Chapter 16

Definability and Interpolation in Philosophy

Johan van Benthem
University of Amsterdam, The Netherlands
Stanford University, USA
Tsinghua University, China

Abstract

This paper offers a light tour of occurrences of Craig's interpolation theorem and Beth's definability theorem in philosophy. There is no coherent research program with a central agenda around this theme in the literature, but extending the discussion in [9], we trace some lines and identify some issues in establishing contacts at this interface between the realms of logic and the philosophy of logic, of science, of language and information, epistemology, and metaphysics. Our presentation can also be viewed as a case study for more general aspects of the logic–philosophy interface.

Contents

1 The Broad Ideas **592**
2 What is a Logical System? **594**
3 Supervenience **598**
4 Determinism **600**
5 Questions **602**
6 Dependence **603**
7 Information and Generalized Consequence **605**
8 Theoretical Terms in Scientific Theories **608**
9 Modularity in Knowledge and Belief **609**
10 A Historical Precursor: Bolzano **611**
11 Conclusion **612**
Acknowledgments **614**
References **614**

© by the author(s)
In Balder ten Cate, Jean Christoph Jung, Patrick Koopmann, Christoph Wernhard, and Frank Wolter, editors. *Theory and Applications of Craig Interpolation*. Ubiquity Press, 2026. To appear.
https://www.doi.org/10.5334/bdg.p

1 The Broad Ideas

Beth's Definability Theorem says that, in first-order theories, implicit definability of vocabulary is equivalent with explicit definability. More precisely, given a first-order theory T in a language with vocabulary $\boldsymbol{L}$ plus some predicate letter P, implicit definability says that, given two models M, M' for T which are identical qua domain of objects and interpretation of the vocabulary of $\boldsymbol{L}$, the predicate interpretations P^M and $P^{M'}$ must also be identical. Put more loosely, fixing the interpretation of $\boldsymbol{L}$ in a model for T automatically also fixes that of P. In contrast, explicit definability says that there exists a first-order formula $\varphi(\boldsymbol{x})$ using only the vocabulary $\boldsymbol{L}$ with a tuple of variables $\boldsymbol{x}$ of the arity of the predicate P such that the theory T implies the formula $\forall \boldsymbol{x}\,(P\boldsymbol{x} \leftrightarrow \varphi(\boldsymbol{x}))$.

▶ **Theorem 1.** *For first-order theories, implicit and explicit definability are equivalent.*

Clearly, explicit definability implies implicit definability, but Beth's achievement was to show the converse, which is not at all trivial for first-order logic [18].

The background to Beth's Theorem was his interest in logic as a joint study of definability and proof, a tandem view of the history of the discipline since Antiquity on which he elaborated in various publications, in particular his magnum opus [19] which also contains his more general philosophical views based on logical foundational research. In particular, Beth arrived at the Definability Theorem by reflecting on Padoa's Method for disproving explicit definability in the foundations of geometry. This combination of logic and philosophy of science was characteristic for the 1950s. Beth founded the "Instituut voor Logica en Grondslagen van de Exacte Wetenschappen" (Institute for Logic and Philosophy of the Exact Sciences), the precursor of the current Institute for Logic, Language and Computation in Amsterdam, and equally significantly, around the same time Beth's correspondent and ally Alfred Tarski founded the still existing Group in Logic and Methodology of the Exact Sciences at the University of California at Berkeley.

For our later discussion, it is convenient to rephrase the earlier formulation of Beth's notion to the following equivalent version of implicit definability:

> If two models M, N both make $T(\boldsymbol{L}, P)$ true and f is an $\boldsymbol{L}$-isomorphism between M and N, then f is also a P-isomorphism.

This isomorphism-lifting version of Beth's Theorem generalizes to weaker structure invariances such as potential isomorphism,[1] a notion that will return in Section 7 below.

Various notions of implicit and explicit determination in the above vein can be found in many areas of philosophy.

---

[1] A *potential isomorphism* between two models M, N is a family $\boldsymbol{F}$ of finite partial isomorphisms between these two models satisfying the following Back-and-Forth conditions. For every $f \in \boldsymbol{F}$ and $a \in M$ there is a $b \in N$ such that $f \cup \{(a, b)\} \in \boldsymbol{F}$, and the same condition holds in the opposite direction.

Next let us consider the Craig Interpolation Theorem for first-order logic, [27, 28], using the following simple formulation displaying just a few predicates for convenience.

▶ **Theorem 2.** *If formula $\varphi(P, Q)$ in predicate vocabulary $\{P, Q\}$ semantically implies formula ψ with vocabulary $\{Q, R\}$, then there exists a first-order formula α using only the shared vocabulary $\{Q\}$ such that $\varphi(P, Q) \models \alpha(Q) \models \psi(Q, R)$.*

More elaborate versions of the Interpolation Theorem, of which there exist quite a few by now, cf. [46] for a sample, give more detailed information about how the shared predicate Q occurs syntactically in α in terms of how it occurs in the formulas φ and ψ. See [24] in this book for more details and in-depth results.

Note. In various chapters of the present volume, a distinction is made between plain and deductive interpolation as versions of the preceding result which are not always equivalent for logical systems. While important, this distinction will not be highlighted in what follows.

Craig Interpolation implies Beth Definability for first-order logic by a well-known simple argument. The converse does not hold. For instance, the natural Guarded Fragment of the first-order language satisfies the latter property, but not the former, cf. [40, 39].[2] For more information on these matters, see also [42] in this book.

By an easy argument, Interpolation also implies the stronger Projective Beth Theorem which says the following, in a simplified notation. If the theory $T(A, B, C) \cup T(A, B', C')$ implies $\forall x\, (Cx \leftrightarrow C'x)$, then $T(A, B, C)$ implies an explicit definition for C by a first-order formula using the vocabulary A only. We will meet this stronger version occasionally.

We will also encounter applications of Interpolation in philosophy, though at first glance it may look like a more technical system property. For greater unity in philosophical applications, it would be nice if one could capture the surplus of interpolation over Beth in a natural and appealing manner. We will not resolve this matter here.[3]

We are going to show how the above two circles of ideas return in philosophy. Section 2 is about the role of the two theorems in defining the notion of a logical system, a live topic in the philosophy of logic. After that, the focus is on Beth's Theorem in Sections 3–6 on supervenience, determinism, and reasoning about questions, culminating in the suggestion that the common logical thread in philosophical applications of Beth's Theorem is the role of implicit and explicit *dependence*.

Then the focus of our narrative shifts to Craig's Theorem and its role in highlighting the importance of *vocabulary* in reasoning and theory construction, in Sections 7–9 on generalized notions of logical inference, the architecture of scientific theories, and the modular structure

---

[2] The first of these papers makes a case that the Guarded Fragment does satisfy the right version of interpolation from a 'modal' perspective – an issue that we do not go into here.

[3] One candidate for a mathematical notion that might throw light on this comparative surplus is the *amalgamation property* in varieties of algebras, perhaps in a generalized category-theoretic setting.

of beliefs. Finally, Section 10 flags some connections with the work of Bernard Bolzano, a creative and somewhat neglected 19th century pioneer in the emergence of modern logic.

2 What is a Logical System?

One immediate road into probing the importance of our two theorems is their relevance to a recurrent issue in the philosophy of logic, and just as well in logic itself.

What is a logical system?, or in a more refined version:
What is a good, or well-designed, logical system?

This is not an abstract quibble, at stake here is the proper view of what modern logic in its prolific system-building guise is about. What do our two theorems have to say here?

Our discussion will focus on the Interpolation Theorem, but the Definability Theorem plays a similar role. Satisfying a Craig-style interpolation theorem is sometimes seen as a key desideratum in the folklore, as it expresses a form of 'expressive completeness' of a logic for its language, as being able to formulate definitions and state crucial 'transitions' from premises to conclusions. Even so, this intuition is not easy to make precise beyond just restating what the theorems say, and among formal approaches that go further, I only know of [33] and an earlier proof-theoretic analysis in [58]. A general semantic way of seeing the centrality of Beth and Interpolation might go via their role in Abstract Model Theory, where Beth's Definability Theorem seems to follow from Lindström's Theorem for first-order logic under quite general conditions, though the case of Craig is less clear – cf. [55] in this volume.[4]

But here is another way into the above question, restated as follows with a flavor of the crucial notion of invariance that we already noted in connection with Beth's Theorem:

When are two logical systems the same?

