THE BULLETIN OF SYMBOLIC LOGIC
Volume 14, Number 1, March 2008

LOGICALITY AND INVARIANCE

DENIS BONNAY

Abstract. This paper deals with the problem of giving a principled characterization of the class of logical constants. According to the so-called Tarski–Sher thesis, an operation is logical iff it is invariant under permutation. In the model-theoretic tradition, this criterion has been widely accepted as giving a necessary condition for an operation to be logical. But it has been also widely criticized on the account that it counts too many operations as logical, failing thus to provide a sufficient condition.

Our aim is to solve this problem of overgeneration by modifying the invariance criterion. We introduce a general notion of invariance under a similarity relation and present the connection between similarity relations and classes of invariant operations. The next task is to isolate a similarity relation well-suited for a definition of logicality. We argue that the standard arguments in favor of invariance under permutation, which rely on the generality and the formality of logic, should be modified. The revised arguments are shown to support an alternative to Tarski's criterion, according to which an operation is logical iff it is invariant under potential isomorphism.

On the traditional semantic account of logical consequence, a sentence ϕ is said to follow from a set Γ of sentences iff, for every uniform reinterpretation of the extra-logical expressions in Γ and ϕ, if all sentences in Γ are true, then ϕ is true. What is nice with this semantic definition of logical consequence is that it is purely extensional. It operates a reduction of logical truth to some kind of general truth, thanks to the quantification over all interpretations, without any need to appeal to metaphysical notions of possibility or necessity. What is more problematic though is that it rests crucially on the distinction between logical and extra-logical expressions. To get the account of logical consequence right, it is thus mandatory to know where the line should be drawn, unless one is ready to accept that we have nothing but a relative concept of logical entailment.

We aim here at giving a principled characterization of the class of (possible denotations for) logical constants. The first section is devoted to a

Received February 2, 2007.

I wish to thank the following for helpful discussions: Johan van Benthem, Serge Bozon, Mikaël Cozic, Jacques Dubucs, Paul Égré, Solomon Feferman, Henri Galinon, John MacFarlane, Gabriel Sandu, Philippe Schlenker, Gila Sher, Jouko Väänänen and Dag Westerståhl. Special thanks to Johan van Benthem, Serge Bozon and Solomon Feferman for very fruitful suggestions.

1079-8986/08/1401-0002/$5.00

presentation of the standard semantic approach to the problem in terms of invariance. In agreement with equally standard objections in the literature, we argue that Tarski's thesis, according to which permutation invariance is a necessary and sufficient condition for logicality, should be revised because it overgenerates by counting too many operations as logical. The second section introduces a general framework to discuss the connection between notions of similarity between structures and the operators which are invariant under these notions of similarity. The problem is to find suitable constraints on the similarity relations to be used for a characterization of logicality. This is done in the third section. We introduce two main constraints, closure under definability and absoluteness, which are shown to yield a nice map of the landscape of all available similarity relations. These two constraints provide the basis of two arguments in favor of a certain revision of Tarski's thesis. These arguments are given in the fourth section, which is devoted to an evaluation of the consequences of such a revision upon the relationship between logic and mathematics.

§1. The semantic road to logicality.

1.1. Characterizing logical constants. Some demarcations of the class of logical constants will clearly not do. The following example by Etchemendy (1999) makes clearly the point:

(1) If Leslie was a president of the US, then Leslie was a man.

Let us assume furthermore that "if ... then ", "president of the US" and "man" belong to the logical vocabulary, and that "Leslie" does not. In this case, (1) is a logical truth iff the following is true:

(2) For all x, if x was a president of the US, then x was a man.

(2) happens to be true, due to very contigent features of the history of the United States. Therefore, (1) is a logical truth, relative to the chosen demarcation. But this choice is clearly inadequate, because it makes the property of being a logical truth depend on substantive facts about the world which have nothing to do with logic. Etchemendy's point is more far-reaching than this example might suggest. He thinks that no choice of logical expressions will do, because it is congenial to the semantic definition of logical consequence that logical truth is defined in terms of some general truths. For example,

(3) If John is tall, then Greg is tall.

is a logical truth iff, for the standard choice of logical constants, the following is true:

(4) $\forall x, y \forall P\ (Px \rightarrow Py)$

Intuitively, (3) is not logically true. But (4) might be true if our background set theory was so weak that it would fail to provide enough extensions for predicates.

Now, this point can indeed be granted to Etchemendy. But, as remarked by Shapiro (1998), this just shows that the adequacy of the demarcation makes sense only with respect to the setting in which the formalization is to take place. In this respect, no absolute foundation of logical truth is achieved, but, inside a given setting, a satisfactory account of the realm of logical truths, as opposed to empirical or mathematical truths, might be nonetheless possible.

What would be a satisfactory account? Something more than "getting the definition of logical consequence right" is at stake. A raw list of expressions may be such that taking exactly these expressions as logical would result in a definition of logical consequence matching our intuitions. It will not be sufficient as an explanation of logical truth. We do want to know both which sentences are logically true and what makes them so. Therefore, the demarcation is to be attained through a conceptual analysis of what it is to be logical. The story should go roughly like this. Logic being what it is, logical expressions have to enjoy some special properties. Then the explanation of the distinctive nature of logical truth hinges on the fact that only expressions with these properties should occur essentially in a sentence for it to be logically true. In this case, something more than extensional adequacy would indeed be achieved: the formalization would provide an explanation of what makes for the difference between the realm of logic and other kinds of truths.

Various strategies have been proposed to tackle this issue. Some of them do not fit into the model-theoretic picture. This applies in particular to proof-theoretic approaches and claims like "an expression is logical if its use is entirely determined by rules of a certain form".[1] Other solutions are pragmatic rather than *stricto sensu* semantic; given some nice model-theoretic properties of logics, the set of logical constants is taken to be the largest set compatible with these.[2] This paper is devoted to a discussion of the purely semantic route. The aim is to provide a characterization of logicality as a property of semantic values available as denotations for logical expressions. To put it bluntly, what is so special about the interpretation of logical constants that forces us to keep them fixed when testing for logical entailment?

1.2. Tarski's thesis. The standard semantic account of logicality is couched in terms of invariance. Let us see what the claim exactly is, how it can be justified, and what are its consequences.

TARSKI'S THESIS. Given a set M, an operation Q_M acting on M is logical iff it is invariant under all permutations.

To make sense of this thesis, quantifiers need to be viewed as higher-order predicates: they yield truth values when applied to predicate extensions. A

[1]Dummett (1991) or Došen (1994) are good examples.

[2]Quine's (1986) insistence on completeness exemplifies such a pragmatic state of mind.

unary monadic quantifier is thus interpreted on a domain M by an operation $Q_M : \wp(M) \to \{T, F\}$. Q_M is said to be **invariant under permutation** iff for all permutations π, for all $A \subseteq M$, $Q_M(\pi(A)) = Q_M(A)$. For example, we can interpret $\exists$ on a domain M by an operation $Q_\exists$ defined by $Q_\exists(A) = T$ if $A \neq \emptyset$, and $Q_\exists(A) = F$ if $A = \emptyset$ (where $A \subseteq M$). Now, let π be a permutation, the image $\pi(A)$ of a non-empty subset A will always be another non-empty subset, and the empty set will always be mapped to the empty set. So $\exists$ is invariant under permutation, hence logical. On the contrary, let a and b be two elements of M, the operation $Q_{a,b}$ defined by $Q_{a,b}(A) = T$ iff $a \in A$ or $b \in A$ grants a special role to a and b. It is not invariant under permutation, and it should not count as logical, because it is sensitive to the difference between a and b on the one hand and the other objects in the domain on the other hand.

As it stands, the claim about logical constants is underspecified: Tarski's thesis is just a thesis about logical operations. We still need a story about the logical symbols themselves. Sher (1991) suggests the following connection between logicality as a property of operations and logicality as a property of interpreted symbols:

TARSKI–SHER THESIS. An interpreted symbol $\overline{Q}$ is logical iff:

- $\overline{Q}$ is of type level at most 2 (*i.e.*, $\overline{Q}$ has the syntactic type of a propositional connective or of a first-order quantifier).
- $\overline{Q}$ is interpreted by an operator Q associating to each domain M an operation $Q(M)$ of the appropriate semantic type.
- Q is invariant under bijections across domains.

The shift from Tarski's thesis about logical operations to the Tarski–Sher[3] thesis about logical symbols is not as trivial as it may seem. The type restriction could be challenged: why should we not deal with higher-order languages and higher-order logical constants? The fact that it is sufficient for a symbol to be logical that its denotation be logical could be challenged as well. First, it has been argued that a logical symbol should have something like a logical *intension*, in a sense which is not so easy to make precise. Second, it has also been argued that a logical symbol should have a *uniform* logical denotation: logical operations exhibiting a wildly heterogenous behavior across domains of different size should be banned. Notwithstanding the interest of these problems,[4] we shall focus here on what we take to be the

[3]"Tarski's thesis" and "Tarski–Sher thesis" are sometimes used synonymously in the literature. It is clear that the core of the Tarski–Sher thesis is Tarski's thesis. Since the idea of permutation invariance was put forward by Tarski as early as 1966, it seems fair to attribute this thesis to Tarski. On the other hand, Tarski does not consider explicitly the question of logical symbols, which was addressed in Sher's book. Since there are many reasons, even maybe some good ones, to accept Tarski's thesis while rejecting Sher's view, it seems equally fair to us to isolate explicitly a "Tarski–Sher thesis" about logical symbols.

[4]See Bonnay (2006) for a more detailed discussion of these issues and how they relate to the one issue on which we focus here, namely overgeneration.

central issue concerning the Tarski–Sher thesis, namely Tarski's thesis itself. We shall first be concerned with *ex ante* justifications for the thesis, before discussing, *ex post*, the adequacy of its consequences with respect to what logic is.

Invariance under permutation can be justified along two different lines. They are often confused in the literature. However, from a conceptual point of view, it is worth distinguishing them. The first one, endorsed by Tarski (1986), is based on a conceptual analysis of the generality of logic. The second one tries to capture the idea that logic is formal. It can be found in Sher (1991) as well as in previous mathematical works on generalized quantifiers (Mostowski 1957).

The idea that a theory can be characterized by the class of transformations under which it is invariant was successfully proposed by Felix Klein in order to unify and classify the field of geometry. The notions of Euclidean geometry are invariant under transformations of space that preserve distance (the isometric transformations) and more generally under those that preserve ratios of distances (the so-called similarity transformations). The point of the *Erlangen Program* is to turn this into a definition: the study of Euclidean geometry is just the study of the notions that are invariant under similarity transformations The class of transformations under consideration can be narrowed down or widened up. For example, if all transformations which preserve collinearity – similarity transformations are just a subclass of these—are allowed, this will characterize a more general geometry, namely affine geometry. Affine geometry is more general in the sense that it can distinguish fewer objects (all triangles are the same) and has fewer theorems. The next move is to extend this beyond geometry and to assume that it is relevant to try such a characterization for all kinds of theories, as has actually also been done in the field of physics. Then, a characterization of logic will come up as a result of the following argument:

The generality argument.

G.1 The distinctive feature of logic among other theories is that it is the most general theory one can think of.

G.2 The bigger the group of transformations associated with a theory, the more general the theory.

G.3 The biggest group of transformations is the class of all permutations.

The logical notions are the notions invariant under permutations.

The idea of G.1 is to characterize logic through its position in our conceptual scheme. Logical notions can be used in any other field, logical laws are used in any axiomatic theory, so that none of these should come with any restriction on the kind of objects to which they apply. We have seen that in the setting of the Erlangen Program, generality is gained by enlarging the family of transformations: this is the point of G.2. Now a characterization of logic is at hand: it ensues from an answer to the question, what is the biggest group of transformations? The answer, provided by G.3, is Tarski's answer (1986): it is just the class of all permutations of the universe onto itself. Note that G.3 is as secure as a mathematical fact can be: it follows from the definition of a transformation group that the biggest group of transformation acting on a set is the group of all one-one functions onto this set. Geometries are characterized by transformations respecting some structure of space, be it measure ratios or collinearity. Abstracting from any kind of special feature of the universe in order to get to the most general notions, we just end up with all permutations.

The other justification by Sher and others rests on different assumptions:

The formality argument.

F.1 Logic deals with formal notions, as opposed to non formal ones.
F.2 Formal notions are those which are insensitive to arbitrary switchings of objects.
F.3 A notion is insensitive to arbitrary switching of objects iff it is invariant under permutation

---

The logical notions are the notions invariant under permutation.

By contrast with G.1, F.1 is an ontological claim. Logic is characterized directly in terms of the kind of objects it deals with. However, logic, in some sense, is not about any special kind of objects; it is topic neutral. Therefore, it can just be said to be about the formal features of reality, those that make sense for every kind of objects or properties. Next, an account has to be provided of what it is to be formal. G.2 suggests that non formal properties,[5] as opposed to formal ones, are sensitive to the identity of objects. For example, "red" is a non formal, empirical predicate, and it distinguishes between things which are red and things which are not. If it applies to some object, say this apple, and if the object under consideration is arbitrarily replaced by another one, say this chair, there is no reason why it should still apply to it (the apple's being red does not force the chair's being red). On the contrary, a formal property should not distinguish among objects. It should be preserved under arbitrary switchings. But then—this is G.3—a permutation precisely performs such switchings. Therefore, formal properties can be singled out as being precisely those that are invariant under all permutations.

