Intersubstitutivity

Jaroslav Peregrin*

Atomism vs. holism in semantics

Atomists explain properties of wholes as compositions of properties of their parts; in particular properties of complex expressions as composed of properties of their parts. Especially, semantic atomists explain meanings of complex expressions as composed of meanings of their parts. Holists deny themselves this way: they insist that at least in some cases properties of wholes are more basic than, or not reducible to, properties of their parts; in particular, semantic holists claim that meanings of (at least some) wholes are more basic than meanings of parts. Now as atomists have *composition* as the way of getting themselves from the meanings of parts to the meanings of wholes, holists need something to get them from the meanings of wholes to those of their parts. What is usually invoked in this context is the concept of *intersubstitutivity*, which is, however, not always wholly clear.

Intersubstitutivity is always w.r.t. an equivalence. Indeed, let E be an equivalence relation over a part-whole structure S; if $z[x/y]$ denotes any element of S which arises from z by replacing some part(s) x of z by y, then the corresponding intersubstitutivity relation I_E is generally defined as follows[1]

$$x \, I_E \, y \equiv_{\text{Def.}} \text{for every } z \in S \text{ and for every } z[x/y]\text{: } z \, E \, z[x/y].$$

With respect to meaning, we face the question what is the equivalence relation we should derive our intersubstitutivity relation from. Sameness of truth values? Or rather sameness of truth *conditions*? Or some kind of inferential equivalence as the inferential semanticists seem to suggest? Let us first thoroughly inspect what the possibilities are, by considering the various kinds of 'implications' or 'modes of entailment' which are on offer and which can possibly ground relevant equivalences.

* Work on this paper was supported by the grant No. 401/04/0117 of the Grant Agency of the Czech Republic.

1 I discussed this in detail elsewhere (Peregrin 2001, Chapter 4).

Modes of entailment

Let us adopt a very general semantic framework.[2] A language will be basically identified with a set S (*statements*) and a set V of mappings of S on the set $\boldsymbol{B} = \{0,1\}$ of the two truth values (*acceptable truth-valuations*). Within this framework, statements may be equivalent either in the sense of being assigned the same truth values by *a particular* valuation, or by *every* valuation. Of the particular valuations, we might assume that one has the distinguished status of being the *actual* one; let us denote it as v_A. Hence let us refine our semantic framework by adding v_A as its third component. Then we define, for an arbitrary two statements A and B,

$$A \Leftrightarrow_I B \equiv_{\text{Def.}} v_A(A) = v_A(B)$$
$$A \Leftrightarrow_{II} B \equiv_{\text{Def.}} \text{for every } v \in V\text{: } v(A) = v(B)$$

We can also define the 'entailment' relations which can be seen as underlying these equivalences

$$A \Rightarrow_I B \equiv_{\text{Def.}} \text{if } v_A(A) = 1\text{, then } v_A(B) = 1 \text{ (i.e. either } v_A(A) \neq 1 \text{ or } v_A(B) = 1)$$
$$A \Rightarrow_{II} B \equiv_{\text{Def.}} \text{for every } v \in V\text{: if } v(A) = 1\text{, then } v(B) = 1$$

More generally, if X is a set of sentences, we define

$$v(X) = 1 \equiv_{\text{Def.}} \text{for every } A \in X\text{: } v(A) = 1$$
$$v(X) = 0 \equiv_{\text{Def.}} \text{there is an } A \in X \text{ such that } v(A) = 0$$

and we further define

$$X \Rightarrow_I B \equiv_{\text{Def.}} \text{if } v_A(X) = 1\text{, then } v_A(B) = 1$$
$$X \Rightarrow_{II} B \equiv_{\text{Def.}} \text{for every } v \in V\text{: if } v(X) = 1\text{, then } v(B) = 1$$

Moreover, it seems that it might be useful, for the sake of generality, to refine our framework in one more respect. It seems that truth valuations may come in 'bundles': we may, for example, want to consider various Kripke structures for modal logic and interpretations related to these structures; so that we might end up with a different set of acceptable truth valuations for each of the structures. Hence let us replace our set V of valuations by a family $\langle V_i \rangle_{i \in I}$ of such sets. We will call the elements of the family *the acceptable clusters* of valuations; and we will keep using the symbol V for the set of all valuations, i.e. for $\cup_{i \in I} V_\mathrm{i}$.

[2] I argued for this kind of framework in Peregrin (1997). Since then, I have discovered that some other authors adopted it too. The first as far as I have been able to find out, was van Fraassen (1971); recently it was elaborated by Dunn and Hardegree (2000).

Hence what we now see as a language is the ordered triple $L = \langle S, \langle V_i \rangle_{i \in I}, v_A \rangle$. Given this last refinement, we may consider a third kind of entailment relation:

$X \Rightarrow_{III} B \equiv_{\text{Def.}}$ for every $i \in I$: if $v(X) = 1$ for every $v \in V_i$, then $v(B) = 1$ for every $v \in V_i$

We will call $\Rightarrow_{III}$ and $\Rightarrow_{II}$, following Fagin et al. (1992), *validity* and *truth entailment*, respectively. We will use the traditional name *material entailment* for $\Rightarrow_I$.

