From *Theoria* **71** (2005), 138–181.

Béziau's Translation Paradox

Lloyd Humberstone,
Monash University

ABSTRACT

Jean-Yves Béziau ('Classical Negation can be Expressed by One of its Halves', *Logic Journal of the IGPL* **7** (1999), 145–151) has given an especially clear example of a phenomenon he considers a sufficiently puzzling to call the 'paradox of translation': the existence of pairs of logics, one logic being strictly weaker than another and yet such that the stronger logic can be embedded within it under a faithful translation. We elaborate on Béziau's example, which concerns classical negation, as well as giving some additional background (especially from intuitionistic logic) to the example. Our interest is more on the logical exploration of the phenomenon Béziau's case exemplifies than on the question of whether that phenomenon is (even *prima facie*) paradoxical, though in Section 5 we do approach the latter question – somewhat obliquely – by considering an analogous phenomenon which it is hard to find puzzling.

KEYWORDS: Connectives, translational embeddings, negation, definition.

1. Introduction

As a prelude to giving the central example of Béziau (1999), we set up some suitable notation and terminology, as well as making some remarks pertinent to adapting Béziau's example from classical to intuitionistic logic. ("Logic" means *propositional* logic throughout.) A (propositional) language comprises the set of formulas obtained from countably many propositional variables (sentence letters) p_1, p_2, p_3, … by the application of connectives in some set which varies from language to language. (We will usually write p_1, p_2, p_3, as p, q, r.) A valuation for a language is any function mapping the set of formulas of the language into the two-element set $\{T, F\}$. For languages amongst whose connectives are to be the 1-ary connective $\neg$ or the binary connective $\rightarrow$ we call a valuation v *$\neg$-boolean* (resp. *$\rightarrow$-boolean*), if it satisfies, for all formulas A, B, the condition (1), (resp. (2)):

$$v(\neg A) = T \text{ if and only if } v(A) = F \tag{1}$$

$$v(A \rightarrow B) = T \text{ if and only if either } v(A) = F \text{ or } v(B) = T. \tag{2}$$

The logic *determined by* a class of valuations for some language is the set of formulas A of that language such that for each valuation v in the class, $v(A) = T$.

The above definition takes logics as sets of formulas, whereas a philosophically preferable identification of logics is with consequence relations $\vdash$. We presume known the defining conditions for a relation between sets of formulas and formulas (of some language) to constitute a consequence relation (on that language), and recall here the notion of when such a relation $\vdash$ is the consequence relation *determined by* a class of valuations V for the language of $\vdash$: namely when for all sets of formulas Γ and formulas B we have $\Gamma \vdash$ B if and only if for each $v \in V$ with v(C) = T for all C $\in \Gamma$, we have v(B) = T. One reason this is generally thought to be a better conceptualization of a logic than the set-of-formulas account is that it enables us to make a nice contrast between logics and theories: a set of formulas Θ is a $\vdash$-*theory* when for all B with $\Theta \vdash$ B, we have B $\in \Theta$. As is well known, for certain technical purposes an even more refined conception of logics, as generalized ("multiple-conclusion") consequence relations, is often preferred.[1] Again we assume known the definition, and using $\Vdash$ to symbolize such relations, recall only the definition of what it takes for such a relation $\Vdash$, between sets of formulas and sets of formulas, to be *determined by* a class of valuations (for the language of the given $\Vdash$): namely, when for all such sets Γ, Δ, we have $\Gamma \Vdash \Delta$ if and only if for each $v \in V$ with v(C) = T for all C$\in \Gamma$, we have v(D) = T for some D $\in \Delta$. This conception of logics affords a particularly convenient framework in which to explicate two logical relations of special significance for what follows, namely contrariety and subcontrariety. Here we use these terms in the rather weak sense of, e.g., Lemmon (1965), for their current convenience, without suggesting that richer notions of contrariety and subcontrariety aren't more convenient for other contexts. (See Humberstone forthcoming *a*, for example.) Formulas A and B are *contraries* according to a generalized consequence relation $\Vdash$ when (3) is satisfied, and *subcontraries* when (4) is satisfied.

$$\text{A, B} \Vdash \varnothing \tag{3}$$

$$\varnothing \Vdash \text{A, B}. \tag{4}$$

If $\Vdash$ is determined by any class of valuations V with $v \in V$ then when (3) is satisfied for a given A, B, we cannot have v(A) = v(B) = T, while if (4) is satisfied, then we cannot have v(A) = v(B) = F. Note that these conditions on v amount respectively to the "only if" and the "if" halves of (1) above for the case in which B is $\neg$A. Instead of relativizing contrariety and subcontrariety to logics (variously conceived) we could instead consider relativizing them to classes of valuations, saying that two formulas are *contraries relative to V* when no $v \in V$ verifies both of them, and *subcontraries relative to V* when no $v \in V$ falsifies both of them; thus relative to the class of $\neg$-boolean valuations (for a given language with A an arbitrary formula thereof) A and $\neg$A are contraries and also subcontraries. (In traditional parlance, an operator – such as $\neg$ when attention is restricted to $\neg$-

---

[1] In Humberstone (2000) these three ways of conceptualizing a logic are described as the logical frameworks FMLA, SET-FMLA, and SET-SET, respectively, and in the case of the latter two we make a sharp distinction between the metalinguistic predicate symbol ("$\vdash$" or "$\Vdash$") symbolizing a (generalized) consequence relation and the symbol used as a sequent separator. Here, and also in Humberstone (in preparation *b*), we shall be rather casual about this distinction and treat conditional requirements like (6) below as though they were sequent-to-sequent rules (rather than, more correctly, statements to the effect that certain such rules are admissible for the (generalized) consequence relations. Humberstone (in preparation *b*) was originally submitted as an Appendix to the present paper, for editorial reasons now hived off as a separate note.

boolean valuations – transforming what it applies to into something which is both a contrary and a subcontrary is said to symbolize *contradictory* negation.) In the degenerate case in which $V = \{v\}$, it is sometimes be convenient to say that the formulas in question are (sub)contraries relative to v when they are so related relative to V.

Let us proceed, however, with our tracing of the corresponding concepts as they apply relative to logics determined by such classes of valuations. If we pass from logics as generalized consequence relations to logics as consequence relations then we find that the closest we can come to (3) and (4) are (5) and (6) respectively

$A, B \vdash C$ for all formulas C (5)

$\Gamma, A \vdash C$ and $\Gamma, B \vdash C$ together imply $\Gamma \vdash C$, for all formulas C, sets Γ (6)

which no longer guarantee that no v in a V determining $\vdash$ assigns T to both A and B, or that no such v assigns F to both of them, though they do guarantee that $\vdash$ is determined by *some* class of valuations each member of which respects these constraints. (Should Γ be required to be $\varnothing$ in (6)? See Humberstone (in preparation *b*), Section 2.) Béziau's discussion is conducted at the level of logics as consequence relations, rather than generalized consequence relations, but we shall sometimes work with the even coarser conception of logics as (certain) sets of formulas, partly for consistency with some examples from Gabbay (1981) which will appear below (Section 3) and partly for the sake of a comparison with first-order theories (in Section 5); thus we need to reformulate these conditions, and an appropriate reformulation on the assumption that we have present the connective $\rightarrow$ with at least the logical properties conferred on it by intuitionistic logic ("positive implication", as it is sometimes called), we can reformulate (3) and (4) for the given A and B as (7) and (8) respectively, which say that A and B are contraries or subcontraries respectively, according to S (writing "$\vdash_{\mathsf{S}} A$" for "$A \in \mathsf{S}$"):

$\vdash_{\mathsf{S}} A \rightarrow (B \rightarrow C)$ for all formulas C (7)

$\vdash_{\mathsf{S}} (A \rightarrow C) \rightarrow ((B \rightarrow C) \rightarrow C)$ for all formulas C (8)

Simple syntactic descriptions are available of the $\rightarrow$-fragments of classical and intuitionistic implication, which we close this section by reminding the reader of. The fragment for classical logic is most simply isolated at the level of generalized consequence relations, in the present instance as the least such relation $\Vdash$ satisfying (9), while the corresponding formulation for intuitionistic logic is in terms of consequence relations, namely as the least such relation satisfying (10):

$\Gamma, A \Vdash B, \Delta$ if and only if $\Gamma \Vdash A \rightarrow B, \Delta$ (9)

$\Gamma, A \vdash B$ if and only if $\Gamma \vdash A \rightarrow B$ (10)

If we want a consequence relation formulation of the implicational fragment of classical logic, we need to require the satisfaction not only of (10) but also of a further condition, such as

$\Gamma, A \rightarrow B \vdash A$ implies $\Gamma \vdash A$ (11)

or equivalently (given (10)) and more revealingly for present purposes, since it displays $A \to B$ and A as subcontraries for any A, B:

$$\Gamma, A \to B \vdash C \text{ and } \Gamma, A \vdash C \text{ imply } \Gamma \vdash C. \tag{12}$$

For the “sets of formulas” conception of logics, one familiar syntactic description is as the formulas deducible from instances of the axiom-schemata (A1) and (A2) by *Modus Ponens*, for the implicational fragment of intuitionistic logic, with (A3) added for the classical case. ((A3) is Peirce’s Law, reformulated above as (11); note that (12) says that an implication and its antecedent are subcontraries.)

(A1) $A \to (B \to A)$

(A2) $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$

(A3) $((A \to B) \to A) \to A$

The most convenient way to think of adding negation in the classical case is, for the generalized consequence relation approach, to impose the conditions (3) and (4) for with B as $\neg A$. If we are using consequence relations instead, then we use the same special cases of (5) and (6). To obtain intuitionistic negation then instead we require (5), securing contrariety of A and its negation, and a principle making $\neg A$ the weakest contrary of A, such as the *Reductio* Rule that if $\Gamma, A \vdash B$ and $\Gamma, A \vdash \neg B$ then $\Gamma \vdash \neg A$. Although in the axiomatic approach a certain homage is customarily paid to economy by adding, for the classical case, an axiom schema such as $(\neg A \to \neg B) \to (B \to A)$ (“De-contraposition”), we prefer again to separate out the contrariety and subcontrariety aspects of the situation and add (7) and (8) with B as $\neg A$, i.e., axiom schemes (A4) and (A5), to be subjoined to (A1)–(A3):

(A4) $A \to (\neg A \to C)$

(A5) $(A \to C) \to ((\neg A \to C) \to C)$

To bring intuitionistic negation into similar line, we add to (A1) and (A2), the schema (A4) and a formula version of *Reductio*, such as

(A6) $(A \to C) \to ((A \to \neg C) \to \neg A)$.

One may check that in these last two axiomatizations (A5) and (A6) can be replaced by the simpler

(A5)′ $(\neg A \to A) \to A$

(A6)′ $(A \to \neg A) \to \neg A$.

As with De-contraposition, however, these fail to observe the principle of the separation of (contrarizing and subcontrarizing) powers, in that getting from them back to the longer forms (A5) and (A6) requires the assistance of (A4). In fact, even with (A6), a perfect separation has not been attained – and here we have in mind not a separation between contrary-formation and subcontrary formation, but between contrary-formation and the “weakestness” of the contrary so formed. Rather than further interrupting our discussion of Béziau, we postpone to Humberstone (in preparation *b*), Section 2, further discussion of the degenerate status of subcontraries in intuitionistic logic, the current failure-of-separation issue, and the (well-known) fact that adding (A4) and (A5) to (A1) and (A2) renders (A3) derivable – so there is some redundancy in the above characterization of classical propositional logic given above.

The classical implication and negation logics isolated in the foregoing discussion are the generalized consequence relation, the consequence relation and the logic-as-set of formulas, determined by the class of all $\to$-boolean and $\neg$-boolean valuations. A semantic description in the intuitionistic case is not given here, requiring the apparatus of the Kripke semantics as explained in, e.g., Gabbay (1981), where the reader will also find an explanation for the absence from the above survey of any reference to the intuitionistic generalized consequence relations, the point being that there is no unique generalized consequence relation deserving such a description (see Gabbay, p.27, and Theorem 6 on p.45 and Theorem 5 on p.49, for two generalized consequence relations agreeing with the consequence relation of intuitionistic logic in the case of $\Gamma \Vdash \Delta$ with Δ a singleton).

Let IL and CL, $\vdash_{IL}$ and $\vdash_{CL}$, be respectively, intuitionistic and classical propositional logic conceived of as sets of formulas, and these two logics conceived of as consequence relations. In both cases we have in mind, for definiteness, the implication-negation fragments of the logics in question (the undefinability of the remaining familiar connectives in the intuitionistic case notwithstanding). For the discussion of Béziau’s example, which concerns specifically the classical case (though there is an intuitionistic analogue, as we shall see in Section 3) we shall need to note that what we will with some *ad hoc* terminology baptize as Weak Classical Logic (WCL, $\vdash_{WCL}$), which can be characterized in either of the following ways: (*i*) $\vdash_{WCL}$ is the consequence relation on the language with connectives $\to$, $\neg$, determined by the class of all valuations satisfying the booleanness condition for $\to$, i.e., (2) above, and one of the two implicational conditions combined into the $\neg$-booleanness condition (1), namely (13)

$$v(\neg A) = T \text{ only if } v(A) = F \tag{13}$$

and (*ii*) $\vdash_{WCL}$ is the least consequence relation $\vdash$ on the present language satisfying (10), (11), and (5) with B taken as $\neg$A:

$$A, \neg A \vdash C \quad \text{for all formulas C.} \tag{14}$$

The formula logic WCL is then variously describable as (i) the set of formulas derivable by *Modus Ponens* from (A1)–(A4), or as (*ii*) the set of formulas A for which $\varnothing \vdash_{WCL} A$. From these descriptions it is evident that Weak Classical Logic as we are temporarily calling it, is indeed weaker than classical logic, i.e., that we have

$$\mathrm{WCL} \subsetneqq \mathrm{CL} \qquad \text{and} \qquad \vdash_{WCL} \subsetneqq \vdash_{CL} \tag{15}$$

with a similar point holding in the case of the corresponding generalized consequence relations, which we have not bothered to define, as well as in the intuitionistic case, for which case we mean by Weak Intuitionistic Logic the following: $\vdash_{WIL}$ is the least consequence relation $\vdash$ on the present language satisfying (10) and (14), while WIL is $\{A \mid \varnothing \vdash_{\mathrm{WIL}} A\}$, which is also, we observe, axiomatized by *Modus Ponens* with (A1), (A2), and (A4). For this case again, the "weak" terminology is justified by comparison of deductive strength:

$$\mathrm{WIL} \subsetneqq \mathrm{IL} \qquad \text{and} \qquad \vdash_{\mathrm{WIL}} \subsetneqq \vdash_{\mathrm{IL}} \tag{16}$$

Finally, we remark that the consequence relations $\vdash_{\mathrm{WCL}}$ and $\vdash_{\mathrm{WIL}}$ can be recovered from WCL and WIL by means of the definition of the existence of a deduction of formula (by successive appeals to *Modus Ponens*) from a set of formulas in the same way that the usual presentation of the Deduction Theorem for CL and IL makes clear.