Both arguments share a common structure: they start with an intuition about what it means to be logical (G.1 and F.1), then they push forward the

[5]Properties which are not formal might be dubbed 'empirical'. However, this does not seem completely correct. 'x is a God' is certainly not a formal property, but one might be reluctant to consider it to be an empirical property. This is the reason why in F.1 we say that 'Logic deals with formal notions, *as opposed to non formal ones*' rather than 'as opposed to empirical ones'.

analysis of this intuition by formalizing it (G.2 followed by G.3, F.2 followed by F.3). Both arguments converge towards permutation invariance, and the convergence is non-trivial. As a matter of fact, the intuitive starting points are different, but both arguments conclude on the basis of the formalization of these intuitions that invariance under permutation is the right criterion to characterize logicality. This agreement suggests that Tarski's thesis is conceptually well-motivated.[6]

1.3. The overgeneration problem. What kind of notions are logical on this account? The standard quantifiers of first-order logic (FOL) are interpreted by such invariant operations. But there are a lot of new quantifiers that turn out to be logical, and the realm of permutation invariant operations goes far beyond FOL. Let $\overline{Q_{\text{most}}}$ be a monadic quantifer such that $\overline{Q_{\text{most}}}x\ \phi(x)$ is true in a model iff most of the objects in the model satisfy $\phi(x)$. The operator Q_{most} is invariant under permutations, but it is not definable in FOL. The same goes for the set-theoretic quantifier $\overline{Q_{>\aleph_0}}$ ("there are uncountably many objects such that"). It is interpreted by an operator $Q_{>\aleph_0}$ which associates with every set M an operation $Q_{>\aleph_0}(M)$ yielding the value "true" iff its argument is an uncountable set. $Q_{>\aleph_0}$ is invariant under permutation, but it is not definable in FOL, and actually, the downward Löwenheim–Skolem property will fail for any logic containing $\overline{Q_{>\aleph_0}}$.

We have just given some examples of logical quantifiers. But it is possible to give a much more accurate account of the nature of logic, once Tarski's thesis has been accepted. Let us draw a comparison with propositional calculus (PC). The functional completeness theorem for PC states that every truth-function is definable using the standard logical constants of PC, say $\wedge$, $\vee$, $\rightarrow$ and $\neg$. Therefore, PC is all there is to truth-functional logic. There is no such standard result for FOL, essentially because there is no standard answer to what FOL is about, which would be similar to the claim that PC is about truth-functions. The Tarski–Sher thesis does bring an answer to this question, and it becomes possible to ask whether a logic—FOL itself or an extension of FOL—is functionaly complete with respect to what FOL is about. We are thus interested in picking a logical language enjoying functional completeness with respect to permutation invariant operations. More precisely, any appropriate choice of syntax and logical expressions should be such that, on the one hand logical expressions are permutation invariant, and, on the other hand, every permutation invariant operation is definable in purely logical terms. This can be done in infinitary logic. We shall use the standard notation $L_{\kappa,\lambda}$, where κ and λ are cardinals. $L_{\kappa,\lambda}$ is the logic with the same basic connectives as FOL in which conjunctions of fewer than κ-sentences are allowed, and quantification over sets of fewer than λ variables are allowed. $L_{\omega,\omega}$ is just FOL. ∞ indicates that sets of

[6] A thorough discussion of the relationships between this kind of arguments and more traditional views about logic can be found in MacFarlane (2000).

arbitrary size are allowed. $L_{\infty,\infty}$ is therefore the strongest infinitary logic, whose sentences are built out of arbitrarily big sets (proper classes are not allowed) of sentences and variables.

THEOREM 1.1 (McGee, 1996). *An operation Q_M acting on a domain M is invariant under permutation iff it is definable in $L_{\infty,\infty}$.*

So on the Tarski–Sher Thesis, logic is quite powerful, at least as powerful as $L_{\infty,\infty}$.[7] And here comes the criticism: it might be actually too powerful, or more powerful than logic really is.

On Tarski's criterion, the quantifier $\overline{Q_{\aleph_1}}$, which tests whether there are exactly $\aleph_1$ objects satisfying a formula, is logical. Intuitively, something has gone wrong. Being of size $\aleph_1$ is a notion which belongs to set theory, not to logic. As pointed out by Feferman (1999), an unfortunate consequence is that logical notions will not be robust with respect to the background set theory. For example, the conditions of application of $\overline{Q_{\aleph_1}}$ highly depend on set-theoretic facts about higher infinites. Being of size $\aleph_1$ is not an absolute notion, its meaning depends on which sets exist: a given set can be uncountable in a countable model $\mathcal{M}$ of ZFC, because there is no bijection in $|\mathcal{M}|$ between this set and $\aleph_0$, and uncountable in an extension of $\mathcal{M}$ in which there is such a bijection. In the same line, Feferman (1999) has also remarked that the restriction to first-order-like operations is only apparent, because "$L_{\infty,\infty}$ accomodates second-order quantification as a logical operation across domains" (p. 37). Therefore it is possible to express in $L_{\infty,\infty}$ the Continuum Hypothesis and other substantial set-theoretic claims.

The situation is even worse than that. Consider a topological space S and the operation $Cont_S\colon \wp(S \times S) \to \{T, F\}$ defined, for $R \subseteq S \times S$ by $Cont_S(R) = T$ iff R is a continuous function. As desired, the operation *Cont* is not invariant under permutation because it takes into account the underlying topological structure of the space S, which does not put all points on a par. But for an arbitrary set M, consider the operation $Cont'_M\colon \wp(\wp(M)) \times \wp(M \times M) \to \{T, F\}$ defined for $\mathcal{O} \subseteq \wp(M)$ and $R \subseteq M \times M$ by $Cont'_M(\mathcal{O}, R) = T$ iff the family $\mathcal{O}$ is a topology on M and R is a continuous function on $\langle M, \mathcal{O}\rangle$. $Cont'_M$ is invariant under permutation and, if the domain set is allowed to vary, $Cont'$ is invariant under bijection. Thus, by making the topological structure explicit, we can turn continuity into a logical notion.

As a matter of fact, every structure or class of structures can now be described up to isomorphism in one shot by a tailor-made logical operation:[8]

[7] Note however that operations *across* domains cannot be defined unless one allows for proper classes in the syntax. We will go back to this point in the last section.

[8] In this section, we try to remain as neutral as possible with respect to the setting we choose. A more detailed presentation of invariance criteria and of the setting in which we discuss them will be given in the next section.

FACT 1.2. *Let K be a class of structures of a given similarity type. If K is closed under isomorphism, there is a language L, whose logical symbols are interpreted by bijection invariant operators, and a sentence ϕ_K of L such that for every model $\mathcal{M}$, $\mathcal{M} \in K$ iff ϕ_K is true in $\mathcal{M}$.*

PROOF. For simplicity, assume that K is a class of relational structures of the form $\langle M, R\rangle$. We define the operation $Q_{K,M}$ over a set M by setting, for $R \subseteq M \times M$, $Q_{K,M}(R) = T$ iff $\langle M, R\rangle$ is in K. This operation can be used as the interpretation of a binary generalized quantifier $\overline{Q_K}$ endowed with the following satisfaction clause:

$$\mathcal{A} \vDash \overline{Q_K}x, y\ \phi(x, y)\ [\sigma] \text{ iff } Q_{K,M}(\|\phi(x, y)\|_{\mathcal{A},\sigma}) = T$$

where σ is an assignment on A and $\|\phi(x, y)\|_{\mathcal{A},\sigma}$ is the interpretation of ϕ over $\mathcal{A}$ according to σ, that is the set of pairs $\langle a, b\rangle$ of elements of $\mathcal{A}$ such that $\mathcal{A} \vDash \phi(x, y)\ [\sigma][x := a][y := b]$.

We consider a first order language L extended with the quantifier $\overline{Q_K}$. Since K is closed under isomorphism, Q_K is bijection invariant. Conversely, the class K is trivially characterized by the sentence $\overline{Q_K}xy\ \overline{R}xy$: $\langle M, R\rangle \vDash \overline{Q_K}xy\ \overline{R}xy$ iff $\langle M, R\rangle \in K$ by definition of $\overline{Q_K}$. ⊣

Every class of structures, like our class K, which is closed under isomorphism thus gives rise to a logical quantifier, like $\overline{Q_K}$. In this sense, every mathematical notion gives rise to a closely related[9] logical notion. The question whether a given class of structures is elementary is trivialized, and the very idea of a difference between logical and mathematical notions becomes elusive.

All this suggests that Tarski's criterion overgenerates and counts too many operations as logical. First, since the aim is to distinguish the realm of logic proper, a proposal which conflates logical notions and mathematical notions does not seem to be on the right track. At least from a methodological standpoint, one should look for alternative proposals to see whether it is possible to do justice to the intuitive idea that logic is more basic than the whole of mathematics, and more basic than set theory. Second, if logic is to be used as a framework to develop alternative mathematical theories, it should be neutral with respect to issues which are not mathematically settled, such as the Continuum Hypothesis. Therefore, permutation invariance, though it is a necessary condition for being logical, does not seem to provide a sufficient condition. This situation is not very comfortable. We have seen that Tarski's thesis was well-motivated through the formality and the generality arguments. But, unfortunately, its consequences are hard to accept. Should we nonetheless accept those consequences, as suggested by Sher, or are there reasons to think that the conceptual analysis has gone wrong somewhere?

[9]This qualification is necessary. 'Being 0' for example, as a notion applying only to the 'genuine' zero is not invariant under permutation. However, the higher-order correlate 'being an empty-set' and the relativized version 'being the smallest element in a structure isomorphic to the natural numbers' come out as invariant.

We choose not to bite the bullet. Both the generality and the formality arguments are actually more problematic than it might seem at first sight. The argument in terms of generality rests on the assumption that invariance under the biggest class of transformations yields maximal generality. The idea is that the group of all permutations is as "big" as one might wish, because in that case the transformations do not respect any extra-structure, such as *e.g.*, the topological structure of the space. Let us have a closer look at this idea. Permutation invariance just says that as soon as there is an automorphism linking $\langle M, A\rangle$ and $\langle M, A'\rangle$, a quantifier Q acting on M has to give A and A' the same value. On the one hand, this is indeed liberal, because no further structure beyond the extensions A and A' on M is taken into account. But on the other hand, this is quite demanding: for $\langle M, A\rangle$ and $\langle M, A'\rangle$ to be similar from a logical point of view, they have to share exactly the same structure—they have to be isomorphic. Now there are a lot of other concepts of similarity between structures which are used in model theory and in algebra which are far less demanding. Instead of requiring the structure to be fully preserved, they lower the requirement to some kind of approximate preservation. Why should we refrain from resorting to these other concepts? To sum up, even if one grants that generality is a good way to approach logicality, there is no evidence that the class of all permutations is the best applicant for the job.

The problem with the justification in terms of formality is not with the adequacy of the formalization, but rather with the adequacy of the intuitive starting point itself. It is reasonable to think of logic as being formal, namely free from 'empirical' content. But why should we take this to be sufficient to characterize logic as opposed to mathematics or set-theory? After all, there is more to content than permutation sensitive content. The problem with a quantifier like $Q_{\aleph_1}$ is precisely that it encapsulates too much "set-theoretical content". Formality is a property of logic that is shared by set-theory and other branches of mathematics: it is not a surprise that taking formality as the starting point of an analysis of logicality yields a collapse of logic into mathematics. A proper analysis of the distinctive feature of logic should take into account the fact that it is even more "content-free" than set-theory.

§2. Generalized invariance.

In this section, we aim at introducing a general setting to discuss invariance criteria for logicality.

2.1. Invariance under a similarity relation. The main idea is to consider invariance under arbitrary definitions of what it means for two structures to be similar, without restricting attention to groups of transformations in Klein's sense.[10] First, instead of looking at the interpretation of generalized

[10] This general view on invariance is implicit in van Benthem (2002). Van Benthem insists on the close connection between the invariance approach and definability results. Moreover, he notes that "Any reasonable invariance analysis must postulate some relevant equivalence

quantifiers as families of boolean-valued operations on sets, we will think of them just as classes of structures. If $\overline{Q}$ is a unary monadic quantifier of type $((e,t),t)$, its denotation Q yields an operation Q_M on each set M which sends subsets of M (objects of type (e,t)) to objects of type t, that is, either T or F. Alternatively, Q will be viewed as the class of structures of the form $\langle M, P\rangle$ such that $Q_M(P) = T$. For example, on this account, the denotation $Q_\exists$ of $\exists$ is the class of structures of the form $\langle M, P\rangle$ such that P is a non-empty subset of M. Note that truth-functions fit into this picture if booleans are part of the structures in one way or the other. One option would be to work with multi-sorted structures and two separate domains, the domain of objects and the domain of truth-values. For simplicity, we prefer to consider booleans as genuine parts of the structure over a set: we will take them to be the interpretation of 0-ary predicates. By a natural generalization of the notion of interpretation of an n-ary predicate for $n \neq 0$, there are only two possible values for a 0-ary predicate, namely $\emptyset$ and $\{\emptyset\}$, that we will construe respectively as false and true. On this account, the denotation of the disjunction $\vee$ is the class of structures of the form $\langle M, T, T\rangle$, $\langle M, T, F\rangle$ or $\langle M, F, T\rangle$. Finally, for any operator Q and for any structure $\mathcal{M}$, we will use the notation $Q(\mathcal{M})$ to indicate that $\mathcal{M}$ is in the class Q. All operators we consider will be **finitary** operators, namely classes of structures with a finite number of constants and a finite number of relations of finite arity.[11]

A **similarity relation** S is a relation between structures respecting signatures (*i.e.*, S is a family of relations S_σ between σ-structures for all signatures σ). The notation is $\mathcal{M}\ S\ \mathcal{M}'$.

We are interested in the **invariance** of operators under similarity relations. We say that an operator Q is S**-invariant** iff, for any structures $\mathcal{M}$, $\mathcal{M}'$, if $\mathcal{M}\ S\ \mathcal{M}'$, then $Q(\mathcal{M})$ iff $Q(\mathcal{M}')$. For example, let us say that $\mathcal{M}$ *Auto* $\mathcal{M}'$ iff there is an automorphism between $\mathcal{M}$ and $\mathcal{M}'$, and that $\mathcal{M}$ *Iso* $\mathcal{M}'$ iff there is an isomorphism between $\mathcal{M}$ and $\mathcal{M}'$. It is clear that *Auto*-invariance is just invariance under permutation in the previous sense, and that *Iso*-invariance is just invariance under bijection. Looking for an alternative to Tarski's thesis, we want to defend a claim of the form:

> An operator Q is logical iff it is S-invariant.

for a certain similarity relation S. The point is therefore to find an S such that it is possible to give conceptual arguments in favor of S in the spirit of the formality and generality arguments, and such that S-invariance does not overgenerate as badly as does *Iso*-invariance.