It is easy to see that the truth entailment is stronger than both the material entailment and the validity entailment in the sense that for every X and A

if $X \Rightarrow_{II} A$, then both $X \Rightarrow_I A$ and $X \Rightarrow_{III} A$

Let us call a language *normal* if none of its acceptable valuations is a constant function; and let us call it *extensional* if it has only one acceptable truth-valuation (which hence must be the actual one). Let us call it *straight* iff V_i is a singleton for every $i \in I$.

Theorem 1. For a normal language, $\Rightarrow_{II}$ coincides with $\Rightarrow_I$ iff the language is extensional.
Proof: Let L be a normal language. As the inverse implication is trivial, we must prove only the direct one. Hence let $X \Rightarrow_I A$ imply $X \Rightarrow_{II} A$ for every X and A. We are going to prove that if $v_A(A) = 1$, then $v(A) = 1$ for every $v \in V$; and that if $v_A(A) = 0$, then $v(A) = 0$ for every $v \in V$; and hence that $v = v_A$ for every $v \in V$. So let first $v_A(A) = 1$. Then $\varnothing \Rightarrow_I A$, and hence $\varnothing \Rightarrow_{II}$ A, and hence $v(A) = 1$ for every $v \in V$. Next, let $v_A(B) = 0$. Then $B \Rightarrow_{II} C$ for every $C \in S$. Hence $B \Rightarrow_{II} C$ for every $C \in S$ and hence for every $C \in S$ and every $v \in V$, it is the case that either $v(B) = 0$ or $v(C) = 1$. As for every $v \in V$ there is a $C \in S$ such that $v(C) = 0$ (for the language is normal), $v(B)$ is bound to be 0. Hence every $v \in V$ coincides with v_A.

Theorem 2. $\Rightarrow_{II}$ coincides with $\Rightarrow_{III}$ iff for every set X of statements and every statement A the following is the case: if there is a $v \in V$ such that $v(X) = 1$ and $v(A) = 0$, then there is an $i \in I$ such that $v'(X)=1$ for every $v' \in V_i$ and $v'(A)=0$ for at least one $v' \in V_i$.
Proof: The inverse implication is obvious; hence let us prove only the direct one. Thus, let $X \Rightarrow_{III} A$ entail $X \Rightarrow_{II} A$ for every X and A; and let $v \in V$ be such that $v(X) = 1$ and $v(A) = 0$. Then it is not the case that $X \Rightarrow_{II} A$ and hence not the case that $X \Rightarrow_{III} A$. This means that not for every $i \in I$ is it the case that if $v(X) = 1$ for every $v \in V_i$, then $v(A) = 1$ for every $v \in V_i$; in other words, there is an $i \in I$ such that $v(X) = 1$ for every $v \in V_i$ and $v(A) = 0$ for at least one $v \in V_i$.

Corollary: $\Rightarrow_{II}$ coincides with $\Rightarrow_{III}$ if the language is straight.

Are these 'modes of entailment' the only ones which are worth considering as underlying equivalences potentially relevant for meaning? Fagin et al. (*ibid.*) point out that if we see the set of statements as structured, we might get more important modes from taking substitutions into account.

Let us assume that our set S of statements is the result of applying statement-forming operations $<O_j>_{j\in J}$ to a set S_{At} of atomic statements. Let *Subst* be the set of all automorphisms from S to S (i.e. substitutions induced by assignments of elements of S to those of S_{At}). Then Fagin et al. (*ibid.*) propose to concentrate on the following variants of our entailment relations which they call *schematic:*[3]

$$X \mid\Rightarrow_{II} A \equiv_{\text{Def.}} \text{for every } s\in Subst_{\text{L}}, \text{ every } i\in I \text{ and every } v\in V_i\text{: if } v(s(X)) = 1, \text{ then } v(s(A)) = 1$$

$$X \mid\Rightarrow_{III} A \equiv_{\text{Def.}} \text{for every } s\in Subst_{\text{L}} \text{ and every } i\in I\text{: if } v(s(X)) = 1 \text{ for every } v\in V_i, \text{ then } v(s(A)) = 1 \text{ for every } v\in V_i$$

This can be generalized to I:

$$X \mid\Rightarrow_{I} A \equiv_{\text{Def.}} \text{for every } s\in Subst_{\text{L}}\text{: if } v_A(s(X)) = 1, \text{ then } v_A(s(A)) = 1,$$

so that we have the general

$$X \mid\Rightarrow_{N} A \equiv_{\text{Def.}} \text{for every } s\in Subst_{\text{L}}\text{: } s(X) \Rightarrow_N s(A).$$

It is clear that if $X \mid\Rightarrow_N A$, then $X \Rightarrow_N A$, for N = I, II, III.

Should we consider these modes of entailment instead of, or in addition to, the above ones? Let us notice that as what we are interested in are *fully interpreted* languages, we will certainly face entailment relations which are *not* schematic. If the statements A, B and C are *Fido is a dog*, *Fido is a cat* and *Fido is an animal*, then A entails C but certainly not B, so the entailment cannot be schematic. Hence while a schematic entailment is suitable if we see a language *as a propositional*[4] *logic* (i.e. with only operators interpreted), they are not suited for the case of (fully interpreted) *languages*.