2. Béziau's Paradox

Béziau's starting point (re-expressed in our terminology notation) is (15) above. What we are calling weak classical logic is a proper sublogic of classical logic (in $\neg$ and $\to$). Now suppose that we define a new 1-ary connective, ~, say, by

(Def.1~) $\qquad {\sim}A = A \to \neg A$

In weak classical logic, recall, $\neg A$ is a contrary of A, but not (in general) a subcontrary of A; with (Def.1~) in force, we still have A and ${\sim}A$ contraries, since A and $A \to \neg A$, *alias* ${\sim}A$, together would give $\neg A$, already a contrary of A. Thus the contrariety of A and $A \to \neg A$ essentially piggybacks on the contrariety of A and $\neg A$, given the strength of the logic governing $\to$. (Only intuitionistically acceptable properties of $\to$ are used here, a fact to which we return in the following section.) But now we have more: since any conditional and its antecedent are subcontraries in weak classical logic (which does not weaken the powers of $\to$, only those of $\neg$), this holds in particular when the consequent is $\neg A$, so A and $A \to \neg A$, *alias* ${\sim}A$, are not only contraries but also subcontraries. Note that here, by contrast, we are exploiting a feature of $\to$ which is specific to classical logic – cf. (12) above. (The interested reader is invited to formalize this either as a proof using one of the axiomatizations of WCL offered in Section 1, or using one of the syntactic characterizations of $\vdash_{\mathrm{WCL}}$.)

But wait a minute! Weak classical logic is supposed to be a *weakening* of the implication-negation fragment of classical logic, and yet it turns out that here we have, albeit in $\to$ and ~ rather than in $\to$ and $\neg$, the whole of this fragment, with a further connective, the somewhat unusual $\neg$, lying around as well. If we take the latter into account, we seem to have an extension, rather than a subsystem, of classical logic on our hands; if we leave it out of account, then we still seem to have a contradiction with our earlier characterization of weak classical logic as *strictly weaker* than classical logic. This anomaly is what we are calling Béziau's paradox. (Béziau refers to it as the *paradox of translation*: we shall see how translation enters the story presently.) Note that while classical logic is determined by the class of $\to$-boolean valuations satisfying both the "if" and the "only if" halves of the $\neg$-booleanness condition (1) from Section 1, weak classical logic is determined by the class of all $\to$-boolean valuations satisfying the "only if" half of the latter condition (= (13) above). We will call the former the class of *boolean* valuations and the latter the class of *weak boolean* valuations. (Thus the $\neg$ in the latter case is a contrary-forming but not necessarily a subcontrary-forming operator.) This is the reason for saying, as the title of Béziau (1999) puts it, that "classical negation can be expressed by one of its halves."[2] (Béziau's discussion is semantically oriented, but we could equally well put the point syntactically: we retain the relevant condition (5), to the effect that $A, \neg A \vdash C$, while dropping the dual condition (6), to the effect that $\Gamma, A \vdash C$ and $\Gamma, \neg A \vdash C$ together imply $\Gamma \vdash C$. Alternatively, on the axiomatic approach, we retain (A4) and drop (A5). This is the reason the contrarizing and subcontrarizing effects of $\neg$ were – albeit imperfectly – separated out in the discussion of Section 1.) One may well at this point consider the possibility of "expressing classical negation" by means of its *other* half, and starting with a subcontrary-forming operator in terms of which to define full classical negation. This issue is treated in the Humberstone (in preparation *b*), Section 4.

There is a complication for they way we have set matters up with (Def.1~) caused by the fact that there are two distinct views of what such a definition should be regarded as doing, distinguished in somewhat loaded terms by Meyer (1974) as the 'dishonest' and the 'honest' views of definitions. On the former, which we prefer more neutrally to call the 'metalinguistic abbreviation' view, with a definition like (Def.1~) we are announcing our intention to refer, in the metalanguage, to formulas of the form $A \to \neg A$, by using the abbreviation "~A".[3] The object language itself still only has two connectives, namely $\to$ and $\neg$. On the latter 'new symbol' view, the effect of a definition is to add to the object language a new connective, namely "~" together with a stipulation that for any formula A, ~A and $A \to \neg A$ are to be inter-replaceable in all theorems (for the 'set of formulas' conception of a logic) or in all statements of consequence (for the consequence relation conception of a logic), in the logic which results from a given logic by adding the definition. We can sidestep this issue for our discussion, and follow Béziau (1999) in considering a translational embedding of one consequence relation (to fix for convenience on this conception of logics) in another. We are concerned here, in particular, with a *faithful* embedding, which means that not only the "$\Rightarrow$" but also the "$\Leftarrow$" half of (17) holds, (17) giving what it takes for a mapping τ from the (set of formulas of) the language of a consequence relation $\vdash_1$ to the language of another such relation $\vdash_2$, to embed $\vdash_1$ faithfully into $\vdash_2$:

$$A_1,\ldots,A_n \vdash_1 B \Leftrightarrow \tau(A_1),\ldots, \tau(A_n) \vdash_2 \tau(B) \tag{17}$$

(17) is to be understood as requiring that we have the equivalence in question for all formulas $A_1,\ldots.A_n$, B, of the language of $\vdash_1$ (and in general there is no need to restrict attention to the case of finitely many formulas on the left, though nothing is lost in the present instance by doing so). A τ for which (17) holds when $n = 0$ is an embedding of one logic in the 'set of formulas' sense into another, with $\vdash_1$ and $\vdash_2$ indicating provability of the formulas to their right in the logics in question. A list of references discussing translational embeddings may be found on p.441 of Humberstone (2000), which should be supplemented by a reference to da Silva *et al.* (1999);

---

[2] Classical logic is called *K* by Béziau, and weak classical logic accordingly called *K*/2.

[3] Strictly speaking the quotation marks here should be quasi-quotes.

faithful embeddings are referred to as *conservative* in the latter work (and as 'unprovability-preserving' in Inoué (1990)).

As we have set things up, the language of the two consequence relations – $\vdash_{\mathrm{CL}}$ and $\vdash_{\mathrm{WCL}}$ in our case – may be taken as being the same, with connectives $\neg$, $\to$. (This departs from the way matters are presented in Béziau 1999.) Many translations from one language to another embed a consequence relation on the one into a consequence relation on the other, and the special features of the τ with which we are concerned here are summed up by describing it as a *definitional* translation, as in Wójcicki (1988), p.70. The two features in question are (*i*) that $\tau(p_i) = p_i$ for each sentence letter p_i, and (*ii*) that for each primitive *n*-ary connective # of the language constituting the domain of the function τ there is an *n*-variable formula $\#^{\tau}(p_1,\ldots,p_n)$ of the codomain of τ, for which τ satisfies: $\tau(\#(\mathrm{A}_1,\ldots,\mathrm{A}_n)) = \#^{\tau}(\tau(\mathrm{A}_1),\ldots,\tau(\mathrm{A}_n))$. Translations τ satisfying (*i*), resp.(*ii*), we shall call *variable-fixed*, resp. *compositional*, for the sake of some comparative remarks made below for translations not necessarily satisfying both conditions. Béziau's discussion of translations suggests a concern with definitional translations because of the role stressed for definitions – the connection of these with definitionality of a translation emerges in the following paragraph – such as (Def.1~), though he does not explicitly impose the variable-fixedness condition. We interpret the reference to "respecting the connectives" in his dismissal (Béziau (1999), p.149) of "translations between logics (as) functions just respecting the consequence relation and not the connectives", as a reference to what we are calling compositionality.

The formulas $\#^{\tau}(p_1,\ldots,p_n)$ figuring in (17) can be regarded as furnishing definitions for # as a new connective (on the 'new symbol' view of definitions) provided that $\#^{\tau}(p_1,\ldots,p_n)$ is not itself a formula constructed with the aid of #, which accounts for the terminology of 'definitional' translation. To satisfy that proviso in the case of the particular translation with which we are concerned, namely (18), the notation of the third clause should be amended to avoid this double usage, as of course we have done above by defining ~A (rather than $\neg$A) as A $\to$ $\neg$A as in (Def.1~):

$$\tau(p_i) = p_i; \qquad \tau(\mathrm{A} \to \mathrm{B}) = \tau(\mathrm{A}) \to \tau(\mathrm{B}); \qquad \tau(\neg\mathrm{A}) = \tau(\mathrm{A}) \to \neg\tau(\mathrm{A}). \qquad (18)$$

The key fact about this translation is given in Lemma 2.1, which allows us to prove the result on which Béziau hangs his discussion, given here as Theorem 2.2 and, in a different notation and prefaced with the words "it can be proved that", on p.148 of Béziau (1999).

LEMMA 2.1 *Let* v_0 *be an assignment of truth-values to the sentence letters,* v *be the unique boolean valuation extending* v_0*, and* v' *be any weak boolean valuation extending* v_0*. Then with* τ *as defined in (18) we have, for any formula* A, $v(\mathrm{A}) = v'(\tau(\mathrm{A}))$.

Proof. By induction on the construction of A. The only case of interest is the inductive case presented by $\neg$. When we use "$\neg$" or "$\to$" not as part of a formula we intend to refer to the truth-function associated with these connectives on boolean valuations. For the case in which A is $\neg$B, then, first suppose $v(\mathrm{A}) = \mathrm{T}$. Thus $v(\mathrm{B}) = \mathrm{F}$, and we must show that $v'(\tau(\neg\mathrm{B})) = \mathrm{T}$. The inductive hypothesis gives us that $v'(\tau(\mathrm{B})) = v(\mathrm{B}) = \mathrm{F}$, since $\tau(\neg\mathrm{B}) = \tau(\mathrm{B}) \rightarrow \neg\tau(\mathrm{B})$,

$$\begin{aligned} v'(\tau(\neg\mathrm{B})) &= v'(\tau(\mathrm{B}) \rightarrow \neg\tau(\mathrm{B})) \\ &= v'(\tau(\mathrm{B})) \rightarrow v'(\neg\tau(\mathrm{B})) \\ &= \mathrm{F} \rightarrow v'(\neg\tau(\mathrm{B})) \\ &= \mathrm{T}, \text{ as required (since } \mathrm{F} \rightarrow x = \mathrm{T} \text{ for } x \in \{\mathrm{T}, \mathrm{F}\}). \end{aligned}$$

Suppose instead that $v(\mathrm{A}) = \mathrm{F}$, in which case $v(\mathrm{B}) = \mathrm{T}$. As before

$$\begin{aligned} v'(\tau(\neg\mathrm{B})) &= v'(\tau(\mathrm{B})) \rightarrow v'(\neg\tau(\mathrm{B})) \\ &= \mathrm{T} \rightarrow v'(\neg\tau(\mathrm{B})), \text{ by the inductive hypothesis, since } v(\mathrm{B}) = v'(\tau(\mathrm{B})), \\ &= v'(\neg\tau(\mathrm{B})) \text{ (since } \mathrm{T} \rightarrow x = x \text{ for } x \in \{\mathrm{T}, \mathrm{F}\}) \\ &= v'(\neg\tau(\mathrm{B})). \end{aligned}$$

As $v'(\tau(\mathrm{B})) = \mathrm{T}$ and v' is a weak boolean valuation, $v'(\neg\tau(\mathrm{B})) = \mathrm{F}$; thus $v'(\tau(\neg\mathrm{B})) = \mathrm{F}$ and again $v'(\tau(\mathrm{A})) = v(\mathrm{A})$, as claimed. ⊞

THEOREM 2.2 (Béziau) *For the* τ *defined by (18) we have, given any formulas* $\mathrm{A}_1,\ldots.\mathrm{A}_n$, B:

$$\mathrm{A}_1,\ldots,\mathrm{A}_n \vdash_{\mathrm{CL}} \mathrm{B} \Leftrightarrow \tau(\mathrm{A}_1),\ldots,\tau(\mathrm{A}_n) \vdash_{\mathrm{WCL}} \tau(\mathrm{B})$$

Proof. $\Rightarrow$: Suppose $\tau(\mathrm{A}_1),\ldots, \tau(\mathrm{A}_n) \nvdash_{\mathrm{WCL}} \tau(\mathrm{B})$, Thus there exists a weak boolean valuation v' verifying each of $\tau(\mathrm{A}_1),\ldots, \tau(\mathrm{A}_n)$ but falsifying $\tau(\mathrm{B})$. If v_0 is the restriction of v' to sentence letters and v is the unique boolean valuation extending v_0 then the conditions of Lemma 2.1 are satisfied by v_0, v and v', so we have $v(\mathrm{A}_i) = v'(\tau(\mathrm{A}_i)) = \mathrm{T}$ for $i = 1,\ldots,n$ and $v(\mathrm{B}) = v'(\tau(\mathrm{B})) = \mathrm{F}$, so $\mathrm{A}_1,\ldots,\mathrm{A}_n \nvdash_{\mathrm{CL}} \mathrm{B}$. A similar argument gives the $\Leftarrow$ direction. ⊞

Thus although weak classical logic is properly included in classical logic, we can translationally embed classical logic faithfully in its weaker cousin. Weak classical logic contains within it a perfect copy of classical logic. In this setting, whether or not we intend to baptize what correspond to the classical formulas $\neg\mathrm{A}$ as $\sim\mathrm{A}$ or instead to leave them in their $\mathrm{A} \rightarrow \neg\mathrm{A}$ form, Béziau's paradox is the apparent anomaly that within a weaker logic we should be able to obtain such a complete replica of the stronger logic. In his own words:

> What we call the translation paradox is, given two logics *L*1 and *L*2, the conjunction of the following facts:
>
> — *L*1 is strictly included (up to language-isomorphism) in *L*2
>
> — *L*2 is translatable into *L*1
>
> This recalls the Galilean paradox, i.e., the fact that the set of even numbers is strictly included in the set of natural numbers and that at the same time there is one-to-one correspondence between these two sets. One can say that this paradox was solved by the framework of set theory, which perfectly explains this difference.

