DIVISION OF COMPUTER SCIENCE

Structural Invariance, Structural Definability and the Galois Theory of Elementarily Invariant Structures

J D Broido

Technical Report No.234

October 1995

Structural Invariance, Structural Definability and the Galois Theory of Elementarily Invariant Structures.

J.D. Broido

Introduction.

Suppose A and B are two adequately described structures— can we decide whether and how A is interpretable in terms of B?

This question is itself in need of clarification, of course. In different contexts, the term *interpretation* admits of different readings that suggest different kinds of operations between the alleged structures; even the term *structure*, popular and precise as it may sound, is already used with somewhat divergent senses, within the range of Pure and Applied Mathematics—the very disciplines that are supposed to focus on structure *per se*. The use of the term *interpretation* is certainly neither restricted to structures that are models of the same first-order theory, nor even to such as are merely "structures for" the very same minimal set of predicates. As we all know, it is possible to envisage "interpretations"—and Science is full of such—where the fundamental individuals and predicates in one structure are *mapped* on totally different types of entities, logically, which may be much more complex and derivative relative to the "interpreting" structure.

If we consider interpretations as *mappings*, however, it is intuitively apparent that we must satisfy at least two kinds of conditions: *First*, that basic entities in the interpreted source structure must be mapped on entities which are in some sense *definable* in the interpreting target structure; and, *secondly*, that truths of the interpreted source must become, under such interpreting mapping, *truths—apparent or derivative—* of the interpreting structure. In this paper we explore some of the ramifications of these minimal constraints on interpretability for *first order strucures* , as understood by *Model Theory.*

This exploration will ultimately set the stage for a theory of structure-concepts in general. Such a theory will adopt a *Nominalistic* approach to general concepts of structure, and elucidate them in terms of *equivalence relations* between *descriptions—* the nominal starting point. From the vantage point of interpretability, however, the significant thing about this approach is that different equivalence relations between "structural" descriptions are to be based on different types of *bilateral interpretability.* The weaker our contraints on what constitutes an "interpretation", the more liberal and general will be our general concepts of structure!

__Definability and Invariance__. Definability is normally understood with respect to*Theories.* Yet since *first-order models* —exemplifying (as we shall see) the simplest and strictest concept of structure—are defined for a given first-order language, it also makes sense to talk of sets of various kinds as being definable or undefinable in a given first order structure.

Thus, any subset S of $|\mathcal{A}|^n$ —where $|\mathcal{A}|$ is the universe of a first order structure $\mathcal{A}$—will be said to be *definable in* $\mathcal{A}$, when and only when there is a formula $F(x_1,\ldots,x_n)\in\mathbb{L}(\mathcal{A})$ such that $F(a_1,\ldots,a_n)$ is true in $\mathcal{A}$ iff $\langle a_1,\ldots,a_n\rangle \in S$. [$\mathbb{L}(\mathcal{A})$ being the *minimal structural language* for which $\mathcal{A}$ is a structure]

For some first order structures, including all such finite structures, there is a perfect correspondence between definale sets and *invariant* (set-theoretical) entities.
A set $S \subseteq |\mathcal{A}|^n$ will be said to be *invariant in* $\mathcal{A}$ **iff** for every automorphism, σ, of $\mathcal{A}$, we have $\sigma S=S$ (assuming always that $\sigma S=\{\sigma x \mid x \in S\}$, for any set S). Now, it is easy to show that any finite structure, $\mathcal{A}$, has the following property
β: A subset of $|\mathcal{A}|^n$ is definable in $\mathcal{A}$ if and only if it is invariant in $\mathcal{A}$.

For finite structures β can be easily derived from **Beth's Definability Theorem**, or more directly and constructively, by other means. Yet for structures with a denumerable infinity of elements, β constitutes an extremely powerful constraint. It does not hold for most of the best known infinite mathematical structures, such as the Standard Model of Number Theory or the Algebraic Numbers. In the case of The Standard Model of Number Theory, for instance, every subset of n-tuples of natural numbers is invariant (there sre no non-trivial automorphisms), but there are clearly more subsets ($2^{\aleph_0}$) than definitions ($\aleph_0$ only). Thus, with respect to β, one can prove only (see §4 later)

THEOREM 4.A. **A countable structure, $\mathcal{A}$, has the property β iff Th($\mathcal{A}$) is $\aleph_0$-categorical.**

The requirement that Invariance and Definability be coextensive therefore restricts us to models of *categorical* and properly $\aleph_0$-*categorical* theories, and it can be shown—in either of these cases—that for each n there must be only finitely many invariant subsets of n-tuples.