More formally, we can show that if the space of acceptable valuations is such that they fit the logic-conception, the schematic modes of entailment reduce to the nonschematic ones. The proof is based on the assumption that if we see a language as a propositional logic, we do not distinguish between individual atomic statements, i.e. we see any assignment of truth values to them as equally good as any other.

In particular, seeing language as a propositional logic clearly involves seeing

[3] More precisely, they call the ones we have considered so far *nonschematic*.

[4] Of course, there is a corresponding notion of schematicity for *predicate* logic.

any atomic statement as capable of playing the role of any statement whatsoever. This means that given an acceptable interpretation, replacing an atomic statement by any other statement must bring us again to an acceptable interpretation. In other words, given an interpretation I of L, there must be an interpretation I' which maps atomic statements on the values of randomly chosen statements as assigned to them by I – hence that a composition of an interpretation with any substitution must be again an interpretation.

In the case when interpretations come in clusters, the same should hold for the whole clusters: given a cluster V_i of interpretations of L there must be a cluster V_j consisting of interpretations which arise out of interpretations of V_i by way of a given substitution – for any substitution there must be a cluster which consists of compositions of elements of V_i with the substitution. (That is, a substitution not only maps interpretations on interpretations, but also clusters on clusters.)

We will say that a language $L = \langle S, \langle V_i \rangle_{i \in I}, v_A \rangle$ is *closed under substitution* iff for every valuation v and every substitution s, $v \circ s$ is again a valuation. We will say that L is *strongly closed under substitution* iff, moreover, for every $i \in I$ and every substitution s, there is a $j \in I$ so that $V_j = \{v \circ s \mid v \in V_i\}$.

Theorem 3: If L is closed under substitution, then $X \Rightarrow_{II} A$ entails $X \mid\Rightarrow_{II} A$, and hence the two relations coincide.

Proof: Let $X \Rightarrow_{II} A$, hence let it be the case, for every valuation v, that $v(X) = 1$ entails $v(A) = 1$. Let v be a valuation, $s \in Subst_L$ and let $v(s(X)) = 1$. Then $v' = v \circ s$ is a valuation and hence as $v'(X) = 1$, $v'(A) = 1$. However $v'(A) = v \circ s(A)$, and hence $v \circ s(A) = 1$.

Theorem 4: If L is strongly closed under substitution, then if $X \Rightarrow_{III} A$, then $X \mid\Rightarrow_{III} A$, and hence the two relations coincide.

Proof: Let $X \Rightarrow_{III} A$, hence let it be the case, for every $i \in I$, that if $v(X) = 1$ for every $v \in V_i$, then $v(A) = 1$ for every $v \in V_i$. Let, moreover, $v(s(X)) = 1$ for every $v \in V_j$ for some $j \in I$ and some $s \in Subst_L$. Then there is a $k \in I$ so that $V_k = \{v \circ s \mid v \in V_j\}$; and hence $v(X) = 1$ for every $v \in V_k$ and hence $v(A) = 1$ for every $v \in V_k$. Hence $v(s(A)) = 1$ for every $v \in V_j$.

We have proved that if a language is strongly closed under substitution, then $X \mid\Rightarrow_{II} A$ coincides with $X \Rightarrow_{II} A$ and $X \mid\Rightarrow_{III} A$ coincides with $X \Rightarrow_{III} A$. This completes our argument for the claim that investigating a fully interpreted language we can disregard the two schematic modes of entailment discussed by Fagin et al.

However, we may further ask: is there a similarly natural condition under which $X \mid\Rightarrow_I A$ coincides with $X \Rightarrow_I A$? This does not seem to be the case, but there is an interesting condition under which $X \mid\Rightarrow_I A$ coincides with

$X \Rightarrow_{II} A$. And though it is unclear whether the relation $|\Rightarrow_I$ is really interesting at all (to my knowledge nobody has proposed investigating it), we will discuss it, for it is connected to the interesting concept of *root* (introduced by Dunn and Hardegree, §5.9): we call a valuation v a *root* iff for every valuation v' there is a substitution s such that $v'=v \circ s$. Now we prove that

Theorem 5: If v_A is a root, then $X |\Rightarrow_I A$ entails $X \Rightarrow_{II} A$.
Proof: Let $X |\Rightarrow_I A$, hence let, for every substitution s, $v_A \circ s(X) = 1$ entail $v_A \circ s(A) = 1$. Let v be a valuation such that $v(X) = 1$. Then, as v_A is a root, there exists a substitution s so that $v = v_A \circ s$. Hence $v_A \circ s(X) = 1$, and consequently $v_A \circ s(A) = 1$. But as $v = v_A \circ s$, this means that $v(A) = 1$.

Corollary. If L is closed under substitution and v_A is a root, then the relations $|\Rightarrow_I$ and $\Rightarrow_{II}$ coincide.
Proof: If L is closed under substitution, then $X \Rightarrow_{II} A$ entails $X |\Rightarrow_{II} A$ and hence $X |\Rightarrow_I A$. If v_A is a root, then, vice versa, $X |\Rightarrow_I A$ entails $X \Rightarrow_{II} A$.