> To solve the translation paradox, we must find a framework according to which inclusion and translatability between logics have intuitive interpretations which are not incompatible.[4]

A few explanatory remarks are in order. The reference to language isomorphism is to be understood thus: Béziau calls two languages isomorphic when there is a one-to-correspondence between their sets of primitive connectives which preserves the arity of the connectives. The first is that whereas explicitly Béziau alludes here only to the translatability of one logic into another, it is clear from his discussion[5] that the fact that we are dealing with a definitional translation plays a significant role in leading him to find the phenomenon illustrated perplexing.

There is another aspect of the situation which is also playing a role, though it is not brought out in the above statement, or in the strict inclusions mentioned at the end of our opening section ((15) and (16)). We can put this informally by saying that it is not just that the inclusion is strict, but that the locus of the strictness. of the conclusion is precisely the behaviour of the connective being translated: $\neg$. In the terminology of Humberstone (1986), given two consequence relations $\vdash_1$ and $\vdash_2$ with and connectives, $\#_1$ and $\#_2$ of the same arity from the languages of these respective consequence relations, with (this being here to simplify the definition) every other connective of the language of $\vdash_1$ being a connective of the language of $\vdash_2$), we say that $\#_1$ *as it behaves according to* $\vdash_1$ is a *subconnective* of $\#_2$ *as it behaves according to* $\vdash_2$ when for all formulas $A_1,\ldots,A_n,B$, if $A_1,\ldots,A_n \vdash_1 B$ then $A_1^*,\ldots,A_n^* \vdash_2 B^*$, in which for any formula A, A* is the result of replacing every occurrences of $\#_1$ in A by $\#_2$. The intention is, then, that one connective, viewed as having certain prescribed logical behaviour, is a subconnective of another when any principle satisfied by the former is satisfied by the latter. (For an account of this relation in terms of a precise explication of the notion of satisfying a principle, see Williamson 1998.) We insert the word "proper" to indicate that the converse implication does not hold. For example, intuitionistic negation (i.e., $\neg$ as it behaves according to $\vdash_{IL}$) is a subconnective of classical negation ($\neg$ as it behaves according to $\vdash_{CL}$). The case of current interest is that $\neg$ as it behaves according to $\vdash_{WCL}$ is a proper subconnective of $\neg$ as it behaves according to $\vdash_{CL}$. (Yet a further aspect of the paradoxicality of the situation from Béziau's perspective will emerge from a quotation we give in the following section.)

Finally, by way of commentary on the inset quotation above, there is the question of "solving" the translation paradox. Béziau has presented a nice example of an interesting phenomenon.[6] But is it sufficiently problematic to merit description as (*prima facie*) paradoxical? In Section 5, we will take some small steps to wind down the appearance of paradoxicality – without, to be sure, developing anything as extensive as Cantor's theory of cardinality, Béziau's suggested analogue for sorting out Galileo's paradox. Before getting to that, there are several further aspects of the

---

[4] In rendering this passage, from Béziau (1999), p.149*f*., we have edited the English slightly: the word "remembers" has been replaced by "recalls", and two occurrences of "into" (after "included") by "in".

[5] See lines 2–6 on p.149 of Béziau (1999).

[6] We should point out that several issues raised by Béziau are not even touched on in our discussion, such as the interesting question of whether or not there is a weakest logic into which classical logic can be translated (i.e., embedded via a faithful definitional translation). The discussion surrounding Figure 1 in Humberstone (in preparation *b*) bears on this question, but does not tell us whether WCL, for instance, can be further weakened and still play this role.

background to his main example that it will be helpful to get out into the open, of some interest in their own right, and airing them will be the business of Sections 3 and 4, as well as the supplementary discussion Humberstone (in preparation *b*).

3. Béziau's Example: Background and Discussion

Here we devote ourselves, first, to describing an alternative and more familiar variant of Béziau's example; secondly, to looking at the analogue of his example in the setting of intuitionistic rather than classical logic. Again, we shall see that the point is a familiar one to be found in various forms in the textbook literature on the subject. Thirdly, we will consider some other examples of definitional translations exhibiting the phenomenon Béziau draws attention to, mostly mentioned by Béziau himself. In the following section we will give a new example closer in certain respects to the actual example of Béziau (1999), reviewed above in Section 2.

Let us recall the use of an all-implying ('bottom' or *Falsum*) constant $\bot$ in some versions of the language of propositional logic (subject to the axiom schema $\bot \to A$, that is, in the 'logics as sets of formulas' tradition, or to the condition $\bot \vdash A$ in the consequence relation approach), with the negation of A then defined as $A \to \bot$. Given classical logic $\to$, this gives us a formula which is both contrary and subcontrary to A, as desired. In Weak Classical Logic, we can simulate this with the following alternative to (Def.1~) – recalling that here $\neg$ only forms a contrary and not in general a subcontrary of what it attaches to, and that we aim, as with (Def.1~), to define ~ with both of these properties:

(Def.2~) $\quad {\sim}A = A \to \neg(A \to A)$

Taking as read the above points about the $A \to \bot$ procedure, we need only check that for any A, $\neg(A \to A)$ is an all-implying formula. (Given conjunction, we could use $A \wedge \neg A$ in place of $\neg(A \to A)$: *cf*. the discussion of Gabbay's Exercise 15(*b*) below.) We do the working only for the axiomatic treatment, noting that if we put $A \to A$ for A in (A4), we may detach the consequent, which says exactly this about $\neg(A \to A)$. One may think to pick a particular formula whose $\neg$-negated self-implication would do duty as a *definiendum* for $\bot$, so we could define the latter as $\neg(p \to p)$, for example. This would raise certain complications there is no need to go into here,[7] with its introduction of an extraneous propositional variable – so, recalling that since we only need $\bot$ as the consequent of an implication, we may as well take the antecedent of that implication in place of *p*, giving (Def.2~). The fact that "negation can be expressed by one of its halves" is thus implicit in the practice of defining the negation of A as implying $\bot$. $\bot$ is a contrary of every formula, whereas our contrary-forming $\neg$ forms, from any given, A specifically a formula which is a contrary of A, but as we have just seen, the hypothesis that we have such 'local contraries' is not actually any weaker than the hypothesis that we have a single all-purpose 'contrary of everything' formula.[8] Interestingly, no one seems to have found this way of introducing classical negation

---

[7] The interested reader will find them discussed in Section 3 of Humberstone (1993).

[8] *Modulo* a background assumption satisfied in the cases we are considering, and best made explicit using the apparatus of consequence relations. Suppose that $\vdash$ provides 'local contraries' in the sense that for every formula A there is a formula B satisfying (5) from Section 1. Then, provided that there is at least one formula D such that $\vdash D$ (i.e. $\varnothing \vdash D$),

paradoxical in the way that Béziau suggests we should – pending a detailed resolution – find the (Def.1~) route. Perhaps this is because the equivalence of having a global and a local contrary forming operator, as explained in the preceding note, is not immediately obvious.[9] (This is not to say that any operator forming global contraries from provable formulas and hence participating in (Def.2~), will itself be a local contrary forming operator and hence be capable of participating in (Def.1~). We address this question in Section 5 of Humberstone (in preparation *b*).)

The "$A \to \bot$" style treatment of negation is perhaps more common in presentations of intuitionistic logic than in the case of classical logic, and we have promised to display some textbook variations on Béziau's theme – minus the suggestion of paradoxicality – in this area. We cite as our text Gabbay (1981), pp.126–7, of which present numerous exercises on negation(s). We warn that the notation roles played by "~" and "¬" are the reverse of those we have chosen for them to play. Exercise 16, consisting of several questions, asks us to consider enriching the language of positive logic (the $\{\wedge, \vee, \to\}$-fragment of intuitionistic logic) with a singulary connective ~ with

(*a*) $\sim p \to (p \to q)$

as an axiom governing the new connective, to be subjoined to an axiomatization of positive logic with *Modus Ponens* and Uniform Substitution as rules. (Representative instances of (A1) and (A2) from Section 1 above will do as axioms. We are using Gabbay's labelling "(*a*)", "(*b*)", etc. here, but not his name for the logic currently under consideration.) Let us call this logic GS (mnemonic for "Gabbay's System"). He then asks us to show (1) that if we add to GS

(*b*) $(p \to \sim p) \to \sim p$

then we obtain IL (intuitionistic logic, with negation ¬), if ~A is read as ¬A, and (2) that (*b*) is not provable in GS itself. We mention this exercise because it leads naturally to the next, which is of greater current interest.[10] Question (3) begins by considering a definition

(*c*) $\neg A = A \to \sim A$

---

if we denote some contrary of D, as promised by (5), by ¬D, it is easy to see that ¬D will do as a 'global' contrary, i.e., we can, for any formula A, choose this same formula as the B which (5) says together with A will imply every formula. Thus given the proviso just stated, the $\forall\exists$ condition that every formula should have a contrary, implies the corresponding $\exists\forall$ condition. Note also that a global contrary is the same thing as an all-implying formula, since if B is all-implying then (by 'Thinning') any A together with B will have every C as a consequence, and if B is a contrary of every formula, then it is a contrary of B itself, so {B, B}, *alias* {B}, has every formula as a consequence. (Thus both directions just established would fail for the axiomatic approaches using a substructural logic without axioms encoding thinning – essentially (A1) from Section 1, and, for the second direction, the principle of contraction, which is implicit in (A2).)

[9] In fact in the preceding note we had a syntactic difference in that the global contrary former is a 0-ary rather than a 1-ary connective. So it may be useful to consider a 1-ary connective $\bot_1$ with the following logical behaviour: for any formula A, the formula $\bot_1 A$ implies every formula. Now that we have matched up the arities, we can say that defining the classical negation ~A, of A, via (Def.1~), as $A \to \neg A$, or via the current variant of (Def.2~), as $A \to \bot_1 A$, do differ in the following respect: behaving as currently envisaged, the local contrary-former ¬ is, while the global contrary former $\bot_1$ is not, a subconnective of classical negation.

[10] We have already mentioned that (b), written with "¬", suffices alongside (a), or a permuted version thereof, for intuitionistic negation, in commenting on (A4) and (A6′) in Section 1. As for the unprovability of (*b*) in GS, just consider ~ as the $\bot_1$ of the preceding note.

and then asks us to show that $\vdash_{\mathrm{IL}}$ A implies $\vdash_{\mathrm{GS}}$ A. This is of course Béziau's (Def.1~), though with the symbols "$\neg$" and "~" interchanged, transplanted into the setting of intuitionistic rather than classical logic. It is surprising that Gabbay doesn't ask instead that we show: for all A in the language of IL, $\vdash_{\mathrm{IL}}$ A if and only if $\vdash_{\mathrm{GS}}$ A. In fact the preceding batch of questions Gabbay's book (Exercise 15, p.126) asked a similar question in this fuller form. So the constructive variant of Béziau's example is close to explicit here. These earlier questions concern positive logic extended by *strong negation* (the strong negation of A being written as –A); the detailed logical behaviour of this connective is immaterial except for one thing: the strong negation of a formula is a contrary of that formula – according to SN, as we shall call the logic of strong negation. Gabbay's Exercise 15(*b*) has us, for some fixed formula f of the language of SN, defining $\neg$A as $\mathrm{A} \to (f \wedge -f)$, prove that for all formulas B of the language of IL, showing that $\vdash_{\mathrm{IL}}$ B if and only if $\vdash_{\mathrm{SN}}$ B. Since $\vdash_{\mathrm{SN}} (p \wedge -p) \to (q \wedge -q)$ there is no need to have a fixed formula playing the role of f here, and we can equally well define $\neg$A as $\mathrm{A} \to (\mathrm{A} \wedge -\mathrm{A})$, *à la* (Def.2~) and thus distributing the $\to$, as $(\mathrm{A} \to \mathrm{A}) \wedge (\mathrm{A} \to -\mathrm{A})$, and so, finally, throwing away the redundant conjunct, as $(\mathrm{A} \to -\mathrm{A})$, as in (Def.1~): the form preferred by Béziau. The remaining questions from the later Exercise 16 provide good practice with the logical properties of intuitionistic negation but need not detain us here.[11] Our point has been: it is a familiar enough fact that given the contrariety schema (A4) governing – to revert to our own choice of notation $\neg$ – instead of strengthening the logic (with (A1) and (A2) as already in the background) to obtain the logic of intuitionistic negation, as by adding (A6) or (A6)′, one can instead define a new connective, ~, say, using either (Def.1~) or (Def.2~), and the defined connective, with no special axioms needed, already embodies in the resulting system, precisely the logical behaviour of intuitionistic negation. The easiest way to see this is that we have the contrariety schema (A4) for the same reasons as in the classical case, and that we also get the schema (A5)′, which suffices to complete the logic of intuitionistic negation, "for free" from the fact that intuitionistic implication satisfies the contraction schema $(\mathrm{A} \to (\mathrm{A} \to \mathrm{C})) \to (\mathrm{A} \to \mathrm{C})$, giving (A5)′, with ~ in place of $\neg$, whichever of (Def.1~), (Def.2~), is used (taking C respectively as $\neg\mathrm{A}$, $\neg(\mathrm{A} \to \mathrm{A})$). We could put this in terms of definitional translations, and follow the lead of (Def.2~) rather than (Def.1~) in the treatment of negation in (18), which would then identify $\tau(\neg\mathrm{A})$ as $\tau(\mathrm{A}) \to \neg(\tau(\mathrm{A}) \to \tau(\mathrm{A}))$. Using the terminology introduced in our introductory section, we can put matters by saying that either the original or this new translation, both of which embed $\vdash_{\mathrm{CL}}$ into $\vdash_{\mathrm{WCL}}$, also embed $\vdash_{\mathrm{IL}}$ into $\vdash_{\mathrm{WIL}}$.