The trouble with β is that it restricts the number of invariants only to those that can be finitely "explicated" by finite first order formulae (defining them)—which is even stricter than what might be required by Intuitionism. Classically, however, we can easily see that in $|\mathcal{A}|$there are bound to be $2^{\mathfrak{c}}$ invariants (including $|\mathcal{A}|$ and the empty set) where $\mathfrak{c}$ is the cardinality of the set of all *minimal invariant subsets*—also known as *orbits*. If there are infinitely many minimal invariants, there will be more invariants than available definitions.
This suggests that we should consider weaker constraints than β (but still sufficient for our purposes). Consider the following property, β_{fin}, restricting the equivalence of definability and invariance to finite sets:
β_{fin}: For any n, a finite subset of $|\mathcal{A}|^n$ is definable in $\mathcal{A}$ iff it is invariant in $\mathcal{A}$.

This property is shared by great many countable structures, including many that were the subject of traditional Mathematics—e.g., the Standard Models of The *Natural Numbers*, the *Rationals* and the *Algebraic Numbers*, and is exactly the kind of property we need in order to study humanly useful interpretability relations between first order structures.

As it turns out (see §5 below), this property too is mirrored by a pure Model Theoretic property of (first order) structures—a property we call *Elementary invariance* :
Elementary Invariance. A structure A (for language L) is *elementarily invariant* when and only when its domain is an invariant in any elementary extension of $\mathcal{A}$.

There are some easily recognisable features of first order models that guarantee their elementary invariance. The most useful property of this kind has each individual in the model's domain belonging to some finite definable subset. This is tantamount to requiring that the *orbits of individuals be all finite*. While trivially satisfied in the case of finite structures, this is not a trivial property for an infinite structure. There is however one general category ensuring such behaviour—that of structures with finitely many symmetries.

<u>Galois Theory of Structures</u>. Classical Galois Theory is sometimes upheld as a paradigm of transforming a seemingly intractable problem, in one mathematical framework, into a relatively simple problem in another. The solvability of Algebraic equations over a given field is transformed by Galois Theory into a decidable question about the structure of certain finite groups. The fundamental mapping behind this miraculous transformation is the one which maps an Algebraic structure on the group of those of its auto-morphisms which leave unmoved the elements of of a certain substructure.

Although Galois Theory was generalised for other Algebraic structures beyond the original fields, such theories are almost never presented as the inevitable highway to interpretability questions of certain kinds (embeddability). In this paper we show that many of the classical percepts and theorems of Galois Theory are naturally applicable to first order structures that have the property β_{fin} —with a near perfect analogy in the case of those structures with finitely many symmetries. Thus, the analogues of Galois Theory will apply in particular to *interpretability relations* between *any finite structures*, and provide for *decidability* in principle.

Such a Galois theory is fundamentally predicated(much like the original one) on *relative* notions of Definability and Invariance (in the original theory the terms used were quite different!). Given a substructure $\mathcal{A}_0$, of $\mathcal{A}$, one may ask which entities in $\mathcal{A}$ are definable by means of the individuals in $|\mathcal{A}_0|$—using them in addition to the structural predicates and functions. In the same vein one may talk of *Invariance-relative-to-*$\mathcal{A}_0$, by which is meant invariance under all those automorphisms which leave *every* individual in $\mathcal{A}_0$ *unmoved* [such automorphisms constitute a subgroup $G(\mathcal{A}/\mathcal{A}_0)$ of the group $G(\mathcal{A})$ of all automorphisms of $\mathcal{A}$]. As an example of a close analogue of a classical Galois theorem consider the following