Should we expect that the actual valuation of every interesting language is a root? Not really; but there is an interesting class of languages for which not only the actual, but every valuation is a root. To characterize them, let us introduce some more terminology. If L is a language, O is a rule of the language and v is its valuation, then we will say that the ordered $(n+1)$-tuple $<t_1, ..., t_n, t_{n+1}>$ of truth values is an *O,v-pattern* iff there are statements $A_1, ..., A_n$ so that $v(A_1) = t_1, ..., v(A_n) = t_n$ and $v(O(A_1, ...,A_n)) = t_{n+1}$.

The collection of *O,v-patterns* for every v characterizes the behaviour of O in L. If this collection is a singleton, then O can be called *truth-functional*: indeed, then the truth value of $O(A_1, ...,A_n)$ is always determined by those of $A_1, ..., A_n$ in an unique way, If every O of L is truth-functional we can call L truth-functional. Similarly, the collections of *O,v-patterns* for every O characterizes v: if each such collection is a singleton, then v can be called *truth-functional*. (Note that it is *not* the case that L is truth-functional iff each of its valuations is truth-functional!) Now we can characterize the concept of root in the following way

Theorem 6. If L is normal and truth-functional, then each of its valuations is a root.
Proof: Let v and v' be acceptable valuations of L; we will find a substitution such that $v'=v \circ s$. As L is normal, there must be a sentence, call it 1_v, such that $v(1_v) = 1$, and another sentence, call it 0_v, such that $v(0_v) = 0$. Let now s be such that for every atomic statement A, $s(A) = 1_v$ if $v'(A) = 1$ and $s(A) = 0_v$ if $v'(A) = 0$ (and, of course, $s(O_j(A_1, ..., A_n)) = O_j(s(A_1), ..., s(A_n)))$ for every $j \in J$ and every n-tuple $A_1, ..., A_n$ from the domain of O_j). We will prove that $v'(A) = v \circ s(A)$ for every statement A by induction. Hence let first A be atomic. Then: (1) if

$v'(A) = 1$, then $v \circ s(A) = v(s(A)) = v(1_v) = 1$; and (2) if $v'(A) = 0$, then $v \circ s(A) = v(s(A)) = v(0_v) = 0$. Now assume that the statement holds for $A_1, ..., A_n$ and let us show that then it also holds for $O_j(A_1, ..., A_n)$. As L is truth-functional $v'(O_j(A_1, ..., A_n)) = O_j^*(v'(A_1), ..., v'(A_n))$. As according to the assumption, $v'(A_i) = v(s(A_i))$, $O_j^*(v'(A_1), ..., v'(A_n)) = O_j^*(v(s(A_1)), ..., v(s(A_n)))$. Then once more, in force of the truth-functionality of L, $O_j^*(v(s(A_1)), ..., v(s(A_n))) = v(O_j(s(A_1), ..., s(A_n)))$. And as s is a substitution and hence $O_j(s(A_1), ..., s(A_n)) = s(O_j(A_1, ..., A_n))$, it is the case that $v(O_j(s(A_1), ..., s(A_n))) = v \circ s(O_j(A_1, ..., A_n))$.

Corollary. If L is normal, truth-functional and closed under substitution, then $X \mid\Rightarrow_I A$ coincides with $X \Rightarrow_{II} A$.

Truth-functionality is not a necessary condition for every valuation being a root. Expressed informally, v is a root if any kind of pattern displayed by any valuation is also displayed by v. We are not going to make this informal statement precise here; let us only note that this characterizes a root as a valuation which realizes all patterns of behaviour of operators which are possible at all. Thus, we may imagine that the concept of root extends the concept of normal valuation – a valuation is normal if it displays a *minimal* amount of variation, if it does not assign the same truth value to all propositions. On the other hand, a valuation is a root if it displays a *maximal* amount of variation, if it instantiates every kind of pattern which is possible at all. Thus we could perhaps say that a root is a valuation which is 'patternally representative'.

In sum: for every normal language strongly closed under substitution the two important schematic modes of entailment can be reduced to the nonschematic ones: $\mid\Rightarrow_{II}$ reduces to $\Rightarrow_{II}$ and to $\mid\Rightarrow_{III}$ to $\Rightarrow_{III}$. (Moreover, for a nontrivial class of languages this holds also for the third schematic mode: for all truth-functional languages – but not *only* for them – $\mid\Rightarrow_{\mathrm{I}}$ reduces to $\Rightarrow_{II}$). This indicates that in our case of fully-interpreted languages, the schematic modes of entailment can be disregarded.[5]

[5] When is the distinction between schematic and nonschematic modes of entailment important? We have seen that for fully interpreted languages what we need is the nonschematic modes, whereas for logics the two modes coincide. But we may want, in the context of a fully interpreted language, to distinguish between the logical rules of inference and other ones: and here the requirement of schematicity might become the touchstone.

Examples

As examples, let us consider four kinds of simple propositional languages:

L1. The class of sentences of the language of this kind is generated from a basic stock of atomic sentences by means of the usual operators $\neg$, $\wedge$, $\vee$, $\rightarrow$. There is only one acceptable valuation, fufilling, for every two sentences, the following requirements:

$$v(\neg A) = 1 - v(A)$$
$$v(A \wedge B) = \min(v(A), v(B))$$
$$v(A \vee B) = \max(v(A), v(B))$$
$$v(A \rightarrow B) = \max(1 - v(A), v(B))$$

(We will call a valuation fufilling them *classical*.) Hence a **L1**-language is extensional; and thus for this language, all our three modes of entailment, $\Rightarrow_I$, $\Rightarrow_{II}$, and $\Rightarrow_{III}$, coincide.

