

# Power Matrices and Dunn–Belnap Semantics: Reflections on a Remark of Graham Priest

LLOYD HUMBERSTONE

Monash University

## Abstract

The plurivalent logics considered in Graham Priest’s recent paper of that name can be thought of as logics determined by matrices (in the ‘logical matrix’ sense) whose underlying algebras are power algebras (a.k.a. complex algebras, or ‘globals’), where the power algebra of a given algebra has as elements *subsets* of the universe of the given algebra, and the power matrix of a given matrix has the power algebra of the latter’s algebra as its underlying algebra, with its designated elements being selected in a natural way on the basis of those of the given matrix. The present discussion stresses the continuity of Priest’s work on the question of which matrices determine consequence relations (for propositional logics) which remain unaffected on passage to the consequence relation determined by the power matrix of the given matrix with the corresponding (long-settled) question in equational logic as to which identities holding in an algebra continue to hold in its power algebra. Both questions are sensitive to a decision as to whether or not to include the empty set as an element of the power algebra, and our main focus will be on the contrast, when it is included, between the power matrix semantics (derived from the two-element Boolean matrix) and the four-valued Dunn–Belnap semantics for first-degree entailment (*à la* Anderson and Belnap) in terms of sets of classical values (subsets of  $\{T, F\}$ , that is), in which the empty set figures in a somewhat different way, as Priest had remarked his 1984 study, ‘Hyper-contradictions’, in which what we are calling the power matrix construction first appeared.

## 1 Priest’s Plurivalent Semantics

We begin by recalling that a *matrix* – more explicitly, a ‘logical matrix’ – for a propositional language consists of an algebra of the same similarity type (or ‘signature’) as the language, together with a subset of the universe of the algebra, whose elements are called the designated elements of the matrix, a formula  $\psi$  being counted as a consequence of a set  $\Gamma$  of formulas according

to the matrix – “ $\Gamma \models_{\mathbf{M}} \psi$ ”, where  $\mathbf{M} = (\mathbf{A}, D)$  – just in case for every  $\mathbf{M}$ -evaluation  $h$  for which  $\{h(\varphi) \mid \varphi \in \Gamma\} \subseteq D$ , we have  $h(\psi) \in D$ . Here an  $\mathbf{M}$ -evaluation is a homomorphism from the language to  $\mathbf{A}$ . In the opening sentence of this paragraph and in the sentence just preceding, we take the language itself to be an algebra, whose universe comprises the formulas of the language, freely generated by a countable set (of ‘propositional variables’ or ‘sentence letters’)  $p_1, p_2, \dots, p_n, \dots$ , and having the primitive connectives of the language as its fundamental operations. In particular, if  $p_1, p_2, p_3$  are the only propositional variables used in the construction of a formula  $\varphi$ ,  $h(\varphi)$  is fixed once  $h(p_1), h(p_2), h(p_3)$ , are given. We follow the convention that any matrix denoted by  $\mathbf{M}$  ( $\mathbf{M}'$  etc.) is of the form  $(\mathbf{A}, D)$  ( $(\mathbf{A}', D')$  etc.) where  $A$  is the universe of  $\mathbf{A}$ .  $\mathbf{M}$  is a *submatrix* of  $\mathbf{M}'$  if and only if  $\mathbf{A}$  is a subalgebra of  $\mathbf{A}'$  and  $D' = D \cap A'$ . We write  $\mathbf{M} \sqsubseteq \mathbf{M}'$  to mean that  $\mathbf{M}$  is a submatrix of  $\mathbf{M}'$ , and recall that whenever  $\mathbf{M} \sqsubseteq \mathbf{M}'$ , we have  $\models_{\mathbf{M}} \subseteq \models_{\mathbf{M}'}$ .