One answer, found in both the mathematical and the philosophical literature, is this. The two logics $\mathfrak{L}_1$, $\mathfrak{L}_2$ should have mutual faithful syntactic *translations* or embeddings, in the sense that the following equivalence holds for all formulas under some syntactic translation τ for the languages of these logics:

$$\varphi \models \psi \text{ in } \mathfrak{L}_1 \quad \text{iff} \quad \tau(\varphi) \models \tau(\psi) \text{ in } \mathfrak{L}_2$$

and vice versa in the opposite direction from $\mathfrak{L}_2$ to $\mathfrak{L}_1$ for some syntactic translation σ.

This is just a start, however, and there is no philosophical consensus on what more might be required for a good notion of system identity. Here is one possible richer view.

[4] One might question this criterion given the ubiquity of failures of interpolation in important systems found in philosophical logic, cf. [34] in this book. This issue will return in Section 3 below.

In Abstract Model Theory, logical systems come with both abstract sentences and abstract semantic structures, connected by a truth relation. Accordingly, the natural notion of translation then also involves a semantic transformation $\mathbb{T}$ of models running in the opposite direction to the above syntactic map τ, leading to the following contravariant equivalence:

$$M \models \tau(\varphi) \quad \text{iff} \quad \mathbb{T}(M) \models \varphi \qquad (\#)$$

This equivalence views translation as having two components: connecting languages and connecting models for two logics. This dual perspective is the basic mechanism of information flow in Situation Theory [5] – see Section 7 below for more on this framework.

The abstract mathematical theory of this setting involves adjoint maps in Chu Spaces [8], an elegant category-theoretic perspective that we will not pursue here.[5]

The broader issue at stake here is the following. Looking under the surface of the myriads of logical systems found in the literature, what are the underlying invariances and connections?[6] In other words, what is the coherence of the field of the logic? And again, do Craig and Beth have a bearing on this?

Without entering into a general philosophical discussion of these issues, here is one connection with our main topic in this paper. If Interpolation is made an integral part of the definition of an abstract logic, *what constraint is induced on translations* between logical systems? Here, to formulate the interpolation property in its proper generality, we need abstract notions of 'sentences', 'consequence', 'vocabulary', and 'occurrence'. Proper translations should now respect this basic structure, and make sure that interpolation is transferred from one logical system to any properly translation-related logical system, a concrete constraint which now becomes a more precise issue.[7]

Still, making Interpolation key to the identity of logical systems and translations between them does not establish just *why* one should consider this property so important. [12] conjectured that Interpolation is the key axiom in a *complete axiomatization of the meta-theory* of first-order logic, viewed as a two-sorted structure consisting of sentences and finite vocabulary sets with relations of logical consequence and 'occurrence'. The thinking behind this conjecture was that Craig interpolation seems the last major property of first-order logic discovered in modern history.[8] However, [43] showed that even the meta-theory of classical

---

[5] Faithfulness of the translation requires extra conditions, such as surjectivity of the transformation T.

[6] One famous connection is the faithful *Gödel translation* from intuitionistic propositional logic into the classical modal logic S4 which has led to a wide variety of fruitful contacts between the two systems.

[7] As a concrete illustration of this transfer, [11] show how a pair consisting of suitable homomorphic translation from the Guarded Fragment into the Modal Dependence Logic LFD mentioned in Section 9 below, plus a converse satisfiability reduction yields the Interpolation Theorem for LFD from the earlier-mentioned Beth Theorem for the Guarded Fragment. More such case studies are needed to see more clearly which constraints are imposed by transfer of Interpolation.

[8] Of course, this historical observation is no longer true when we consider the remarkable probabilistic properties of first-order logic discovered later, such as the Zero-One Law, [31].

propositional logic construed in this way is *non-axiomatizable*, and in fact admits of a faithful embedding of True Arithmetic. Thus, the central role of Interpolation does not show in this particular construal of the structure of logical systems – but perhaps all this shows is that we have made the notion of 'meta-theory' too concrete, importing peculiarities of the natural numbers and the complexity of their theory.

We conclude this section with another link to our main theme that may be of interest. The topic of translations also raises a nice connection with the Beth Definability Theorem, and in fact, it suggests a more general take on what this theorem said about definability.

Translation as Generalized Definability. One can think of an explicit *syntactic translation* τ satisfying the above-mentioned equivalence (#) as providing a *generalized explicit definition* for the basic notions of the logic $\mathfrak{L}_1$, not inside of its own vocabulary (the special case of definability at issue in Beth's Theorem) but in terms of formulas in another language: namely, that of the logic $\mathfrak{L}_2$. Accordingly, this time, the corresponding notion of *implicit definability* is not in just one model of the same similarity type, but it runs across possibly different models related by the semantic model-transforming component of the translation.

Here is a first intuitive formulation of this idea:

> Fixing the $\mathfrak{L}_2$-structure of a model M suffices for fixing the denotation of the basic predicates of the logic $\mathfrak{L}_1$ in all $\mathbb{T}$-matching models $\mathbb{T}(M)$ for transformations $\mathbb{T}$ satisfying some suitable conditions.

More precisely, let the transformation $\mathbb{T}$ in the above definability schema (#) be defined as a surjective function from $\mathfrak{L}_2$-models to $\mathfrak{L}_1$-models which also maps the domain of a $\mathfrak{L}_2$-model onto the domain of the matching $\mathfrak{L}_1$-model, i.e., $\mathbb{T}$ also maps objects to objects. *Please note:* this is surjectivity for objects in domains of $\mathbb{T}$-correlated models, not for $\mathbb{T}$ viewed as a map from models to models themselves.

It is easy to see that the schema (#) then implies the following cross-model condition, which we state here for unary predicates only, purely for convenience:

> If $\mathbb{T} : M \to N$ with $\mathbb{T}(b) = a$, M is $\mathfrak{L}_2$-isomorphic to M' with b matching b', and $M' \to N'$ with $\mathbb{T}(b') = a'$, then for any $\mathfrak{L}_1$-predicate P, $P^M(a)$ implies $P^{M'}(a')$.

Technical Elaboration. The setting of generalized defintitions invites a use of the Projective Beth Theorem, or rather, the Interpolation Theorem behind it. It applies to yield an explicit translation from an implicit one, provided the above transformation $\mathbb{T}$ is defined in first-order terms. Let this definition take the form of a first-order theory Σ with vocabulary $\{\boldsymbol{L}_1, \boldsymbol{L}_2, T, M, N\}$.[9] Here M, N stand for disjoint unary predicates, the $\boldsymbol{L}_1$-predicates (the

[9] For convenience, we use the earlier symbols M, N that denoted whole models as unary predicate letters for domains of models now, and the binary predicate letter T as a name for the transformation $\mathbb{T}$. The latter local notation will not clash with our use of T for theories in other passages in this paper.

vocabulary of the logic $\mathfrak{L}_1$) can only hold inside the subset defined by N, the $\boldsymbol{L}_2$-predicates (from the logic $\mathfrak{L}_2$) can only hold inside the subset defined by M, and where Σ only makes assertions about objects in $M \cup N$, even if a model itself is larger.[10]

Now we state the extended implicit definability introduced here as a first-order consequence in an enriched language, where the symbol P is a predicate from $\boldsymbol{L}_1$ (taken to be unary again, just for simplicity), a, a', b, b' are individual constants, Z stands for a binary relation:

$\Sigma(\boldsymbol{L}_1, \boldsymbol{L}_2, T, M, N)$, $Tb = a$, Pa (i), and
"Z is an $\boldsymbol{L}_2$-isomorphism between M and M'", $b\,Z\,b'$ (ii), and
$\Sigma(\boldsymbol{L}'_1, \boldsymbol{L}'_2, T', M', N')$, $T'b' = a'$ (iii)
$\models\ P'a'$.

This says that, for two pairs of $\mathbb{T}$-related models, with one source model $\boldsymbol{L}_2$-isomorphic to the other, the target models will agree on the interpretation of the $\boldsymbol{L}_1$-predicate P.

By a standard appeal to the Compactness Theorem for first-order logic, the total set of formulas occurring in the clauses (i), (ii) and (iii) here can be taken to be finite, and with some abuse of notation we can rewrite this consequence to one that holds between formulas:

$\Sigma(\boldsymbol{L}_1, \boldsymbol{L}_2, T, M, N) \wedge Tb = a \wedge Pa\ \models$ ("Z is an $\boldsymbol{L}_2$-isomorphism between M and M'"
$\wedge\, b\,Z\,b' \wedge \Sigma(\boldsymbol{L}'_1, \boldsymbol{L}'_2, T', M', N') \wedge T'b' = a') \rightarrow P'a'$.