---

relation on models" (2002, p. 431). Our generalization basically consists in letting any kind of relation come into play. As we will see, equivalence relations do play a special role though. Abstract equivalence relations have also been considered in abstract model theory: for an overview, see Makowsky and Mundici (1985).

[11] In principle, generalization to the infinite case should be straightforward.

2.2. The Galois connection for invariance. Before considering particular similarity relations, it will be useful to gain a better understanding of the framework and to look into the duality between classes of "quantifiers" (classes of operators) and similarity relations. Given a similarity relation S, there is a class of operators which is naturally associated with S, namely the class of all operators which are S-invariant, that we shall note $Inv(S)$. In the other direction, given a class of operators K, there is a similarity relation which is naturally associated with K, namely the biggest similarity relation leaving all operators in K invariant, that we shall note $Sim(K)$ and formally define by $\mathcal{M}\ Sim(K)\ \mathcal{M}'$ iff for all $Q \in K$, $Q(\mathcal{M})$ iff $Q(\mathcal{M}')$.

Let $\mathcal{K}$ be the class of all classes of operators. We shall look at $\mathcal{K}$ as an ordered class, by setting $K \leq K'$ iff $K \subseteq K'$. In the same way, let $\mathcal{S}$ be the class of all similarity relations. We shall also look at $\mathcal{S}$ as an ordered class, by setting $S \leq S'$ iff $S' \subseteq S$ (*i.e.*, iff for all $\mathcal{M}$, $\mathcal{M}'$, if $\mathcal{M}\ S'\ \mathcal{M}'$ then $\mathcal{M}\ S\ \mathcal{M}'$). Intuitively, S is lower than S' iff S is less fine-grained than S' (S identifies more structures than S').

$Sim\colon \mathcal{K} \to \mathcal{S}$ and $Inv\colon \mathcal{S} \to \mathcal{K}$ are two "symmetric" functions which allow us to go from classes of operators to similarity relations and the other way around. How symmetric are they? It is easy to check that Inv and Sim are both monotone: if $K \leq K'$, then $Sim(K) \leq Sim(K')$, and if $S \leq S'$ then $Inv(S) \leq Inv(S')$. However, Inv and Sim are not inverses: in general, it is not the case that $Sim(Inv(S)) = S$ and $Inv(Sim(K)) = K$. Therefore, Inv and Sim do not provide an isomorphism between our two ordered structures. But one can show the following fact:

FACT 2.1. *For any class K of operators and any similarity relation S, $Sim(K) \leq S$ iff $K \leq Inv(S)$.*

PROOF. Only if. Assume $Sim(K) \leq S$. It follows that $Inv(Sim(K)) \leq Inv(S)$. But $K \leq Inv(Sim(K))$. Therefore $K \leq Inv(S)$.

If. Assume $K \leq Inv(S)$. It follows that $Sim(K) \leq Sim(Inv(S))$. But $Sim(Inv(S)) \leq S$. Therefore $Sim(K) \leq S$. ⊣

$\langle \mathcal{K}, \mathcal{S}, Sim, Inv \rangle$ is thus a **Galois connection** (or, in terms of category theory, Sim and Inv are adjoint functors). This makes precise the "symmetry" between Sim and Inv. It follows immediately from the Galois connection that $Inv(S) = Inv(Sim(Inv(S)))$ and $Sim(K) = Sim(Inv(Sim(K)))$. Moreover, we have an isomorphism between $Inv(Sim(\mathcal{K}))$ and $Sim(Inv(\mathcal{S}))$. Finally the so-called **kernel operator** $Sim \circ Inv\colon \mathcal{S} \to \mathcal{S}$ is of special interest:

FACT 2.2. *$Sim(Inv(S))$ is the reflexive, transitive and symmetric closure of S.*

PROOF. Let $\mathcal{M}$ and $\mathcal{M}'$ be two structures such that $\mathcal{M}\ Sim(Inv(S))\ \mathcal{M}'$. We assume for contradiction that there is no sequence $\mathcal{M}_1, \dots, \mathcal{M}_n$ of structures such that $\mathcal{M} = \mathcal{M}_1$, $\mathcal{M}' = \mathcal{M}_n$ and for all i, $1 \leq i \leq n$, $\mathcal{M}_i\ S\ \mathcal{M}_{i+1}$ or $\mathcal{M}_{i+1}\ S\ \mathcal{M}_i$. Let $Orb(S, \mathcal{A})$ be the operator defined by $Orb(S, \mathcal{A})(\mathcal{M})$

iff there is such a sequence from $\mathcal{M}$ to $\mathcal{A}$. Clearly, $Orb(S, \mathcal{A})$ is S-invariant, therefore $Orb(S, \mathcal{A})$ is in $Inv(S)$. By hypothesis, we have that $Orb(S, \mathcal{A})(\mathcal{M})$ but not $Orb(S, \mathcal{A})(\mathcal{M}')$. This contradicts $\mathcal{M}\ \ Sim(Inv(S))\ \mathcal{M}'$. $\dashv$

Thus, $Inv(S) = Inv(Sim(Inv(S)))$ tells us that a similarity relation and the smallest equivalence relation containing it induce the same class of operators.

There are two main differences between generalized invariance and Klein's idea of invariance by a group of transformation. The first one is that similarity relations are not equivalence relations, so that we loose the "group" idea. However, we have just seen that this is not a substantial point, because we can always look at $Sim(Inv(S))$, which is an equivalence relation, instead of looking at S. The second difference is that elements of a group of transformations were transformations, namely one-one function from a set onto itself. Two structures were considered as similar if one could be transformed into the other by a transformation in the group: this means that similarity is always induced by a bijection. Generalized invariance is much more liberal with respect to that: any kind of relation between structures, be it grounded in a bijection or not, can be used as a similarity relation.

The nice thing is that we now have a wide range of possible choices of S, so that we might hope that solving the overgeneration problem while sticking to the invariance approach is possible. On the other hand, any choice of S must be philosophically motivated, and we need therefore to organize the landscape of invariance and classes of operators on a conceptual basis.[12] The next section will be devoted to this task. We will look for plausible constraints on S when S is to be used as a criterion for logicality.

2.3. Invariance under homomorphism. Before closing this section, we should remark that S-invariance admit as particular cases some of the alternatives to Tarski's thesis that have been put forward in the literature.

Feferman's proposal (1999) is to move from isomorphisms to strong homomorphisms:[13]

Definition 2.3 (Strong Homomorphism). Let $\mathcal{M}$ and $\mathcal{M}'$ be two structures of the same signature σ, a strong homomorphism from $\mathcal{M}$ onto $\mathcal{M}'$ is a surjective function $f : |\mathcal{M}| \mapsto |\mathcal{M}'|$ such that:

[12]We are especially indebted to Solomon Feferman and Philippe Schlenker for suggesting that we should provide a detailed account of conceptually meaningful constraints on similarity relations and for helping us to do so.

[13]Casanovas (2007) recently proposed a detailed analysis of the various versions of invariance under homomorphisms that could be used. In particular, he makes it clear that the criterion is quite sensitive to the type-theoretic setting (functional or relational) in which the operations are expressed. Casanovas also puts forward a notion of invariance under homomorphism of its own that cuts across first-order logic itself. This notion of invariance could be analyzed as invariance under a similarity relation in our sense (see Bonnay, 2006, for a detailed discussion). However, we do not consider as very attractive an invariance criterion which bans operations which are definable in pure first-order logic.

- For any constants a in σ, $f(a^{\mathcal{M}}) = a^{\mathcal{M}'}$.
- For any function symbol g of arity n in σ, for any n-uplet $\overrightarrow{a}$ of elements of $|\mathcal{M}|$, $f(g^{\mathcal{M}}(\overrightarrow{a})) = g^{\mathcal{M}'}(f(\overrightarrow{a}))$.
- For any relation symbol R of arity $n > 0$ in σ, for any n-uplet $\overrightarrow{a}$ of elements of $\mathcal{M}$, $\overrightarrow{a} \in R^{\mathcal{M}}$ iff $f(\overrightarrow{a}) \in R^{\mathcal{M}'}$.
- For any predicate symbol p of arity 0 in σ, $p^{\mathcal{M}} = p^{\mathcal{M}'}$.

We shall say that $\mathcal{M}$ *SHom* $\mathcal{M}'$ iff there is a strong homomorphism from $\mathcal{M}$ onto $\mathcal{M}'$. Note that *SHom* is not an equivalence relation, but that, once again, $Inv(SHom) = Inv(Sim(Inv(SHom)))$, $Sim(Inv(SHom))$ being the smallest equivalence relation containing *SHom*. The motivation for choosing $S = SHom$ has to do with the overgeneration problem. Because cardinality quantifiers should not all count as logical, it is tempting to suggest that logical operations should be insensitive to cardinality issues. This is precisely what is achieved by the shift from *Iso* to *SHom*. The injectivity requirement is left out, so that similarity relations are allowed which identify different objects. These intuitive motivations fit in the conceptual analysis which has been set in the previous section. In terms of generality, Feferman's choice for S fares indeed better than *Iso*, because every isomorphism is a strong homomorphism, while the converse is false. A reason for not looking below *SHom* is nevertheless still to be provided. In terms of lack of content, *Iso*-invariance neutralizes empirical content, and, as we have just seen, *SHom*-invariance neutralises both empirical content and "numerical" content, at level of finite and infinite cardinalities. This makes sense if logic is taken to be both formal and blind to numbers. The claim might however seem a bit too strong: the (set-theoretic) reasons to reject $Q_{\aleph_1}$ do not necessarily constitute reasons to reject $\exists!5$ as well.

An other independent reason for choosing *SHom* is the following. *SHom* allows for comparisons between all domains, in contrast with *Iso* which might connect only domains of the same size. Because of this special feature, one can hope that a *SHom*-invariant operation really means the same thing on every domain. In particular, a quantifier acting like $\exists$ on finite domains and like $\forall$ on infinite ones is *Iso*-invariant, but it is not *SHom*-invariant.

The following claim is now lurking around:

SHom THESIS FOR LOGICALITY. An operation Q is logical iff it is *SHom*-invariant.

Actually, this is not the view that Feferman (1999) upholds. He rather supports the following:

MONADIC *SHom* THESIS FOR LOGICALITY. An operation Q is logical iff it is λ-definable from monadic *SHom*-invariant operations.

Judging from the consequences, The monadic *SHom* thesis looks far more plausible than the simpler *SHom* thesis:

THEOREM 2.4. *An operation Q_M acting on a domain M is SHom-invariant iff it is definable in $L^-_{\infty,\infty}$.*[14]

THEOREM 2.5 (Feferman, 1999). *An operation Q is λ-definable from monadic SHom-invariant operations iff it is definable in FOL^-.*

Theorem 2.4 shows that *SHom* overgenerates nearly as much as *Iso*. It characterizes the same infinitary language, except that equality has to be dropped from the stock of logical constants. Now, this is not so significant a change, because, as shown by Quine (1986, p. 63), identity can be simulated in a quantificational language. It can be defined in a Leibnizian spirit as the greatest congruence relation for the language under consideration. In FOL, this works only if the number of extra-logical expressions is finite, but in $L^-_{\infty,\infty}$, even the qualification is unnecessary.

The interest of *SHom* as used in the *SHom* thesis for logicality is thus quite limited. Things are different with Feferman's thesis. Feferman's claim yields a nice characterization result, but only at the cost of putting additional constraints on the characterization. First, invariance concerns a restricted class of operations, the monadic ones. Then, some more machinery is needed to get polyadic operations from these, which is why typed λ-calculus comes into play. As a consequence, it is not sufficient to motivate the shift from *Iso* to *SHom*. Feferman needs two more arguments to support his thesis: one to justify the restriction to monadic operations, another one in favor of the logicality of the class of λ-terms which is used. Feferman does not explictly discuss the second point, but the lack of argument here is relatively harmless. Since *SHom*-invariance is preserved by λ-definability, if one is to accept *SHom* for monadic types, one should accept the closure under λ-definability. But the first point is more tricky: if *SHom* enables one to disregard exactly the features of objects which logic disregards, why should one demand more than that and restrict the application of *SHom*-invariance? Feferman's argumentation is based on linguistic datas. He considers that "natural language evidence supports the view of quantifiers as, first and foremost, monadic operators" (1999, p. 47). Following ideas put forward by Montague, Barwise and Cooper (1981) have shown that monadic generalized quantifiers provide very efficient tools for a formal semantics of natural language sentences. Sentences are basically formed by combining a verb phrase and a noun phrase. In the sentence

Most critics are pedantic.

the semantic interaction of the noun phrase "Most critics" with the verb phrase "are pedantic" is commanded by the determiner of the noun phrase. "Most" is construed as a quantifier of type $\langle 1, 1\rangle$ which applies to the two

[14]This answers an open question raised by Feferman (1999, question 6.1 p. 47). L^- is the equality free version of L. The proof is omitted here.

predicates provided by “critics” and “pedantic”. In more complex sentences, like

> Most critics reviewed just four films.

a polyadic quantifier of type $\langle\langle 1, 1\rangle, 2\rangle$ has to be constructed out of the monadic quantifiers “Most” and “Four”. The centrality of monadic quantification boils down to the fact that polyadic quantifiers are obtained in general by lifting monadic quantifiers in one way or the other.

This line of argument raises two questions. First, why should natural language evidence come into play in a characterization of logic? Second, does the fact about lifting really support the restriction to λ-definability from genuine monadic logical quantifiers?

Answers to these questions have to take into account another linguistic fact. This fact is stated by Keenan and Westerståhl: “On [...] natural notions of expressive power, monadic quantifiers are *not* sufficient to express quite common constructions [...] involving polyadic quantifiers ” (1997, p. 890). Relevant cases include branching quantification:

> Quite a few of the boys in my class and most of the girls in your class have dated each other.

or reciprocals, such as exemplified by

> Most of the boys in your class like each other.