Priest [42] investigates what happens if we start with a matrix  $\mathbf{M}$  and go ‘plurivalent’ in the sense of replacing the single-valued  $\mathbf{M}$ -evaluations of the preceding paragraph (‘univalent matrix semantics’) with binary evaluating *relations* between formulas and elements of the algebra of  $\mathbf{M}$ , first with the background assumption that every formula stands in relation in question to at least one value (element of the universe of the algebra of  $\mathbf{M}$ , that is) – the ‘positive’ version of the semantics – and then without this background assumption, allowing evaluation relations which formulas bear to none of the values – the ‘non-negative’ version. Compositionality is effected by saying for any evaluation relation that a compound  $f(\varphi_1, \dots, \varphi_n)$  stands in the relation to  $b \in A$  just in case there are  $a_1, \dots, a_n \in A$  with  $\varphi_i$  standing in the relation to  $a_i$  ( $1 \leq i \leq n$ ) and  $f^{\mathbf{A}}(a_1, \dots, a_n) = b$ , where  $f^{\mathbf{A}}$  is the operation of  $\mathbf{A}$  corresponding to the  $n$ -ary connective  $f$ . We define  $\Gamma \models_{\mathbf{M}} \psi$  to hold when every evaluation relation between the set of formulas and  $A$  which relates each  $\varphi \in \Gamma$  to an element of  $D$ , not necessarily the same element for different formulas, also relates  $\psi$  to such an element.

As Priest recognises, we can streamline – though he might not be happy with that choice of verb – the above description by throwing away the evaluation *relations* in favour of traditional evaluation *functions* (the  $h$  of our opening paragraph), and working instead, univalently, in a matrix whose elements are sets of elements of the original matrix – arbitrary sets for the general non-negative case and non-empty sets for the positive case, with  $h(\varphi)$  being the set of elements to which on the description of the preceding paragraph, the formula is related. (Indeed, for the case in which the original matrix is the two-element classical one, this had been the approach taken from the outset in Priest [38].) Let us write  $\widehat{\mathbf{A}}$  for the algebra whose elements

are the subsets of  $A$  and whose fundamental operations are given by:

$$f^{\widehat{\mathbf{A}}}(A_1, \dots, A_n) = \{f^{\mathbf{A}}(a_1, \dots, a_n) \mid a_1 \in A_1, \dots, a_n \in A_n\} \quad (1.1)$$

Thus now for an evaluation  $h$ ,  $n$ -ary connective  $f$ , and formulas  $\varphi_1, \dots, \varphi_n$ , we have:

$$\begin{aligned} h(f(\varphi_1, \dots, \varphi_n)) &= f^{\widehat{\mathbf{A}}}(h(\varphi_1), \dots, h(\varphi_n)) \\ &= \{f^{\mathbf{A}}(a_1, \dots, a_n) \mid a_1 \in h(\varphi_1), \dots, a_n \in h(\varphi_n)\}. \end{aligned}$$

The structure  $\widehat{\mathbf{A}}$  here is called in the literature the *power algebra* of  $\mathbf{A}$ , though other names also have considerable currency ('complex algebra', 'global') and those who use any of this terminology are apt to differ over whether or not to include  $\emptyset$  in the universe of  $\widehat{\mathbf{A}}$ ,<sup>1</sup> which corresponds to the difference over whether to model the non-negative ( $\emptyset$  included) or the positive version ( $\emptyset$  excluded) of Priest's construction. When the algebra  $\mathbf{A}$  is clear from the context, we will usually just write  $f$  for  $f^{\mathbf{A}}$ , and  $\widehat{f}$  for  $f^{\widehat{\mathbf{A}}}$ . (This notation is from [21].) Note that in Priest [42] the positive and non-negative versions of the power algebra construction are referred to as *positive* and *general*, respectively.

But we want a matrix and not just an algebra: it remains to make a decision as to which elements of the power algebra are to be designated. The path followed mostly in Priest [38] and [42], takes an element of the power algebra, i.e., a subset of the universe of the original algebra, to be designated just in case at least one of its elements is designated in the matrix one started with. We can wrap all this up in the following definition: The *power matrix*  $\widehat{\mathbf{M}}$  of a matrix  $\mathbf{M} = (\mathbf{A}, D)$  is the matrix  $(\widehat{\mathbf{A}}, \widehat{D})$  where  $\widehat{\mathbf{A}}$  is, as above, the power algebra of  $\mathbf{A}$ , and  $\widehat{D} = \{A_0 \subseteq A \mid A_0 \cap D \neq \emptyset\}$ .<sup>2</sup> The notations " $\widehat{\mathbf{M}}$ ", " $\widehat{\mathbf{A}}$ ", remain ambiguous of course as between the non-negative and the positive versions of the power construction; generally we have the non-negative version in mind because we are especially interested in the fate of  $\emptyset$