Here the only overlap in vocabulary between the antecedent and consequent formulas is the set $\{M, \boldsymbol{L}_2, b\}$. Therefore, there is an interpolant $\delta(M, \boldsymbol{L}_2, b)$ in this vocabulary which just describes a property of the object b in M-type models.[11]

By some standard manipulation, we then see that the following two assertions hold:

(a) $\Sigma(\boldsymbol{L}_1, \boldsymbol{L}_2, T, M, N)\ \models\ Tb = a \rightarrow (Pa \rightarrow \delta)$.
(b) "Z is an $\boldsymbol{L}_2$-isomorphism between M and M'" $\wedge\, b\,Z\,b' \wedge \Sigma(\boldsymbol{L}'_1, \boldsymbol{L}'_2, T, M', N')\ \models$ $Tb' = a' \rightarrow (\delta \rightarrow P'a')$.

Now clearly, the identity relation on M satisfies all conditions for an $\boldsymbol{L}_2$-isomorphism between M and $M' = M$. Substituting this for the predicate Z in clause (b), while renaming some predicate variables, implies

$\Sigma(\boldsymbol{L}_1, \boldsymbol{L}'_2, T, M, N)\ \models\ Tb = a \rightarrow (\delta \rightarrow Pa)$.

Thus, in all, the theory Σ implies the explicit definition

$Ta = b \rightarrow (Pa \leftrightarrow \delta(a, b))$.

---

[10] This effect can be achieved by assuming that all quantifiers in Σ are bounded by $M \vee N$. The 'book-keeping conditions' stated here do not seem overly restrictive for defining a function between models.

[11] Reducing to this syntactic form may take some first-order manipulation, noting that quantification over N in the formula δ has no essential effect since no $\boldsymbol{L}_2$-predicates can be true there.

Finally, given $\boldsymbol{L}_2$-definitions for the predicates in $\boldsymbol{L}_1$, a standard inductive proof lifts this to an explicit *compositional translation* for all formulas of $\mathfrak{L}_1$. *Note:* Here the quantifier step of the formula induction uses the above assumption that the transformation $\mathbb{T}$ comes with a surjective function from the domain of M onto that of N.[12]

Of course, the preceding construal does not yet analyze the notion of 'translatability' as a relation between logical systems, which does not start from a given semantic transformation. For an abstract analysis of the latter sort, cf. [16].

Note that our discussion in this section has focused on isomorphism as the central semantic invariance. This is just one choice, and for instance, in Section 7, we will also use the weaker notion of 'potential isomorphism' introduced earlier. But in principle, our discussion would also apply to a whole spectrum of invariance notions, including still weaker ones such as *bisimulation*, in which case the results we cited would be talking about definability in *modal languages* rather than in full-fledged first-order languages.

In summary, this section has established some connections between our key theorems and issues in the philosophy of logic that may be worth reflecting on, but no definite answers were provided that could not be challenged. This sort of delicate contact is typical for the logic-philosophy interface, and we will encounter further instances as we proceed.

3 Supervenience

Our next topic starts from semantic invariance relations and their liftings. In philosophy, the notion of *supervenience* of a predicate Q on one or more predicates P is ubiquitous. Examples with broad sweep are the supervenience of our mental states on our physical/biological states, of the moral aspects on the factual aspects of a scenario, and so on.

For an excellent survey, we refer to the Encyclopedia article [44], from which we take just a few notions to make our main points. For a start, let us say that Q *supervenes on* P if

Whenever two situations are P-equivalent, they are also Q-equivalent.

We state this for situations, but we can also think of supervenience as comparing objects, or whatever relevant entity we want to consider. One reason why supervenience is attractive is that it expresses a 'harmony' qua properties between two different realms, but not necessarily brute-force *reducibility* of Q to P via some explicit definition 'explaining Q away'.

Of course, formulations like this require further precision to get more traction: which realm of situations or objects is considered, which notion of semantic equivalence is meant?

[12] Another construal of implicit definability would fix the semantic transformation $\mathbb{T}$ in the assertions about M, N and M', N'. Interpolation still applies in the above style, but we now also get undesirable definitions for the predicate P that do not just refer to $\boldsymbol{L}_2$, but also to the function $\mathbb{T}$, as in "Pa holds iff exactly 2 objects in M get $\mathbb{T}$-mapped onto a".

But even without specifying these parameters further, it should be obvious that supervenience has the same flavor as implicit definability in the sense of Beth's Theorem.[13]

In fact, more can be said, since several varieties of supervenience, weaker and stronger, have been proposed in the philosophical literature. Here are just two samples.

In a *local version*, the formulation that we started with compares equivalent or similar items inside one given world. Transposing this to a model-theoretic formulation, one way of expressing this more formally is that in a given model, P-automorphisms are Q-automorphisms – though other similarity relations could also be used, such as varieties of bisimulation.

But in a more *global version* across different worlds, the requirement might be as follows:

If object x in world w is P-similar to object y in world v, then x, y are also Q-similar.

In model-theoretic terms, this has precisely the flavor of the notion of implicit definability as introduced in Section 1, with similarity cashed out as isomorphism or potential isomorphism.

One recurrent issue in philosophy has been this: When does supervenience imply *reduction* in the sense of explicit definability? This is indeed what Beth's Theorem would say, but to apply this result to supervenience, one must first introduce a formal first-order language plus a background theory into the philosophical discussion. It might be debatable whether this methodological constraint fits the spirit of the philosophical literature, or at least, the felicity of this formalization move needs argumentation. As we said, failure of explicit definition is in fact one reason why supervenience is philosophically attractive as a non-reductionist position positing merely harmony between two realms without brute force reduction.

Two technical points can be made here. First, some logics with the Beth property might still be harmless from a reductionist perspective. For instance, in second-order logic an explicit definition can be extracted routinely from an implicit one by means of an impredicative existential quantification over the irrelevant vocabulary, trivializing Beth's Theorem. Such a 'reduction' restating the implicit definition seems harmless.

But more interestingly, one might try to take a fresh look at the relevant philosophical literature using non-higher-order non-classical logics that *lack the Beth property* as a way of avoiding reductionism. One example are the relevant logics lacking the Beth property discussed in [53], cf. also [34] in this book. But of course, this weakening of the base logic should also have some independent motivation: mere avoidance of the Beth property is hardly a positive virtue. After all, the function is there, but the vocabulary in logics of this kind does not allow us to define it explicitly. One immediate test for this way-out would be if the known technical counterexamples (sometimes produced by computer-assisted search) have any independent conceptual interest.[14]

---

[13] As we shall see in Section 9, supervenience is a form of abstract *dependence*, and in fact some authors have analyzed its general logic in these terms, cf. [32].

[14] Even so, somewhat perversely, could it be that undefinability (and once on this track, unprovability), i.e., keeping some things implicit only, might sometimes be regarded as positive features of a logical system?

Finally, there is also an alternative way of thinking that calls for different distinctions. By the above definition, supervenience as implicit semantic definability already says that there is an *abstract functional dependence* of the supervening predicate Q as the value of some function on the equivalence classes for its underlying predicate P. Beth's theorem then adds that, under special circumstances, this abstract functional dependence can be expressed generically across models, by some explicit first-order definition stated in the P-language. What does this further feature really imply?

Returning to the philosophical concern of brute-force 'reducibility', even such an explicit definition need not be a burdensome form of reduction if that definition is so unwieldy as to be inaccessible to bounded human agents. In this sense, one might say that explicit reduction only hurts and constrains us if we (or others) can *grasp and manipulate* the dependence.

What such finer distinctions would call for are more computational analyses of the powers of bounded human agents, and in terms of Beth's Theorem, stricter bounds on the *complexity of the definitions* produced. In this way, some of the work in computational logic on Beth-type theorems presented in [52] and [56] of this book might become relevant to philosophical discussions.

Finally, interestingly, the earlier local version of supervenience is not directly the content of Beth's Definability Theorem, and rather falls under this well-known theorem from [51]:

> If $\boldsymbol{L}$-automorphisms are also P-automorphisms in all models for a first-order theory T, then T implies a disjunction of explicit definitions for P in terms of $\boldsymbol{L}$, not necessarily one single definition.[15]

Supervenience is just one instance where a more detailed study of a rich philosophical literature might match up with a greater variety of notions and results in existing logical theory. We now turn to other examples of such interfaces.

4 Determinism

The notion of supervenience also relates to another pervasive philosophical theme: that of *determinism*. This notion occurs in classical metaphysics, where Aristotle's famous Sea Battle argument investigated a sense in which the logical law of Excluded Middle for future truths seems to imply determinedness of our present behavior placing it outside of our control. Similar themes have kept returning, all the way to causal determinism in the empirical sciences, though the determination there usually runs in the opposite forward temporal direction, from what is true now to what will become true later.