As a consequence, the claim that monadic quantification is all there is to natural language quantification can be sustained only if quite powerful ways of lifting monadic quantifiers are used. In particular, λ-definability, which is used in the monadic *SHom* thesis, will not do the job. As a consequence, any defender of the thesis faces a dilemma:

- Either natural language evidence is relevant to the project of a characterization of logicality. But then, logic should be expressive enough to account for how polyadic quantifiers can be defined from monadic ones. This is not compatible with Feferman’s account of logic, as shown by theorem 2.5.
- Or the semantics of complex natural language sentences is not relevant to this projet. But then, there is no good reason for the restriction to monadic quantification.

One could try to escape the dilemma by arguing that natural language evidence is relevant as long as monadicity is concerned, but that liftings do not have to be provided by purely logical means. In this case, the argument from natural language unfortunately appears quite *ad hoc*. There is still hope that one could come up with a characterization of logicality which would be purely in terms of invariance and would not rely on type restrictions. The next section is devoted to a further exploration of this route.

§3. **Constraints on similarity relations.** Now we go back to generalized invariance and introduce two constraints on similarity relations suitable to characterize logical operations. The first one, closure under definability is meant to qualify the idea of maximal generality in our framework. The second one, absoluteness, is meant to complement the idea that logical notions are (nearly) devoid of content.

3.1. Some interesting similarity relations. First, we shall introduce some similarity relations that will be used below, in order to give an idea of the kind of proposals that can be made. We shall actually highlight two kinds of similarity relations: a first family of very liberal similarity relations, a second family of similarity relations which are "local" variations on the idea of "being isomorphic". In what follows, $\mathcal{M} = \langle M, R_1, \dots, R_n, a_1, \dots, a_m \rangle$ and $\mathcal{M}' = \langle M', R'_1, \dots, R'_n, a'_1, \dots, a'_m \rangle$ are arbitrary structures (without functions, for the sake of simplicity), equipped with n relations and m constants.

Definition 3.1 (*Univ*). The similarity relation *Univ* is defined by $\mathcal{M}$ *Univ* $\mathcal{M}'$ iff $\mathcal{M}$ and $\mathcal{M}'$ have the same signature.

Intuitively, *Univ* is the universal similarity relation, which identify any two structures to which an operator of a given type could be applied to. It is the smallest element of $\mathcal{K}$.

Definition 3.2 (*Bool*). The similarity relation *Bool* is defined by $\mathcal{M}$ *Bool* $\mathcal{M}'$ iff for all $i \in \{1, \dots, n\}$ such that R_i in $\mathcal{M}$ and R'_i in $\mathcal{M}'$ are 0-ary relations, we have $R_i = R'_i$.

Intuitively, *Bool* is the boolean similarity relation, which identify any two structures whose boolean part is identical, that is any two structures which agree on extensions of 0-ary relations.

Definition 3.3 (*App*). The similarity relation *App* is defined by $\mathcal{M}$ *App* $\mathcal{M}'$ iff for all $i \in \{1, \dots, n\}$, $j_1, \dots, j_k \in \{1, \dots, m\}$, we have that $(a_{j_1}, \dots, a_{j_k}) \in R_i$ iff $(a'_{j_1}, \dots, a'_{j_k}) \in R'_i$ where m is the arity of R_i and R'_i.

Intuitively, *App* is the smallest similarity relation respecting functional application. Two structures which are *App*-similar satisfy exactly the same atomic formulas of a language whose signature is the signature of the structures, and also exactly the same Δ_0 formulas of that language. *App* respects booleans as a special case of functions applied to zero arguments.

We now define some less liberal similarity relations which refine on *Iso* by requiring the identity of formal features only at a local level. Two structures $\mathcal{M}$ and $\mathcal{M}'$ are locally similar if there is an isomorphism between two substructures of $\mathcal{M}$ and $\mathcal{M}'$. For example, a structure $\mathcal{M}$ containing a copy of $\mathbb{Z}$ will be locally similar in this sense to another ordered structure containing a copy of $\mathbb{Z}$ as well. This idea cannot make sense as a criterion of similarity, because the other parts of $\mathcal{M}$ and $\mathcal{M}'$ can be very different. One has to guarantee that any local similarity can be extended to other parts of the structures as well. This leads to similarity relations which are far less demanding than *Iso*.

DEFINITION 3.4 (Partial isomorphism). Let $\mathcal{M}$ and $\mathcal{M}'$ be two structures, and $f : |\mathcal{M}| \to |\mathcal{M}'|$ a function. f is a partial isomorphism between $\mathcal{M}$ and $\mathcal{M}'$ iff there are two substructures $\mathcal{A}$ and $\mathcal{A}'$ of $\mathcal{M}$ and $\mathcal{M}'$ such that f is an isomorphism between $\mathcal{A}$ and $\mathcal{A}'$.

The idea of local resemblance can be captured by requiring the existence of families of partial isomorphisms satisfying certain properties. In particular, we can require that it is always possible to extend the partial isomorphisms in any direction *a finite number of times*. We get the following definition where the ordinal α is a parameter precising what is meant by "always possible":[15]

DEFINITION 3.5 (α-isomorphism). Let α be an ordinal, $\mathcal{M}$ and $\mathcal{M}'$ two structures, an α-isomorphism I from $\mathcal{M}$ to $\mathcal{M}'$ (notation $I : \mathcal{M} \approx_\alpha \mathcal{M}'$) is a sequence

$$I_0 \supseteq I_1 \supseteq \cdots \supseteq I_\beta \supseteq \cdots \supseteq I_\alpha$$

s.t. for any $\beta \leq \alpha$, I_β is a non-empty set of partial isomorphisms, and if $\beta + 1 \leq \alpha$ and $f \in I_{\beta+1}$, then for any $a \in |\mathcal{M}|$ (resp. $b \in |\mathcal{M}'|$), there is a partial isomorphism $g \in I_\beta$ such that $f \subseteq g$ and $a \in dom(g)$ (resp. $b \in rng(g)$).

By requiring that partial isomorphisms can be *infinitely* extended, we get:

DEFINITION 3.6 (Potential isomorphism). A potential isomorphism I between two structures $\mathcal{M}$ and $\mathcal{M}'$ (notation $I : \mathcal{M} \approx_p \mathcal{M}'$) is a non empty set of partial isomorphisms such that:

for all $f \in I$ and $a \in |\mathcal{M}|$ (resp. $b \in |\mathcal{M}'|$), there is a $g \in I$ with $f \subseteq g$ and $a \in dom(g)$ (resp. $b \in rng(g)$).

Here are corresponding similarity relations. For an ordinal α, the similarity relation Iso_α is defined by $\mathcal{M}\ Iso_\alpha\ \mathcal{M}'$ iff there is an α-isomorphism between $\mathcal{M}$ and $\mathcal{M}'$. The similarity relation Iso_p is defined by $\mathcal{M}\ Iso_p\ \mathcal{M}'$ iff there is a potential isomorphism between $\mathcal{M}$ and $\mathcal{M}'$. These similarity relations are standard tools in model theory; they are used as semantic correlates of elementary equivalence for certain languages. Note however that the shift from results about elementary equivalence to results about invariants is not trivial. For example, consider Iso_ω. Two structures are Iso_ω-similar iff they satisfy exactly the same sentences of FOL. However, if we look at the logic of Iso_ω -invariants operators, we will not get FOL: there are Iso_ω-invariant operators (*e.g.*, "being infinite") which are not definable in FOL.

Finally, note that these similarity relations are linearly ordered by $\leq$:

[15] We use Feferman's (1972) definitions.

FACT 3.7. *Univ* $\leq$ *Bool* $\leq$ *App* $=$ *Iso*$_0$ $\leq$ *Iso*$_\alpha$ $\leq$ *Iso*$_{\alpha'}$ $\leq$ *Iso*$_p$ $\leq$ *Iso*, *where* α *and* α' *are ordinals such that* $\alpha \leq \alpha'$.

We also have:

FACT 3.8. *App* $\leq$ *SHom* $\leq$ *Iso*.

But $\leq$ is only a partial ordering. In particular, *Iso*$_p$ and *SHom* are unrelated. First, *SHom* $\not\geq$ *Iso*$_p$: let $\mathcal{A} = \langle\{1,2,3\},\{1,3\}\rangle$ and $\mathcal{B} = \langle\{1,2\},\{1\}\rangle$ be two structures, we have $\mathcal{A}$ *SHom* $\mathcal{B}$ but we do not have $\mathcal{A}$ *Iso*$_p$ $\mathcal{B}$. Second, *Iso*$_p$ $\not\geq$ *SHom*. If we look at $\mathcal{M} = \langle\mathbb{Q},\leq\rangle$ and $\mathcal{M}' = \langle\mathbb{R},\leq\rangle$, we have $\mathcal{M}$ *Iso*$_p$ $\mathcal{M}'$ but we do not have $\mathcal{M}$ *SHom* $\mathcal{M}'$. This highlights the fact that *Iso*$_p$ and *SHom* correspond to two really different weakenings of *Iso*. By shifting to *SHom*, one keeps a global similarity relation, but liberalizes over *Iso* by discarding equality. By shifting to *Iso*$_p$, one goes from a global to a local idea of similarity, and there is no special treatment for equality. It would be possible to define and study "equality-free" versions of *Iso*$_p$ and the *Iso*$_\alpha$.

3.2. Closure under definability and generality. This section is devoted to an enquiry into the limits of generality in the framework of generalized invariance, and it will rely upon the particular similarity relations we have just introduced as special landmarks.

3.2.1. *Generality in our framework.* The generality argument characterizes logic as the most general theory in our conceptual scheme. In Klein's setting, there is a correspondence between groups of transformation and the generality of invariants: the bigger the group (*i.e.*, in our setting, the lower the associated similarity relation), the more general the invariants. Because all considered groups are groups of transformations, the minimal element, which corresponds to maximal generality, is the groups of all one-one and onto transformations, corresponding in our setting to the similarity relation *Iso* (or *Auto* if the domain is kept fixed). By generalizing the idea of invariance under a group of transformations to the idea of invariance under a similarity relation, we have opened up a domain of new similarity relations which do not correspond to any group of transformation in Klein's sense. As a consequence, *Iso* is no longer the lowest element with respect to the ordering $\leq$ on $\mathcal{S}$: we have seen that *Iso*$_p$ $\leq$ *Iso*, *Iso*$_\omega$ $\leq$ *Iso*, *App* $\leq$ *Iso*, etc. What happens to the formalization of the idea of maximal generality in this setting? Two questions should be answered:

- Why do similarity relations provide a more liberal notion of similarity than Kleinian groups of transformations?
- How can the requirement of generality be applied in the new setting? Which is the similarity relation whose invariants are the most general notions, can it be used as a logicality criterion?

Concerning the first point, it is important to distinguish two parameters that determine the liberality of a similarity relation:

(i) the amount of extra-structure which has to be preserved.
(ii) the degree to which structure is to be preserved.

What do we mean by "extra-structure"? Invariance under bijection (*Iso*) is more liberal than invariance under homeomorphism, because invariance under bijection allows for arbitrary switchings of elements of the domains, whereas homeomorphisms are continuous transformations which preserve some implicit structures of the domains thought of as sets of points, namely their topological structure. In this sense, homeomorphisms preserve some extra-structure which is not explicit on the structures considered, whereas the similarity relation *Iso*, corresponding to invariance under bijection, is only sensitive to the explicit formal features of the structures to which the operators are applied. For this reason, *Iso* is indeed the lowest similarity relation with respect to the first parameter: no extra-structure is preserved under *Iso*.

However, by thinking in terms of similarity relations instead of groups of transformation, we discover the significance of an other parameter: the degree to which structure (implicit "extra-structure" or explicit structure) is to be preserved. *Iso* reflects full formal similarity: two structures which are isomorphic are exactly identical to one another as structures, the only difference consisting in the nature of the elements of their domains. On the contrary, we have seen that two structures which are potentially isomorphic are not fully identical: they are just similar in the sense that they are always, so to speak, locally isomorphic. Iso_p is thus lower than *Iso*, because it is more liberal than *Iso* with respect to the second parameter of liberality, namely the degree to which structure preservation is required. Note that the two parameters are independent. One could think of a potential version of homeomorphisms for example, in which partial preservation of the implicit topological structure would be required.

Now which is the smallest similarity relation in $\mathcal{S}$, *i.e.*, which is the similarity relation which is the most liberal with respect to both parameters? By definition, it is *Univ*. But of course, in some sense, *Univ* is a trivial relation of similarity, which discards all the formal features of structures. It comes therefore as no surprise that *Inv*(*Univ*) is a very meager class of operators: it will contain only trivial operators which contain either all the structures of a given signature, or none of them.

There is an important conceptual lesson to be drawn from this trivial mathematical fact. In the setting of generalized invariance, it does not make sense to require full generality without qualification, because the most general notions are as much useless as they are general. We do not think that this is an unwelcome feature of the framework. On the contrary this is as it should be; there is no reason why requiring utmost generality should not result in a total loss of content. Logical notions are the most general notions which deal with certain kind of features; but in order to make sense of the generality of logic, one has to say what matters to logic. For example, the difference between true and false is central to the project of defining what logical consequence is: in this respect, it seems fair to restrict one's attention to similarity relations which do not conflate true and false, *i.e.*, to similarity relations which preserve the boolean parts of structures (these are the S such that $Bool \leq S$). The class of operators $Inv(Bool)$ is just the class of all truth-functions: this is as expected. One can try to go a bit higher than *Bool*. It seems fair as well to consider only similarity relations S such that $App \leq S$: functional application is the semantic correlate of syntactic concatenation, and it seems a very basic semantic feature of languages, nearly as basic as the difference between being true and being false. Logic should therefore be sensitive to functional application as well. $Inv(App)$ is the class of operators which are λ-definable from $Inv(Bool)$, and, again, this is not a surprise.

3.2.2. *Introducing closure under definability.* All this seems to be fairly limited: we do not go really further than propositional logic, and the correspondence between similarity relations and classes of operators is so immediate that the conceptual gain of the invariance approach seems to be dangerously close to zero. We now wish to introduce two constraints on generality that will prove to be much more substantial.