We turn now to some other examples of the phenomenon Béziau illustrates (more simply) with his own example. Béziau (1999, p.146) mentions the discussion in Suszko (1975) of Lukasiewicz's three-valued logic and also (p.148) the case of intuitionistic logic, for which we quote also some preparatory remarks conveying a further aspect of what makes Béziau consider all these cases paradoxical. Recall from our note 2 that by K and $K/2$ Béziau means what we are calling $\vdash_{\mathrm{CL}}$ and $\vdash_{\mathrm{WCL}}$ respectively. The words "this result" refer to the result recorded in our discussion as Theorem 2.2, and as in our earlier quotation from Béziau, we have changed "into" to "in" at two points:

---

[11] We should warn that Gabbay's Question (4) asks us to show that $\neg p \to {\sim}p$ and ${\sim}p \to \neg p$ are respectively provable and unprovable in GS, but this is the wrong way round. The first implication is just a rewritten form of (*b*) above, a demonstration of whose GS-unprovability was already asked for as part of Question (1). The second implication unpacks by the definition of $\neg$ to ${\sim}p \to (p \to {\sim}p)$ and so is provable without recourse to any special features of ~.

> One can interpret this result as saying that *K* is translatable into *K*/2. This suggests that *K*/2 is at least as strong as *K*. We will see that *K*/2 cannot be translated in a similar way into *K*. In view of the strict inclusion of *K*/2 in *K* this seems paradoxical.
>
> A situation quite similar happens with intuitionistic logic.(…)
>
> Even if nowadays the various translations of classical logic into intuitionistic logic are well known, the fact that classical logic can be translated into intuitionistic logic, which is strictly included in it, is still a paradox because it is against intuition and has not yet been properly explained.

The ellipsis in this quotation deals with classical and intuitionistic arithmetic and so is not straightforwardly to the point in the present discussion. The opening paragraph of this passage strengthens the appearance of paradoxicality we paraphrased and quoted Béziau as giving voice to in Section 2: if it seems strange that Weak Classical Logic can be faithfully embedded in Classical Logic by a definitional translation, so that the deductively weaker logic contains within itself a perfect copy (with ~ playing the role of negation rather than $\neg$) of the stronger logic, and so in some perhaps other sense is at least as strong as the latter, then it must seem stranger still that in view of the non-existence of a translation in the reverse direction, the former is even stronger than the latter in whatever the sense in question is. (A brief exploration of what this sense of comparative strength might be will be found in Section 5.) If we had a definitional translation in this other direction, it would be a translation τ satisfying

$$A_1,\ldots,A_n \vdash_{WCL} B \Leftrightarrow \tau(A_1),\ldots,\tau(A_n) \vdash_{CL} \tau(B) \qquad (19)$$

and in particular there would have to be some formula $\neg^{\tau}(p)$ for which $\tau(\neg A)$ was $\neg^{\tau}(\tau(A))$, for all A, which would imply, since for $\vdash_{CL}$ there are to within equivalence only four pairwise non-equivalent formulas in a single propositional variable – p, say – namely p, $\neg p$, $p \to p$, and $\neg(p \to p)$, that $\neg^{\tau}(p)$ was classically equivalent to one of these.[12] Each candidate can be ruled out, as we content ourselves with indicating in the case of $\neg p$. Here we have the fact, for instance that $\neg\neg p \vdash_{CL} p$ to contend with, since (19) would then require that $\neg\neg p \vdash_{WCL} p$, which is not the case. (Note that we do not use subcontrariety directly, arguing that since $p \to q, \neg p \to q \vdash_{CL} p$, we should have to have – what is again not the case – $p \to q, \neg p \to q \vdash_{WCL} p$; this argument supposes that $\to$ is translated by itself, whereas we need grounds that are independent of any particular choice of $\to^{\tau}$.) By a similar argument, Wójcicki showed that even the negation-implication fragment of $\vdash_{CL}$ cannot be faithfully embedded by a definitional translation into $\vdash_{IL}$. (Calling $\vdash_{CL}$ and $\vdash_{IL}$ respectively *K* and *J*, this is recorded in Theorem 2.6.9 of Wójcicki (1988) as "*K* is not definable in *J*". Wójcicki first published this result in 1970. It can be strengthened to the claim which results from dropping the "faithfully", i.e., to the claim that there is no definitional translation τ satisfying even only the unidirectional:

$$A_1,\ldots,A_n \vdash_{CL} B \Rightarrow \tau(A_1),\ldots,\tau(A_n) \vdash_{IL} \tau(B). \qquad (20)$$

---

[12] In speaking of A and B as equivalent with respect to a consequence relation $\vdash$ we mean of course that A $\vdash$ B and B $\vdash$ A (sometimes written as "A $\dashv\vdash$ B" for brevity).

See Section 5 of Humberstone (2000) for details.) Thus we should give a very cautious reception to Béziau's reference to "the fact that classical logic can be translated into intuitionistic logic": he is perhaps thinking either of the formula logics CL and IL (in which case Gödel's device of translating the CL connectives into negation and conjunction can be used) or else discarding the restriction – which is, as already noted, not quite explicit in Béziau (1999) – to *definitional* translations (in which case various translations associated with the names of Glivenko and Kolmogorov come to mind, in which the compositionality condition is violated when double negations are inserted in front of whole formulas, rather than constructing their translations from the translations of their subformulas).[13] A definitional embedding is impossible in either direction for the present case, a fact which we may as well put explicitly on record.

THEOREM 3.1 *Neither* $\vdash_{\mathrm{IL}}$ *nor* $\vdash_{\mathrm{CL}}$ *can be faithfully embedded in the other by a faithful definitional translation.*

Proof. We have already mentioned Wójcicki's proof that $\vdash_{\mathrm{CL}}$ cannot be embedded in $\vdash_{\mathrm{IL}}$ (by a faithful definitional translation) so it remains only to observe that $\vdash_{\mathrm{IL}}$ cannot be embedded in $\vdash_{\mathrm{CL}}$ (by such a translation). This follows from the fact that while in $\vdash_{\mathrm{IL}}$, taken as the implication-negation fragment of the full intuitionistic consequence relation, there are to within equivalence, six non-equivalent formulas in one variable while for $\vdash_{\mathrm{CL}}$ there are four. (Of course the situation is even worse for 'full' $\vdash_{\mathrm{IL}}$.) ⊞

We should also recall that for the pure implicational fragment of $\vdash_{\mathrm{IL}}$, C. A. Meredith in the 1950's managed to come up with a definitional translation which embeds the implicational fragment of classical logic faithfully into this consequence relation, namely that given by (21):

$$\tau(p_i) = p_i; \qquad \tau(\mathrm{A} \to \mathrm{B}) = ((\tau(\mathrm{B}) \to \tau(\mathrm{A})) \to \tau(\mathrm{A}))) \to \tau(\mathrm{B}). \tag{21}$$

Meredith used the definition of a new connective which we shall write as "$\supset$", with A $\supset$ B being defined to be ((B $\to$ A) $\to$ A) $\to$ B, and pointed out that the logical behaviour of formulas constructed from sentence letters by means of $\supset$ is exactly that of classical implication.[14] As a useful way of thinking of this, we recall the definability in classical logic of A $\vee$ B by either of the CL-equivalent (A $\to$ B) $\to$ B, (B $\to$ A) $\to$ A, the latter two being equivalent in IL neither to each other nor to A $\vee$ B, the implicational forms being termed the intuitionistic pseudo-disjunction of A with B and of B with A, respectively, in Humberstone (2001). In both IL and CL we have, as equivalent to A $\to$ B the formula (A $\vee$ B) $\to$ B, and in IL this is equivalent again to the result of

---

[13] Our earlier reference to Gödel was intended to call to mind the translation of a classical formula into a $\neg$, $\wedge$ compound, with variables translated as themselves, which works for the formula logics CL and IL but (as is well known) not for the consequence relations in view of the failure of $\neg\neg p \vdash p$ for $\vdash = \vdash_{\mathrm{IL}}$. What is called the Gödel-Gentzen negative translation at p.57 and Troelstra and van Dalen (1988) translates p_i as $\neg\neg p_i$, and faithfully embeds $\vdash_{\mathrm{CL}}$ into $\vdash_{\mathrm{IL}}$ – though this is not a definitional translation since it is not variable-fixed.

[14] Several interesting embeddings of $\vdash_{\mathrm{CL}}$ and $\vdash_{\mathrm{IL}}$ into classical and intuitionistic linear logic may be found in Chapter 5 of Troelstra (1992), which, though faithful, do not quite qualify as definitional.

replacing the disjunction by the pseudo-disjunction of A with B. What Meredith does is to define (what we are writing as) A $\supset$ B by replacing this instead with the pseudo-disjunction of B with A. References to Meredith's work and some subsequent rediscoveries of the result may be found in Section 7 of Humberstone (2001) – to which we add here Meredith and Prior (1968), p.214 – a remote further rediscovery being discernible in Theorem 3.3 of Guzmán (1994), used here in the other direction: to derive Meredith's result. (Readers with no interest or background in the area of *BCK*-algebras may like to skip forward to (26) below.[15])

Guzmán's theorem concerns two varieties of *BCK*-algebras, the class of positive implicative *BCK*-algebras (also called Hilbert algebras) and – a proper subvariety of the latter – the class of implicative *BCK*-algebras. Guzmán refers to these classes as *pBCK* and *iBCK* respectively; they provide algebraic analogues of the implicational fragment of intuitionistic and classical logic (respectively). Though Guzmán uses a dual notation we translate into the format which makes these logical correspondences clearer, so that here $\rightarrow$ is the fundamental *BCK*-algebra operation. The only aspect of these correspondences we need is that a propositional formula A is intuitionistically (resp. classically) provable if and only if the identity $t_{\mathrm{A}} \approx 1$ holds in every algebra in *pBCK* (resp. *iBCK*), where t_{A} is the term corresponding to A (e.g., if A is the formula $p \rightarrow ((q \rightarrow p) \rightarrow q)$, then t_{A} is the term $x \rightarrow ((y \rightarrow x) \rightarrow y)$). Transposed into this notation, Guzmán's Theorem 3.3 becomes:

THEOREM 3.2 (Guzmán) *Suppose* $\mathbf{A} \in$ *pBCK, where* $\mathbf{A} = \langle A, \rightarrow, 1\rangle$. *Define a new binary operation* $\supset$ *on A by setting, for* $a, b \in A$, $a \supset b = ((b \rightarrow a) \rightarrow ((a \rightarrow b) \rightarrow b)) \rightarrow b$. *Then, where* $\mathbf{A}^* = \langle A, \supset, 1\rangle$, $\mathbf{A}^* \in$ *iBCK.*

A proof may be found in Guzmán (1994, p.403). The connection with Meredith's translation may not be immediately clear. For this, one needs to observe that the identity (22) is satisfied by all *pBCK*-algebras

$$((y \rightarrow x) \rightarrow ((x \rightarrow y) \rightarrow y)) \rightarrow y \approx ((y \rightarrow x) \rightarrow x) \rightarrow y \tag{22}$$

which is most easily seen for those with a propositional logic background by checking the intuitionistic equivalence of the corresponding implicational formulas (putting distinct propositional variables for the distinct individual variables in (22)[16]). Thus Guzmán could more simply have defined $a \supset b$ to be $((b \rightarrow a) \rightarrow a) \rightarrow b$, and we shall assume that $\mathbf{A}^*$ has been defined this way, with a matching transformation on the terms:

$$1^* = 1; \qquad x_i^* = x_i; \qquad (t \rightarrow u)^* = ((u^* \rightarrow t^*) \rightarrow t^*) \rightarrow u^* \tag{23}$$

---

[15] Those in search of such a background will find a textbook treatment in Meng and Jun (1994), with a richer discussion of the connections with logic than is given here to be found in Blok and Pigozzi (1989).

[16] Thus we obtain the formulas $((q \rightarrow p) \rightarrow ((p \rightarrow q) \rightarrow q)) \rightarrow q$ and $((q \rightarrow p) \rightarrow p) \rightarrow q$. We have already remarked that the antecedent of the latter is the pseudo-disjunction of q with p, and add here that the antecedent of the former is what is sometimes, after Church (1951), referred to as the 'Church disjunction' of p and q. (This derived connective is commutative in IL.) With q fixed as the consequent, these two antecedents (not IL-equivalent taken by themselves) give equivalent implications.