[15] To see the difference, think of a theory T consisting of just the statement $\forall x\,(Px \leftrightarrow Lx) \vee \forall x\,(Px \leftrightarrow \neg Lx)$. This does not satisfy the implicit definability for Beth's Theorem. Two models for T with the same domain and denotation of the predicate L can still disagree on the denotation of P. But what is true is that in each of these models, the automorphisms that leave L in place will also leave P in place.

A lot is at stake in debates around this notion: ranging from the (non-)existence of free will to the role of causality in Artificial Intelligence or the nature of what physics can provide.

As with supervenience, many proposals have been made for making the notion of determinism more precise including some by prominent philosophers such as Bertrand Russell.

As it happens, in this area, there is a landmark attempt by a well-known logician to sort out the issues and apply Beth's and Craig's results in the philosophy of science. [45] formalizes physical theories such as Newtonian mechanics in first-order terms as having a mathematical part referring to the natural numbers $\mathbb{N}$ and real numbers $\mathbb{R}$ plus tuples of functions on reals that encode relevant physical system properties such as position and momentum of particles, and other notions – including abstract, non-observable ones postulated by the theory.

In Montague's analysis, a physical theory T describes a set of possible histories, where we think of models as having the standard natural and real numbers (here the latter also encode the passage of time), while functions defined on the reals give the current state of, say, a mechanical system as a function of the time. In this setting, different definitions can be given of what is a 'deterministic theory', reflecting various proposals in the philosophical literature which we will not reproduce here.[16]

Here is one typical 'implicit' version which occurs widely in the literature, going back to a well-known historical statement of determinism in physics formulated by Laplace:

> *If two histories allowed by T have equal states for the physical system at time t_0, and $t_0 < t$, then the two histories also have the same total system state at time t.*

Modulo quite a few syntactic details in Montague's treatment which are irrelevant to the intended illustration here, this semantic condition on models for physical systems can be formulated as a form of implicit definability in a first-order theory. Craig Interpolation and Beth's Theorem then apply to conclude that the theory also provides strong determinism in the form of an explicit definition in the language of the theory for the system state at time t in terms of the state-at-time-t_0, t_0 and t, a definition which can be used for physical reasoning and empirical calculation.

At the same time, linking to our earlier discussion of various perspectives on supervenience, Montague's formal apparatus also allows us to see more precisely what is required for the Beth property, and thus, indirectly also, how one might try to avoid it.

It has to be said that there is a lot of fine-print to this analysis, and Montague's paper may not have had the impact it deserves due to an overload of notions, results and conclusions pointing in various directions. However, this may just illustrate an earlier general point:

[16] Montague notes that theories can easily be deterministic in terms of arbitrary functions for the dependence of one state on others, while the more practical issue is generic definability of the functional dependence, fit for computational uses. This distinction between model-dependent definitions and generic ones for all models echoes an earlier point made about supervenience and reductionism in Section 3.

bringing logic to bear on philosophical notions and discussions need not lead to unique or definitive outcomes that can lay matters to rest forever. Also, our earlier point applies that the logical results can only be made to work if one first formalizes the general philosophical discussion of determinism in terms of formal languages and theories.

There are many further types of formalization of empirical theories by logical means, on which there exists a vast literature, cf. [50, 6]. Of the many classical themes treated in this literature, one will return in Section 8 below, viz. the different vocabularies that go into the making of an empirical theory, where the relevant connection is now with Interpolation.

For now, we move to another broad topic in philosophy in the spirit of our themes so far.

5 Questions

New information from experiments in scientific theories or through more mundane channels of communication, often arises from asking *questions*. Accordingly, questions are ubiquitous in the philosophy of language, or in general epistemology – where knowledge can be seen as knowing the answer to some relevant question which controls the direction of inquiry.

Ever since the early pioneering work documented in [1] it has been proposed by philosophers to let logic deal with, not just consequence relations among propositions or statements that give information, but also between questions that ask for information.

In a very common setting for implementing this, cf. [37], logical semantics for questions models them in terms of their complete answers as forming a partition of a relevant state space. One often thinks of the partition as a current 'issue' to be resolved by identifying one of the partition cells as the actual one.

Valid consequence from a set of questions then becomes the following notion, which has been around for a long time in the semantic literature:

If we have all the answers to the 'premise' questions $?P_1, \ldots, ?P_k$, *we automatically also have an answer to the 'conclusion' question* $?C$.

For instance, in this sense, (i) $?A$ implies $?\neg A$, (ii) $?A, ?B$ imply $?(A\&B)$, and so on, suggesting a connection between consequence for questions and definability from premises.

A semantic definition for the notion underlying consequence between questions runs as follows on modal state spaces where formulas can be true or false:

If two states s, t *agree on all* P_i*-truth values* $(1 \leq i \leq k)$, *then they also agree on the* C*-truth value.*

Stated in other terms, the joint refinement of the P_i-partitions is included in the C-partition.

Of course, there is much more to the actual behavior of questions in natural language, as they raise or modify issues and steer the course of conversation and inquiry. In addition, questions interact with other informational events, and for instance, the above notion of

question-consequence can be made conditional on events $!Q$ of receiving the indubitable information that Q by restricting the total state space only to states where Q holds.

The preceding notion is clearly in the same spirit as the invariance and similarity notions discussed earlier in connection with Beth-style implicit definability.[17] Using this analogy, a complete logic for consequence between questions suitably defined in a predicate-logical setting was given in [23], who use the slightly strengthened projective version of Beth's Theorem to prove the following key equivalence for conditional implication between questions:

▶ **Theorem 3.** $!R, ?P \models ?Q$ *iff there is a first-order formula* $\alpha(P)$ *with* $R \models Q \leftrightarrow \alpha(P)$.

This result can be seen as a sort of expressive completeness of the usual logical constants for the purpose of giving explicit answers, typing logic of questions closely to logic of propositions.

Also, interestingly, the cited paper connects Uniform Interpolation (see [36] in this book) with the existence of a 'best answer' to a question given some background information.

Thus, the philosophically important notion of questions can be brought within the scope of logic, and the Beth and Craig theorems are key to obtaining more definite results.[18]

Again, this is just a start, and many issues remain for discussion, and here is just one. The preceding account will work in a first-order setting, but it is of course a moot point whether natural language is first-order. In particular, do the above results go through for extended languages with non-first-order *generalized quantifiers* such as "most" or "many"? In this area of Generalized Quantifier Theory, as elsewhere, formal semantics of natural language meets with issues in mathematical logic concerning the properties of logical systems, [48].

6 Dependence

It is time to step up the abstraction level of our analysis. Consequence between questions is a special case of semantic *functional dependence* between variables whose values can be truth values for Yes/No-questions, or other objects in the case of more general *wh*-questions, cf. [26]. But functional dependence also lies behind many notions discussed earlier in this paper, with our analysis of determinism involving variables taking values in state space, or supervenience as a case where the values of variables are even just equivalence classes of some relevant invariance relation.

In modern dependence logics, [54], functional dependence runs between sets of variables X and variables y in *state spaces* M where not all simultaneous assignments of values to these

[17] In fact, analogies run much wider, and similar 'equal answer' equivalence relations have been used to study the widespread phenomenon of 'ceteris paribus' reasoning [13].

[18] For much more on the still growing area of logics of questions, we refer to [29].

variables may be available at states. These 'gaps' are precisely what induce correlations between the values of variables, of which functional dependence is one important species.[19]

The dependence setting also makes sense in other areas of philosophy, such as the Situation Theory to be discussed in Section 7, where information flows and knowledge arises because of various types of informational correlation between situations in the world, which need not come from their structural similarities, but be enforced by global constraints on the behavior of the overall system.

The above semantic definition of functional dependence is one of *implicit semantic determination*, where we say that $s =_X t$ if s, t give the same values to all variables in the set X. Here is the general notion of functional dependence of a variable y on a set of variables X:

For any two states s, t in state space model M, $s =_X t$ implies $s =_y t$.

This abstract framework applies to questions with truth values for the relevant propositions, but also to knowing objects as in constructions like "knowing who killed Mr. T", and even to the earlier notion of supervenience. It also applies to implicit definability in Beth's sense when values of variables can be whole equivalence classes of some relevant similarity relation.

The earlier-quoted completeness theorem for the basic logic of propositional questions can then be seen as special case of a general minimal propositional modal logic of functional dependence proposed in [2] with questions as variables and truth values as their values.