We can see an **L1**-language as a language within the classical predicate calculus. (Let me remind the reader that we are dealing with *fully interpreted* languages, hence I speak about a language *within* a logic, i.e. a language consisting of the logical constants of the logic plus some extralogical constants.) It is a kind of language which might occur within the realm of mathematics, where sentences do not change their truth values – hence the only acceptable valuation.

L2. Languages of this kind are like **L1**-languages with the only difference that their spaces of acceptable truth-valuations consist of more than one (classical) valuation. Hence an **L2**-language is not extensional and its relations $\Rightarrow_I$ and $\Rightarrow_{II}$ need not coincide (though $\Rightarrow_{II}$ and $\Rightarrow_{III}$ still do). It is, however, truth-functional.

Like **L1**-languages, **L2**-languages can be seen as languages within the classical predicate calculus, but now as languages the sentences of which are likely to change their truth values; prototypically *empirical* languages (hence the non-trivial space of valuations).

L3. A language of this kind can be seen as an extension of an **L2**-language – it arises out of it by means of the addition of the operator $\Box$ plus the additional requirement on the acceptable valuations:

$$(\Box)\ v(\Box A) = 1 \text{ iff for every } v' \in V\text{: } v'(A)=1$$

Hence, the set of sentences of the resulting **L3**-language is an extension of that of the original **L2**-language and every acceptable valuation of the latter gets uniquely extended to an acceptable valuation of the former. Hence, as in the case of **L2**-languages, the relations $\Rightarrow_I$ and $\Rightarrow_{II}$ of **L3**-languages need not coincide, but $\Rightarrow_{II}$ and $\Rightarrow_{III}$ do.

L3-languages can be seen as languages within a logic reminiscent of Carnap's mostly forgotten modal logic **C** (see Schurz, 2001).

L4. Languages of this kind are like **L3**-languages except that their acceptable valuations come in clusters (at least some of which have more than one element) and the condition (□) is replaced by

(□′) if $v \in V_i$, then $v(\Box A) = 1$ iff for every $v' \in V_i$: $v'(A)=1$.

Hence an **L4**-language is not straight and so $\Rightarrow_{II}$ and $\Rightarrow_{III}$ need not coincide.

L4-languages can be seen as languages within the modal logic **S5**.

Modes of equivalence

We saw that we may concentrate on the equivalences induced by our three non-schematic modes of entailment. $\Leftrightarrow_I$, *material equivalence*, is simply sameness of (actual) truth values; whereas $\Leftrightarrow_{II}$, *truth equivalence*, may be seen as amounting to the sameness of truth conditions.[6] As for $\Leftrightarrow_{III}$, *validity equivalence*, it can be seen as equivalence w.r.t. validity (in a structure).

Now we are finally approaching the ultimate topic of the paper, intersubstitutivity. Let us call a formula *D* an *A/B-variant* of a formula *C* iff *D* arises out of replacing zero or more occurrences of *A* by *B*. Let us introduce the sign ≈ in the following way:

A $\approx_N$ B $\equiv_{\text{Def.}}$ for every sentence *C* and every *A*/*B*-variant $C[A/B]$ of *C*:
$C \Leftrightarrow_N C[A/B]$

Thus our three modes of entailment yield us three modes of intersubstitutivity. A $\approx_I$ B, meaning that *A* and *B* are intersubstitutive *salva veritate*; A $\approx_{II}$ B, amounting to intersubstitutivity w.r.t. truth conditions, i.e. intersubstitutivity *salva conditione veritatis*; and

A $\approx_{III}$ B, which we will call intersubstitutivity *salva analyticitate*. There are some obvious relations between the three modes: it is clear that if A $\approx_{II}$ B then both A $\approx_{III}$ B and A $\approx_I$ B. Intersubstitutivity *salva conditione veritatis* is thus the strongest of the three.

[6] More precisely, it amounts to sameness of truth conditions *insofar as those are expressible within the language*.

In the beginning of the paper we hinted at inferentialism: the doctrine that the meaning of an expression is its role w.r.t. inferential rules. This may appear to lead to the relation of intersubstitutivity directly w.r.t. entailment

> A $\sim_N$ B $\equiv_{\text{Def.}}$ for every (n+1)-tuple $C_1,...,C_n,C$ of sentences and every every (n+1)-tuple of A/B-variants $C_1[A/B],...,C_n[A/B],C[A/B]$ of $C_1,...,C_n,C$: $C_1,...,C_n \Rightarrow_N C$ iff $C_1[A/B],...,C_n[A/B] \Rightarrow_N C[A/B]$)

However, it is easy to see that $\sim_N$ is the same relation as $\approx_N$:

Theorem 7. A $\approx_N$ B iff A $\sim_N$ B.
Proof. Let A $\approx_N$ B. Then $C_1[A/B] \Leftrightarrow_N C_1,...,C_n[A/B] \Leftrightarrow_N C_n$ and $C \Leftrightarrow_N C[A/B]$ and hence in view of the transitivity of $\Rightarrow_N$, $C_1,...,C_n \Rightarrow_N C$ iff $C_1[A/B],...,C_n[A/B] \Rightarrow_N C[A/B]$) and thus A $\sim_N$ B. Let conversely A $\sim_N$ B. Then, as $C \Rightarrow_N C$, $C \Rightarrow_N C[A/B]$ and $C[A/B] \Rightarrow_N C$ and hence A $\approx_N$ B.