Note that there is a certain amount of use-mention laziness in our presentation here. For instance, in (23) “1” is a term, whereas in the statement of Theorem 3.2 it represents the fundamental nullary operation in an algebra (denoted by that term), where more explicitly it should be replaced by “$1^{\mathbf{A}}$” to indicate that we are dealing with the value of this term in the algebra **A**. The relationship between the two uses of “*” is recorded in (24):

For any terms t, u, and any $\mathbf{A} \in pBCK$, *we have* $\mathbf{A} \models t^* \approx u^* \Leftrightarrow \mathbf{A}^* \models t \approx u$. (24)

Then – reversing the direction of derivation implicit in our description of Guzmán’s theorem as a rediscovery of Meredith’s observation – we can derive Meredith’s result from Guzmán’s. Here $\vdash_{\mathrm{IL}\to}$ and $\vdash_{\mathrm{CL}\to}$ are the implicational fragments of $\vdash_{\mathrm{IL}}$ and $\vdash_{\mathrm{CL}}$; in the course of the proof we appeal also to the following fact:

For any formulas A, B, *we have* $((\mathrm{B} \to \mathrm{A}) \to \mathrm{A}) \to \mathrm{B} \vdash_{\mathrm{IL}} \mathrm{A} \to \mathrm{B}$ (25)

COROLLARY 3.3 (Meredith) *For the translation* τ *given by (21), we have, for all pure implicational formulas* $\mathrm{A}_1,\ldots,\mathrm{A}_n$, B: $\mathrm{A}_1,\ldots,\mathrm{A}_n \vdash_{\mathrm{CL}\to} \mathrm{B} \Leftrightarrow \tau(\mathrm{A}_1),\ldots,\tau(\mathrm{A}_n) \vdash_{\mathrm{IL}\to} \tau(\mathrm{B})$.

Proof. We first show this for the case in which $n = 0$. The $\Leftarrow$ direction is easy, for if $\vdash_{\mathrm{IL}\to} \tau(\mathrm{B})$, then (as IL $\subseteq$ CL) $\vdash_{\mathrm{CL}\to} \tau(\mathrm{B})$, so since B and $\tau(\mathrm{B})$ are classically equivalent, $\vdash_{\mathrm{CL}\to}$ B. For the $\Rightarrow$ direction, suppose $\nvdash_{\mathrm{IL}\to} \tau(\mathrm{B})$. Let t_{B} be the *BCK*-algebraic term corresponding to the formula B, in which case t_{B}^* is the term corresponding to the formula $\tau(\mathrm{B})$, so for some $\mathbf{A} \in pBCK$, we have $\mathbf{A} \not\models t_{\mathrm{B}}^* \approx 1$, and so by (25) and the fact that 1* is 1, for the *iBCK*-algebra **A*** promised by Theorem 3.2, we have $\mathbf{A}^* \not\models t_{\mathrm{B}} \approx 1$ and therefore $\nvdash_{\mathrm{CL}\to}$ B. To extend the result to $n > 1$ we can argue by induction (on n), with the inductive step being the passage to the conclusion that

$$\mathrm{A}_1,\ldots,\mathrm{A}_n, \mathrm{A}_{n+1} \vdash_{\mathrm{CL}\to} \mathrm{B} \Leftrightarrow \tau(\mathrm{A}_1),\ldots,\tau(\mathrm{A}_n), \tau(\mathrm{A}_{n+1}) \vdash_{\mathrm{IL}\to} \tau(\mathrm{B})$$

on the hypothesis that this holds for all B when there are n formulas on the left of the turnstiles. The $\Leftarrow$ direction does need this inductive hypothesis, the reasoning being the same as for the $\Leftarrow$ direction of the $n = 0$ case already given. For the $\Rightarrow$ direction, suppose that $\mathrm{A}_1,\ldots,\mathrm{A}_n,\mathrm{A}_{n+1} \vdash_{\mathrm{CL}\to}$ B; in that case (by (10) of Section 1) $\mathrm{A}_1,\ldots,\mathrm{A}_n \vdash_{\mathrm{CL}\to} \mathrm{A}_{n+1} \to \mathrm{B}$, so by the inductive hypothesis, $\tau(\mathrm{A}_1),\ldots,\tau(\mathrm{A}_n) \vdash_{\mathrm{IL}\to} \tau(\mathrm{A}_{n+1} \to \mathrm{B})$, where formula on the right here is (by (21))

$$((\tau(\mathrm{B}) \to \tau(\mathrm{A}_{n+1})) \to \tau(\mathrm{A}_{n+1}))) \to \tau(\mathrm{B}).$$

Therefore, by (25), we have $\tau(\mathrm{A}_1),\ldots,\tau(\mathrm{A}_n) \vdash_{\mathrm{IL}\to} \tau(\mathrm{A}_{n+1}) \to \tau(\mathrm{B})$, and hence (by (10) again, in the other direction), $\tau(\mathrm{A}_1),\ldots,\tau(\mathrm{A}_n), \tau(\mathrm{A}_{n+1}) \vdash_{\mathrm{IL}\to} \tau(\mathrm{B})$, as required. ⊞

A few remarks are in order. Note that this case exhibits Béziau’s phenomenon, in exhibiting a stronger logic (implicational classical logic) as duplicated in disguise within a weaker logic (implicational intuitionistic logic). As already noted, one could introduce the defined connective ⊃ into intuitionistic logic and see in the logical relations amongst ⊃-formulas exactly those holding amongst the implicational formulas of classical logic. (By a ⊃-formula here we mean what on the ‘new symbols’ view of definition would be a formula in which the only connective to appear is ⊃, and on the ‘metalinguistic abbreviation’ view would be a formula all of whose occurrences of $\rightarrow$ were limited to those (as here exhibited) in subformulas of the form $((C \rightarrow B) \rightarrow B) \rightarrow C$.) Somewhat surprisingly perhaps, given the occasional informal talk one hears of the intuitionistic conditional as being stronger than its classical namesake, (25), together with the failure of its converse, tells us that the direction of relative strength between $\rightarrow$ and $\supset$ is the other way around. (In fact in Section 5 we will suggest that such relative strength comparisons – which should not be confused with the subconnective relation – are of little significance in this context.) Finally, we add the reminder that there is no way of extending the embedding so as to bring negation within its ambit, in view of Theorem 3.1. In particular, the idea of simulating classical negation using an all-implying $\bot$ and the definition of $\sim A$ (say) as $A \supset \bot$ just makes $\sim A$ equivalent to (the familiar intuitionistic negation) $\neg A$, which by (25) gives the formula $\sim\sim p \supset p$ the IL-unprovable consequence $\neg\neg p \rightarrow p$. In describing $\bot$ as all-implying, we had in mind the provability of $\bot \rightarrow A$ for all A; then $A \supset \bot$ unpacks into $((\bot \rightarrow A) \rightarrow A) \rightarrow \bot$, the antecedent of whose antecedent, being provable, can be dropped, leaving $A \rightarrow \bot$ (*alias* $\neg A$). A thought which might occur to the reader is that the “implying” in “all-implying” should be understood instead in terms of the provability of $\bot \supset A$ for all A. But to postulate this turns out, on unpacking the “$\supset$”, to amount to a version of double negation elimination, $((A \rightarrow \bot) \rightarrow \bot) \rightarrow A$, which even in the absence of any further conditions on $\bot$, is enough to turn IL into CL.

Digression. The Wójcicki-style argument we gave above for the nonexistence of a (definitional) τ satisfying (19) was not the same argument as Béziau (1999) gives for this conclusion, and since his argument is interesting in its own right, we recapitulate it here. The following terminology derives from Makinson and Segerberg (see Segerberg (1980)). We call an *n*-ary connective # in the language of some consequence relation $\vdash$ *congruential* according to $\vdash$ if whenever $A_i \vdash B_i$ and conversely for each i $(1 \leq i \leq n)$, then we have $\#(A_1,\ldots,A_n) \vdash \#(B_1,\ldots,B_n)$ (and so also conversely), and we call $\vdash$ itself congruential when every connective in the language of $\vdash$ is congruential according to $\vdash$. (Thus a congruential consequence relation is one that supports an unrestricted replacement – or ‘substitutivity’ – principle for provably equivalent formulas.) One readily sees that no definitional translation can embed a non-congruential consequence relation faithfully into a congruential consequence relation. In the present instance, for example, there is no candidate to play the role of $\sim^{\tau}$, since while $\sim$ fails to be congruential according to $\vdash_{\mathrm{WCL}}$, $\vdash_{\mathrm{CL}}$ is congruential. We prefer the argument given above because it applies not only to $\vdash_{\mathrm{WCL}}$ but also, unlike this one, to the least congruential extension of $\vdash_{\mathrm{WCL}}$. Congruential logics with an operator forming contraries but not subcontraries occur naturally enough – an example is the impossibility operator in normal (or indeed arbitrary congruential) modal logics formalized with this operator and, say, implication, as primitives. Such logics make an appearance in Section 5 of Humberstone (in preparation *b*). ***End of Digression.***

In the case of one final logic alluded to by Béziau, namely Lukasiewicz’s three-valued logic (with $\rightarrow$ and $\neg$ as primitive), we do have a translational embedding with the definitional translation τ given by (26)

$$\tau(p_i) = p_i; \qquad \tau(A \rightarrow B) = \tau(A) \rightarrow (\tau(A) \rightarrow \tau(B)); \qquad \tau(\neg A) = \tau(A) \rightarrow \neg\tau(A). \tag{26}$$

and in this case, the embedding works for the consequence relations (i.e., satisfies (17)) and not just for the formula logics (i.e., (17) for $n = 0$) when $\vdash_1$ is the three-valued Lukasiewicz consequence relation (understood in terms of preserving the designated value) and $\vdash_2$ is $\vdash_{CL}$. (Here $\vdash_1$ and $\vdash_2$ are meant as in (17) – that is, as the source and target of the embedding. On the present point, *cf*. Church (1956), Exercise 19.8.) Although Béziau cites in this connection the paper Suszko (1975), Suszko himself expresses a reliance on the work of several other Polish logicians, including Wójcicki, and the detailed proof of the claim just made for the τ of (26) is most conveniently accessible in Wójcicki (1988) – see Theorem 1.8.6 there – which also provides some philosophical discussion roughly in the same vein as Béziau's own: at least the occurrence of the word "nevertheless" in the final sentence of this passage, which seems to acknowledge a certain unexpected contrast, even if it does not go as far as Béziau does in suggesting that we have on our hands a paradox in need of resolution; we have kept Wójcicki's notation intact, in particular with variously subscripted "$\mathcal{C}$" for consequence operations, and in particular $\mathcal{K}^{\vDash}$ and $\mathcal{L}_3^{\vDash}$ for the consequence operations corresponding to the consequence relations $\vdash_{CL}$, in the former case,[17] and to the consequence relation determined by Lukasiewicz's three-valued tables. Faithful embeddability by means of a definitional translation, Wójcicki simply calls the definability (of one consequence operation in another); $C_i \le C_j$ means, where $\vdash_i$ and $\vdash_j$ are the corresponding consequence relations, that $\vdash_i \subseteq \vdash_j$:

> From an intuitive standpoint, to say that C_1 is definable in C_2 (…) is to say that C_1 can be viewed as (implicitly perhaps) 'contained in' C_2 (…), and thus that the expressive power of C_1 (…) is not greater than that of C_2 (…) – whatever can be said in terms of the connectives of the former calculus (…) can also be said in terms of the latter.
>
> It is worth noticing that the expressive power of C_2 can be greater than that of C_1 even if C_2 is weaker than C_1, $C_2 \le C_1$. This, for instance, is the case of classical logic $\mathcal{K}^{\vDash}$ and its three-valued Lukasiewiczian counterpart $\mathcal{L}_3^{\vDash}$; $\mathcal{L}_3^{\vDash}$ is weaker than $\mathcal{K}^{\vDash}$, nevertheless $\mathcal{K}^{\vDash}$ is definable in $\mathcal{L}_3^{\vDash}$ (see 1.8.6). (Wójcicki 1988, p.67.)

The ellipses in this quotation are references to the 'logics as sets of formulas' conception. The word "calculus" in the first paragraph may occasion some surprise, as though it is particular proof systems that are at issue; in fact Wójcicki means by a *calculus* here simply the ordered pair of a formal language and a consequence operation on that language, and refers to such a pair by the same label as is used to denote the consequence operation concerned (from which the language can be recovered in any case).

In spite of the "nevertheless" in the closing sentence of the above passage, which may be there simply to alert the reader to a possible trap, we should consider whether Wójcicki here provides something of what Béziau wanted: at least the beginnings of a satisfying theoretical account (analogous to set theory for the case of Galileo's Paradox) for making what seemed paradoxical no longer appear in that light. If so, they are the merest beginnings, since the crucial notion of expressive power remains unexplicated. Notions going under this name appear in many logical

---

[17] In our earlier quotation from Wójcicki, preceding (20) above, "K" was used for this consequence operation; the significance of the notational change was not clear to this reader.

contexts. There is the absence of a formal rendering of this or that natural language construction, such led people in the 1970s to suggest that the expressive power of the language of conventional modal logic needed to be boosted by the addition of an operator corresponding to (some uses of) the word “actually”.[18] Such considerations can be hardened into model-theoretic proofs of the relative expressive weakness of the unaugmented language.[19] Compare also the modal definability of various classes of frames in the conventional language of modal language as opposed to that of tense logic with two operators interpreted with mutually converse accessibility relations; the latter language has greater expressive power in that more classes of frames become modally definable. At the level of truth-functional logic, we have the idea of approximating more closely to functional completeness as a similar measure of expressive power. But these ideas all require particular semantical settings, and it is not clear what the idea comes to in the absence of any such setting. Wójcicki is happy to write that “whatever can be said in terms of the connectives of the former calculus (…) can also be said in terms of the latter” but this remains merely metaphorical at the level of abstraction (arbitrary consequence operations) at which his discussion is located: there is nothing at all “said in terms of” this or that set of connectives. One way of cashing out the metaphor would be thus: “whatever logic can be embedded (via a faithful definitional translation) into the former can be embedded into the latter” – but this would then hardly count as any kind of explanation of Béziau’s phenomenon, or anything that might be expected to diminish any reaction to this phenomenon as *prima facie* paradoxical.