In this setting, an earlier point returns. The stated semantic equivalence suffices for the existence of an abstract function F on value domains for variables so that for each state s,

$$s(y) = F(s[X]).$$

While such a function can always be found ad-hoc, locally in a given state space model M, Beth's Theorem then gives a special definable function F, under the right conditions, which works generically across models. This suggests a more general perspective on Beth-type results as an interesting companion to dependence logics, as providing conditions under which general abstract dependencies can be made explicit and thus fit for computation and deduction.

Given these connections, it is a natural question whether current dependence logics satisfy Beth Definability or Craig Interpolation theorems with respect to variables, and not just shared predicate vocabulary. For second-order dependence logic in the style of [54], the answer is that even Uniform Interpolation (see [36] for a definition and results) holds in the variable-sharing sense, as is explained in [55] in this book. For the decidable modal dependence logic of [2], only the predicate version of interpolation is known, [11], but it seems likely that a shared-variable version holds as well.

[19] Note that, in this semantic setting, variables are no longer the anonymous binding devices used in standard first-order logic, but named functions from states to values, displaying an individual behavior reflecting the uses of the notion of 'variable' in empirical science or probability theory.

Dependence seems to be the common thread behind most of our philosophical topics in this paper. It clearly applies to our discussions of supervenience and determinism. But dependence logic also applies to our earlier discussion of translations between logical systems, since as we noted, the existence of an explicit syntactic translation matches an implicit semantic dependence in the model classes for the two logics being connected. The reason why this works is, of course, that the variables in dependence logics can stand for anything, and likewise, their values can be anything from numbers or objects to equivalence classes of models or even information states of agents.

Given this ubiquity, basic dependence reasoning is an interesting candidate for extending the core calculus of logic beyond the usual repertoire of 'logical constants' in standard logics.

Going further, functional dependence of y on X is one end on a spectrum, where fixing the values of the variables in X leaves only one value for y. At the opposite end lies *independence*, where the current X-values place no constraint on the values that y can take at the current state. Independence, too, might play a role as a philosophical notion, say, as a sort of 'anti-definability' or 'unlearnability'. And in between on the spectrum lie notions of *correlation* between X and y which can 'soften' the force of functional dependence. This might make sense in various areas considered in this paper.

This correlation perspective again suggests interesting mathematical issues as to whether Beth's Theorem and Craig's Theorem admit of 'softer' versions producing definitions or interpolants up to exceptions, in either qualitative or even probabilistic senses.

We will now leave the invariance/dependence thread running through the first part of this paper, and look at how definability and interpolation highlight the crucial role of *vocabulary* in shaping both everyday beliefs and scientific knowledge.

7 Information and Generalized Consequence

Let us start with a core notion studied in logic, viz. *valid consequence*. A common philosophical view holds that drawing logical consequences has to do with *information*, in particular, with elucidating information that is implicitly present in given premises or data. The handbook chapter [15] discusses this view of consequence, while also identifying other broad notions of information at work in logic.

A broader focus on information is found in Situation Theory, [4], a logic-friendly formal paradigm inspired by the philosophical view of the notion of knowledge as based on information flow proposed in [30]. In this philosophical system the world consists of partial situations (at least if one takes the theory as one of metaphysics rather than epistemology), and information flows because of the existence of *correlations* between the behavior of these situations, mediated by abstract or concrete physical channels.

Logical consequence may now be thought of accordingly as acting typically *across different situations* or, more formally, different models. Say, when looking at a radar screen on the

ground, we observe light dots changing position, and conclude to movement of planes far away, and this is correct because of a correlation between their actual flight paths and what is visible locally on the screen that we are looking at.[20]

What the relevant connections are between the two situations in such inferences can of course vary in practice, since many kinds of correlation can carry information.

This situational view suggests a new notion of entailment, not just in one model as with ordinary logical consequence, but across models. Here are some notions and results from [3] developing this theme. We say that φ *entails* ψ *along a relation* R if,

for any two models M, N, if $M \models \varphi$ and $M\,R\,N$, then $N \models \psi$.

To make this a bit more concrete, in this perspective, a model M does not just support truths about itself which can be elucidated by inference from what we already know about M, the model is also a beacon 'broadcasting' further information to other models along various relations. These relations range from very strong analogies, like isomorphism, to weaker ones, including homomorphism or modal bisimulation. The further away one gets from M qua structural similarity, the weaker the information transfer.

This setting offers a fresh look at the Interpolation Theorem. We already have a valid consequence $\varphi \models \psi$. What is the further benefit of having an interpolant α in the shared vocabulary $\boldsymbol{L}$? What this really gives us is the following form of cross-model transfer:

if $M \models \varphi$ and N is $\boldsymbol{L}$-potentially isomorphic with M, then $N \models \psi$.[21]

That is, thanks to the interpolant, φ entails ψ along $\boldsymbol{L}$-potential isomorphism.[22]

This analysis of what interpolants do inspires a new version of the Interpolation Theorem.

▶ **Theorem 4.** *The following are equivalent for formulas as described just now:*
(i) Existence of an $\boldsymbol{L}$*-interpolant, (ii) Entailment along potential* $\boldsymbol{L}$*-isomorphism.*

For first-order logic, this new version is equivalent to the usual Craig Interpolation Theorem, using the downward Löwenheim–Skolem theorem plus the fact that, on countable models, potential isomorphism and isomorphism amount to the same thing. But in general, the new formulation of Interpolation is weaker, and for instance, for the fundamental infinitary

[20] This cross-situational function of inference was already present in ancient Indian logic with its standard example of 'smoke means fire', drawing conclusions about fire on an inaccessible mountain top from observations made here on the ground, [20].

[21] There are also other interpretations of the surplus. Say, one could view interpolation as stating that a logical consequence can always be related to a *relevant* implication in the sense of shared topics.

[22] Formally speaking, this notion has a modal form $\varphi \to [p.i._{\boldsymbol{L}}]\psi$. This also suggests looking at other modal notions such as $\langle p.i._{\boldsymbol{L}}\rangle\varphi$ which defines all pointed models that are $\boldsymbol{L}$-potentially-isomorphic to the current model. The latter would define a uniform interpolant in the present sense for the formula φ.

version of first-order logic $\mathfrak{L}_{\infty\omega}$ (see [55]), often touted as a typical example of failure of the standard interpolation theorem, this new version turns out to hold.[23]

These findings bring us back to a general issue raised in Section 2. When saying that a good logical system might have Interpolation as a characteristic feature, we took the standard formulation of Craig's Theorem for granted. But with our two options on the table, this becomes more delicate.

What is the natural formulation of the Interpolation property? And more generally, what are natural stable versions of known meta-properties of logical systems that we would like to transfer, discounting peculiarities of their original formulation when first discovered?

Similar generalized-interpolation versions exist for the classical model-theoretic *preservation theorems* tying important types of semantic behavior to syntactic form, whose content already involves crossing between models. For instance, instead of the usual Los–Tarski theorem characterizing preservation under submodels in terms of definability in special universal syntactic form, we also have the following version:

▶ **Theorem 5.** *φ entails ψ along $\boldsymbol{L}$-submodels iff there exists a syntactically universal $\boldsymbol{L}$-formula α which interpolates as follows: $\varphi \models \alpha \models \psi$.*

The original model-theoretic preservation theorem is retrieved by taking the special case that $\varphi = \psi$. A similar case is the following characterization theorem for modal logic:

A first-order formula φ entails a first-order formula ψ along $\boldsymbol{L}$-bisimulation iff there exists a modally $\boldsymbol{L}$-definable interpolant.

The proofs of such results often proceed by inspecting standard proofs of the preservation result, and finding that they actually also establish the more informative interpolation version.

There is more to the logical correlation view of information flow between situations, since it need not arise from structural similarity, but can also result from global dynamical system behavior. This phenomenon was already identified in our discussion of dependence logics in Section 6 above, but we will not pursue it here.

Finally, what is the philosophical import of the preceding discussion? Inspired by early work on Situation Theory, this section has addressed some fundamental issues in the philosophy of logic but also in the philosophy of information and one might say, even in epistemology.

We now turn to other aspects of the role of vocabulary in philosophy: in particular, the structuring of theories consisting of organized data and regularities that we reason with.

---

[23] The proof of this result uses a Lindström-type argument with instead of the usual reliance on the Compactness Theorem, the still very powerful model existence property encoded by the well-known *projective undefinability of the notion of well-ordering* in the infinitary logic $\mathfrak{L}_{\infty\omega}$.