First, if we accept that the operations of FOL are indeed logical operations, we have to choose a similarity relation that is sensitive to basic set-theoretic features of the structures. In particular, it is clear that if we think that the existential quantifier is logical, two structures $\langle M, P\rangle$ and $\langle M', P'\rangle$ such that P is a non-empty subset of M and P' is the empty set should not be considered as logically similar. One way to capture this very minimal kind of structural similarity would be to say that for two structures to be logically similar, it must be impossible to see that they are different just by looking at an arbitrary individual in one of the structures. Or, that if two structures $\mathcal{M}$ and $\mathcal{M}'$ are similar, then if we pick an arbitrary individual say a in $|\mathcal{M}|$, it must be possible to pick another individual a' in $\mathcal{M}'$ such that a and a' share the same properties. A similarity relation satisfying this property is a relation S such that $Iso_1 \leq S$.

Second, the invariants of a similarity relation should be closed under definability. Let us explain this. Given a similarity relation S, $Inv(S)$ is the class of S-invariant operators. We are interested in the class $Inv(S)$ as the putative class of logical operators. This means that we want to use the operators in $Inv(S)$ as building blocks for the logical part of a language. Now, given a language, it is possible to define in it certain operators in a purely logical way. For example, let us consider the class K of operators containing just the existential and universal quantifiers, the operator for equality and the boolean operators. The logic associated with K is just FOL. Now, in FOL, it is possible to define new logical operators. For example, the purely logical formula "$\exists x, y, z\ ((Px \wedge Py \wedge Pz) \wedge (x \neq y \wedge x \neq z \wedge y \neq z))$" defines the operator $Q_{\geq 3}$ ("there are at least three"), which is the class of all structures of the form $\langle M, P\rangle$ where P is a subset of M containing at least three elements. Even if Q was not in K, it was "implicitly" there, because it is definable in a language based on K.

Operators which are definable in a purely logical manner are logical. We just do not see how a non-logical element could creep in the logical elements of the definition and make the defined operator non logical. This is what we might call the principle of closure under definability:

Principle of closure under definability. An interpreted symbol definable only by means of logical constants is a logical constant.

Le us make this mathematically precise. Let K be a class of operators. The **logic L_K associated with K** consists in first-order variables and logical constants interpreted by operators in K. For any signature σ, we thus obtain a language $L_K(\sigma)$ with extra-logical symbols corresponding to σ, whose interpretation varies freely, and logical symbols whose interpretation is taken from K and is kept fixed. For example, let $Q \in K$ be a class of structures of the form $\langle M, R\rangle$ where $R \subseteq M \times M$, L_K contains a logical symbol $\overline{Q}$ which is interpreted by Q. To recall the point made in the proof of Fact 1.2, the clause for Q in the recursive definition of satisfaction for L_K is the following one:

$$\mathcal{M} \vDash \overline{Q}x, y\ \phi(x, y)\ [\sigma] \text{ iff } Q(M, \|\phi(x, y)\|_{\mathcal{M},\sigma})$$

where $\|\phi(x, y)\|_{\mathcal{M},\sigma}$ is the interpretation of ϕ over $\mathcal{M}$ according to σ, that is the set of pairs $\langle a, b\rangle$ of elements of $\mathcal{M}$ such that

$$\mathcal{M} \vDash \phi(x, y)\ [\sigma][x := a][y := b].$$

Conversely, given a logic L and an operator Q—for simplicity, we assume the type of Q is the same as before—we shall say that **Q is definable in L** iff there is a sentence ϕ_Q of $L(\overline{R})$ such that:

$$Q(\langle M, R\rangle) \text{ iff } \langle M, R\rangle \vDash_L \phi_Q.$$

Closure under definability can be construed as an operation on classes of operators which adds to a given class K of operators all the operators which are definable from operators in K in this sense. More precisely, we define a function $C\colon \mathcal{K} \to \mathcal{K}$ by taking $C(K)$ to be the class of operators definable in L_K. $C \circ \mathit{Inv}\colon \mathcal{S} \to \mathcal{K}$ is the composition of *Inv* with C. To simplify notations, we shall denote this function by *CInv*.

We are interested in similarity relations S which induce classes of operators $\mathit{Inv}(S)$ providing a stock of denotations for logical constants. If we accept the principle of closure under definability, we should require that these similarity relations be such that $CInv(S) = Inv(S)$.[16] To put in another way, as long as invariant operators are used as building blocks of the logical part of a language, we should not consider similarity relations whose invariants are not closed under definability. This requirement is far from being trivial. In particular, it is not satisfied by Iso_ω, the similarity relation corresponding to elementary equivalence in FOL.

FACT 3.9. $CInv(Iso_\omega) > Inv(Iso_\omega)$.

The proof is an elementary exercise in model theory:

PROOF. We shall consider the operator $Q_{\geq\aleph_0}$. $Q_{\geq\aleph_0}$ is Iso_ω-invariant: let $\langle M, P\rangle$ and $\langle M', P'\rangle$ be two structures such that we have $Q_{\geq\aleph_0}(\langle M, P\rangle)$ but not $Q_{\geq\aleph_0}(\langle M', P'\rangle)$ (thus, $|P| \geq \aleph_0$ where as $|P'| < \aleph_0$). There is an integer n such that $|P'| = n$, but then it is not the case that $\langle M, P\rangle\, Iso_{n+1} \langle M', P'\rangle$, hence it is not the case that $\langle M, P\rangle\, Iso_\omega \langle M', P'\rangle$. We shall now consider the operator Q' defined by the sentence "$\overline{R}$ is an equivalence relation $\wedge$ $\exists x \overline{Q_{\geq\aleph_0}} y\ x\overline{R}y$" ($Q'$ picks out the relational structures $\langle M, R\rangle$ such that R is an equivalence relation with an infinite equivalence class). Since $Q_{\geq\aleph_0} \in Inv(Iso_\omega)$, we have that $Q' \in CInv(Iso_\omega)$. It is now sufficient to show that Q' is not Iso_ω-invariant. We construct two $L(\overline{R})$-structures $\mathcal{M} = \langle M, R\rangle$ and $\mathcal{M}' = \langle M, R'\rangle$ such that:

- The interpretation of $\overline{R}$ on both models is an equivalence relation.
- $\mathcal{M}$ contains an infinite number of R-equivalence classes of arbitrarily big finite cardinality, but no infinite equivalence class.
- $\mathcal{M}'$ is just as $\mathcal{M}$ but it contains also an infinite equivalence class.

It is clear that $\mathcal{M}\ \ Iso_\omega\ \ \mathcal{M}'$. But we have that $Q'(\mathcal{M}')$, whereas $\mathcal{M}$ is not in Q'. $\dashv$

Here is now the theorem that tells us what are the consequences of our requirements on the range of available similarity relations:[17]

THEOREM 3.10. *Iso_p is the lowest similarity relation S such that $Iso_1 \leq S$ and $CInv(S) = Inv(S)$.*

If S is an equivalence relation, we have that $CInv(S) = Inv(S)$ if and only if $Sim(CInv(S)) = S$. Thus, Theorem 3.10 tells us that Iso_p is the smallest similarity relation extending Iso_1 which is a fixed point for $Sim \circ CInv$.[18] Or, in more vernacular terms, Iso_p is the lowest similarity relation extending Iso_1 which is closed under definability. As a consequence, if we accept both the preceding argument according to which a good similarity relation S for logicality must satisfy $Iso_1 \leq S$, and the argument in favor of the principle of closure under definability as well, then we should only be interested in

[16]Note that $CInv(S) = Inv(S)$ is stronger than the principle of closure under definability itself, because it encapsulates the assumption that logical operations are to be characterized in terms of invariance.

[17]The proof is given below.

[18]We are indebted to Henri Galinon for the suggestion to think in terms of fixed points.

similarity relations S such that $Iso_p \leq S \leq Iso$. Note also that Iso is a fixed point for $Sim \circ CInv$. Since Iso is the greatest similarity relation we consider, Iso is trivially the greatest fixed point for $Sim \circ CInv$. Because of the conceptual interest of closure under definability, it would be nice to know more about the structure of fixed points for $Sim \circ CInv$ between Iso_p and Iso.

3.2.3. *What happens to the Galois connection.* At a conceptual level, it seems thus quite natural to close $Inv(S)$ under definability. But, on the mathematical level, things get much more complicated. As a matter of fact, the shift from Inv to $CInv$ breaks the symmetry between classes of operators and similarity relations. The Galois connexion is lost. We had that $Inv(S) \leq Inv(Sim(Inv(S)))$. This property is preserved:

FACT 3.11. *For any similarity relation S, $CInv(S) \leq CInv(Sim(CInv(S)))$.*

PROOF. $Inv(S) \leq CInv(S)$, therefore $Sim(Inv(S)) \leq Sim(CInv(S))$ by monotonicity of Sim. Moreover $CInv(S) = CInv(Sim(Inv(S))$, since $Inv(S) = Inv(Sim(Inv(S)))$. Therefore $CInv(S) \leq CInv(Sim(CInv(S)))$. ⊣

We had also the other direction, namely $Inv(Sim(Inv(S))) \leq Inv(S)$. This property is lost:

FACT 3.12. *There are similarity relations S such that $CInv(Sim(CInv(S))) \not\leq CInv(S)$.*

PROOF. Take S to be Iso_1. $CInv(Iso_1)$ is just the class of operators definable in FOL. Therefore, $Sim(CInv(Iso_1))$ is Iso_ω. We have seen earlier an example of operator, $Q_{\geq\aleph_0}$ which is in $Inv(Iso_\omega)$ (and therefore, *a fortiori*, in $CInv(Iso_\omega)$). But $Q_{\geq\aleph_0}$ is not definable in FOL, hence it does not belong to $CInv(Iso_1)$. ⊣

The lesson to be drawn is that closure under definability is a very strong requirement in the context of the invariance approach. It breaks the natural symmetry between similarity relations and classes of operators. This symmetry is recovered for a limited class of similarity relations, namely the fixed points of $Sim \circ CInv$, for which we do have that $CInv(Sim(CInv(S))) \leq CInv(S)$. Among those fixed points, Iso_p plays a special role as the smallest one for which $Inv(Iso_p)$ is not too meager (but note that App is also a fixed point).

3.2.4. *Closure under level 1 projection.* At the heart of Theorem 3.10, there is a more elementary fact[19] which has its own conceptual interest. In order to force $Iso_p \leq S$, we do not need the full power of closure under definability. Closure under projection, which is the restriction of closure definability to existential quantification, is actually sufficient.

We shall label Pr the function which, given an operator Q consisting in a class of structures of the form $\langle M, R_1, \dots R_n, a_1, \dots a_m, b\rangle$ yields the operator $Pr(Q)$ defined by $Pr(Q)(\langle M, R_1, \dots R_n, a_1, \dots, a_m\rangle)$ iff there is a

[19]We are indebted here to Johan van Benthem for showing to us what the core of our previous somewhat inelegant proof was.

$b \in M$ such that $Q(\langle M, R_1, \dots R_n, a_1, \dots, a_m, b\rangle)$. A class of operators K is **closed under level 1 projection** iff for all $Q \in K$, $Pr(Q) \in K$. We then define what it means for a similarity relation to commute with projections (this is the main idea involved in potential isomorphisms and bisimulations).

DEFINITION 3.13 (commutation with level 1 projections). A similarity relation commutes with level 1 projections iff, for any structures $\mathcal{A}$, $\mathcal{B}$ such that $\mathcal{A}\ S\ \mathcal{B}$, for any expansion $\mathcal{A}, a$ of $\mathcal{A}$ by a constant $a \in A$, there is a $b \in B$ such that $\mathcal{A}, a\ S\ \mathcal{B}, b$.

The main fact behind 3.10 is precisely that closure under definability with respect to existential quantification is equivalent to the back and forth mechanism expressed by commutation with projections:

FACT 3.14. *For any similarity relation S, $Inv(S)$ is closed under level 1 projection iff S commutes with level 1 projections.*

Before proving Fact 3.14, let us establish the link with potential isomorphisms:

FACT 3.15. *For any similarity relation S such that $App \leq S$, if $Inv(S)$ is closed under level 1 projection, then $Iso_p \leq S$.*

PROOF. Let S be a similarity relation such that $App \leq S$ and $Inv(S)$ is closed under level 1 projection. We want $Iso_p \leq S$. Let $\mathcal{A}$, $\mathcal{B}$ be two structures such that $\mathcal{A}\ S\ \mathcal{B}$. We need to show that $\mathcal{A}\ Iso_p\ \mathcal{B}$. By definition of Iso_p, this amounts to finding a non empty set I of partial isomorphisms between $\mathcal{A}$ et $\mathcal{B}$ satisfying the back and forth property. We set $I = \{f : A \rightarrow B / \mathcal{A}, a_1, \dots, a_n\ S\ \mathcal{B}, f(a_1), \dots, f(a_n)\}$. By hypothesis, $\mathcal{A}\ S\ \mathcal{B}$, therefore I is non-empty, since it contains the empty function. Now I has the back and forth property, since S commutes with level 1 projections, (by fact 3.14). Finally, since $App \leq S$, the f in I are partial isomorphisms. ⊣

Theorem 3.10, that we recall here, follows from fact 3.15:

THEOREM. *Iso_p is the lowest similarity relation S such that $Iso_1 \leq S$ and $CInv(S) = Inv(S)$.*

PROOF. Let S be a similarity relation such that $Iso_1 \leq S$ and $CInv(S) = Inv(S)$. We show first $Iso_p \leq S$. Since $Iso_1 \leq S$, $Q_\exists \in Inv(S)$. As a consequence, for all Q, if $Q \in Inv(S)$, then $Pr(Q) \in CInv(S)$. But $CInv(S) = Inv(S)$, therefore $Inv(S)$ is closed under level 1 projection. $Iso_1 \leq S$ implies trivially $App \leq S$, therefore fact 3.15 tells us that $Iso_p \leq S$.