<sup>1</sup>A further difference is that some authors – beginning with Gautam [21] – include additional set-theoretic operations ( $\cup$  etc.) among the fundamental operations of the power algebras while others do not. Clearly we are following the latter here, wanting to keep the signature of an algebra and its power algebra the same. The former path leads in the direction of the Boolean-algebras-with-operators of Jónsson and Tarski and thereby connects power algebras with the Kripke-style model theory for various intensional logics: see Brink [9] and Goldblatt [23].

<sup>2</sup>Thus a set is designated if it contains at least one designated element. An alternative, requiring all of the elements to be designated, is explored in the Appendix to Priest [42]; Section 3 of Priest [38] also explores a variation with anti-designated elements and validity as requiring forward preservation of designated status and backward preservation of anti-designation. Such variations are not considered here.

in this setting, in particular by contrast to its behaviour in the Dunn–Belnap semantics, an issue raised in the remark quoted from Priest in the following paragraph.<sup>3</sup>

Inspired by variations on the Liar Paradox, Priest [38] investigates what happens when we iterate the transition from matrix to power matrix starting from the three-element matrix he associates with the ‘Logic of Paradox’ (‘LP’), in which the values are thought of as representing true(-and-not-false), both-true-and-false, and false(-and-not-true), the first two of these being designated.<sup>4</sup> He also considers as a starting point for such an iteration, the four element matrix associated with Dunn and Belnap in which a further undesignated element is added to represent the neither-true-nor-false (or ‘truth-value gap’) case. He finds that the consequence relation determined is not affected by passage from a matrix to its power matrix in the cases of either progression – either starting from the three-element matrix or from the four-element matrix – and makes the following remark, that referred to in our title, about the consequence relation determined in the latter case:

This logic is not, what it might at first be thought to be, first degree entailment.<sup>5</sup> Rather the point-wise definitions give the extension of the LP functors according to the rule: gap-in, gap-out.<sup>6</sup>

If we start with the matrix based on the two-element Boolean algebra<sup>7</sup> with the top element designated, the three-element LP matrix arises by using the positive version of the power matrix construction and the four-element Dunn–Belnap matrix arises by application of the non-negative version of the construction. In what follows we will not be concerned, as [38] is, with further iterations of either process, but just with what happens in one application.<sup>8</sup>

---

<sup>3</sup>Though he sets things up slightly differently, essentially Priest [42] uses “ $\dot{\mathbf{M}}$ ” for “ $\widehat{\mathbf{M}}$ ” construed positively and “ $\check{\mathbf{M}}$ ” for “ $\widehat{\mathbf{M}}$ ” construed non-negatively.

<sup>4</sup>We are not concerned here with the philosophical motivation behind [38]; those wanting to see some debate on this can find it in Everett [20] and Priest [39].

<sup>5</sup>Here Priest appends a footnote mentioning the source of such an expectation as being an acquaintance with the four-valued semantics of Dunn [17].

<sup>6</sup>At this point there is another footnote: “Essentially as in Smiley [48].” This reference might be supplemented with Dale [15].

<sup>7</sup>For definiteness, take any reference to this algebra to have meet, join and complement as fundamental non-nullary operations. Sometimes, as indicated, we consider expansions with additional operations – like that corresponding to the biconditional – as primitive (for the sake of forming a power algebra while sticking to the letter of the definition given with (1.1) above).