THEOREM (see §6 , theorems **6.D and 6.E): Let $\mathcal{A}$ be any Elementarily Invariant Structure and let $\mathcal{A}_0$ be an invariant substructure thereof, with corresponding subgroup $G(\mathcal{A}/\mathcal{A}_0)$. Then (1) $G(\mathcal{A}/\mathcal{A}_0)$ is a *normal subgroup* of $G(\mathcal{A})$; (2) $\mathcal{A}_0$ will be *functionally closed* in $\mathcal{A}$** [i.e., $|\mathcal{A}_0|$ includes any $\mathcal{A}_0$-definable singleton in $\mathcal{A}$] **<u>if and only if</u> $|\mathcal{A}_0|$is the set of all elements in $\mathcal{A}$ unmoved by $G(\mathcal{A}/\mathcal{A}_0)$** [in which case one can call $\mathcal{A}_0$ a *Galois substructure* of $\mathcal{A}$]; **(3) if $G(\mathcal{A}/\mathcal{A}_0)$ is *finite*, then there is a finite subset $S \subseteq |\mathcal{A}|$,with K elements, where $0 \leq K \leq \log_2(\text{order}(G(\mathcal{A}/\mathcal{A}_0)))$, and where $\mathcal{A}$ is the functional closure of $|\mathcal{A}_0| \cup S$; and (4) $\mathcal{A}$ can be viewed as the *"splitting" structure* over $\mathcal{A}_0$ for some monadic formula in $S_1(\mathbf{L}(\mathcal{A}))$ with a finite number of solutions** [i.e., $\mathcal{A}$ is functionally generated by these solutions over $|\mathcal{A}_0|$].

Interpretability. The last theorem only illustrates the degree to which a general Galois theory of Elementarily Invariant structures mimics the classical theory. It does not tell us how we are to use such tools to decide on the existence of interpretability-mappings between structures—especially when we allow such transformations to map basic individuals onto *complex* entities, constructed by means of the host-structural language and the normal set-theoretic aparatus.

Here again the direction is pointed out by the classical theory: Just as we can understand classical Galois Theory to be dealing with the existence of a monomorphic embedding from a given field into what is obtained by repeated radical-extensions, so can we reduce the problem of interpreting one structure in an extension of another to the existence of a suitable monomorphic embedding of one structure in some set-theoretic extension of another—for which we can generate *necessary and sufficient group theoretic criteria.* However, we know how to do this, in general, only for injections of *finite* structures in set-theoretic extensions of *elementarily invariant* host structures.

To have the flavour of such results, we introduce now a few simple definitions and notations:

Let $\mathbf{C}_{set}(\mathcal{A})$ denote the union of $|\mathcal{A}|$ with the class of all sets *constructible* (by normal set-theoretic operations) from the structure $\mathcal{A}$ [Set Theoretically this means starting with the sets $\{\varnothing,|\mathcal{A}|,\ldots,\mathfrak{R}^{n}{}_{i},\ldots\}$—where $\mathfrak{R}^{n}{}_{i}$ models in $\mathcal{A}$ the n-ary predicate-symbol R_i—and repeatedly applying the set-construction tools provided by standard ZF set theory]. We extend the original $\mathbf{L}(\mathcal{A})$ to $\mathbf{L}_{set}(\mathcal{A})$, which includes set membership symbol '$\in$' and the symbols '{','}' and '|'. A *finite entity* in $\mathbf{C}_{set}(\mathcal{A})$ is constructed by forming only finite sets, at each set-theoretic stage, using only a finite number of stages. With the exception of $\varnothing$, the type of an entity will be defined as the set of types of its members, where the type of individuals in $\mathcal{A}$ is set to 0. Thus the type of any non-empty subset of $|\mathcal{A}|$is {0}, while a non-empty set of ordered pairs of elements in $|\mathcal{A}|$ will be of type $\{\{0,\{0\}\}\}=\{<0,0>\}$.

We expect of any interpretation $\psi:\mathcal{A}\rightarrow\mathbf{B}$ that the individuals of $\mathcal{A}$ should be mapped on entities in which are all of the same (arbitrary) type τ, and that any definable subset of the type$\{<0^{\leftarrow},.^{n}.\,.^{\rightarrow},0>\}$ should be then mapped *onto* an entity in $\mathbf{C}_{set}(\mathbf{B})$ of type $\{<\tau^{\leftarrow},.^{n}.\,.^{\rightarrow},\tau>\}$. The definitions of $\mathcal{A}$-*Invariance* and $\mathcal{A}$-*Definability* of entities in $\mathbf{C}_{set}(\mathcal{A})$ are obvious generalisations of our previous definitions [the only difference being that we allow for formulae in $\mathbf{L}_{set}(\mathcal{A})$ to serve in definitions]. It is easy to show that if β_{fin} is true of $\mathcal{A}$ then very *finite* $\mathcal{A}$-invariant entity is $\mathcal{A}$-definable by a formula in $\mathbf{L}_{set}(\mathcal{A})$.