A more daring response aimed at disarming such a reaction might proceed along the following lines. It is a common – and by and large correct[20] – observation that the weaker a logic is in the sense of what is provable in it, the stronger it is in its ability to draw distinctions. That is, roughly speaking, *deductive* strength and (let’s say) *discriminatory* strength vary inversely. The daring response would then pick up on the “Nevertheless” in Wójcicki’s formulation, quoted above, with something along the following lines: “There is no ‘nevertheless’ about it – the increase in expressive power noted a propos of the Lukasiewicz logic over classical logic is just a manifestation of this general phenomenon that deductively weaker logics exhibit greater discriminatory strength, and so is only to be expected.” In Section 5, we shall see that this daring response fails.

This concludes our examination of the parallels in pairs of logics cited in Béziau (1999) and some close relatives as presenting analogues of the main example (as presented in our Section 2) of that paper. In Section 5 we shall meet a case of Béziau’s phenomenon in modal logic as the natural analogue of a corresponding phenomenon in the context of various first-order theories. Before that, we devote a section to a case of the phenomenon which is rather more similar in a certain respect to the central (negation) example of Béziau (1999) than are those which we have been reviewing lately.

---

[18] Hazen (1976), Crossley and Humberstone (1977).

[19] Hodes (1984).

[20] Various counterexamples to the unrestricted claim that strengthening a logic deductively weakens the distinctions it makes are collected in Humberstone (in preparation *a*).

4. An Example Closer to Béziau's Own

Condition (1) from Section 1 associates a truth-function with ¬ on every valuation satisfying the condition. Identifying n-ary functions with (certain) sets of ordered n-tuples in the usual manner, this function is the set {⟨T, F⟩, ⟨F, T⟩}, while condition (2) associates with → the function {⟨T, T, F⟩, ⟨T, F, F⟩, ⟨F, T, T⟩,⟨F, F, T⟩}. Let us call the elements of these sets *determinants* of the function in question. In the case of negation the we have the contrariety determinant ⟨T, F⟩ and the subcontrariety determinant ⟨F, T⟩, and Béziau's example proceeded by throwing away the second of these determinants and looking at (the logic determined by) the class of all valuations satisfying the first (for ¬). (We have promised to consider retaining the second while throwing out the first: see Humberstone (in preparation *b*), Section 4) Here a valuation v is understood as satisfying a determinant $\langle \xi_1,\ldots,\xi_n,\zeta\rangle$ (ξ_i, $\zeta \in$ {T,F}) for an n-ary connective # when for all formulas $A_1,\ldots,A_n$, if $v(A_i) = \xi_i$, for $i = 1,\ldots,n$, then $v(\#(A_1,\ldots,A_n)) = \zeta$. (In fact, in this terminology, the logic of interest as the target of Béziau's embedding, Weak Classical Logic, was determined by – or as we might less confusingly say here, was sound and complete with respect to – the class of weak boolean valuations: those satisfying ⟨T, F⟩ for ¬ and also satisfying all four determinants listed above for → – the latter being summarised in Section 1 with the label "→-boolean".) Béziau then found that we could define another 1-ary connective – we wrote it as "~" – in terms of the → and ¬ and that the logic of this connective was sound and complete (a phrase we prefer here to avoid a potentially double usage of the "determin-" vocabulary[21]) w.r.t. the class of →-boolean valuations satisfying for ~ not only the determinant ⟨T, F⟩ but also the determinant discarded in the case of ¬, namely ⟨F, T⟩. When we single out a class of valuations as those satisfying for some n-ary connective some but not all of the 2^n determinants of any candidate n-ary truth-function, we shall say that the connective in question is *partially determined* with respect to the class of valuations. To indicate such partial determination, we draw a dash, indicating a blank in what would otherwise be the tabular specification of a truth-function. This convention has us representing ¬ as it behaves in Weak Classical Logic by the 'partial' table on the left of Figure 1:

¬	A
F	T
–	F

A	∧′	B
T	T	T
T	F	F
F	F	T
F	–	F

Figure 1

---

[21] The current use of 'partially determined' may be found in Humberstone (1997), and is a descendant of the talk of type determination in Segerberg (1982).

It is to be emphasized that the blanks in such tables are not intended to indicate the absence of a truth-value (T or F) for the compound in question (by contrast with a similar notation in Blamey (1986)): we are concerned only with completely defined bivalent valuations here. It is rather that the class of valuations, *V*, let's say, such tables single out are those satisfying the determinants indicated, so that the blank means the absence of a determinant (satisfied by all valuations in the class of interest) for that combination of values for the components of the compound rather than the absence of a value for the compound. Nor should it be supposed that each valuation in *V* satisfies some determinant, the blank being there only because there is no one determinant satisfied by all of them alike. (That would make the connective in question, not truth-functional, but still extensional – "in sentence position" – relative to *V*, in the terminology of Humberstone (1986), (1997)[22].) There is simply nothing said for all A, to take the case of $\neg$ in Figure 1, about $v(\neg A)$ for $v \in V$ when $v(A) = F$. Before we get on to the other table in Figure 1, we pause to look at the mechanism whereby ~ emerges with both determinants even when defined in terms of the partially determined $\neg$. We recall that ~A is defined to be $A \to \neg A$:

A	$\to$	$\neg$A
T	F	F T
F	T	– F

Figure 2

Labouring the obvious perhaps, Figure 2 details the truth-value assignment to ~A over the class of valuations satisfying the contrariety determinant for $\neg$ (weak boolean valuations, as we called them in Section 2). In the second line of the table, we get the blank under $\neg$A telling us we can draw no conclusion about its truth-value on such valuations. But on this line A takes the value F and so (since we consider only $\to$-boolean valuations) we are still able to draw a conclusion about the truth-value of $A \to \neg A$, and this conclusion matches what we would have for the whole compound had we stipulated from the outset that valuations were to satisfy both of the determinants for negation.

We shall now use the second table from Figure 1 to see another case of 'partially determined connectives' can be used to define 'fully determined connectives'. We intend the "$\wedge'$" to be taken as a single symbol, and have chosen something reminiscent of the usual symbol for conjunction because the table shows us three of the four determinants for the truth-function associated with conjunction. But it should be noted that it equally depicts three of the four determinants for the truth-function associated with $\leftrightarrow$, on which more presently. (Does this mean that the table tells us we are interested in those valuations on which $\wedge'$ behaves like $\wedge$ together with those on which it

---

[22] Where the term "pseudo-truth-functional" is also used for this. On p.39 it is suggested that contrariety be studied as the hybrid (= greatest common subconnective) of the negation and the constant false connectives (in classical logic), which amounts to enforcing the extensionality condition we are here not imposing. (The word 'product' is there used for hybrids in this sense, which invites confusion with the product-matrix sense of 'product'.)

behaves like $\leftrightarrow$? Certainly not: that is the imputation of extensionality we were at pains to disavow. The class of valuations includes these any many more besides.) Let us suppose that we have the connective $\to$ in our language to start with, as well as $\wedge'$, and we are considering the logic determined by the class of all $\to$-boolean valuations satisfying for $\wedge'$ the three determinants of the second table in Figure 1. This is like the initial situation in Béziau's example in which we are considering the logic determined by the class of all $\to$-boolean valuations satisfying for $\neg$ one of the two familiar boolean determinants. As Béziau defined a connective for which the missing determinant was recovered, we shall define a connective, naturally enough to be written as “$\wedge$” for which the missing conjunction determinant (namely $\langle$F, F, F$\rangle$) is recovered. As in Béziau's case, we need a little help from $\to$:

(Def.$\wedge$) $A \wedge B = A \wedge' (A \to B)$

Because, on $\to$-boolean valuations, an implication and its antecedent are subcontraries the blank in the table on the right of Figure 1 never engages our attention in calculating the value of the $\wedge'$-compound on the right of (Def.$\wedge$). The problematic case for $\wedge'$, when both A and B are false, does not arise because we are now replacing B by $A \to B$, which is true in that case, and duly combines with the false A to produce the value F as required by the $\langle$F, F, F$\rangle$ for $\wedge$. So what we have here is similar to Béziau's example in the following respect. When we throw out some determinants for a truth-function and consider the valuations satisfying those that remain, we always obtain (in the resulting logic) a subconnective of a connective for which all determinants remain in force. So just as $\neg$ in Weak Classical Logic is a subconnective of $\neg$ in Classical Logic, so if we were to have written $\wedge$ rather than $\wedge'$ in the logic lately under consideration, it would have been a proper subconnective of its classical namesake. But in each case we can introduce by a definition a connective with exactly the same powers as the connective which was reduced to a partially determined version of its former self (alternatively: we can exhibit a definitional translation embedding the first logic into the second). Our exposition has been slightly different here, in that we used a different symbol ($\wedge'$) for the subconnective and then introduced the expected symbol ($\wedge$) by the definition, whereas in exposition of Béziau we opted for the reverse procedure, keeping “$\neg$” for the subconnective in the weaker logic and using a novel symbol (“~”) for that introduced by a definition effecting the desired classical rapprochement. But the structure of the example remains the same in the respects just summarized.

There is, however, a certain difference at a greater level of detail in that whereas (Def.1~) defines ~ with a *definiens* in which the partially determined $\neg$ appears within the scope of the fully determined $\to$, with (Def.$\wedge$) it is the other way around: the partially determined $\wedge'$ has the fully determined $\to$ within its scope. The result of this is that if a table such as that in Figure 2 is constructed for the right-hand side of (Def.$\wedge$), our gap-indicating “–” does not make an appearance, rather than making an appearance which is, as in Figure 2 itself, immaterial to the calculation of the resulting truth-value.

This last feature also applies in the case of another definition we could make using $\wedge'$ and $\to$. Since, as we have just seen, $\wedge$ can be defined with their aid, we could provide the usual definition of the biconditional, defining $A \leftrightarrow B$ to be the conjunction (with $\wedge$) of $A \to B$ and its converse. More economically, however, we could equally well define A $\leftrightarrow$ B to be the $\wedge'$-compound of A $\rightarrow$ B with B $\rightarrow$ A, since a conditional and its converse, no less than a conditional and its antecedent, are subcontraries on $\rightarrow$-boolean valuations. Thus we never face the “both components false” case in dealing with (A $\rightarrow$ B) $\wedge'$ (B $\rightarrow$ A), and so can use this as a simple alternative to the previous example of $\wedge'$ as a strict subconnective of $\wedge$, the latter nevertheless definable: $\wedge'$ as a strict subconnective of $\leftrightarrow$, the latter again definable. (Here we omit the “as it behaves according to such-and-such logic” part of the official habitat of talk of subconnectives, taking the relevant logics as clear from the context.)

We should not end this discussion without emphasizing that throwing out determinants from truth-functions is far from the only way of obtaining subconnectives of the connectives of classical logic. For example, if we have a binary connective, $\circ$, say, all of whose behaviour in a certain logic follows from its being commutative, then while determinants $\langle$T, T, T$\rangle$ and $\langle$F, F, F$\rangle$ remain from the truth-tables of the various commutative boolean connectives, we cannot represent a restriction of attention to those valuations v with $v(\mathrm{A} \circ \mathrm{B}) = v(\mathrm{B} \circ \mathrm{A})$ by a partial truth-table like those of Figure 1 but with gaps in the second and third rows, since this does not embody the condition that those gaps must be filled by the same value on any valuation in the class.[23]

5. How Puzzling Should We Find All This?

Our intention in this section is to reduce any feeling of paradoxicality about the Béziau phenomenon – the translatability of a stronger logic into a weaker logic – by pointing out that there is a similar phenomenon, one which is easily illustrated and well known, in the case of first order theories, with a stronger theory translatable into a weaker one. We discuss first order theories, i.e., classes of formulas of a first order language which are closed under the consequence relation of classical predicate logic, because their semantic treatment via traditional model theory is so transparent. It is always open to the reader who insists on finding the phenomenon alluded to problematic to say that this is the appropriate response in the case of theories no less than in the case of logics. We do not expect this reaction to be widely shared, however.

Suppose we have on our hands the first order theory – call it T_0 – of an arbitrary binary relation. Its non-logical vocabulary consists simply of the dyadic predicate symbol R, and it can be axiomatized, with classical predicate logic as the background logic, by taking the set of non-logical axioms as the empty set. We now become interested in studying a more restricted range of relations: those which are symmetric. There are two obvious ways in which we may extend T_0 to reflect this narrowing of interest. Most obviously, we may pass from T_0 to an *axiomatic* extension, T_1, say, by adding a new axiom, namely

$$\forall x \forall y(Rxy \rightarrow Ryx) \qquad (27)$$

---

[23] Cf. Rautenberg (1989), (1991) for detailed investigation of similar issues.

forcing the relation assigned to R to be symmetric in any model of T_1. Alternatively, we may instead consider a *definitional* extension of T_0, call it $T_0(S)$, by adding to the vocabulary of T_0 a new two-place predicate symbol S and with it a definition

$$\forall x \forall y(Sxy \leftrightarrow (Rxy \wedge Ryx)). \tag{28}$$

Two points before proceeding. First, the contrast between axiomatic and definitional extensions of a theory may be found objectionable, or rather, objectionably so described: the definition (28) is after all a new axiom. Perhaps so, but some way of marking the contrast between a principle constraining existing vocabulary and one introducing new vocabulary is desirable, and this is how we mark the distinction here. Relatedly, Béziau (1999) uses (again *à propos* of the logics involved in his example) the phrase *definitional expansion* rather than *definitional extension*.[24] However, we prefer to reserve the term 'expansion' for a relation between structures interpreting the languages under consideration (whether in the case of logics of theories). The second point is that while in the case of logics there are the two approaches to definition which we distinguished in Section 2 as the 'metalinguistic abbreviation' and the 'new symbol' approach, in the context of definitions in first order theories only the analogue of the latter approach has much of a following, and it is this that we are assuming here. (Otherwise $T_0(S)$ would just be the same theory as T_0 and there would be no point in having both labels; we take this style of notation from Rantala (1991), incidentally.) The reason for this preference is that it makes for the most convenient semantic description of the situation, enabling us to track the presence and absence of the defined predicate symbol by considering expansions and reducts of models supplying or omitting a relation to serve as the interpretation of that symbol.