<sup>8</sup>Priest finds that further iterations make no difference to the logic determined. Whether this is so for a similar iteration on the Dunn–Belnap front is a different matter: see Shramko and Wansing [47], which extends the lattice–bilattice relationship – an aspect of the Dunn–

The passage just quoted from [38] may suggest that differences between the Dunn–Belnap treatment and the power matrix treatment arise only when the ‘gap’ value, corresponding to  $\emptyset$  as a subset of  $\{T, F\}$  (or  $\{1, 0\}$ ) is taken into consideration – though this is claimed only for the “LP functors” in the above passage (which are defined at p. 238 of [38] to be  $\wedge$ ,  $\vee$  and  $\neg$ ,<sup>9</sup> as treated on the positive version of the power matrix semantics with the underlying matrix being the Boolean two-element matrix). When we consider arbitrary bivalent truth-functions, we shall see that the situation is not quite so straightforward, in particular because, as was emphasized in Dale [16], there is no unique Dunn–Belnap treatment for a given bivalent truth-function – whereas there is a unique power operation induced by a given term function on the two-element Boolean algebra; Dale’s observation accordingly sounds an *a priori* warning about the feasibility of obtaining an alignment between the two semantic approaches even if  $\emptyset$  is set aside.

**Digression.** Several attempts at obtaining such an alignment can be found in Brink [8] and [12] (see also [10]), though these make heavy use the ‘power relations’ alluded to in note 10, rather than validity as designation preservation in a power matrix, and when describing the four-element Dunn–Belnap matrix ([8], p. 80), Brink gives  $\{T\}$  as the only designated element, instead of  $\{T\}$  and  $\{T, F\}$  ( $\mathbf{t}$  and  $\mathbf{b}$  in the notation below); this determines a different consequence relation, on which see Pietz and Rivieccio [36]. The error, which may be caused by a confusion with the Smiley matrix, which includes a table

---

Belnap semantics we are not going into here. There is a discussion of Priest [38]’s power matrix semantics at p. 82 of [47], and Priest includes some remarks on [47] in note 10 of [42]. In the former discussion Shramko and Wansing cite an unpublished paper by Pragati Jain as explaining the “gap (=  $\emptyset$ ) in, gap out” feature of Priest’s treatment, in terms of the fact that this semantics (quoting the summary in [47]) “treats truth-functions in terms of the members of each argument, but  $\emptyset$  has no members.” With  $f$  binary by way of example, we defined  $\hat{f}(X, Y)$  to be  $\{f(x, y) \mid x \in X, y \in Y\}$ , and this notation itself of course abbreviates “ $\{z \mid \exists x \in X, \exists y \in Y . z = f(x, y)\}$ ”, so the real explanation even a single  $\emptyset$  argument yields  $\emptyset$  as value is the existential quantifiers in the defining condition (after the “|”), which make it a condition that cannot be satisfied if either  $X$  or  $Y$  is empty. (A similar definition but with the existential quantifiers negated would also be one “in terms of the members of each argument,” but would not have the effect in question.) In any case, as will be clear from the literature cited below, this behaviour of  $\emptyset$  is a familiar feature of what we are calling the non-negative version of the power algebra construction, one underlying the failure of linear equations which are not also regular – this terminology explained below – to be preserved on passage to the power algebra so conceived. The relevant behaviour is captured by describing  $\emptyset$  as an *absorbing* element in the power algebra (as defined in note 13 below).

<sup>9</sup>We use “ $\neg$ ” throughout even for discussing the principal and favoured negation (‘De Morgan negation’) of the Anderson–Belnap tradition, which often reserves “ $\neg$ ” for something else: see [9] and references.

for  $\rightarrow$  (see [1], p. 161*f.*, or [15] p. 323, both of which use ‘matrix’ in the older sense for each individual table) is incidental to Brink’s exposition, though, which mainly works with power algebras and power orderings, and does not actually isolate the concept of a power matrix as Priest does (though not using that terminology, in Priest’s case). There is some overlap with Priest’s discussion in [38], though, in particular over connecting the misalignment with the presence of  $\emptyset$ , saying for instance, after an initial stab ([8], p. 82), “the troublesome differences in detail are all due to the empty set”. The treatment in Brink [12] is different again, applying the power algebra construction to a ‘partial’ version of the Boolean algebra of sets of truth-value assignments; as with [8], a full exposition and discussion would take us too far afield here. *End of Digression.*

We return to the matter of Dale’s observation after reviewing the basic outlines of the Dunn–Belnap semantics in Section 2. But we close these introductory remarks by placing Priest’s “gap in, gap out” comment in the broader perspective of the (non-negative) power-algebraic construction. This is particularly appropriate because Priest does not actually present his plural-valent semantics using the terminology of power algebras and matrices based thereon, in either [38] or [42].