It remains to check that $CInv(Iso_p) = Inv(Iso_p)$. It is sufficient to show that $\equiv_{L_{Inv(Iso_p)}}$, the relation of elementary equivalence for the logic $L_{Inv(Iso_p)}$ whose logical constants denote Iso_p-invariant operators, is Iso_p. The proof is by induction on the complexity of formulas. The induction hypothesis is the following one:

$$\text{If } \mathcal{M}, \overrightarrow{a}\ Iso_p\ \mathcal{M}', \overrightarrow{a'} \text{ then } \mathcal{M} \vDash \phi(\overrightarrow{a}) \text{ iff } \mathcal{M}' \vDash \phi(\overrightarrow{a'}).$$

The crucial step is the case where ϕ is of the form $\overline{Q}\overrightarrow{x}\ \psi(\overrightarrow{x}, \overrightarrow{y})$.

We assume $\mathcal{M}, \overrightarrow{a}\ Iso_p\ \mathcal{M}', \overrightarrow{a'}$. We have to show that $\mathcal{M} \vDash \overline{Q}\overrightarrow{x}\ \psi(\overrightarrow{x}, \overrightarrow{a})$ iff $\mathcal{M}' \vDash \overline{Q}\overrightarrow{x}\ \psi(\overrightarrow{x}, \overrightarrow{a'})$.

Now assume $\mathcal{M} \vDash \overline{Q}\overrightarrow{x}\ \psi(\overrightarrow{x}, \overrightarrow{a})$. We want $\mathcal{M}' \vDash \overline{Q}\overrightarrow{x}\ \psi(\overrightarrow{x}, \overrightarrow{a'})$. Since Q is Iso_p-invariant, it is sufficient to show that

$$\langle M, \|\psi(\overrightarrow{x}, \overrightarrow{a})\|_{\mathcal{M}}\rangle\ Iso_p\ \langle M', \|\psi(\overrightarrow{x}, \overrightarrow{a'})\|'_{\mathcal{M}}\rangle.$$

We take as potential isomorphism I between these two structures the set of finite partial isomorphisms between $\mathcal{M}, \overrightarrow{a}$ and $\mathcal{M}', \overrightarrow{a'}$. I is non-empty and has the back and forth property. We need to show that the $f \in I$ are indeed partial isomorphisms for $\langle M, \|\psi(\overrightarrow{x}, \overrightarrow{a})\|_{\mathcal{M}}\rangle$ and $\langle M', \|\psi(\overrightarrow{x}, \overrightarrow{a'})\|'_{\mathcal{M}}\rangle$.

But this is precisely what the induction hypothesis says: by definition of potential isomorphism, we have that $\mathcal{M}, \overrightarrow{a}, \overrightarrow{Dom(f)}\ Iso_p\ \mathcal{M}', \overrightarrow{a'}, \overrightarrow{Rng(f)}$, and therefore for any sequence $\overrightarrow{c}$ of elements of $Dom(f)$ and its image $\overrightarrow{f(c)}$, a sequence of elements of $Rng(f)$, we do have that $\mathcal{M} \vDash \psi(\overrightarrow{c}, \overrightarrow{a})$ iff $\mathcal{M}' \vDash \psi(\overrightarrow{f(c)}, \overrightarrow{a'})$. ⊣

To conclude on the mathematical background of Theorem 3.10, here is the proof for Fact 3.14:

Proof. Let $\mathcal{M}$ be a structure and S a similarity relation. We note $Orb(S, \mathcal{M})$ the orbit of $\mathcal{M}$ with respect to S, *i.e.*, the class of structures $\mathcal{M}'$ such that there is a S-path from $\mathcal{M}$ to $\mathcal{M}'$. $Orb(S, \mathcal{M})$ is an operator.

There are two directions:

(i) If S commutes with level 1 projections, $Inv(S)$ is closed under projection. We assume $Q \in Inv(S)$, we want to show $Pr(Q) \in Inv(S)$, *i.e.*, for any structures $\mathcal{A}$ and $\mathcal{B}$ such that $Pr(Q)(\mathcal{A})$ and $\mathcal{A}\ S\ \mathcal{B}$, we have $Pr(Q)(\mathcal{B})$.
   (a) $Q \in Inv(S)$.
   (b) $\mathcal{A}\ S\ \mathcal{B}$.
   (c) $Pr(Q)(\mathcal{A})$.
   (d) $Q(\mathcal{A}, a)$ for an $a \in A$ by (c) and the definition of Pr.
   (e) $\mathcal{A}, a\ S\ \mathcal{B}, b$ for a $b \in B$ by (b) and commutation for S.
   (f) $Q(\mathcal{B}, b)$ by (a) and (e).
   (g) $Pr(Q)(\mathcal{B})$ by definition of Pr and (f).

(ii) If $Inv(S)$ is closed under level 1 projection, S commutes with level 1 projections. Let $\mathcal{A}$ and $\mathcal{B}$ be two structures such that $\mathcal{A}\ S\ \mathcal{B}$. Let $a \in A$, we need to find a $b \in B$ such that $\mathcal{A}, a\ S\ \mathcal{B}, b$.
   (a) $\mathcal{A}\ S\ \mathcal{B}$.
   (b) $Orb(S, \mathcal{A}, a) \in Inv(S)$ by definition of Orb.
   (c) $Pr(Orb(S, \mathcal{A}, a)) \in Inv(S)$ by closure under projection for $Inv(S)$ and (b).
   (d) $Pr(Orb(S, \mathcal{A}, a))(\mathcal{A})$ by definition of Pr.

(e) $Pr(Orb(S, \mathcal{A}, a))(\mathcal{B})$ by (a), (c) and (d).
(f) $Orb(S, \mathcal{A}, a)(\mathcal{B}, b)$ for a $b \in B$, by definition of *Pr* and (e).
(g) $\mathcal{A}, a\ S\ \mathcal{B}, b$ by definition of *Orb* and (f). $\dashv$

The discussion of the philosophical significance of Theorem 3.10 for the revision of Tarski's thesis will be postponed until the last section. But it shall be already clear that the generality approach can be up to some point rescued from trivialization.

3.3. Formality revisited.

3.3.1. *Formality and lack of content.* Is the formality argument bound to fall into the trap of overgeneration? Or is it possible to use the setting of generalized invariance to prevent it from falling into that trap? Following the diagnosis proposed at the end of the previous section, the problem with the formality argument is that it fails to distinguish between logic and mathematics. The point is not that any account of logicality should lead to a refutation of logicism; this assumption would be as *ad hoc* as the opposite one. The point is rather that it is not legitimate to seek a characterization of logicality on the basis of a property—namely formality—which is obviously common to logic, set-theory and mathematics in general. Logicism might come as a consequence of the conceptual analysis of what it is to be logical. But it should not be the starting point of such an analysis. Which kind of strengthening of formality should one consider?

Let us go back to the intuition relating logicality and formality. Logic is expected to be non committing. Logical truths are tautologies which do not put any constraint on how the world should be. Logical notions have no empirical basis. The concept "cat" is an empirical concept; mastering this concept involves special recognitional abilities, like being able to distinguish a meowing animal from a barking animal. On the contrary, the concept associated with existential quantification does not require any mundane abilities: it is sufficient to be able to tell the difference between an empty set and a non empty set.

The basic intuition underlying the formality argument could then be rephrased as: "Logical notions are (nearly) deprived of content", or "logical notions are non substantial". But the shift from lack of content to formality might not be as innocuous as it seems: being formal means being deprived of any kind of content *which is sensitive to the identity of objects*. But the overgeneration problem stems from the fact that formal notions are far from being deprived of content, full stop. "Being of size $\aleph_1$" is a formal notion conveying a rich *mathematical content* or a rich set-theoretical content. Now the problem is that this might be a dead-end. If one construes the concept of "content" in a sufficiently liberal way, no notion is absolutely deprived of content. "Being non-empty", which is synonymous with "being of cardinality at least 1" also conveys a bit of set-theoretical content, just like "being of cardinality at least $\aleph_1$". Still, there is a difference between "being of cardinality at least 1" and "being of cardinality at least $\aleph_1$": the content of the second notion is clearly more problematic than the content of the first notion. In particular, as we explained earlier, the second notion depends on the exact extent of the set-theoretical universe (again, see Feferman, 1999, p. 38), whereas the first notion does not. A non-empty set living in a given model $\mathcal{M}$ of ZFC will remain non-empty, no matter how you shrink or extend $\mathcal{M}$ into a smaller or bigger model of ZFC.

To sum up, we do not wish to argue that logical notions are fully deprived of set-theoretic content, but we would like them to be free from *problematic* set-theoretic content.

3.3.2. *Absoluteness and lack of problematic set-theoretic content.* How could we make sense of the idea of *problematic* set-theoretic content, in the setting of generalized invariance? The idea is that if the only difference between two formally identical structures is set-theoretically problematic, these two structures should be logically similar. *Iso* itself does not pass the test. Given a model $\mathcal{M}$ of ZFC and a structure $\langle A, P\rangle$ in $\mathcal{M}$, the question whether $\langle A, P\rangle$ belongs to $Q_{\aleph_1}$ is the question whether $\mathcal{M} \vDash \langle A, P\rangle$ *Iso* $\langle card(A), \aleph_1\rangle$. But this similarity is problematic from a set-theoretic point of view. It might be the case that the answer is negative in $\mathcal{M}$—because $\mathcal{M}$ is not rich enough to contain a bijection between the set P and the cardinal $\aleph_1$—and positive in a generic extension of $\mathcal{M}$, in which such a bijection exists. Therefore, *Iso* is a typical example of a set-theoretically problematic similarity relation, whose extension depends on the specific features of the model of set theory one is working with.

So, basically, what we expect from a notion of similarity suitable for a characterization of logicality is that it does not suffer from the same defects as *Iso*. How can we capture this? As remarked by Feferman (1999), the notion of absoluteness, introduced by Gödel (1940), gives at least a necessary condition for being insensitive to problematic set-theoretical features:

DEFINITION 3.16 (persistent formula). Let T be a theory in the language of set theory, a formula $P(x_1, \dots, x_n)$ is persistent with respect to T iff for any models $\mathcal{M} = \langle M, \in\rangle$ and $\mathcal{M}' = \langle M', \in'\rangle$ of T such that $\mathcal{M}$ is a submodel of $\mathcal{M}'$ and if $a \in' b$ and $b \in M$ then $a \in M$, if $\mathcal{M} \vDash P(a_1, \dots, a_n)$ then $\mathcal{M}' \vDash P(a_1, \dots, a_n)$ (where the $a_1, \dots, a_n$ are elements of M).

DEFINITION 3.17 (absolute formula). Let T be a theory in the language of set theory, a formula $P(x_1, \dots, x_n)$ is absolute with respect to T iff both $P(x_1, \dots, x_n)$ and $\neg P(x_1, \dots, x_n)$ are persistent.

Typically, the formulas "x is transitive", "x is an ordinal", "x is a limit ordinal" are absolute with respect to ZFC, whereas "x is a cardinal" or "x is of size $\aleph_1$" are not absolute with respect to ZFC. We now apply the idea[20] to

[20]We are indebted here to Gabriel Sandu, who draw our attention to Barwise's papers on absoluteness and suggested that we should apply the idea to our similarity relations.

the case in point, namely to similarity relations:[21]

DEFINITION 3.18 (absolute similarity relation). A similarity relation S is absolute with respect to T iff S is definable in T by a formula which is absolute with respect to T.

Iso_p is absolute with respect to ZFC (essentially because it depends on the existence of *finite* partial isomorphisms), but *Iso* is not (*Iso* is persistent, but not absolute, because an isomorphism can appear when the model grows bigger).

On the syntactic side, Kreisel and Feferman proved (see Feferman, 1968) that absolute formulas with respect to a theory T are Δ_1 with respect to T. This provides another argument for absoluteness. Our definition of what it means to be logically similar has to be cashed out in set-theoretic terms. But even if this point it granted, it is desirable that the definition remains as simple as possible, so that the distinction between the more elementary realm of logic and set-theory is preserved. Requiring that the definition of the notion of logical similarity is Δ_1 amounts precisely to requiring that we do not rely too heavily on unbounded quantification over sets.

Since our plan is to use absoluteness as a requirement on similarity relations, we should therefore ask which kind of similarity relations meet this requirement. Before answering the question, we shall introduce a few notations. When Q is an absolute operator, $\phi_Q(x)$ is an absolute formula defining it. Similarly, for an absolute similarity relation S, we note $\phi_S(x, y)$ an absolute formula defining S. $\phi_{Iso}(x, y)$ is a (persistent) formula of ZFC for "x and y are isomorphic". Similarly, ϕ_{Iso_p} will be an (absolute) formula of ZFC for "x and y are potentially isomorphic". We use the notation "$S \leq Iso$" for the formula $\forall x, y\ \phi_{Iso}(x, y) \rightarrow \phi_S(x, y)$. By "model of ZFC" we mean a transitive set in which the axioms of ZFC are true. A CTM is a countable (and transitive) model of ZFC. Here is now the answer to our question, provided by the following theorem:

THEOREM 3.19 (Barwise, 1973, p. 31). *Iso_p is the greatest similarity relation S such that S is absolute with respect to* ZFC *and* ZFC $\vdash$ *"$S \leq Iso$".*

Theorem 3.19 provides a characterization of Iso_p which is dual to the one provided by theorem 3.10. Iso_p was the lowest similarity relation closed under definability, it appears now to be the greatest absolute similarity relation.

Note that if S is an absolute similarity relation with respect to ZFC, defined by a formula $\phi_S(x, y)$, then for any transitive model $\mathcal{M}$ of ZFC and for all structures $\mathcal{A}$ and $\mathcal{B}$ which are elements of $\mathcal{M}$, we will have that $\mathcal{M} \vDash \phi_S(\mathcal{A}, \mathcal{B})$ iff $\mathcal{A}\ S\ \mathcal{B}$. So in particular, the difference between *Iso* and Iso_p which matters to us is that being potentially isomorphic is determinately true or false in all transitive models, whereas being isomorphic is not. From a conceptual of

[21]Similarity relations are relations on the set-theoretic universe V. The theories T we consider are supposed to be true in V.

view, the absoluteness requirement will be all the more cogent if transitive models of set theory are granted a special status.[22] After all, consider a straightforward realist view about sets, according to which there is one and only one intended model of ZFC, V, the universe of all sets. Maybe then absoluteness is not as natural a requirement as we said it is. What seems 'problematic' with operations like $Q_{\aleph_1}$ might be nothing more than a side-effect of unfortunate limitations on our knowledge of set theory, rather than a problem with $Q_{\aleph_1}$ itself. A dual point can be made if there are no intended models of set theory at all. Consider non transitive models of ZFC. A set can be infinite in one model and have a non standard integer as its cardinality in another model, so that $Q_{\geq\aleph_0}$ would look just as problematic as $Q_{\aleph_1}$. Now consider the following view. Our intentions are unable to pick out exactly one model of ZFC as the intended model of set theory, so that the naive realist picture is wrong. But it would be equally wrong to consider that all models fit our intentions. All well-founded models fit our intentions, none of which doing it better than the others. In this case, we shall be particularly happy with the absoluteness requirement: it expresses the idea that 'being logically similar' should have a determinate truth-value in all models of set theory that fit our intentions.