8 Theoretical Terms in Scientific Theories

A key theme in the philosophy of science is the difference between (a priori) mathematical theories and theories in the empirical sciences. The latter contain *observable terms* whose values can be observed or measured directly, but also crucial are *theoretical terms* whose values cannot be observed directly, but which organize and streamline the theory. Where one draws this boundary in any given theory has been a matter of much debate, but the observational/theoretical distinction makes intuitive sense, not just in scientific settings.

Worth noting historically here is the seminal paper [17], again by E. W. Beth, which for the first time introduced the idea that empirical theories in science could be studied model-theoretically as combining syntactic formulations of laws with defining collections of structures representing actual physical systems that the theory is about.

In these logical terms, consider a syntactic theory $T(\boldsymbol{L}_O, \boldsymbol{L}_t)$ with an observational and a theoretical vocabulary. If Beth's Theorem applies, the observable facts stated in $\boldsymbol{L}_O$ will uniquely fix the denotations of the theoretical terms in $\boldsymbol{L}_t$, and the latter will be explicitly definable in T. However, this is seldom the case, except for stipulative definitions for mathematical convenience without deeper physical significance.

So, let us ask a more general question with no explicit definability assumed:

What is the *empirical content* of a theory T, i.e., what does it say
about the observable world without theoretical superstructure?

A common syntactic answer here has run as follows: that content is the theory $T|\boldsymbol{L}_O$ of all consequences in the sublanguage $\boldsymbol{L}_O$ that follow, or are derivable, from T.

However, another, and highly influential view, due to [49], is semantic: the empirical content of T (its concrete range of application where the theory 'works') consists of all $\boldsymbol{L}_O$-structures that can be *expanded* to a model for T in its full language, or alternatively, the model class $\mathsf{MOD}(T)|\boldsymbol{L}_O$.[24] At least for finitely axiomatized theories T, the empirical content is then defined by the existential second-order *Ramsey sentence* $\exists L_t \cdot T(\boldsymbol{L}_O, \boldsymbol{L}_t)$.

We always have that the above expandable structures satisfy all principles in the syntactic theory $T|\boldsymbol{L}_O$. When the model class $\mathsf{MOD}(T)|\boldsymbol{L}_O$ in fact equals $\mathsf{MOD}(T|\boldsymbol{L}_O)$, the theory T is called 'Ramsey-eliminable'. This is not always the case.

For a counterexample, take the first-order theory of the natural numbers $\mathbb{N}$ with $\boldsymbol{L}_O = \{<\}$, and add a unary theoretical predicate A for which it is stipulated that it is non-empty without a lower bound. The $\boldsymbol{L}_O$-part of this theory is just the first-order theory of $<$ on $\mathbb{N}$, but the latter standard model cannot be expanded to a model for the whole theory.

What is always true, however, is that every structure in $\mathrm{MOD(T)}|\boldsymbol{L}_O$ has an *elementary extension* to a model for the whole theory T. However, allowing this move in an empirical

[24] This fits the earlier view of abstract logics as having sentences and models on a par as components.

theory would amount to not just introducing new theoretical predicates in an empirical situation, but also additional new 'theoretical objects'.[25]

A natural specialization of this setting are finitely axiomatized empirical theories T. In that case, the first notion of empirical content asks for the *strongest consequence* of T in its empirical subvocabulary, which would require *Uniform Interpolation* as defined in [36].[26]

On a more refined view, there are more layers of vocabulary in a scientific theory: Philosophers sometimes distinguish 'observational' (fact-describing), 'dispositional' (low-level empirical generalizations) and 'theoretical'. But a similar analysis to the one above applies.[27]

There is much more to the interface of logic and philosophy of science than the few issues sketched here. For instance, notions of invariance and definability are equally fundamental in both fields, cf. [22] on the role of symmetries in physics.

Another crucial aspect of logic that we have left out completely in our brief discussion is the crucial computational role of perspicuous syntax in allowing us to express regularities and laws in scientific theories and compute practical predictions.

9 Modularity in Knowledge and Belief

Moving from scientific theories to bodies of knowledge and opinion, we now turn to epistemology as the philosophical study of what it means for humans to know and believe. In this broader field, the crucial role of language returns, if only because what we know or believe is usually expressed in natural language. Various new issues now emerge, and Craig Interpolation applies in new ways.

For a start, our knowledge and belief are about certain facts and notions, and the notion of the *topic* of a proposition, what it is 'about', has been a long-standing philosophical theme.

Now one natural logical approach to the topic of a formula defines it as the vocabulary of the smallest sublanguage in which the formula has a logical equivalent. The following result, due to [47], says that this is well-defined.

▶ **Theorem 6.** *For any formula φ stated in a first-order language $\boldsymbol{L}$, there is a unique smallest sublanguage of $\boldsymbol{L}$ under inclusion where φ has an equivalent.*

Here Craig Interpolation is at work again. Clearly there are inclusion-minimal sublanguages of the finite vocabulary $\boldsymbol{L}$ where an equivalent can be found. But suppose that there were two different ones, say, $\boldsymbol{L}_1, \boldsymbol{L}_2$, with equivalents φ_1, φ_2. Then $\varphi_1 \models \varphi_2$, and by interpolation,

[25] These notions have also been refined. The empirical content of T might also just be the observational facts and universal laws about observables that follow from T, a universal subtheory of $T|\boldsymbol{L}_O$.

[26] Interestingly, Craig, p.c., started his seminal work trying to solve the 'strongest consequence problem' for all of first-order logic, and then proving Craig Interpolation as the next best thing.

[27] The distinction also makes sense in natural language, where observable predicates such as hair color or temperature co-exist with more theoretical dispositional predicates such as being "friendly" or "breakable", or even highly theoretical predicates such as a person's "destiny".

there exists a formula α in $\boldsymbol{L}_1 \cap \boldsymbol{L}_2$ with $\varphi_1 \models \alpha \models \varphi_2$. But then α would be an equivalent for φ in a still smaller sublanguage: yielding a contradiction.

Here is another basic language-related issue concerning the modular architecture of first-order theories. It is related to the following equivalent version of the Craig Interpolation Theorem (cf. also the well-known Robinson Joint Consistency Theorem):

Suppose that a theory $T = T_1 \cup T_2$ merging two subtheories is inconsistent, then there is a formula α in the shared language $\boldsymbol{L}_1 \cap \boldsymbol{L}_2$ with $T_1 \models \alpha$ and $T_2 \models \neg\alpha$.

Now it is a commonplace that our beliefs are compartmentalized into different topics that can be disjoint with formulations that do not interact. However, most current logics of knowledge and belief just work with one cover-all language without subdivisions. Therefore, it is of interest to *split* the total language into disjoint sublanguages, and ask for the effects of this on the structure of theories. In particular, questions arise about information flow between the various components of a theory T in the total language.

Here Craig's Interpolation Theorem has been used in [41, 47] to obtain the following result on the fine-structure of bodies of beliefs.[28]

The relevant notion of modularity is the following. We say that a theory T is *split* by a partition $\boldsymbol{L}_1, \boldsymbol{L}_2$ of its total language $\boldsymbol{L}$ if there are formulas φ in $\boldsymbol{L}_1$, ψ in $\boldsymbol{L}_2$ such that T is $\{\alpha \mid \{\varphi, \psi\} \models \alpha\}$. There is a straightforward generalization to larger splittings.

Next, it is easy to see by two successive applications of Interpolation that the following statement holds whenever T is split into $\{A, B\}$ by a partition $\boldsymbol{L}_1, \boldsymbol{L}_2$:

If $T \models \varphi$, then there is an $\boldsymbol{L}_1$-formula α and an $\boldsymbol{L}_2$-formula β such that
(i) A $\models \alpha$, (ii) B $\models \beta$, and (iii) $\alpha, \beta \models \varphi$.

A Parallel Interpolation Theorem in this spirit can also be proved for larger finite splittings by suitably iterating Craig Interpolation. Using this, one obtains the following general result, where 'largest' refers to the number of partition cells in a splitting:

▶ **Theorem 7.** *Every theory with a finite vocabulary has a unique largest splitting.*

As a third example, we just mention a further major theme in formal epistemology, viz. *information dynamics*. Theories as bodies of knowledge and belief typically *change* when update takes place with new information produced by observation, reasoning, or communication – three major processes that govern science as well as ordinary life. With the above results in place, the usual logical accounts of knowledge update and belief revision for rational epistemic agents, [35, 10], can be considerably refined to introduce more topic-sensitivity and modular structure into the representation of what agents know and believe at any stage of the process.