Deductors

An important part of logical vocabulary are constants making the relation of entailment explicit – i.e. allowing us to *say* that something follows from something else. Let us call a logical constant $\rightarrow$ a *deductor* if $A \rightarrow B$ says that A entails B, i.e. iff

$$A \Rightarrow B \text{ iff } \Rightarrow(A \rightarrow B),$$

However, as we have three versions of $\Rightarrow$ and as there are two occurrences of it within the definition, we have in general nine different versions of the deductor: we will call $\rightarrow$ the *M*,*N*-deductor iff

$$A \Rightarrow_M B \text{ iff } \Rightarrow_N(A \rightarrow B).$$

In view of the fact that $\Rightarrow_{II} C$ if and only if $\Rightarrow_{III} C$, $\rightarrow$ is a 2,2-deductor iff it is a 2,3 deductor and it is a 3,2-deductor iff it is a 3,3 deductor. Hence the different kinds of deductor we have reduce to seven. Moreover, as the 1,1-deductors, 1,2-decuctors and 1,3-deductors do not seem to be really interesting, we will concentrate on the remaining four kinds. We will dub the 2,2-deductor and the 3,2-deductor the *strong* and the *weak deductor*, respectively; and we will dub the 2,1-deductor and the 3,1-deductor the *strong* and the *weak counterfactual*, respectively. Hence

> $\rightarrow$ is a *strong counterfactual* $\equiv_{\text{Def.}}$ $A \Rightarrow_{II} B$ iff $\Rightarrow_I (A \rightarrow B)$
> for all sentences A and B

$\rightarrow$ is a *weak counterfactual* $\equiv_{\text{Def.}}$ $A \Rightarrow_{III} B$ iff $\Rightarrow_{I} (A \rightarrow B)$
for all sentences A and B
$\rightarrow$ is a *strong deductor* $\equiv_{\text{Def.}}$ $A \Rightarrow_{II} B$ iff $\Rightarrow_{II} (A \rightarrow B)$
for all sentences A and B
$\rightarrow$ is a *weak deductor* $\equiv_{\text{Def.}}$ $A \Rightarrow_{III} B$ iff $\Rightarrow_{II} (A \rightarrow B)$
for all sentences A and B

It is easy to see that our previous results entail that if L is straight, then the concepts of weak and strong counterfactual coincide, and similarly for those of weak and strong deductor; and if L is extensional, then strong deductors coincide with strong counterfactuals and similarly for the weak ones.

What is important is that the presence of some of these operators means a further reduction of our different modes of intersubstitutivity:

Theorem 8: If a language has a strong deductor, then $\approx_{II}$ coincides with $\approx_{III}$.
Proof: As clearly $\approx_{II} \subseteq \approx_{III}$, it is enough to prove the inverse inclusion. Suppose, then, that $A \approx_{III} B$. As it is obviously the case that $C \Leftrightarrow_{II} C$ and as $\rightarrow$ is a strong deductor, it is the case that $\Rightarrow_{II}(C \leftrightarrow C)$, which is the same as $\Rightarrow_{III}(C \leftrightarrow C)$. Thus, as $A \approx_{III} B$, $\Rightarrow_{III}(C \leftrightarrow C)[A/B]$, and in particular $\Rightarrow_{III}(C \leftrightarrow C[A/B])$, which is the same as $\Rightarrow_{II}(C \leftrightarrow C[A/B])$. But this means that $C \Leftrightarrow_{II} C[A/B]$, and hence that $A \approx_{II} B$.

Theorem 9: If a language has a strong counterfactual, then $\approx_{II}$ coincides with $\approx_{I}$.
Proof: Again, it is clear that $\approx_{II} \subseteq \approx_{I}$, so let us prove only the inverse inclusion. Suppose that $A \approx_{I} B$. Obviously $\Rightarrow_{I}(C \leftrightarrow C)$; Hence as A $\approx_{I}$ B, $\Rightarrow_{I}(C \leftrightarrow C)[A/B]$, in particular $\Rightarrow_{I}(C \leftrightarrow C[A/B])$. Hence, as $\rightarrow$ is a strong counterfactual, $C \Leftrightarrow_{II} C[A/B]$; and hence A $\approx_{II}$ B.

The most common operator of the kind of our deductors is the well-known *material implication*; can it be placed somewhere on our scale? The fact is that although material implication is necessarily a strong deductor, a strong deductor is not necessarily a material implication.

However, return, for a moment to the relation $\Rightarrow_{I}$. It concerns exclusively the actual truth-valuation. But the point of the non-actual valuations is usually in that they can *become* actual – that the role of the actual valuation is an itinerant one. And given this, we might want that something does not cease to be the kind of deductor it is if the actual valuation changes. This may bring us to the redefinition of our concept of N,1-deductor

for every $v \in V$: $A \Rightarrow_{N} B$ iff $\Rightarrow_{I} (A \rightarrow B)$.