We now define a *strong interpretation* to be one , $\psi:\mathcal{A}\rightarrow\mathbf{B}$, which satisfies—in addition to preserving relative type differences—the following conditions:

(1) ψ when restricted to $|\mathcal{A}|$ is an injection (monomorphism) into a definable entity of $\mathbf{C}_{set}(\mathcal{A})$, of type $\{\tau\}$, for some τ; **(2)** if $(\exists z\in X)$ & $X\in\mathbf{C}_{set}(\mathcal{A})$ then $\psi X=\{\psi w|\ w\in X\}$; and **(3)** If R is any predicate symbol in $\mathbf{L}(\mathcal{A})$ then $\psi\mathfrak{R}_{\mathcal{A}}$ is $\mathcal{A}$-definable.

(We take $\psi\mathfrak{R}_{\mathcal{A}}$ or its definition to be the interpretation of R in $\mathbf{B}$).

A simple example of an interpretability result is the following

THEOREM (see §7, theorem 7.F). When $\mathfrak{B}$ is *elementarily invariant* and $\mathcal{A}$ is *finite,*
A *necessary and sufficient condition* for $\psi{:}\mathcal{A}\to\mathfrak{B}$ to constitute a *strong interpretation* of $\mathcal{A}$ in $\mathfrak{B}$ is

$$\psi^{-1} \otimes G(\mathfrak{B}|_{\psi|\mathcal{A}|}) \otimes \psi \subseteq G(\mathcal{A}) \ ,$$

where by $\mathfrak{B}|_{\psi|\mathcal{A}|}$ we mean the substructure of $\mathfrak{B}$ determined by $\psi\mathcal{A}$.

An equivalent formulation of this condition is that $\psi \otimes G(\mathcal{A}) \otimes \psi^{-1}$ must contain a subgroup isomorphic to the quotient group $G(\mathfrak{B})/ G(\mathfrak{B}|_{\psi|\mathcal{A}|})$ [where, of course, $G(\mathfrak{B}|_{\psi|\mathcal{A}|})$ must be a *normal subgroup* of $G(\mathfrak{B})$].

(Note: in the above conditions '$\otimes$' signifies *composition of mappings* and 'ψ' in the context of such composition stands for what should be properly denoted by '$(\lambda x)\psi x$' —if we were to use the λ-notation).

This is directly proved on the basis of the following instructive
Lemma (see 7.1 in §7): When $\mathfrak{B}$ is *elementarily invariant* and $\mathcal{A}$ is *finite,*
an injection ψ will constitute a *strong interpretation* of $\mathcal{A}$ in $\mathfrak{B}$ iff

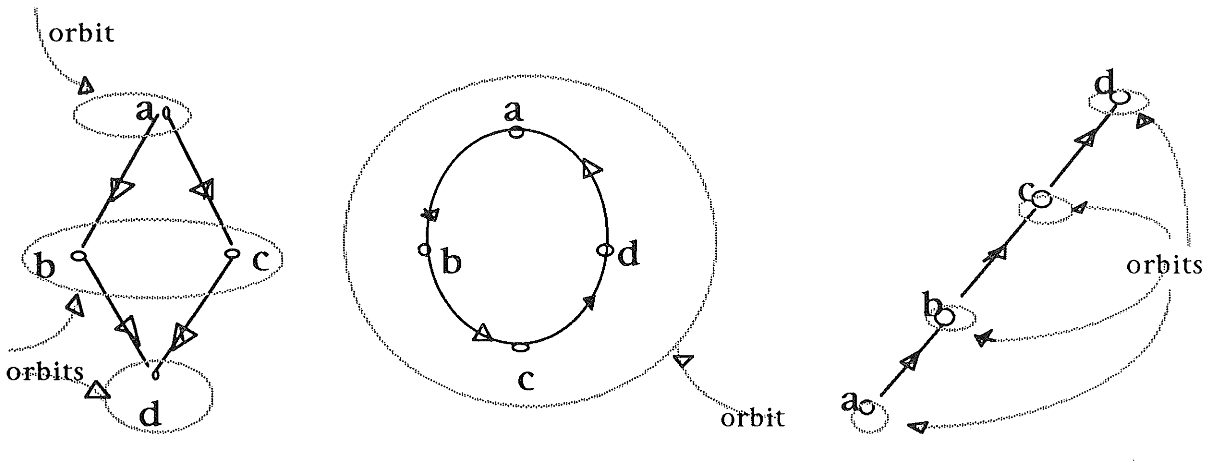
for any $\vec{a}\in|\mathcal{A}|^n$ $(n\geq 1)$, $\psi(\text{orbit}(\vec{a}))$ is $\mathfrak{B}$-invariant.

