupposition is $(F \vee G)$ resulting in the addition of (say) **F** to **T**. Then by hypothesis there is an interpolation formula **I** for the resulting derivation of length **n**. This **I** satisfies conditions (i)-(iii) for the derivation of length **n+1**, and thus is a suitable interpolation formula for this derivation.

Sample Case 5: the first step in the derivation is an application of the rule for wh-questions where the presupposition of the question is $\exists xS[x]$. The result is the addition of $S[b]$ to the left column. By hypothesis, there is an interpolation formula $I[b]$ for the resulting derivation of length **n**. This $I=I[b]$ satisfies conditions (i)-(iii) for the derivation of length **n+1**, as one can easily show.

Since the only case in which **I** contains nonlogical constants not shared by **T** and **C** is this last case, it follows that any such constant is a name introduced by answers to wh-questions. This completes our sketch of the proof of the interpolation theorem.

Next we shall consider interrogative counterparts to definability, i.e., different concepts of identifiability. A one-place predicate **P** is said to be *implicitly identifiable* in **M** on the basis of **T** just in case

$$\mathbf{M{:}T\ P(b)} \quad \text{or}$$
$$\mathbf{M{:}T\text{-}P(b)}$$

for each **b** naming a member of **Do(M)** by means of questions and answers which do not contain **P**. A one-place predicate **P** is said to be *explicitly identifiable* just in case

(2) $\mathbf{M{:}T \vdash \forall x(P(x) \leftrightarrow D[x,\ a_1, a_2, \ldots, a_n])}$

Likewise where **P** is not found in **D** and $\mathbf{a_1, a_2, \ldots a_n}$ are names of members of **Do(M)**.

In order to study these concepts, we shall first prove the following lemma:

Lemma: Let **T'** be like **T** except that occurrences of **P** in **T** are everywhere replaced by **P'** in **T'**. Let **M'** be like **M** except that **P'** is interpreted in it in the same way as **P**. Furthermore, let the tautological formula

$$\mathbf{V = \ \forall x(P(x) \leftrightarrow P'(x)) \lor \text{-} \forall x(P(x) \leftrightarrow P'(x))}$$

be available to be added to the left column. Then if **P** is implicitly identifiable in **M** on the basis of **T**, i.e., if

(3) $\mathbf{M{:}TP \vdash (b)}$ or $\mathbf{M{:}T \vdash \text{-}\ P(b)}$

for each **b**, we have

(4) $\mathbf{M{:}TUT' \vdash \forall x(P(x) \leftrightarrow P'(x))}$.

Comment: Both in (3) and in (4) it is assumed that Nature answers only questions that do not contain **P**. We shall also assume that Nature will answer all wh-questions with a quantifier-free answer.

Proof: Construct a *tableau* for (4). Add **V** to the left column and apply the disjunction rule to it, adding $\mathbf{\forall x(P(x) \leftrightarrow P'(x))}$ to one *subtableau* and $\mathbf{\text{-}\forall\ x(P(x) \leftrightarrow P'(x))}$ to the other. The first closes immediately. In the other we in effect have $\mathbf{\exists x((P(x)\ \&\ \text{-}P'(x)) \lor (\text{-}P(x)\ \&\ P'(x)))}$. This can serve as the presupposition of a wh-question which by assumption yields an answer of the form

(5) $\mathbf{(P(b)\ \&\ \text{-}P'(b)) \lor (\text{-}P(b)\ \&\ P'(b))}$.

Apply the disjunction rule to (5). By symmetry, it suffices to consider only the *subtableau* containing **P(b) & -P'(b)**, hence **P(b)** and **-P'(b)**. Now because of the implicit identifiability of **P** in **M**, either **P(b)** and **P'(b)** or **-P(b)** and **-P'(b)** can both be derived interrogatively in **M'** from **T U T'**. This closes the *tableau* and proves the lemma.

Extended Beth's Theorem: Assume that Nature will answer all wh-questions with quantifier-free answers. If either $\mathbf{M{:}T \vdash P(b)}$ or $\mathbf{M{:}T \vdash \text{-}P(b)}$ for each **b** naming a member of **Do(M)**, then $\mathbf{M{:}T \vdash \forall x(P(x) \leftrightarrow D[x,\ a_1,\ a_2, \ldots, a_n]}$, where **P** does not occur in **D** and $\mathbf{a_1,\ a_2, \ldots, a_n}$ are a finite number of names of members of **Do(M)**, in both cases without questions that contain **P**.

In brief, if a predicate is implicitly identifiable, it is explicitly identifiable. In the proof of the extended Beth's theorem, M:T⊢C will express interrogative derivability by means of questions and answers that will not involve P or P'.

Comment: We have to make here the auxiliary assumption that tautological disjunctions (S ∨ -S) can be added to the left column if S or -S already occurs in it.