The first systematic study of power algebras was made in Gautam [21], where they are called complex algebras, and the non-negative version of – a minor variant (see note 1) of – the construction is employed.<sup>10</sup> Essentially the main result is that the identities of an algebra which are guaranteed to be identities of its power algebra, other than those of the form  $t \approx t$  (for some term  $t$ ) are precisely those in which every variable occurring occurs exactly once on each side – the so-called *regular linear* identities, where ‘linear’ means that in  $t \approx u$ , each of  $t, u$ , is a linear term, i.e., a term in which each variable to occur occurs only once, and calling  $t \approx u$  ‘regular’ means that the same variables occur in  $t$  and  $u$ .<sup>11</sup> Thus for example, the

---

<sup>10</sup>The subsequent literature on the topic has become voluminous. Frequently cited papers include Bleicher et al. [5], Shafaat [46] and Grätzer and Lakser [25]. There has also been extensive discussion of power algebras in specific varieties; for example, apropos of semigroups, relatively early instances include Tamura and Shafer [50] and Mogiljanskaja [33]. (The former paper has a semigroup-specific reference to work in French by P. Dubreil antedating Gautam [21].) Several authors have also considered extending the power construction to general structures in which relations as well as functions (operations) are present: see Grätzer and Whitney [26], Brink [11]. A useful overview is given in Bošnjak and Madarász [7].

<sup>11</sup>The notion of linearity of terms carries over naturally to formulas of propositional logic (if indeed one wants to distinguish these cases at all – something done here only by switching from  $x, y, z, \dots$  to  $p, q, r, \dots$ ) a linear formula then being one in which no propositional variable appears more than once. If classical propositional logic is developed in a

power algebra of a semigroup will be a semigroup because associativity is such an identity, while the power algebra – whether conceived positively or non-negatively – of an idempotent semigroup (alias ‘band’) is not guaranteed to be another such semigroup because the idempotent law ( $x \cdot x \approx x$ ) is not a linear identity. We may call this non-preservation of such identities in passing to power algebras the Gautam effect in view of the seminal role of [21], and illustrate it with the idempotent law just mentioned. that where  $*$  is an idempotent operation on the set  $A$ , and  $\widehat{*}$  the corresponding power operation, we always have  $X \subseteq X\widehat{*}X$  (here  $X \subseteq A$ ), since  $a \in X$  can be written as  $a * a$ , putting it into  $X\widehat{*}X$  but we need not have the converse inclusion since with  $a, b \in X$  we will get  $a * b$  in  $X\widehat{*}X$  though  $a * b$  need not have been an element of  $X$ .<sup>12</sup> We return to one aspect of this issue in the paragraph after next. Note that while some identities may be lost on passage to the power algebra, no new identities arise, because the power algebra – whether in its positive or in its non-negative incarnation – has a subalgebra isomorphic to the original, namely that whose universe comprises the unit sets of the original algebra’s elements. (This was noted in [21], and it is also exploited in Priest [38].)

Essentially, the difference between the non-negative and the positive power algebra constructions is that the latter yields algebras preserving all linear identities, and not just the regular ones. (Some pertinent literature is cited in note 24.) In the non-negative case, the presence of  $\emptyset$  as an element of the power algebra means that we have an absorbing (‘absorbent’ or ‘annihilating’) element which plays that role for all the operations,<sup>13</sup> so that in any equation  $t \approx u$  with some variable occurring only on one side or other means that taking  $\emptyset$  as the value of that variable and non-empty sets to the remaining variables will fail as an identity of the power algebra since whichever of  $t, u$  has the extraneous variable concerned will denote  $\emptyset$  while the other side will not. It is this absorptive behaviour on the part of  $\emptyset$  that

---

language in which there are no primitive connectives which are constant in the sense that they form equivalent compounds regardless of variation in the components (this condition being vacuously satisfied by any nullary connectives, though of course the definition covers the general  $n$ -ary case), then one has the simple observation that no linear formulas in this language are classical tautologies – mentioned, for example, as Corollary 9.27.4 in Humberstone [30].