[28] For convenience, we talk about finitely axiomatized theories though the cited results hold more generally by a simple appeal to the Compactness Theorem for first-order logic.

This concludes our survey of uses of Interpolation in various areas of philosophy. Compared to our first theme of semantic invariance and dependence, we do not offer one crisp denominator covering all of these, though the pervasive influence of particular vocabularies being used in reasoning and theory construction is an obvious red thread.[29] Whether these two themes unify further than this is not resolved in this paper, and we leave finding a more integrated perspective as a desideratum.

10 A Historical Precursor: Bolzano

Beth's Theorem and the Interpolation Theorem highlight the importance of language as crucial ingredient in the central notion of logical consequence. But actually, this theme in logic may be less new than we have made it out to be.

A historical precursor to this linguistic focus is Bernard Bolzano whose "Wissenschaftslehre", [21], has the first systematic semantic notion of consequence, and the first explicit study of different notions of consequence for different purposes (an early form of 'logical pluralism'), with varying properties depending on the standards that one adopts for strictness of styles of reasoning. The following discussion is taken from [7].

For Bolzano, whether we call a consequence valid depends on which parts of the premises and conclusion we consider as having a *fixed meaning*, and which parts are considered *variable*, and hence amenable to substitutions of other expressions in the same category.

For instance, a step from premises {John is taller than Mary, Mary is taller than Paul} to "John is taller than Paul" is valid in this setting when we consider "John, Paul, Mary" as variable terms, but fix the meaning of "taller". But this consequence is not valid if we also consider "taller" as a variable term for which we allow substitutions of other binary predicates that need not be transitive.

The inference from "x is P-er than y, y is P-er than z" to "x is P-er than z" will even remain valid in Bolzano's sophisticated sense if we just fix the meaning of the general comparative construction "-er" in natural language which turns adjectives into comparatives.

Thus, we can view Bolzano's notion of consequence as being *ternary* rather than binary:

$$P_1, \ldots, P_k \models_{\boldsymbol{L}} C,$$

where the third argument, the vocabulary set $\boldsymbol{L}$, explicitly records which subexpressions of the formulas involved we see as variable.

This means that a logical calculus must now keep explicit track of the relevant vocabulary, leading to questions, e.g., about how to combine consequences with different vocabularies.

[29] Our focus here has been the *non-logical vocabulary* in formulas. It should be noted, however, that *logical vocabulary* also plays a key role in constructing interpolants, and this feature may go into the intuition that an interpolation theorem requires some sort of expressive completeness for a logical system.

For instance, let $P_1, \ldots, P_k \models_{\boldsymbol{L}_1} C$ and $Q_1, \ldots, Q_k \models_{\boldsymbol{L}_2} D$ be given valid consequences. Should we now intersect the two vocabulary sets $\boldsymbol{L}_1, \boldsymbol{L}_2$, or rather take their union when drawing the natural conjunctive conclusion $C \wedge D$ from the union of the two sets of premises? We leave an answer to the reader.

All this enriches the usual modern way of thinking about the basic 'structural rules' governing different styles of inference, where the language aspect is not an explicit parameter: this time, it has become crucial.

This explicit manipulation of vocabulary seems in the spirit of the Interpolation Theorem, and ternary sequent calculi are in fact a standard tool in proof-theoretic arguments establishing its truth. But beyond this auxiliary use, we might also come to see ternary sequents that highlight vocabulary as the primary vehicles for inference.[30]

11 Conclusion

This short and light survey has revolved around one circle of general ideas, namely, dependence, lifting, vocabulary, transfer, and definability as these occur in various areas of philosophy. What we have tried to show is that both the *notions* and the *results* found in logic show strong analogies with themes in philosophy, and can, under certain circumstances, be applied to yield broader insights.

It should be noted here that our methodology at the logic-philosophy interface can be broadened. The discussion in this chapter has mainly concentrated on the semantic side of logic and on model-theoretic analysis. It would also be of interest to revisit all topics of the preceding sections in a proof-theoretic perspective. For instance, the standard proof-theoretic analysis of Craig Interpolation delivers a recursive procedure which, given a proof of an implication $\varphi \to \psi$ outputs an interpolant and separate proofs for $\varphi \to \alpha$ and $\alpha \to \psi$. This changes the discussion from pure truth or validity to 'reasons' for these semantic facts, and this, too, can be connected up with various themes in philosophy.[31]

But even as it is, our case study also raises issues concerning the fit of logical analysis to philosophy. Logic needs a strictly defined formal *language*, both for its model theory and proof theory, and while this language dependence helps for precision and proving definite results, it may not always fit the spirit of philosophical discourse and thinking, except for those areas which have undergone the twentieth-century 'linguistic turn'. And even when

[30] Not just our theme of language awareness has roots going back to pioneering work in the 19th century. The same is true for our analysis of invariance and definability, which can be traced back as far as the seminal studies in [38] on the emergence of language as a medium for describing notions that are invariant under relevant transformations. See also [57] for further history.

[31] In arithmetical terms, switching from validity to provability and coding up formulas and proofs as natural numbers, the complexity of the usual formulation of Craig Interpolation is Π^0_2. The proof-theoretic analysis provides a stronger Π^0_1 version, so there are also interesting technical issues here.

language matters, it is not clear whether the restriction to a first-order language (the original habitat of the Beth and Craig theorems) is always meaningful in philosophy, just as this restriction does not always fit practice in mathematics and the sciences.

In fact, except for technical subfields of formal epistemology, philosophy of science, and philosophy of language, professional philosophy largely makes do with *natural language*, perhaps enriched with some notations and regimented with technical terminology. What the meta-theorems of logic say in such a setting is clearly a legitimate matter for debate. And we have also suggested occasionally that the properties expressed by these meta-theorems, usually considered a virtue or even a requirement on logic design, might not always be desirable.

Having said this, one can also view the use of logic as lying at a greater distance, offering a model for a philosophical practice (with 'model' now in its broader colloquial sense), in which case application just means providing attractive new perspectives and suggestions. Also, the formal similarities noted in this paper could be a force for seeing unifying themes and perspectives across philosophical subdisciplines.

Continuing with the presuppositions of logical analysis, the reliance of Beth's Theorem on a specified *theory* can be questioned, since many philosophical discussions (and perhaps even working practices in sciences such as physics) are about One World where everything of interest takes place.[32] On the other hand, one might also argue that theory reliance is in line with the axiomatic approach in abstract mathematics, and this is definitely the broader context in which Beth saw his own work.

Despite these caveats, we believe that our survey has offered a wide variety of natural points of contact between logic and philosophy along our main themes, and indeed, several of the authors cited exemplify the union of the two fields, being logicians who were at the same time serious philosophers. More specifically, it might be a rewarding project to put the relevant literature in the areas of philosophy mentioned in this survey paper side by side with current logical studies of definability and interpolation in a much more detailed and systematic manner. Here, our selection of topics provided just one sample of course, while even the topics selected could profit from broadening the logic interface to include themes from computational logic such as complexity, from interfaces with quantitative approaches, [14], or from the vast area of non-classical logics.

Finally, as to what is at stake in a chapter like this, the real benefit of the juxtaposition that we have pursued may not be so much logical solutions to philosophical problems as *mutual inspiration* – since, conversely, logic can also learn from the philosophical imagination and wide-ranging theorizing in the areas highlighted here.

[32] Say, in Padoa's method, implicit definability might just be determination in the single model of Euclidean Space that we focus on, the way we think, say, about how a triangle determines its centre of gravity, and to match this, we would need localized versions of Beth's Theorem in suitable structures.

Acknowledgments

I thank Alexandru Baltag, Balder ten Cate, Sören Brinck Knudstorp, Martin Otto, Ian Pratt-Hartmann, Pablo Rivas Robledo, Martin Stokhof, Albert Visser, Frank Wolter, and Urszula Wybraniec-Skardowska, as well as audiences for presentations of this material in Amsterdam, Beijing, Berkeley, and Dortmund for their comments on topics in this paper. I also thank Christoph Wernhard for his generous help in optimizing the look of this paper.

References

1 K. Ajdukiewicz. Interrogative sentences. In J. Giedymin, editor, *The Scientific World Perspective and Other Essays; 1931–1963*, volume 108 of *Synthese Library*, pages 155–164. Springer, 1978. Originally published as: Zdania pytajne, in Logiczne podstawy nauczania, Nasza Księgarnia, Warszawa (1934).

2 A. Baltag and J. van Benthem. A simple logic of functional dependence. *Journal of Philosophical Logic*, 50:939–1005, 2021.