In particular, $\rightarrow$ would be the 1,1-deductor iff

for every $v \in V$: ($v(A)=0$ or $v(B)=1$) iff $v(A \rightarrow B) = 1$.

And it is easy to see that it is this new concept of the 1,1-deductor which coincides with the concept of material implication.

Examples continued

L1: It is clear that $\rightarrow$ embodies all kinds of deductors: it is both a strong and a weak deductor, and both a strong and a weak counterfactual.

L2: $\rightarrow$ is still a strong and weak deductor, but it is no longer a counterfactual. In general, languages of this kind need not have a counterfactual. Therefore it seems natural to extend languages of this kind to languages of the **L3**-kind, the modal operator of which allows for a formation of a counterfactual.

L3: $\rightarrow$ is still a strong and weak deductor; and a (strong and weak) counterfactual can be defined:

$A \Rightarrow B \equiv_{\mathrm{Def.}} \Box(A \rightarrow B)$

L4: $\rightarrow$ is only a strong deductor. The week deductor can be defined as follows:

$A \Rightarrow B \equiv_{\mathrm{Def.}} \Box A \rightarrow \Box B$

In general, languages of this kind need not have a counterfactual.

Compositionality

It is clear that if A $\approx_N$ B, then A $\Leftrightarrow_N$ B; but of course not necessarily *vice versa*. We will now consider a characterization of the languages for which also the inverse holds, but we will approach the matter via a detour through the all-important concept of compositionality.

An assignment of values to sentences of a language is called *compositional* if the value of a complex sentence is uniquely determined by the values of its sentential parts (and the way of their combination). As the compositionality is usually associated with *meanings* and we have so far made no official pronouncement as to what, if anything, is to correspond to meaning within our framework, we will avoid the term in our following definitions and return to it explicitly only later.

Let $L=\langle\langle S_{At},\langle O_j\rangle_{j\in J}\rangle,\langle V_i\rangle_{i\in I},v_{\mathrm{A}}\rangle$ be a language. We will say that L is *quasiextensional* iff the actual valuation is truth-functional, i.e. iff for every $j \in J$

there is a function O_j^* so that for every n-tuple $A_1, ..., A_n$ from the domain of O_j it is the case that $v_A(O_j(A_1, ..., A_n)) = O_j^*(v_A(A_1), ..., v_A(A_n)))$. For every sentence A, let $\| A \|$, the *range of* A, be the set of all valuations which verify it; hence let $\| A \| = \{v \in V \mid v(A)=1\}$. We will say that L is *at most intensional* iff for every $j \in J$ there is a function O_j^* so that for every n-tuple $A_1, ..., A_n$ from the domain of O_j it is the case that $\| O_j(A_1, ..., A_n) \| = O_j^*(\| A_1 \| , ..., \| A_n \|))$. We will say that L is *intensional* iff it is at most intensional and not quasiextensional. We will say that it is *hyperintensional* iff it is not at most intensional.

Suppose that what we are after is (actual) truth, and that we are to account for it in a compositional way. If a language is quasiextensional, then the only thing we must consider is the actual truth valuation. Hence we can restrict ourselves to the extensional language which arises out of L by omitting all valuations save the actual one (hence the term "quasiextensional" – the language can be identified with an extensional language). If L is not quasiextensional, we cannot do this. However, if it is intensional, we can make do taking into account *all* valuations – we can have a compositional theory of ranges; and hence of truth, though not merely of the actual valuation. If a theory is hyperintensional, then the valuations cannot yield us *any* compositional theory of the actual valuation.

Theorem 10. A language is quasiextensional iff for all statements A and B, $A \Leftrightarrow_I B$ implies $A \approx_I B$. A language is at most intensional iff for all statements A and B, $A \Leftrightarrow_{II} B$ implies $A \approx_{II} B$.
Proof: Both direct implications are obvious; hence we will prove the inverse ones. First, suppose that $A \Leftrightarrow_I$ B implies A $\approx_I$ B, i.e. that $v_A(A) = v_A(B)$ implies $v_A(C) = v_A(C[A/B])$ for every C and every A/B-variant $C[A/B]$ of C. Then obviously for every O_j and every two n-tuple $A_1, ..., A_n$ from its domain $v_A(O_j(A_1, ..., A_n)) = v_A(O_j(A_1', A_2, ..., A_n))$ if $v_A(A_1) = v_A(A_1')$, and hence $v_A(O_j(A_1, ..., A_n)) = v_A(O_j(A_1', ..., A_n'))$ if $v_A(A_1) = v_A(A_1')$, ..., $v_A(A_n) = v_A(A_n')$. But if this is the case, then we can define O_j^* simply by stipulating $O_j^*(v_A(A_1), ..., v_A(A_n))) = v_A(O_j(A_1, ..., A_n))$.