<sup>12</sup>Here we use the abbreviative notation of  $f$  and  $\widehat{f}$  for what appear in (1.1) as  $f^{\mathbf{A}}$  and  $f^{\widehat{\mathbf{A}}}$ , except that of course we also write binary  $f$  as  $*$  and use infix rather than prefix notation.

<sup>13</sup>With a bias toward multiplicative terminology: an all-purpose zero-element; i.e., an element  $b \in A$ , to put the definition in terms of an arbitrary algebra  $\mathbf{A}$  (with universe  $A$ ) such that for each fundamental operation  $f^{\mathbf{A}}$  of  $\mathbf{A}$ , we have  $f^{\mathbf{A}}(a_1, \dots, a_n) = b$ , whenever  $b \in \{a_1, \dots, a_n\}$ , for all  $a_1, \dots, a_n \in A$  (where  $f^{\mathbf{A}}$  is  $n$ -ary).

Priest's 'gap in, gap out' remark refers to.

Returning to the example of idempotence, we observe that there is a reprieve from the Gautam effect as illustrated above – the failure of  $X \widehat{*} X$  to be included in  $X$  (in the power algebra of an idempotent groupoid) in the special case in which  $A$  (and hence each of its subsets  $X$ ), has no more than two elements. To say what is special about working with a set of no more than two elements – especially significant for present purposes because the set  $\{T, F\}$  (or, if preferred,  $\{1, 0\}$ ) of classical truth-values is such a set – we recall the concept of a *conservative* operation as an operation  $f$  on a set  $A$ , here taken as having arity  $n$ , for which  $f(a_1, \dots, a_n) \in \{a_1, \dots, a_n\}$  for all  $a_1, \dots, a_n \in A$ . The point is that while any conservative binary operation on a set is idempotent (take  $a_1 = a_2$ ), if  $\|A\| = 2$ , then any idempotent operation on  $A$  is conservative.<sup>14</sup> For we may reason: let  $f$  be such an operation and consider  $f(a, b)$  with a view to showing that this is either  $a$  or  $b$ . If  $a = b$ , then by idempotence  $f(a, b) = f(a, a) = a$ . If  $a \neq b$  then since  $\|A\| = 2$  by the pigeonhole principle,  $f(a, b)$  must either equal  $a$  or  $b$  (or else  $f(a, b)$ ,  $a$ , and  $b$ , would be three different elements. To press this observation into service in connection with the generally unavailable inclusion mentioned above,  $X \widehat{*} X \subseteq X$  for  $X \subseteq A$  when  $\|A\| = 2$  (returning to the  $*$  notation, still retaining the assumption of idempotence), the only case of interest is that of  $X = A = \{a, b\}$ , so  $X \widehat{*} X$  is  $\{a, b, a * b\}$ . But as just observed,  $*$  is conservative and so this set is included in  $X = \{a, b\}$  as claimed.

Now, not only the idempotent law but several other identities of special significance for propositional logic are non-linear, including the (lattice  $\wedge/\vee$ ) distribution and absorption laws, these last not even being regular, so the question arises as to their preservation on passing from the two-element Boolean algebra to its power algebra. We return to his question at the end of Section 4.