(Note: In the above it is enough to consider $\vec{a}$-s of dimensions corresponding to the n-arities of the predicate symbols in $\mathbf{L}(\mathcal{A})$).

__**Structural Information Theory**__. How much information does a structure $\mathcal{A}$ provide about one of its individuals or, more generally, about any particular element , e, in $\mathbf{C}_{set}(\mathcal{A})$?
Define the *minimal neighbourhood of* e *relative to* $\mathcal{A}$, as the smallest $\mathcal{A}$-invariant set containing e, (or *the orbit of* e). It is obvious that this minimal neighbourhood, $S_{\mathcal{A}}(e)$, contains only elements of the same type as e, and that in the case of a basic individual of the structure it is none other than its classical *orbit*. If e is of type τ and $\mathbf{m}$ is a measure on the subsets of $\mathcal{A}^{\tau} =\{x|x\in\mathbf{C}_{set}(\mathcal{A})\ \&\ \text{type}(x)=\tau\}$, then the specific information provided by $\mathcal{A}$ about e, $\text{Inf}_{\mathcal{A}}(e)$, will be an appropriate function of $\mathbf{m}(S_{\mathcal{A}}(e))/\mathbf{m}(\mathcal{A}^{\tau})$. If $\mathcal{A}$ is finite we can take $\mathbf{m}$ to be the *cardinality-function* for any finite subset of $\mathcal{A}^{\tau}$, and we can choose

$$\text{Inf}_{\mathcal{A}}(e) = -\log_N(\text{cardinality}(S_{\mathcal{A}}(e))/N), \text{ where } N=\text{cardinality of } \mathcal{A}^{\tau}.$$

Thus, In the case of the following three structures



Δ Δ_1 Δ_2

where the arrows represent a binary relation, we have, in Δ, $Inf_{\Delta}(a)=1$,$Inf_{\Delta}(b)=0.5$ and an average information per individual (type 0), $I^0(\Delta)$, of 0.75.
In Δ_1 on the other hand we have $Inf_{\Delta_1}(x)=0$ for any *individual* x (so $I^0(\Delta_1)=0$).
The picture is different for *orderd pairs*, however! In Δ_1 we have

$$Inf_{\Delta_1}(<x,y>)= -\log_{16}(4/16)=0.5 \quad (\text{or } I^{<0,0>}(\Delta_1)=0.5).$$

Finally, in the *linear* structure Δ_2 we have $Inf_{\Delta_2}(e)=1$, for e of any type !

While such measures of the Semantic Information that is contained in a structural description give us a handle on the structure's capacity to encode specific information (about whatever is described when one uses it), the same measures do not reflect the true relative value of various structures in their general use! We may have other reasons to use a specific structure, which far out-weigh its *"Internal Informability"*. Culturally and Scientifically we gravitate towards structures that have for us a high *metaphorical suggestiveness*—the potential to serve as a metaphorical vehicle in describing and representing many different types of data and phenomena. While some of the reasons for such a high metaphorical value are historical and cultural, others are certainly grounded in the nature of the structures themselves. The Intrinsic "Metaphorical" qualities of a structure have to do with its intrinsic *simplicity* and *symmetricity*, since it is a higher value of these, according to our Galois analysis, which will be positively correlated with a greater chance of successfully serving a (strong enough) *interpretative role* vis à vis other, arbitrarily provided, structural descriptions.

It is thus easy to see that when "universes" of the same nominal size are organised by different structures , *Internal Informability* and *Metaphorical Power* are *inversely correlated*, and a decision may have to be made as to how much specific Information should be sacrificed for the sake of simplicity of description and of analogy to structured descriptions of other data. *This balancing act between Information and Metaphorical power is at the heart of both Science and Art.* The above tools allow us to develop precise measures of the intuitive cost-effectiveness of such multi-faceted activities, which integrate various *explanatory desiderata*, in Science, and which correspond to vital aspects of intuitive evaluation of metaphors in the Arts.