Next suppose that for all statements A and B, $A \Leftrightarrow_{II} B$ implies $A \approx_{II} B$, i.e. that if $v(A) = v(B)$ for every $v \in V$, then $v(C) = v(C[A/B])$ for every C, every A/B-variant $C[A/B]$ of C and for every $v \in V$. Then obviously for every O_j and every two n-tuple $A_1, ..., A_n$ from its domain $\| O_j(A_1, ..., A_n) \| = \| O_j(A_1', A_2, ..., A_n) \|$ if $\| A_1 \| = \| A_1' \|$, and hence $\| O_j(A_1, ..., A_n) \| = \| O_j(A_1', ..., A_n') \|$ if $\| A_1 \| = \| A_1' \|$, ..., $\| A_n \| = \| A_n' \|$. But if this is the case, then we can define O_j^* simply by stipulating $O_j^*(\| A_1 \| , ..., \| A_n \|) = \| O_j(A_1, ..., A_n) \|$.

Returning to the question of meaning: elsewhere[7] I have argued that compositionality is a *constitutive* feature of meaning; hence that we can reasonably talk about meanings only w.r.t. a level of semantic values which is composi-

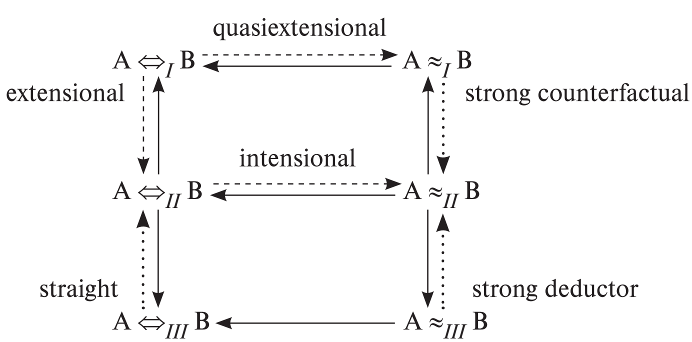
[7] See Peregrin (2005).

tional. For a quasiextensional language, this condition is fulfilled already by truth values – hence can we say that here *meaning = truth value*? Well, I think that though there is nothing else than truth values to play the role of sentence meanings within an extensional language, this is not bacause a truth value could be reasonably seen as meaning in the intuitive sense of the word (after all, who would want to see all true sentences as synonymous?), but rather because there is nothing whatsoever like meaning in the intuitive sense whatsoever within (quasi)extensional languages.

The space and the need for anything like meaning in the common sense of the world emerges only with languages which are *not* extensional. Only in this case there emerges the separation of the level of meanings (which one must know in order to *understand* a sentence) from the level of truth values (which one may be ignorant of despite knowing meaning).[8]

Conclusion

Below is the schema of the interdependencies of the relations we have studied. An arrow with a solid line represents unconditional inclusion, whereas the others kinds of arrows represent an inclusion on the condition specified above or beside; a dotted line means that the condition is only sufficient, whereas the dashed one means that it is also sufficient.



Natural language is clearly not extensional, hence $\Leftrightarrow_I$ is different from $\Leftrightarrow_{II}$. (Is it straight, or is it reasonable to see its interpretations as coming in some nontrivial clusters? This is an interesting question, especially in view of the fact that there are well-established logical regimentations of natural language of both kinds. But we will not address the question here.)

However, it seems that any natural language would have a strong counterfactual ("if [it were the case] ..., then [it would be the case] ..."). This means that $\approx_I$ will be the same relation as $\approx_{II}$ – hence we need not decide whether we

[8] See Peregrin (2001, §8.9) for more details.

can make do with intersubstitutivity *salva veritate*, or we need intersubstitutivity *salva conditione veritatis*. Both would come down to the same. This also means that intersubstitutivity *salva veritate* will be stronger than intersubstitutivity *salva analyticitate*, for intersubstitutivity *salva conditione veritatis* is. (In fact, the two relations will again be identical, for a natural language is likely to have also a strong deductor.) And as we saw that all conceivable varieties of intersubstitutivity *salva consequentia* are again reducible to our three modes, the moral appears to be that within natural language we can make do with good old intersubstitutivity *salva veritate*. The upshot then is that *in natural language*, intersubstitutivity *salva veritate* is likely to accomplish as much as any other of the commonly considered kinds of intersubstitutivities.

Department of Logic, Institute of Philosophy
Academy of Sciences of the Czech Republic
Jilská 1
110 00 Praha 1
peregrin@ff.cuni.cz
http://jarda.peregrin.cz

References

Dunn, J. M. & Hardegree, G. M. (2000). *Algebraic Methods in Philosophical Logic*, Oxford: Clarendon Press.

Fagin, R., Halpern, J. Y. & Vardi, M. Y. (1992). What is an Inference Rule? *Journal of Symbolic Logic*, *57*, 1018–1045.

Peregrin, J. (1997). Language and its Models. *Nordic Journal of Philosophical Logic*, *2*, 1–23.

Peregrin, J. (2001). *Meaning and Structure*. Aldershot: Ashgate.

Peregrin, J. (2005). Is compositionality an empirical matter? In M. Werning, E. Machery & G. Schurz (Eds.). *The Compositionality of Concepts and Meanings* (pp. 135–150), Frankfurt: Ontos.

Schurz, G. (2001). Carnap‘s Modal Logic. In W. Stelzner & M. Stöckler (Eds.). *Zwischen traditioneller und moderner Logik* (pp. 365–380), Paderborn: Mentis.

van Fraassen, B.C. (1971). *Formal Semantics and Logic*, New York: Macmillan.