3 J. Barwise and J. van Benthem. Interpolation, preservation, and pebble games. *Journal of Symbolic Logic*, 64(2):881–903, 1999.

4 J. Barwise and J. Perry. *Situations and Attitudes*. Bradford Books, The MIT Press, 1983.

5 J. Barwise and J. Seligman. *Information Flow. The Logic of Distributed Systems*. Cambridge University Press, 1997.

6 J. van Benthem. The logical structure of science. *Synthese*, 51:431–472, 1982.

7 J. van Benthem. The variety of consequence, according to Bolzano. *Studia Logica*, 44(4):389–403, 1985.

8 J. van Benthem. Information transfer across Chu spaces. *Logic Journal of the Interest Group for Propositional Logics*, 8(6):719–731, 2000.

9 J. van Benthem. The many faces of interpolation. *Synthese*, 164(3):451–460, 2008.

10 J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge UK, 2011.

11 J. van Benthem, B. ten Cate, and R. Koudijs. Local dependence and guarding. In *Proceedings Advances in Modal Logic*, pages 135–154, 2022.

12 J. van Benthem and K. Doets. Higher-order logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume I, pages 275–329. Reidel, Dordrecht, 1983.

13 J. van Benthem, P. Girard, and O. Roy. Everything else being equal: a modal logic for ceteris paribus preferences. *Journal of Philosophical Logic*, 38(1):83–125, 2009.

14 J. van Benthem and T. Icard. Interleaving logic and counting. *Bulletin of Symbolic Logic*, 29(4):503–587, 2023.

15 J. van Benthem and M.-C. Martinez. The logical stories of information. In P. Adriaans and J. van Benthem, editors, *Handbook of the Philosophy of Information*, pages 217–280. Elsevier, Amsterdam, 2008.

16 J. van Benthem and D. Pearce. A mathematical characterization of interpretation between theories. *Studia Logica*, 43(3):295–303, 1984.

17 E. W. Beth. Towards an up-to-date philosophy of the natural sciences. *Methodos*, 1:178–185, 1949.

18 E. W. Beth. On Padoa's method in the theory of definition. *Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings*, 56:330–339, 1953. Also in Indagationes Mathematicae, 15, 330–339, 1953.

19 E. W. Beth. *The Foundations of Mathematics.* North-Holland, Amsterdam, 1959.

20 J. Bochenski. *A History of Formal Logic.* University of Notre Dame Press, Notre Dame Indiana, 1961.

21 B. Bolzano. *Wissenschaftslehre.* Seidelsche Buchhandlung, Sulzbach, 1837.

22 J. Butterfield and S. de Haro. *The Philosophy and Physics of Duality.* Oxford University Press, 2025.

23 B. ten Cate and C. Shan. Axiomatizing Groenendijk's logic of interrogation. In M. Aloni, A. Butler, and P. Dekker, editors, *Questions in Dynamic Semantics*, pages 63–82. Brill, Leiden, 2007.

24 Balder ten Cate and Jesse Comer. Interpolation in first-order logic. In ten Cate et al. [25], chapter 2. Preprint: `https://arxiv.org/abs/2510.03822`. `doi:10.5334/bdg.b`.

25 Balder ten Cate, Jean Christoph Jung, Patrick Koopmann, Christoph Wernhard, and Frank Wolter, editors. *Theory and Applications of Craig Interpolation.* Ubiquity Press, 2026. To appear; preprints accessible from `https://cibd.bitbucket.io/taci/`. `doi:10.5334/bdg`.

26 I. Ciardelli. Dependency as question entailment. In S. Abramsky, J. Kontinen, J. Väänänen, and H. Vollmer, editors, *Dependence Logic: Theory and Applications*, pages 129–181. Birkhäuser/Springer, Cham, 2016.

27 W. Craig. Linear reasoning. a new form of the Herbrand-Gentzen theorem. *Journal of Symbolic Logic*, 22:250–268, 1957.

28 W. Craig. Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic*, 22(3):269–285, 1957.

29 C. Cross and F. Roelofsen. Questions. Stanford Encyclopedia of Philosophy, 2022.

30 F. Dretske. *Knowledge and the Flow of Information.* The MIT Press, 1983.

31 H. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer, New York, 1995.

32 J. Fan. A modal logic of supervenience. *Notre Dame Journal of Formal Logic*, 60(2):283–309, 2019.

33 S. Feferman. Harmonious logic: Craig's interpolation theorem and its descendants. *Synthese*, 164:341–357, 2008.

34 Wesley Fussner. Interpolation in non-classical logics. In ten Cate et al. [25], chapter 3. Preprint: `https://arxiv.org/abs/2512.01600`. `doi:10.5334/bdg.c`.

35 P. Gärdenfors. *Knowledge in Flux. On the Dynamics of Epistemic States.* The MIT Press, 1987.

36 Sam van Gool. Uniform interpolation. In ten Cate et al. [25], chapter 9. Preprint: `https://arxiv.org/abs/2512.15391`. `doi:10.5334/bdg.i`.

37 J. Groenendijk and M. Stokhof. Questions. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 1059–1131. Elsevier, Amsterdam, 1997.

38 H. von Helmholtz. Über die Thatsachen, welche der Geometrie zu Grunde liegen. In *Wissenschaftliche Abhandlungen*, volume II, pages 618–639. Johann Ambrosius Barth, Leipzig, 1883.

39 E. Hoogland and M. Marx. Interpolation and definability in guarded fragments. *Studia Logica*, 70:373–409, 2002.

40 E. Hoogland, M. Marx, and M. Otto. Beth definability for the guarded fragment. In H. Ganzinger, D. McAllester, and A. Voronkov, editors, *Logic for Programming and Automated Reasoning*, volume 1705 of *Lecture Notes in Artificial Intelligence*, pages 273–285. Springer, 1999.

41 G. Kourousias and D. Makinson. Parallel interpolation, splitting, and relevance in belief change. *Journal of Symbolic Logic*, 72(3):994–1002, 2007.

42 Agi Kurucz, Frank Wolter, and Michael Zakharyaschev. From interpolating formulas to separating languages and back again. In ten Cate et al. [25], chapter 10. Preprint: `https://arxiv.org/abs/2508.12805`. `doi:10.5334/bdg.j`.

43 I. Mason. The metatheory of the classical propositional logic is not axiomatizable. *Journal of Symbolic Logic*, 50(2):451–457, 1985.

44 B. McLaughlin. Supervenience. Stanford On-Line Encyclopedia of Philosophy, 2023.

45 R. Montague. Deterministic theories. In R. Thomason, editor, *Formal Philosophy*, pages 303–359. Yale University Press, New Haven, 1974.

46 M. Otto. An interpolation theorem. *Bulletin of Symbolic Logic*, 6(4):447–462, 2000.

47 R. Parikh. Beth definability, interpolation and language splitting. *Synthese*, 179:2011–2021, 2011.

48 S. Peters and D. Westerståhl. *Quantifiers in Language and Logic*. Oxford University Press, Oxford, 2006.

49 F. Ramsey. Theories. In R. B. Braithwaite, editor, *The Foundations of Mathematics and Other Logical Essays*. Kegan Paul, Trench, Trubner and Co., London, 1931.

50 F. Suppe, editor. *The Structure of Scientific Theories*. The University of Illinois Press, Chicago, 1974.

51 L. Svenonius. A theorem on permutations in models. *Theoria*, 3:173–178, 1959.

52 Amirhossein Akbar Tabatabai. Proof complexity and feasible interpolation. In ten Cate et al. [25], chapter 6. Preprint: `https://arxiv.org/abs/2505.03002`. `doi:10.5334/bdg.f`.

53 A. Urquhart. Failures of Beth's theorem in relevance logics. *Review of Symbolic Logic*, 18(3):949–962, 2025.

54 J. Väänänen. *Dependence Logic. A New Approach to Independence-Friendly Logic*. Cambridge University Press, 2009.

55 Jouko Väänänen. Interpolation in model theory. In ten Cate et al. [25], chapter 7. To appear; preprints accessible from `https://cibd.bitbucket.io/taci/`. `doi:10.5334/bdg.g`.

56 Christoph Wernhard. Interpolation with automated first-order reasoning. In ten Cate et al. [25], chapter 12. Preprint: `https://arxiv.org/abs/2507.01577`. `doi:10.5334/bdg.l`.

57 H. Weyl. *Philosophy of Mathematics and Natural Science*. Atheneum, New York, 1963. Reprint of German version from the 1930s.

58 J. Zucker. The adequacy problem for classical logic. *Journal of Philosophical Logic*, 7:517–535, 1978.