## 2 Dunn–Belnap Semantics

Let us recall the Dunn–Belnap semantics using assignments,  $v$ , not of *elements*, but of *subsets*, of the set  $\{T, F\}$  to all formulas (in the language with connectives  $\wedge, \vee, \neg$ ), subject to the three constraints:

$$T \in v(\varphi \wedge \psi) \text{ if and only if } T \in v(\varphi) \text{ and } T \in v(\psi); \quad [\wedge]_T$$

$$F \in v(\varphi \wedge \psi) \text{ if and only if } F \in v(\varphi) \text{ or } F \in v(\psi). \quad [\wedge]_F$$

$$T \in v(\varphi \vee \psi) \text{ if and only if } T \in v(\varphi) \text{ or } T \in v(\psi); \quad [\vee]_T$$

---

<sup>14</sup>This observation is made, and some of its implications pursued, in [31].



$$\begin{aligned}
F \in v(\varphi \vee \psi) \text{ if and only if } F \in v(\varphi) \text{ and } F \in v(\psi). & \quad [\vee]_F \\
T \in v(\neg\varphi) \text{ if and only if } F \in v(\varphi); & \quad [\neg]_T \\
F \in v(\neg\varphi) \text{ if and only if } T \in v(\varphi). & \quad [\neg]_F
\end{aligned}$$

This apparatus provides a semantics for the system of first-degree entailment  $\mathbf{E}_{\text{fde}}$  in that  $\varphi \rightarrow \psi$  is a correct first-degree entailment in the sense of pp. 150–162 from of Anderson and Belnap [1] just in case for every  $v$  as above, we have  $T \in v(\psi)$  whenever  $T \in v(\varphi)$ . (See Dunn [17].) More generally we shall call  $\psi$  an *FDE*-consequence of  $\varphi_1, \dots, \varphi_n$  – writing this as  $\varphi_1, \dots, \varphi_n \vdash_{\text{fde}} \psi$  when  $(\varphi_1 \wedge \dots \wedge \varphi_n) \rightarrow \psi$  is a correct first-degree entailment.

What we may call an *n-ary Dunn–Belnap condition* is a condition the form

$$\Phi = \Phi \left( \begin{array}{c} T \in X_1, \dots, T \in X_n \\ F \in X_1, \dots, F \in X_n \end{array} \right) \quad (2.1)$$

where the more explicit notation on the right indicates the  $2n$  distinct (metalinguistic) atomic components from which  $\Phi$  is constructed, taking at least one of from each vertically aligned pair, using conjunction or disjunction (in the metalanguage). The variables  $X_i$  range over subsets of  $\{T, F\}$ . The Dunn–Belnap condition  $\tilde{\Phi}$ , *complementary* to  $\Phi$ , is obtained from  $\Phi$  by replacing each such atomic component, interchanging the upper and lower conditions at each point – i.e., replacing “ $T \in X_i$ ” by “ $F \in X_i$ ” and conversely – and replacing conjunctions (disjunctions) used in the construction of  $\Phi$  with disjunctions (resp. conjunctions). Thus  $[\wedge]_T$  and  $[\wedge]_F$  above are complementary conditions, as are  $[\vee]_T$  and  $[\vee]_F$ , and  $[\neg]_T$  and  $[\neg]_F$ . Semantics in the spirit of the Dunn–Belnap treatment, which is to say, following the precedent set by the conditions  $[\#]_T$ ,  $[\#]_F$ , with  $\# = \wedge, \vee, \neg$ , identifies the condition under which  $F \in v(\#(\varphi_1, \dots, \varphi_n))$  (for an  $n$ -ary connective  $\#$ ) with the condition complementary to that given for  $T$ ’s belonging to  $v(\#(\varphi_1, \dots, \varphi_n))$ . . . In view of this, it is surprising to see Dale write as follows ([16], p. 441, lettering changed to match ours) on the subject of the material biconditional (here “ $\equiv$ ”):

In conformity with Dunn’s semantics and using his notation the following should be added to his recursive definition of *relevance valuation*:

$T \in v(\varphi \equiv \psi)$  iff either  $T \in v(\varphi)$  and  $T \in v(\psi)$  or  $F \in v(\varphi)$  and  $F \in v(\psi)$ ;

$F \in v(\varphi \equiv \psi)$  iff either  $F \in v(\varphi)$  and  $T \in v(\psi)$  or  $F \in v(\varphi)$  and  $T \in v(\psi)$ .

This seems to be the most appropriate definition, bearing in mind the motivation behind the definition for the other connectives.