Proof: Let P', T' and M' be as before. By the lemma above, (4) holds, therefore also

(6) $M':T \cup T' \vdash \forall x(P(x) \leftrightarrow P'(x))$. Hence

(7) $M':T \cup \{P(x)\} \vdash (T' \rightarrow P'(x))$.

(Moving formulas between the two columns in this way is made possible by the auxiliary assumption mentioned above.) But by the interpolation theorem (7) implies that there is an interpolation formula I[x] with x as its only free variable such that

(8) $M':T \cup \{P(x)\} \vdash I[x]$ and

(9) $I[x] \vdash (T' \rightarrow P'(x))$.

Here I[x] may also contain names of individuals introduced by Nature's answers to wh-question in establishing (6). But (8) entails

(10) $M:T \vdash (P(x) \rightarrow I[x])$

and (9) entails

(11) $I[x] \vdash (T \rightarrow P(x))$, hence also

(12) $T \vdash (I[x] \rightarrow P(x))$.

Together (10) and (12) entail

(13) $M:T \vdash (P(x) \leftrightarrow I[x])$ and hence also

(14) $M:T \vdash \forall x(P(x) \leftrightarrow I[x])$.

The interpolation formula I[x] therefore serves as the desired definiens, proving our Extended Beth's Theorem.

Our extension of Beth's theorem is easily seen to hold for other constants than one-place predicates.

The interest of the interpolation theorem is illustrated by the way in which the complexity of the interpolation formula I reflects the complexity of the process of showing that M:T⊢C. As in the deductive case, the quantificational complexity of I (roughly, the number of layers of quantifiers in I) is determined by the number of those applications of deductive instantiation rules (cf. sample cases 2-3 above) which introduce a new individual symbol to one of the columns, as can be seen by reviewing the proof of the interpolation theorem. Now we can likewise see that the interrogative complexity of the questioning process which establishes that M:T⊢C determines the number of the parameters (names of members of Do(M)) occurring in I. Similar remarks can be addressed to our extension of Beth's theorem.

It is important to notice that in the interpolation theorem the model consequence that **M**:T$\vdash$**I** can be established by a questioning process with the same restrictions on allowable questions as were used for the derivation corresponding to the proof that **M**:T$\vdash$**C**. Likewise, in the theorem concerning implicit and explicit identifiability (14) can be established by the same kind of questioning as was needed to establish (6).

Our main results are relative to a given model **M**, but they can be generalized to the entire space of models e.g., as follows:

Absolute Interpolation Theorem: Assume that the conditions of the interpolation theorem hold for **T** and **C** in each model **M** of **T**. Then there exists a deductive interpolation formula **I** of the form $\exists x_1, \exists x_2, \ldots \exists x_n$ $I[x_1, x_2, \ldots, x_n]$ between **T** and **C**. Again, the interest of this result stems from the fact that the structure of the interpolation formula reflects the structure of the interrogative argument which establishes **C**.

This version of the theorem is not trivial. For reasons of space, its proof cannot be presented here. It constitutes an important link between the relations of deductive consequence and model consequence.

Likewise, we have the following absolute version of Beth's theorem:

If **P** is implicitly identifiable on the basis of **T** in each model **M** of **T**, then there are **k** and D_i $(i=1, 2, \ldots, m)$ such that

$$(8)\quad T \vdash \exists x_1, \exists x_2, \ldots \exists x_n \bigvee_{i=1}^{m} \forall y(P(y) \leftrightarrow D_i[y, x_1, x_2, \ldots, x_n].$$

This result is in effect known from the theory of definability.[7]

The precise consequences of our results for the philosophy of science will have to be spelled out separately. A few examples may nevertheless be in order. Our interrogative counterpart to definability, viz. the notion of identifiability, is an especially promising concept. The identifiability of (say) a one-place predicate **P** in **M** on the basis of **T** is clearly very close to what philosophers of science have meant by its observability or measurability.[8] The dependence of this concept on **T** vindicates the views of those philosophers who have argued that observability is relative to an underlying theory. In a sense, it may be suggested, one can in this way prove the famous thesis of the theory-ladenness of observations.

Moreover, it is not hard to see that identifiability is what many philosophers and methodologists have meant by definability.[9] For instance, in the onetime controversy between H. A. Simon and Patrick Suppes concerning the definability of concepts occurring in certain scientific theories (especially in classical mechanics), the parties were literally speaking of different things, in that Simon was speaking of identifiability while Suppes was speaking of ordinary definability.

In other respects, too, earlier discussions of the definability of various scientific concepts are put in a new light by the introduction of an explicit concept of identifiability.