To return to (28): this definition makes S behave in $T_0(S)$ as R behaves in T_1, so we can study the symmetric relations without restricting the class of models. Instead, we exploit the fact that every model of T_0 has an expansion to a model of $T_0(S)$ in which S is symmetric, and moreover, every symmetric relation can be represented as the interpretation of S in such an expansion. While we should denote the interpretation of S in a model $\mathbf{M}$ by some such notation as $S^{\mathbf{M}}$, let us here be casual – as in the *BCK*-algebraic excursus in Section 3 (in particular *à propos* of (23) there) – and take the superscript as understood. Then the familiar point with (28) is that (*a*) whenever $S = R \cap R^{-1}$, S is symmetric, and moreover, (*b*) whenever S is symmetric we can represent S in the form $R \cap R^{-1}$ for some R (e.g., take $R = S$). Thus a relation is symmetric if and only if it is the intersection of some relation and its converse. (So (28) can be regarded as inducing a function from models of T_0 to their expansions – supplying an interpretation for S – satisfying (28), or again as inducing a function from the former models to the S-reducts of the latter, the latter then constituting precisely, to within isomorphism, the models of T_1.)

While certain aspects of the relation between logics and theories may be problematic, the parallel with examples like Béziau's is clear enough. In the case of the two ways of extending the theory of

---

[24] And – returning from theories to logics – Blok and Pigozzi (1989), p.5, define "extension" (in connection with consequence relations) in such a way as to presume that the language remains the same. ("Axiomatic extension" for them then means an extension, in that 'same-language' sense, from $\vdash_1$, to $\vdash_2$, say with some Δ – containing the axioms in question – such that for all Γ, A, we have $\Gamma \vdash_2$ A iff $\Gamma, \Delta \vdash_1$ A.)

an arbitrary binary relation just reviewed, we have the same effect whether we introduce a new symbol by a definition, or keep the language the same and strengthen the principles governing the existing vocabulary. In Béziau's example, and other like it reviewed in Sections 3 and 4, we get the same effect whether we extend our logic by adding suitably chosen defined vocabulary (boosting the basic contrariety logic – WCL – with the definition of ~A as $A \to \neg A$) or instead by strengthening the principles governing the existing vocabulary (adding a subcontrariety principle for $\neg$: here we are considering the original "classical $\to$" form of the example as in Béziau, rather than its constructive analogue, mentioned in Section 3). We can put the parallel in another way: in the one case we have a stronger theory which can be embedded via a translation into a weaker theory, while in the other a stronger logic which can be embedded via a translation into a weaker logic. We return to this way of describing matters below. For the moment, let us note that there is also the matter of comparative inferential strength within a theory or a logic. $\neg$A is stronger than ~A in the definitionally extended Béziau logic in that all instances of $\neg A \to {\sim}A$ are provable, while the same cannot be said for the converse schema. In the above case of the two theories, by contrast, what we have is the reverse: the defined expression is "weaker than" the original, in the universal closure of the implication $Sxy \to Rxy$ is provable in T_1, while this does not hold for the converse. This aspect of the situation is of no significance, however, since we could equally well have given a definitional extension of T_0 in which matters went the other way. We could just as well have replaced the "$\wedge$" in (28) by "$\vee$" since we have in this case the fact that a relation is symmetric if and only if it is the union of a relation with its converse. This time we end up with *Rxy* provably implying *Sxy* instead of the other way around.

We pause to note that while replacing the "$\wedge$" in (28) by any commutative connective[25] will guarantee that the *S* so defined is symmetric, not every such replacement has the additional feature we have seen in the conjunction and disjunction cases that an arbitrary symmetric relation can be regarded as having the form in question. For example, if we instead put "$\leftrightarrow$" for "$\wedge$" in (28) then the result is that *S* is not only symmetric but also reflexive. (In fact, one can show that a relation is a similarity relation – i.e., is reflexive and symmetric – if and only if it can be obtained from a relation *R* in this way.)

Now there was no need to venture outside of monadic predicate logic to make the above point. We could have considered the "null" theory of a 1-ary relation, with predicate letter *F*, say, and the axiomatic extension of this theory with the axiom $\neg\exists x(Fx)$, on the one hand, or the definitional extension of the theory with definition $\forall x(Gx \leftrightarrow (Fx \wedge \neg Fx))$, as two ways of passing from the original theory to the – not very interesting, to be sure – theory of an empty 1-ary relation (*alias* the theory of the empty set). Our reason for going *via* binary relations was their connection with modal logic *via* the accessibility relations of the Kripke semantics. We can forge such a link by considering how to turn the theory (with non-logical vocabulary *R*, as above) of an arbitrary binary relation – T_0 again – into the theory of a reflexive relation. There is the axiomatic extension route: we add the axiom (29) – a new candidate for the role of T_1; and there is the definitional extension

---

[25] Since we are considering first order theories here, the underlying propositional logic is classical and the reference to commutativity is accordingly to be understood relative to $\vdash_{CL}$.

route: we add instead the definition (30) to obtain a theory we again call $T_0(S)$. (The idea of reusing the same names for the theories is so that we can deploy them in a more schematic capacity below.)

$$\forall x(Rxx) \tag{29}$$

$$\forall x \forall y(Sxy \leftrightarrow (Rxy \vee x = y)). \tag{30}$$

Since a relation is reflexive if and only if it is the union of some relation with the identity relation, we have the same situation as before: A formula $\varphi(S)$ in which the only non-logical vocabulary to appear is S is provable in the definitional extension just in case the corresponding formula $\varphi(R)$ with R in place of S is provable in the axiomatic extension.

Corresponding, via the Kripke semantics for modal logic, to the contrast just reviewed between two ways of focussing our attention on reflexivity, are two ways of extending the basic normal modal logic **K**, determined by (= sound and complete with respect to) the class of all frames, to the logic determined by the class of all reflexive frames. The most familiar route is to make an axiomatic extension, adding the schema **T**, here appearing as (31), thereby obtaining the logic **KT**, while only slightly less familiarly[26] we may instead add to the language of **K** with its non-boolean primitive 1-ary connective $\Box$, a new such connective, $\Box^+$, say, by means of the definition (32):

$$\Box \mathrm{A} \to \mathrm{A} \tag{31}$$

$$\Box^+\mathrm{A} \leftrightarrow (\Box \mathrm{A} \wedge \mathrm{A}) \tag{32}$$

Then for any formula B, B is provable in **KT** if and only if the result of replacing all occurrences of $\Box$ by $\Box^+$ is provable in **K**: in other words, we have an embedding of **KT** into **K** via the translation τ with $\#^\tau = \#$ for boolean # and $= \Box^+$ for $\# = \Box$. This is well known and obvious from the Kripke semantics, since the effect of (32) is to assign as an accessibility relation to $\Box^+$ the reflexive closure of the accessibility relation assigned to $\Box$ (where the reflexive closure of a relation is just the union of that relation with the identity relation).[27] Pelletier (1984) in effect asks whether there is also a translation, which (like the in the τ just described) is the identity map for all connectives other than

---

[26] Such a definition is frequently encountered in modal provability logic, appearing, for example, on p.8 of Boolos (1993); it is clear from Boolos's wording that he takes the 'metalinguistic abbreviation' approach to definition here, whereas in suggesting the addition of (32) we are taking the 'new symbol' approach. Since we want the left and right hand sides of (32) to be, for any formula A, interchangeable in all contexts, we must understand (32) as an axiom-schema to whose instances the rule of necessitation (by $\Box$) is to apply.

[27] Thus one could recast the usual truth-definition and stipulate that $\Box$A is to be true at a point in a model just in case A is true there and also at all accessible points. The logic determined by the class of all frames with this definition of truth in force would be **KT** rather than **K**; see Fact 1.9 and surrounding discussion in van Benthem (1984). Van Benthem illustrates this idea of changing the clause for $\Box$ in the truth-definition in connection with symmetry and the considerations aired above *à propos* of (28) and its variant with disjunction in place of conjunction. We could have used that variant (though not the original (28)) to make the kind of point we have chosen instead to make using (30) if we had taken as our target logic for the embedding the minimal tense logic $\mathbf{K}_t$, whose two $\Box$-operators we write as A. N. Prior did, 'G' and 'H'. Exploiting the fact that precisely the symmetric relations are given by variant of (28), we proceed to embed the normal monomodal **KB** (determined by the class of symmetric frames) into $\mathbf{K}_t$ by a τ with $\tau(\Box \mathrm{A}) = \mathrm{G}\tau(\mathrm{A}) \wedge \mathrm{H}\tau(\mathrm{A})$. If we want to embed instead **KTB** (the original 'Brouwersche' system) then we must throw in $\tau(\mathrm{A})$ as a third conjunct.

$\Box$, in opposite direction, faithfully embedding **K** in **KT**.[28] A fortunate fact pertaining to this pair of logics enables us to return a negative answer to this question: the former logic is Halldén-incomplete while the latter is Halldén-complete.[29] This rules out such an embedding, for suppose we had one, τ, say, and for some A, B, with no sentence letters in common $\vdash_{\mathbf{K}}$ A $\vee$ B. Then $\vdash_{\mathbf{KT}} \tau(\mathrm{A} \vee \mathrm{B})$ so $\vdash_{\mathbf{KT}} \tau(\mathrm{A}) \vee \tau(\mathrm{B})$, since τ is the identity except on $\Box$, so by the Halldén-completeness of **KT**, either $\vdash_{\mathbf{KT}} \tau(\mathrm{A})$ or $\vdash_{\mathbf{KT}} \tau(\mathrm{B})$. (Here we use the fact that τ is a definitional translation, so C and τ(C) have the same sentence letters.) By the faithfulness of τ, we then get that $\vdash_{\mathbf{K}}$ A or $\vdash_{\mathbf{KT}}$ B – which would establish, contrary to fact, that **K** was Halldén-complete.

Before leaving modal logic, we pause to consider the question of whether **KT** can be translationally embedded into an arbitrary logic weaker than **K**. This would have to be the case if the idea floated at the end of Section 3 were correct, namely that the weaker a logic is, the more logics could be embedded in it because of the correlation between deductive weakness and discriminatory strength. Here we consider only (definitional) translations which are the identity on all connectives other than $\Box$, but this restriction is just imposed in the interests of expository simplicity. For parity with the above example, based on (32), we keep the language of the target logic for the embedding the same. Take as a candidate meeting this condition the smallest modal logic – call it **L** – where a modal logic is understood as a set of formulas $\supseteq$ CL and closed under *Modus Ponens* and Uniform Substitution. By a refined version of the kind of argument attributed to Wójcicki in Section 3, we can see that no faithful embedding via a definitional translation exists from **KT** to **L**. The refinement is called for by the fact that there are infinitely many non-equivalent formulas in a single propositional variable in **KT**, and the refinement called for is provided by the observation that for any fixed modal degree (= maximal extent of embedding of $\Box$ in the scope of $\Box$ within a given formula), there are only finitely many non-equivalent formulas of any given modal degree in a single variable in **KT**. So consider a hypothetical translation τ from **KT** to **L**, and in particular let n be the modal degree of $\tau(\Box p)$. Then there are only finitely many equivalence classes of formulas in p of degree n in **KT**, while there are infinitely many in **L** (recalling that, for instance, no two of $\Box p$, $\Box(p \wedge p)$, $\Box((p \wedge p) \wedge p)$,… are equivalent in **L**), so τ cannot after all be a faithful embedding from **KT** into **L**. Of course the same argument would work with **K** itself rather than **KT**, so we could put the objection to the above "the weaker the better" line of thought by saying that while **K** can be embedded into **K** (the identity translation), it cannot be embedded into

---

[28] See the middle of p.426 and also p.428 of Pelletier (1984). (Though it appeared too late to be taken into account for the present paper, we note for the reader's interest the sequel article Pelletier and Urquhart 2003.)

[29] Chapter 15 of Chagrov and Zakharyaschev (1997) reviews the relevant facts on Halldén-completeness. The present discussion shows that the second inset claim two thirds the way down p.428 of Pelletier (1984) is not correct, and so his example fails to illustrate (p.428, base) the fact that "mutual modeling is insufficient for translational equivalence". This is Pelletier's way of marking what for theories is usual called the distinction between mutual interpretability and definitional equivalence – *cf.* Corcoran (1980), (1981), (1983). Indeed (30) is perhaps most commonly encountered in showing the definitional equivalence of certain pairs of theories, such as that of an irreflexive relation and that of a reflexive relation, which are definitionally equivalent via the reflexivizing (30) and its irreflexivizing 'inverse' $\forall x \forall y(Rxy \leftrightarrow (Sxy \wedge \ x \neq y))$. A simpler pair of invertible definitions puts $\forall x \forall y(Rxy \leftrightarrow \neg Sxy)$ and similarly for S, but the definitions with '=' have greater currency, since they allow us to modulate between partial orders and strict partial orders (irreflexive transitive relations). An interesting treatment of the invertibility of definitions can be found in Williamson (1987). We are not particularly concerned with definitional equivalence for purposes of the present discussion, though the relation of equal strength, $\equiv$, introduced below, is more general, and the example just given shows that that the latter relation can obtain between consistent theories which are not jointly consistent.