Before closing this section, let us say a word about the relationship between absoluteness as a property of similarity relations *versus* absoluteness as a property of operators. Absoluteness has been studied in the literature on set-theoretic logics (see the presentation in Väänänen, 1985). In this framework, absoluteness is applied to the satisfaction relation of a logic defined in set-theory. If the syntax is finitary, this amounts to requiring that the operators interpreting the logical constants are themselves absolute. What is the connection with our approach, in which absoluteness is applied to similarity relations? Is this equivalent, *i.e.*, is the class of *Iso*-invariant absolute operators the same as the class of Iso_p-invariant operators? Using the ideas of the proof of theorem 3.19, one can show that Iso_p-invariance is a necessary condition for an *Iso*-invariant operator to be absolute:

THEOREM 3.20. *Let Q be an operator absolute with respect to* ZFC, *if*

$$\text{ZFC} \vdash \forall x, y\ (\phi_{Iso}(x, y) \to (\phi_Q(x) \leftrightarrow \phi_Q(y)))$$

then

$$\text{ZFC} \vdash \forall x, y\ (\phi_{Iso_p}(x, y) \to (\phi_Q(x) \leftrightarrow \phi_Q(y))).$$

However, the converse is not true, there are operators Q definable in ZFC with ZFC $\vdash \forall x, y\ (\phi_{Iso_p}(x, y) \to (\phi_Q(x) \leftrightarrow \phi_Q(y)))$ such that Q is not absolute (Väänänen, p.c.). Using the notation Abs for the class of absolute and provably *Iso*-invariant operators, this means that Abs $\subsetneq Inv(Iso_p)$, *i.e.*, Iso_p does not provide a characterization in terms of invariance of Abs. Does

[22]We are grateful to an anonymous reviewer for making this point.

this mean that we should look for a better candidate than Iso_p? Or should absoluteness be cashed out as a constraint which applies directly to invariant operators? What happens if we choose a weaker set theory? We leave these questions open for future research.

§4. A new thesis for logicality. To sum up, the previous section has been devoted to a critical examination of generality and formality as intuitive characterizations of logicality. As a result, two constraints on similarity relations, closure under definability and absoluteness, have been proposed. In both case, Iso_p, the similarity relation corresponding to potential isomorphism, plays a special role as the lowest or the greatest similarity relation satisfying these constraints *modulo* some mild additional assumptions. Does this justify a revision of Tarski's thesis in which Iso_p would replace *Auto* or *Iso*?

4.1. Logical operators and invariance under potential isomorphism. First, we present a modified version of the generality argument, in which we take into account the fact that maximal generality would lead to triviality:[23]

The mild generality argument.

MG.1 Logic deals with very general notions, but not only with trivial notions.

MG.2 The truth-functions, functional application and first-order existential quantification are logical operators.

MG.3 The good notion of invariance for logicality is to be provided by a similarity relation S such that S is closed under definability.

MG.4 The good notion of invariance for logicality is to be provided by the lowest similarity relation compatible with MG.2 and MG.3.

The logical notions are the Iso_p-invariant notions.

MG.1 corrects G.1 by demanding that some non-trivial notions be recognized as logical. This takes into account the fact that G.1 escaped triviality just because of the lack of generality of the Kleinian setting in which the generality argument took place. MG.2 and MG.3 provide the needed constraints. Some notions which are commonly assumed to be logical have to be invariants of the similarity relation to be chosen. Moreover, this similarity relation should abide by the principle of closure under definability. MG.4 captures the idea of maximal generality, because the ordering on the class of similarity relations mirrors the generality of the associated theories. The conclusion of the mild generality argument follows then from its premisses by theorem 3.15. Note that, as remarked in paragraph 3.2.1, the shift from *Iso* to Iso_p is perfectly in line with the Kleinean picture Tarski had in mind. Invariance under isomorphism as we defined it is Tarski's invariance under

[23]Our revised arguments have lost some of the elegant simplicity of the formality and the generality arguments. We think that they could be rephrased in more vernacular terms, but at the cost of either of clarity or brevity.

bijection: what is preserved by isomorphism is the structure explicit in logical operations and extra-structure (*e.g.*, topological structure) is disregarded. The same holds with Iso_p-invariance: what is preserved under potential isomorphism is the structure explicit in logical operations and extra-structure (*e.g.*, topological structure) is disregarded.

The formality argument should also be modified, to take into account the fact that logical notions should not encapsulate any problematic set-theoretical content:

The lack of content argument.

LC.1 Logic deals with notions which are deprived of non formal content and of problematic set-theoretic content.

LC.2 The good notion of invariance for logicality is to be provided by a similarity relation S such that $S \leq Iso$.

LC.3 The good notion of invariance for logicality is to be provided by a similarity relation S such that S is absolute with respect to ZFC.

LC.4 The good notion of invariance for logicality is to be provided by the greatest similarity relation S satisfying LC.2 and LC.3.

The logical notions are the Iso_p-invariant notions.

LC.1 supplements the premiss F.1 of the formality argument by dismissing contents which are properly set-theoretical. LC.2 captures the first part of LC.1, it is the counterpart in our setting of F.2 and F.3. LC.3 captures the second part of LC.1. We have seen in the previous section that absoluteness is a natural way to formalize independence with respect to which sets exists. LC.3 thus sets the limits of the identification of logic and mathematics: the logician might use the resources provided by set theory, but he should refrain from using notions whose extension is problematic. The requirement of absoluteness is relative to ZFC, because we take it to be the standard background set theory. This point is debatable. If the aim is really to minimize the dependency of logic on set theory, it would be wiser to choose a weaker set theory (*e.g.*, KP). LC.4 claims that the constraints stated by LC.2 and LC.3 are jointly not only necessary but also sufficient to determine which is the good similarity relation for logicality. The formality argument has it that LC.2 is sufficient on its own. But LC.2 is unable to do justice to the idea that logic is not the same as mathematics. It is therefore more plausible that LC.2 is sufficient if is supplemented by a criterion reflecting the specific feature of logic inside the formal sciences. This is precisely the role of the absoluteness requirement put forward by LC.3, and this gives us hope that the overgeneration problem might be solved. The conclusion of the lack of content argument follows from its premisses by theorem 3.19.

Starting from different intuitions, we still recover the convergence of the generality and the formality argument. The mild generality argument and the lack of content argument both support the following thesis:

Iso_p THESIS FOR LOGICALITY. An operator Q is logical iff Q is Iso_p-invariant.

Tarski's thesis and the Iso_p thesis disagree on the similarity relation to be chosen to characterize the invariance properties of logical constants. They do not completely disagree though.[24] The two theses yield the same result on structures which are not too big. As a matter of fact, if $\mathcal{M}$ and $\mathcal{M}'$ are at most countable structures, $\mathcal{M}$ Iso $\mathcal{M}'$ if and only if $\mathcal{M}$ Iso_p $\mathcal{M}'$. The nature of the disagreement between a proponent of Tarski's thesis and an advocate of our revised version of it could be put in the following way. Both agree that, concerning small "surveyable" structures, two structures are logically similar iff they are formally similar. This accounts for the generality of logic, which does not take into account any special feature of the domains of the structure, and for the formality of logic, which is insensitive to the identity of the objects. When it comes to bigger structures, the proponent of Tarski's thesis thinks that the natural way to extend this similarity relation is to require again full formal similarity: two structures are logically similar iff they are isomorphic. On the contrary, the proponent of the Iso_p thesis thinks that this is going to grant a significance to logically unsignificant differences. For him, what is essential in the similarity relation acting on small structures is the idea of an arbitrary ω-long inspection of the structures. Two big structures should be considered as logically similar as long as they pass this test, and differences coming only from cardinality issues involving higher infinites should not come into play. This is the reason why two structures are logically similar if and only if they are potentially isomorphic.

4.2. The logic of invariance under potential isomorphism. The last paragraph has been devoted to an *ex ante* justification of the Iso_p-thesis for logicality. *Ex post*, we have to look into the consequences of the thesis, in order to see whether the conflict with some of our intuitions regarding logicality has been solved, or if the overgeneration problem is still there. To give some examples, the quantifier $Q_{\aleph_0}$ ("there are exactly $\aleph_0$ x such that ... "), $Q_{>\aleph_0}$ ("there are uncountably many x such that ... ") are not Iso_p-invariant. $Q_\exists$ of course, and $Q_{\geq\aleph_0}$ ("there are infinitely many x such that ... ") are Iso_p-invariants. Q_{WF}, the unary relational quantifier testing whether a relation is well-founded, is Iso_p-invariant as well. More generally, what is the logic of Iso_p-invariance? In abstract model theory, Iso_p-invariance, as a relation of elementary equivalence, is known as the Karp property:[25]

DEFINITION 4.1 (Karp property). A logic $\mathcal{L}$ has the Karp property iff, for all structures $\mathcal{M}$, $\mathcal{M}'$, if $\mathcal{M}$ Iso_p $\mathcal{M}'$, then $\mathcal{M} \equiv_{\mathcal{L}} \mathcal{M}'$.

[24] We are indebted here to Gila Sher for the suggestion to discuss the common part of the two theses.

[25] We are indebted here to Dag Westerståhl for suggesting us to make more explicit the link with standard results in abstract model theory.

Since Iso_p is closed under definability, the Karp property provides a bound on the expressivity of any logic whose logical constant abide by the Iso_p thesis:

FACT 4.2. *Let $\mathcal{L}$ be a logic whose logical constants denote Iso_p invariant operators, $\mathcal{L}$ has the Karp property.*

There is an analogon to the "functional completeness" results for $L_{\infty,\infty}$ with respect to invariance under bijection and $L^{-}_{\infty,\infty}$ with respect to invariance under homomorphism:

THEOREM 4.3 (Barwise, 1973). *An operator Q is Iso_p-invariant iff for any set M, Q_M is definable in $L_{\infty,\omega}$.*

Note however that this kind of result is weaker than Feferman's result for FOL^{-} and invariance under homomorphism restricted to monadic operators. His result establishes the global definability of invariant operators: there is one formula of FOL which defines the operation Q_M of an operator Q, whatever M. By contrast, the three results for *Iso*, *SHom* and Iso_p bear only upon the local definability of invariant operators: for every set M, there is a (different) formula defining the operation Q_M. Obviously, global definability implies local definability, but the converse is not true in general. Q_{WF} is Iso_p-invariant. It is locally definable in $L_{\infty,\omega}$ (every order type is definable in $L_{\infty,\omega}$), but it is not globally definable in $L_{\infty,\omega}$: every formula $\phi(R, \dots)$ which has arbitrarily big well-founded models admits a non well-founded model.

How could we recover global definability for Iso_p-invariance? First note that, similarly to what happens with Feferman's result, the restriction to monadic operators would make things easier:

THEOREM 4.4. *The class of operators λ-definable from monadic Iso_p-invariant operators is globally definable in $L_{\aleph_1,\omega}$.*

PROOF. We shall just sketch the proof, which relies on the same idea as Feferman's proof for his Theorem 5.14 (1999, p. 42). It is sufficient to show that each monadic Iso_p-invariant operator is definable in $L_{\aleph_1,\omega}$. Let Q be such an operator. Q is a class of structures of the form $\langle M, P_1, \dots, P_n\rangle$ where each P_i for $i \in \{1, \dots, n\}$ is unary. Since Q is *Iso*-invariant, the fact that a given model $\mathcal{M}$ is or is not in Q is fully determined by the cardinality of each boolean compound of the $P_1, \dots, P_n$. But since Q is Iso_p-invariant, we have more than that: its behavior must be uniform with respect to boolean compounds of infinite cardinality. Therefore the behavior of Q is fully determined by the value belonging to $\{0, \dots, n, \dots, \aleph_0\}$ associated with each boolean compound—this values correspond to the cardinality of the set *up to potential isomorphism*. Each value is described by a sentence of $L_{\aleph_1,\omega}$ (either "there are exactly n objects" or "there are infinitely many objects"). There are 2^n different boolean compounds, so there are only $\aleph_0^{2^n} = \aleph_0$

different possibilities. Hence, Q can be described by a disjunction of size $\aleph_0$ of $L_{\aleph_1,\omega}$-formulas, which is again a formula of $L_{\aleph_1,\omega}$. ⊣

However, for the reasons mentioned in section 2, we do not think that the restriction to monadic operators is conceptually motivated. Is there a natural logic L extending $L_{\infty,\omega}$ such that Iso_p-invariant operators are globally definable in L? This is an open question. We have seen that global definability in $L_{\infty,\omega}$ failed for Q_{WF}. In a sense, it would be sufficient to drop Q_{WF} and related operators to recover global definability. Let us see this. In abstract model theory, a logic L is said to be **bounded** if and only if, for any L-sentence $\phi(\overline{R}, \dots)$ having only models with well-ordered R, there is an ordinal α such that the order type of R is always less than α. By extension, we shall say that a class K of operators is bounded iff the logic L_K associated with it is bounded. The following is a mere rephrasing of Lindström's theorem:

THEOREM 4.5. *If K is a bounded class of operators such that $Inv(Iso_1) \leq K \leq Inv(Iso_p)$, K is globally definable in $L_{\infty,\omega}$.*

PROOF. It is sufficient to check that the standard proof of Lindström's theorem can be adapted. We give a sketch, following the proof by Flum (1985, Theorem 3.1, p. 106). One assumes for contradiction that there is a quantifier Q in K such that Q cannot be defined in $L_{\infty,\omega}$. For each ordinal α, there is a sentence χ_α of $L_{\infty,\omega}$ which is the disjunction of the formulas $\phi^\alpha_{\mathcal{M}}$ describing the α-isomorphism type of the structures $\mathcal{M}$ such that $Q(\mathcal{M})$. By hypothesis, none of the χ_α defines Q. This means that for each α, there are α-isomorphic models $\mathcal{A}_\alpha$ and $\mathcal{B}_\alpha$ such that $Q(\mathcal{A}_\alpha)$ but not $Q(\mathcal{B}_\alpha)$. By coding partial isomorphisms (this is doable in L_K because $Inv(Iso_1) \leq K$) and by using the logical constant $\overline{Q}$ denoting Q, one gets an L_K-sentence ψ which is true exactly in the models $\mathcal{M} = \langle V, W, R, I, \dots \rangle$ (where V and W are unary, R and I binary), such that:

- the V-part of $\mathcal{M}$ is in Q, the W-part of $\mathcal{M}$ is not in Q.
- R is an ordering.
- I is a function and for each a in the field of R, the image $I(a)$ of a by I is a non-empty set of partial isomorphisms between the V-part and the W-part of $\mathcal{M}$.
- The sequence $I(a)$ with a in the field of R has the back and forth property.