It is also of interest to note that the presence of a disjunction in our absolute version of Beth's Theorem for identifiability. What it means is that in different models (applications) of an explicit unambiguous theory, a certain concept may have to be identified differently. This shows that there is a fallacy in the arguments of those philosophers of science who have argued from a difference between the ways a given concept is identified in two theories to the incommensurability of those theories.

But are not our results made unrealistic by the fact that we are restricting our attention to first-order languages? Most philosophers of science would apparently think so, but they are proved wrong by the format of such results as our interpolation theorem. It concerns *each* model of a certain kind. Now what strengthening first-order logic means is eliminating somehow some of the former models. But such an elimination usually makes little difference to results concerning each model of a certain sort. Thus there even seems to be a general methodological moral here for philosophers of science: don't underestimate the uses of the model theory of first-order logic.

Notes

[1]See e.g., Hintikka (1984) and (1988b). Important applications are presented *inter alia* in Hintikka (1988c and in (1988a).

[2]For the *tableau* method, see Beth (1955).

[3]The admissibility of tautological premises of the form (F v -F) to the left column (and analogously for the right column) makes a difference to the available model consequences, unlike the deductive case. For the interpretation of these apparently tautological premises, see Hintikka (1988b).

[4]Craig (1957). In 1960, J.W. Addison wrote that "the Craig interpolation theorem begins to emerge as the most important result in pure logic since 1936." (See Addison, 1962.)

[5]Beth (1953).

[6]This notion of identifiability plays a major role *inter alia* in econometrics and systems theory. Cf., e.g., Koopmans (1949); Hsiao (1983). Our extension of Beth's theorem to identifiability instead of definability is a first step toward a general logical theory of identifiability.

[7]It is closely related to the Chang-Makkai theorem. See Rantala (1977).

[8]Cf. here Raimo Tuomela (1973).

[9]See here Simon (1947, 1959 and 1970); Suppes (1957); McKinsey, Sugar and Suppes (1953); Jammer (1961) especially chapter 9.

References

Addison, J.W. (1962). "The Theory of Hierarchies." In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*. Edited by E. Nagel, P. Suppes and A. Tarski. Stanford: Stanford Universtiy Press. Pp. 26-37.

Beth, E.W. (1953). "On Padoa's Method in the Theory of Definition." *Indigationes Mathematicae* 15: 330-339.

______. (1955). "Semantic Entailment and Formal Derivability." *Mededelingen van de Koninklijke Nederlandse Akademie van Wetenschappen, Afdeling Letterkunde.* Volume 18, 13: 309-342.

Craig, W. (1957). "Linear Reasoning." *Journal of Symbolic Logic* 22: 250-268.

Hintikka, J. (1984). "The Logic of Science as Model-Oriented Logic." *PSA 1984*, Volume 1. Edited by P. Asquith and P. Kitcher. East Lansing, MI: Philosophy of Science Association, pp. 177-185.

________. (1988a). "The Concept of Induction in the Light of the Interrogative Model of Inquiry." *Proceedings of the Pittsburgh Colloquium in the Philosophy of Science.*

________. (1988b). "Knowledge Representation and the Interrogative Approach to Inquiry." Edited by M. Clay and K. Lehrer.

________. (1988c). "What is the Logic of Experimental Inquiry?" *Synthese.*

Hsiao, C. (1983). "Identification." In *Handbook of Econometrics*. Edited by Z. Griliches and M.D. Intriligator. Volume 1. Amsterdam: North Holland

Jammer, M. (1961). *Concepts of Mass in Classical and Modern Physics*. Cambridge: Harvard University Press.

Koopmans, T.C. (1949). "Identification Problems in Economic Model Construction." *Econometrica* 17: 125-144.

McKinsey, J.C.C., Sugar, A.C. and Suppes, P. (1953). "Axiomatic Foundations of Classical Particle Mechanics." *Journal of Rational Mechanics and Analysis* 2: 253-272.

Rantala, V. (1977). "Aspects of Definability." *Acta Philosophica Fennica* 29: 2-3. Helsinki: *Societas Philosophica Fennica.*

Simon, H.A. (1947). "The Axioms of Newtonian Mechanics." *Philosophical Magazine*, Series 7, Volume 33, pp. 888-905

________. (1954). "The Axiomatization of Classical Mechanics." *Philosophy of Science* 21: 340-343.

________. (1959). "Definable Terms and Primitives in Axiom Systems." In *The Axiomatic Method.* Edited by L. Henkin, P. Suppes and A. Tarski. Amsterdam: North Holland, pp. 443-453.

________. (1970). "The Axiomatization of Physical Theories." *Philosophy of Science* 37: 16-26.

Suppes, P. (1957). *Introduction to Logic.* Princeton: Van Nostrand.

Tuomela, R. (1973). *Theoretical Concepts.* Vienna and New York: Springer-Verlag.