L, even though **L** is (deductively) weaker than **K**. (A similar argument to this conclusion proceeds *via* Theorem 3.1: since according to that result $\vdash_{\mathrm{CL}}$ cannot be embedded (by a faithful definitional translation) in $\vdash_{\mathrm{IL}}$, but it can be embedded in $\vdash_{\mathrm{CL}}$, deductively weakening the target logic does not preserve the embeddability of the source logic.)

We will not concern ourselves further with definitional translations in modal logic. The point has been, aside from the digression on Pelletier's question, to illustrate the continuity of the translational embedding issue as it applies to sentential logics with its application in the case of first order theories, to which we now return. With the theories T_0, $T_0(S)$ and T_1 arising in our discussion of the (27)–(28) case and the (29)–(30) case above, and in any analogous case, there is also a fourth theory to consider, obtained by cutting back the definitionally extended theory to the language of the defined symbol, since it is this theory that looks like no more than a re-notated version of the T_1 in question. In more detail – in these cases, what we have are the following: T_0, its definitional extension $T_0(\mathrm{S})$, the restriction of the latter theory to its R-free part $T_0(S) \cap L(S)$, and T_1, an axiomatic extension of T_0. (Here $L(S)$ is the set of sentences constructed using only S as non-logical vocabulary.) Reasoning explicitly about these four theories will enable us to consider one way in which the claim that the Béziau phenomenon – insofar as it concerns theories, though the case of logics is parallel[30] – is indeed paradoxical. To do this, we need to make explicit some notion of the strength of theories according to which the phenomenon presents us with conflicting judgments of the relative strength of T_0 and T_1 in view of the simultaneous existence of T_1-theorems that are not T_0-provable (apparently making T_1 stronger) and of a translational embedding of the latter into a proper subtheory of the former (apparently making T_0 stronger).

To articulate these ideas, then, we need a notion of relative strength of theories, $\lneq$, where $T \lneq T'$ when T' is strictly stronger than T, presumed to be irreflexive and transitive, together with a relation of equal strength, here asserted to hold between T and T' by saying that $T \equiv T'$, where $\equiv$ is an equivalence relation which 'respects' $\lneq$ in the sense that for all theories T, T', and T'', if $T \equiv T'$ then $T \lneq T'' \Leftrightarrow T' \lneq T''$ and $T'' \lneq T \Leftrightarrow T'' \lneq T'$. We refer to the conditions imposed on our relations here as background postulates, and now introduce into the foreground, with explicit labelling, some postulates governing the way these comparative strength relations interact with the logical relations between theories construed as sets of sentences. The asterisk on the first label indicates that we are dealing with a provisional formulation.

(P1)* *If* $T \subsetneq T'$ *then* $T \lneq T'$

(P2) *If* T' *is a notational variant of* T, *then* $T \equiv T'$

(P3) *If* T' *is a definitional extension of* T, *then* $T \equiv T'$.

(P1)* is motivated by the consideration that a theory which is deductively weaker than another in the traditional sense of having its theorems be a proper subset of the latter's, then it should count as

---

[30] Indeed, though we have given it no prominence in our discussion, as well as the K and $K/2$ notation mentioned in note 2, Béziau (1999) also uses the labels $EK/2$ and $REK/2$ for the definitionally extended version of Weak Classical Logic (adding ~, that is, as we are notating matters) and for the linguistic restriction of the latter to $\neg$-free formulas. These two correspond in the present discussion to $T_0(S)$ and $T_0(S) \cap L(S)$, respectively.

having lower strength in the sense at issue in Béziau's Paradox. The phrase 'notational variant' in (P2) means that the theories involved can be mapped by a one-to-one correspondence to each other by a change of primitive (non-logical) vocabulary; the plausibility of (P2) is evident: how can such a change of notation affect the strength of the theory, in the sense pertinent to Béziau's Paradox?[31] (P3) reflects the traditional idea that extending a theory by adding a definition is not supposed to do anything but render what was already expressible more concisely so.[32]

Now we argue, provisionally:

$T_0 \lneqq T_1$, since $T_0 \subsetneq T_1$, by (P1)* (33)

$T_0 \equiv T_0(S)$, by (P3), therefore (34)

$T_0(S) \lneqq T_1$, from (33) and (34) (35)

$T_0(S) \cap L(S) \lneqq T_0(S)$, by (P1)* (36)

$T_0(S) \cap L(S) \lneqq T_0$, from (34) and (36) (37)

$T_0(S) \cap L(S) \lneqq T_1$, from (33) and (37); (38)

but

$T_0(S) \cap L(S) \equiv T_1$, by (P2): contradicting (38). (39)

(38) and (39) contradict each other because we can never have $T \lneqq T_1$ and also $T \equiv T_1$. This is clear from the intended significance of these relations: a theory cannot be strictly weaker than one to which it is equivalent in strength; it emerges from our background postulates because $\equiv$ is an equivalence relation which respects $\lneqq$, so if $T \equiv T_1$ and $T \lneqq T_1$, then $T_1 \lneqq T_1$, contradicting the irreflexivity of $\lneqq$.

The problem with the above reconstruction of the idea that the Béziau phenomenon, at least as it applies in the case of non-logical theories, is paradoxical, lies in its appeals to our provisional postulate (P1)*. For (P1)* and (P3) are already incompatible. Any definitional extension adds more sentences to the theory so whenever T' is a definitional extension of T, (P1)* tells us that $T \lneqq T'$ while (P3) tells us that $T \equiv T'$. Let us accordingly adjust (P1)* to avoid this conflict:

(P1) *If T and T' are theories in the same language and $T \subsetneq T'$ then $T \lneqq T'$*

There were two appeals to (P1)* in the derivation (33)–(39) above. The first, at (33), can be replaced by an appeal to the emended version (P1), since the languages involved are the same. But the second appeal, at (36), is aimed at the case in which the theories, $T_0(S) \cap L(S)$ and $T_0(S)$ are in different languages, since the latter has R in its vocabulary and the former does not.

---

[31] Sometimes notational variance is referred to as isomorphism of theories, but we prefer to reserve the latter terminology for a relation between structures interpreting theories (models, algebras,…); the phrase 'notational copy' is used for the analogous relation in the case of logics in Wójcicki (1988).

[32] More detail on this idea can be found in Rantala (1991).

Although this revision of our first postulate blocks the derivation (33)–(39), it may be that another paradoxicality argument can be constructed relying only on the revised version (together with (P2) and (P3)). Having promised not to theorise extensively about the question of paradoxicality of Béziau's phenomenon, however, we shall not enquire further into such possibilities here. We prefer to rest our case with the points already made: that the phenomenon is sufficiently transparent when first order theories are considered, as our examples with binary relations have perhaps shown, and the analogy between that case and the logical connectives is sufficiently close, that at least some of the puzzlement occasioned by Béziau's example should dissipate when it is compared with these homely examples from the arena of simple first-order theories.[33]

References

J. van Benthem, 'Correspondence Theory', pp.167–247 in D. M. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic, Vol. II: Extensions of Classical Logic*, Reidel, Dordrecht 1984.

J. van Benthem, *Modal Logic and Classical Logic*, Bibliopolis, Naples 1985.

J.-Y. Béziau, 'Classical Negation can be Expressed by One of its Halves', *Logic Journal of the IGPL* **7** (1999), 145–151.

S. R. Blamey, 'Partial Logic', pp.1–70 in D. M. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic III: Alternatives to Classical Logic*, Reidel, Dordrecht. 1986.

W. J. Blok and D. Pigozzi, 'Algebraizable Logics', *Memoirs of the American Math. Soc.* **77**, #396, 1989.

G. Boolos, *The Logic of Provability*, Cambridge University Press, Cambridge 1993.

A. V. Chagrov and M. Zakharyaschev, *Modal Logic*, Clarendon Press, Oxford 1997.

A. Church, 'The Weak Theory of Implication', pp.22–27 in A. Menne, A. Wilhelmy, and H. Angstl (eds.), *Kontrolliertes Denken, Untersuchungen zum Logikkalkül und zur Logik der Einzelwissenschaften*, Munich 1951.

A. Church, *Introduction to Mathematical Logic, Vol I*, Princeton University Press, Princeton NJ 1956.

J. Corcoran, 'On Definitional Equivalence and Related Topics', *History and Philosophy of Logic* **1** (1980), 231–234.

---

[33] Excerpts of this material were presented, under the title 'Béziau's Paradox', to the annual conference of the Australasian Association for Logic, in Canberra November 30 – December 2, 2002. I am grateful to those in the audience on that occasion for their comments, to Matthew Spinks for alerting me to the use of "extension" commented on in note 24, and to a *Theoria* referee for numerous detailed corrections. The original paper, as remarked in note 1, included also an appendix discussing aspects of negation bearing on Béziau's main example, now separated off as Humberstone (in preparation *b*).

J. Corcoran, ‘Definitional Equivalence ≠ Mutual Interpretability’, *Journal of Symbolic Logic* **46** (1981), 430.

J. Corcoran, ‘Mutual Interpretability, Definitional Equivalence, and Elongations’, (Abstract) *Journal of Symbolic Logic* **48** (1983), 516–517.

J. N. Crossley and I. L. Humberstone, ‘The Logic of “Actually”’, *Reports on Mathematical Logic* **8** (1977), 11–29.

D. M. Gabbay, *Semantical Investigations in Heyting’s Intuitionistic Logic*, Reidel, Dordrecht 1981.

F. Guzmán, ‘The Poset Structure of Positive Implicative *BCK*-Algebras’, *Algebra Universalis* **32** (1994), 398–406.

A. P. Hazen, ‘Expressive Completeness in Modal Language’, *Journal of Philosophical Logic* **5** (1976), 25–46.

H. Hodes, ‘Some Theorems on the Expressive Limitations of Modal Languages’, *Journal of Philosophical Logic* **13** (1984), 13–26.

I. L. Humberstone, ‘Extensionality in Sentence Position’, *Journal of Philosophical Logic* **15** (1986), 27–54; with correction *ibid*. **17** (1988), 221–223.

I. L. Humberstone, ‘Zero-Place Operations and Functional Completeness (and the Definition of New Connectives)’, *History and Philosophy of Logic* **14** (1993), 39–66.

I. L. Humberstone, ‘Two Types of Circularity’, *Philosophy and Phenomenological Research* **57** (1997), 249–280.

I. L. Humberstone, ‘Contra-Classical Logics’, *Australasian Journal of Philosophy* **78** (2000), 437–474.

I. L. Humberstone, ‘The Pleasures of Anticipation: Enriching Intuitionistic Logic’, *Journal of Philosophical Logic* **30** (2001), 395–438.

I. L. Humberstone, ‘Logical Discrimination’, in preparation *a*.

I. L. Humberstone, ‘Contrariety and Subcontrariety: The Anatomy of Negation (with Special Reference to an Example of J.-Y. Béziau)’, in preparation *b*.

I. L. Humberstone, ‘Modality’, forthcoming in Frank Jackson and Michael Smith (eds.), *The Oxford Handbook of Contemporary Analytic Philosophy*, Clarendon Press Oxford (to appear).

T. Inoué, ‘A Note on Unprovability-Preserving Sound Translations’, *Logique et Analyse* **33** (1990), 243–257.

E. J. Lemmon, *Beginning Logic*, Nelson, London 1965.

J. Meng and Y. B. Jun, *BCK-Algebras*, Kyoon Moon Sa Co., Seoul 1994.

C. A. Meredith and A. N. Prior, ‘Equational Logic’, *Notre Dame Journal of Formal Logic* **9** (1968), 212–226, *Corrigendum ibid*. **10** (1969), p.452.

R. K. Meyer, ‘Entailment is not Strict Implication’, *Australasian Journal of Philosophy* **52** (1974), 211–231.

F. J. Pelletier, ‘Six Problems in “Translational Equivalence”’, *Logique et Analyse* **27** (1984), 423–434.

F. J. Pelletier and A. Urquhart, ‘Synonymous Logics’, *Journal of Philosophical Logic* **32** (2003) 259–285.

V. Rantala, ‘Definitions and Definability’, pp.135–159 in J. H. Fetzer, D. Shatz, and G. Schlesinger (eds.), *Definitions and Definability: Philosophical Perspectives*, Kluwer, Dordrecht 1991.

W. Rautenberg, ‘Axiomatization of Semigroup Consequences’, *Arch. Math. Logic* **29** (1989), 111–123.

W. Rautenberg, ‘Common Logic of 2-Valued Semigroup Connectives’, *Zeitschr. für math. Logik und Grundlagen der Math.* **37** (1991), 187–192.

G. F. Schumm, ‘A Henkin-Style Completeness Proof for the Pure Implicational Calculus’, *Notre Dame Journal of Formal Logic* **16** (1975), 402–404.

K Segerberg, *Classical Propositional Operators*, Clarendon Press, Oxford 1982.

J. J. da Silva, I. M. L. D’Ottaviano and A. M. Sette, ‘Translations Between Logics’, pp.435–448 in X. Caicedo and C. H. Montenegro (eds.), *Models, Algebras, and Proofs*, Marcel Dekker, New York 1999.

T. J. Smiley, ‘Relative Necessity’, *Journal of Symbolic Logic* **28** (1963), 113–134.

R. Suzko, ‘Remarks on Lukasiewicz’s Three-Valued Logic’, *Bulletin of the Section of Logic* **4** (1975), 87–90.

A. S. Troelstra, *Lectures on Linear Logic*, CSLI Lecture Notes #29, Stanford University 1992.

A. S. Troelstra and D. van Dalen, *Constructivism in Mathematics: An Introduction, Vol. I*, North-Holland, Amsterdam 1988.

T. Williamson, ‘Invertible Definitions’, *Notre Dame Journal of Formal Logic* **28** (1987), 245–258.

T. Williamson, ‘Iterated Attitudes’, *Procs. British Academy* **93** (1998), 85–133.

R. Wójcicki, *Theory of Logical Calculi*, Kluwer, Dordrecht 1988.