For each ordinal α, ψ has a model in which R is well-ordered of order type α. Since L_K is bounded by hypothesis, ψ has a non-well-ordered model $\mathcal{M}'$. Its V-part is in Q, but its W-part is not. Because $\mathcal{M}'$ is not well-ordered, there is an infinite descending sequence with respect to R. Hence, the V-part of $\mathcal{M}'$ and the W-part of $\mathcal{M}'$ are potentially isomorphic. This contradicts $K \leq Inv(Iso_p)$. ⊣

This theorem says that any not too meager class K below $Inv(Iso_p)$ which does not contain Q_{WF} and similar operators is globally definable in $L_{\infty,\omega}$. In the other direction, would it be sufficient to add something like Q_{WF} or a fixed point operator to $L_{\infty,\omega}$ in order to get global definability for the full class $Inv(Iso_p)$?

4.3. Logic, mathematics and the overgeneration problem. The overgeneration problem presented in the first section follows from the collapse of logic onto mathematics that results from Tarski's thesis. The incentive to revise Tarski's thesis comes from the need to account for the intuitive difference between logic and mathematics. Now that we have put forward a conceptually motivated alternative to Tarski's thesis, the question is: what is the picture of the relationship between logic and mathematics according to the Iso_p thesis for logicality, and does it look right? Here are a few facts following from the Iso_p thesis which are relevant to the discussion:[26]

(i) The issue of the logicality of cardinality quantifiers is addressed in a selective way: only quantifiers which do not distinguish among infinite cardinals pass the test.
(ii) All arithmetical truths are logical truths.
(iii) Not all mathematical truths are logical truths.
(iv) Second-order logic is not a genuine logic.[27]

Even if $L_{\infty,\omega}$ is much more powerful than FOL, the shift from *Iso* to Iso_p seems more significant than the shift from *Iso* (and $L_{\infty,\infty}$) to *SHom* (and $L^{-}_{\infty,\infty}$). The Iso_p thesis for logicality sets the boundary between logic and mathematics somewhere between arithmetic and set theory. Regarding the "natural logic" underlying natural language, this seems to be a reasonable place to draw the line. On the one hand, the language of arithmetic indeed belongs to natural language. Number words are regular words, on a par with other determiners and adjectives. The mastery of logical consequence in natural language requires some arithmetical abilities. For example, the inference from "three men and two women came" to "five men and women came" (we assume the group reading for the second sentence) seems to be a logical inference that any competent speaker should accept. The difference between the finite and the infinite is also in the lexicon: "finitely many" or "an infinite number of" are perfectly standard determiners, which do not belong only to the highly specialized language of mathematicians. Moreover, the inferences associated with the distinction between the finite and the infinite seem also to be part of the semantic competence of speakers, at

[26]We omit here the precise meaning of these statements and their mathematical justifications.

[27]As a matter of fact, Second-order logic can express cardinality properties which are not invariant under potential isomorphism, and, for that reason, is out of the picture of logical operations that we drew here. However, a systematic development of generalized invariance to higher-order operations is still missing.

least up to some point. For example, a competent speaker should recognize that the sentence "many French movies favor introspection" does not imply the sentence "infinitely many French movies favor introspection". On the other hand, set theory proper, and its specific questions, like the Continuum Hypothesis or Large Cardinals axioms clearly lack a counterpart in everyday discourse. The notions they involve do not belong at all to our basic semantic competence. As a matter of fact, even if the notions of "set" or "belonging to" can be conveyed in ordinary English, it seems highly implausible that a competent speaker should be able, on the basis of its semantic competence, to accept or reject the Continuum Hypothesis.

As a consequence, even if Iso_p-invariance is quite liberal, specially if one has in mind FOL as the standard target, it seems that the overgeneration problem is at least eased, if not solved, by the shift from *Iso*-invariance to Iso_p-invariance. Moreover, the fact that we do not get exactly FOL does not seem to be a problem to us. After all, linguists know that a lot of arguments which are intuitively valid in English cannot be adequately formalized in FOL. Consider the following one:

> Most French movies favor introspection.
> Most French movies are commercial failures.
> ---
> There are French movies which favor introspection and which are commercial failures.

The determiner "Most" is not definable in FOL, but (at least on countable models, which seem to be sufficient for the kind of examples we consider) it is Iso_p-invariant and definable in $L_{\infty,\omega}$. Because the greater expressive power of $L_{\infty,\omega}$ compared to FOL is indeed useful to formalize natural language arguments, it seems to us that we should not be too quick in considering that any criterion yielding a logic more expressive than FOL overgenerates. In this respect, at least some of the extra expressive power provided by Iso_p invariant operators seems to be a reason to accept the Iso_p thesis, rather than a reason to reject it.

As a consequence of the adoption of the Iso_p thesis, arithmetic turns out to be a part of logic. In this sense, logicism might seem to be partly vindicated. But this "logicization" of arithmetic is not different from the "logicization" of mathematics that ensues from Tarski's thesis. Arithmetical truths are logical truths, but this does not come as the result of the kind of successful reduction of arithmetical truths to a class of more elementary logical truths that Frege tried to accomplish. What we have seen is rather that, according to a certain characterization of what logic is, arithmetic *de facto* belongs to the realm of logic. Even if there is no reduction here, the result is not totally deprived of interest. We have shown that arithmetical notions have in common with the more elementary logical notions a number of properties that grant them a special place in our conceptual scheme. Not only are they devoid of non formal content, but they are devoid of problematic set-theoretical content (they are absolute) and they belong to the smallest (closed under definability) family of notions which are extremely general without being trivial.

The interpretation of Theorem 3.10 seems particularly interesting to us. What the theorem tells us is that, if we try to characterize logicality in terms of invariance and if we accept the principle of closure under definability, we are "automatically" going to embrace the arithmetical notions as soon as we embrace the most elementary logical notions. Negatively, this could raise suspicion: after all, why should the logicality of the existential quantifier and the seemingly innocuous principle of closure under definability have something to do with the logicality of arithmetical notions? More precisely, the suspicion would be that similarity relations and closure principles do not interact well together, and that the strength of Theorem 3.10 mirrors the inadequacy of the framework of generalized invariance to capture logicality. Positively, if we think that characterizing logicality in termes of neutrality with respect to certain types of differences between structures is a good idea, Theorem 3.10 is a consequence that has to be accepted. In this case, the lesson would be something like: as soon as the elementary logical notions are there, the arithmetical notions are "potentially" there. This idea receives its precise mathematical formulation from the principle of closure under definability.[28]

Do all Iso_p-invariants match our intuitions regarding which operations are logical? The well-foundedness quantifier is a case in point. As we said, it is invariant under potential isomorphism, but it does not seem to belong to the 'natural logic' underlying our semantic competence. The same thing goes with 'ordinal quantifiers' such as 'being isomorphic to α', for a given ordinal α, which are invariant under potential isomorphism as well (two well orderings are potentially isomorphic iff they are isomorphic). Again, this reflects the power of the principle of closure under definability which gives us not only standard arithmetic but all of ordinal arithmetic as well. We shall not try to argue that such notions are necessary to formalize natural language arguments. Still, there might be a sense in which ordinal arithmetic is a natural extension of arithmetic which is not plagued by the uncertainties of cardinal arithmetic. In any case, the fact that ordinal arithmetic comes in one piece with arithmetic and that arithmetic itself comes in one piece with existential quantification on behalf our the principle of closure seems to us to be an interesting consequence of the approach, even though one might object that the ghost of overgeneration is still with us.

[28]Is this relationship relevant from a cognitive perspective? We shall not enter into this discussion here, but the question is an important one. The program that consists in characterizing logicality in terms of invariance would be even more legitimate and promising if it was backed up by empirical evidence showing that there is a connection between logical and arithmetical abilities, and that invariance plays a role in this connection. This would constitute some kind of cognitive counterpart to the purely conceptual analysis of the most basic part of our conceptual scheme that we tried to provide.

Conclusion. Tarski's thesis for logical operators provides a necessary criterion of logicality. However, because of the overgeneration problem, this criterion does not seem to be sufficient. Our aim has been to find an alternative criterion, based on a strengthening of the requirement of invariance. At the conceptual level, the two arguments in favor of Tarski's thesis were shown to be flawed. We have suggested that the generality argument should be relocated within the setting of generalized invariance, and supplemented with a constraint of closure under definability. The formality argument would be fine as it stands, if 'empirical' (non formal) content was all there is to content. Our proposal has been to turn the argument into a more general "lack of content argument" in which absoluteness plays a crucial role. The revised arguments support an alternative view on logicality, according to which logical operations are invariant under potential isomorphism.

REFERENCES

J. Barwise [1973], *Back and forth through infinitary logic*, ***Model theory*** (M. D. Morley, editor), Studies in Mathematics, vol. 8, Mathematical Association of America, pp. 5–34.

J. Barwise [1985], *Background and aims*, ***Model-theoretic logics*** (J. Barwise and S. Feferman, editors), Springer, pp. 25–76.

J. Barwise and R. Cooper [1981], *Generalized quantifiers and natural language*, ***Linguistics and Philosophy***, vol. 4, pp. 159–219.

J. van Benthem [2002], *Logical constants, the variable fortunes of an elusive notion*, ***Reflections on the foundations of mathematics: Essays in honor of Solomon Feferman*** (W. Sieg, R. Sommer, and C. Talcott, editors), Lecture Notes in Logic, vol. 15, ASL and AK Peters, Ltd., Natick, MA, pp. 420–440.

D. Bonnay [2006], ***Qu'est-ce qu'une constante logique ?***, Ph.D. Dissertation, University Paris 1.

E. Casanovas [2007], *Logical operations and invariance*, ***Journal of Philosophical Logic***, vol. 36, no. 1, pp. 33–60.

K. Došen [1994], *Logical constants as punctuation marks*, ***Notre Dame Journal of Formal Logic***, vol. 30, pp. 362–381.

M. Dummett [1991], ***The logical basis of metaphysics***, Harvard University Press.

J. Etchemendy [1999], ***The concept of logical consequence***, CSLI Publications, Stanford.

S. Feferman [1968], *Persistent and invariant formulas for outer extensions*, ***Compositio Mathematica***, vol. 20, pp. 29–52.

S. Feferman [1972], *Infinitary properties, local functors, and systems of ordinal functions*, ***Conference in Mathematical Logic, London '70*** (W. Hodges, editor), Lecture Notes in Mathematics, vol. 255, Springer, pp. 63–97.

S. Feferman [1999], *Logic, logics, and logicism*, ***Notre Dame Journal of Formal Logic***, vol. 40, no. 1, pp. 31–54.

J. Flum [1985], *Characterizing logics*, ***Model-theoretic logics*** (J. Barwise and S. Feferman, editors), Springer, pp. 77–120.

K. Gödel [1940], ***The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory***, Annals of Mathematical Studies, vol. 3, Princeton University Press, reprinted in ***Kurt Gödel, Collected Works***, vol. II, (S. Feferman *et alii*, editors), Oxford University Press, Oxford, 1990, pp. 33–101.

E. Keenan and D. Westerståhl [1997], *Generalized quantifiers in linguistics and logic*, ***Handbook of logic and language*** (J. van Benthem and A. ter Meulen, editors), Elsevier, pp. 837–893.

J.G. MacFarlane [2000], ***What does it mean to say that logic is formal***, Ph.D. Dissertation, University of Pittsburgh.

J. A. Makowski and D. Mundici [1985], *Abstract equivalence relations*, ***Model-theoretic logics*** (J. Barwise and S. Feferman, editors), Springer, pp. 717–746.

V. McGee [1996], *Logical operations*, ***Journal of Philosophical Logic***, vol. 25, pp. 567–580.

A. Mostowski [1957], *On a generalization of quantifiers*, ***Fundamenta Mathematicae***, vol. 44, pp. 12–36.

W. V. O. Quine [1986], ***Philosophy of logic***, second ed., Harvard University Press.

S. Shapiro [1998], *Logical consequence: Models and modality*, ***Philosophy of Mathematics Today: Proceedings of an International Conference in Munich*** (Schirn, editor), Oxford University Press, Oxford, pp. 131–156.

G. Sher [1991], ***The bounds of logic***, MIT Press, Cambridge.

A. Tarski [1986], *What are logical notions?*, ***History and Philosophy of Logic***, vol. 7, pp. 143–154.

J. Väänänen [1985], *Set-theoretic definability of logics*, ***Model-theoretic logics*** (J. Barwise and S. Feferman, editors), Springer, pp. 599–644.

DÉPARTEMENT D'ÉTUDES COGNITIVES & IHPST
ÉCOLE NORMALE SUPÉRIEURE
29, RUE D'ULM
75005 PARIS, FRANCE
E-mail: denis.bonnay@ens.fr