Since the  $F$  condition here is not complementary to the  $T$  condition, this proposal can hardly be described as being in “conformity with Dunn’s semantics”, though Dale is correct to continue to remark that the proposal does not provide a Dunn–Belnap semantic treatment for either of the two candidate translations into the language of  $\{\wedge, \vee, \neg\}$  he has been considering, and to which we shall return in Section 3. For the moment, let us pause to observe that formulas in negation normal form, i.e., constructed using  $\wedge$ ,  $\vee$ , and  $\neg$  but which  $\neg$  occurring only before sentence letters, when  $n$  such letters are involved altogether are in a one-to-one correspondence with  $n$ -ary Dunn–Belnap conditions *à la* (2.1) in an obvious way, illustrated here by the case (with  $n = 3$ ) of  $p \vee (\neg q \wedge (q \vee \neg r))$  of  $\Phi$  being as follows, with  $X, Y, Z$  for  $X_1, X_2, X_3$ :

$$T \in X \text{ or } (F \in Y \text{ and } (T \in Y \text{ or } F \in Z)).$$

The propositional formula is true on a Dunn–Belnap valuation  $v$  (i.e.,  $T$  is an element of the image of the formula under  $v$ ) just in case condition is satisfied when  $X, Y, Z$  are taken as  $v(p), v(q), v(r)$ , respectively. Of course both the formula and the corresponding condition can be manipulated into more tractable subcases of negation normal form, such as CNF or DNF, as desired. With this in mind, let us consider what happens when the  $T$  and  $F$  conditions for a proposed new connective, as with Dale’s suggested  $\equiv$  conditions, are not complementary. For a simple example (closest to the simplest possible), consider a proposed 1-ary connective  $\circ$  with conditions:

$$\begin{aligned} T \in v(\circ\varphi) \text{ if and only if } T \in v(\varphi); & \quad [\circ]_T \\ F \in v(\circ\varphi) \text{ if and only if } T \in v(\varphi). & \quad [\circ]_F \end{aligned}$$

While we are at it, we may as well add a dual example of mismatch, this time with 1-ary  $\star$ :

$$\begin{aligned} T \in v(\star\varphi) \text{ if and only if } F \in v(\varphi); & \quad [\star]_T \\ F \in v(\star\varphi) \text{ if and only if } F \in v(\varphi). & \quad [\star]_F \end{aligned}$$

Several anomalies are evident immediately in the presence of such additions to the  $\{\wedge, \vee, \neg\}$  vocabulary; one is that we lose contraposition in the form:

$$\text{If } \varphi \vdash_{\text{fde}} \psi \text{ then } \neg\psi \vdash_{\text{fde}} \neg\varphi,$$

where  $\vdash_{\text{fde}}^\circ$  is the extension of  $\vdash_{\text{fde}}$  to the language with the added connective  $\circ$  interpreted by means of the  $[\circ]_T$  and  $[\circ]_F$ , since for example  $\circ p \vdash_{\text{fde}}^\circ p$  while  $\neg p \not\vdash_{\text{fde}}^\circ \neg\circ p$ . (Put  $v(p) = \{F\}$  and we have  $v(\neg p) = \{T\}$  while  $v(\neg\circ p) = \emptyset$ .)

This anomaly could be rephrased without reference to the behaviour of  $\neg$  by saying that while for  $\vdash_{\text{fde}}$  itself the two conditions (1)  $\varphi \vdash_{\text{fde}} \psi$ , taken as meaning that for any  $v$  with  $T \in v(\varphi)$  we have  $T \in v(\psi)$ , and (2) for any  $v$  with  $F \in v(\psi)$ , we have  $F \in v(\varphi)$  – see Dunn [17] esp. top of p.165, or [19], p.10 – the corresponding equivalence would longer hold for the envisaged extension  $\vdash_{\text{fde}}^\circ$ . Secondly, while on the Dunn–Belnap semantics for the language with  $\wedge, \vee, \neg$  and any formula  $\varphi(p_1, \dots, p_n)$  in the sentence letters exhibited, any valuation on which  $v(p_i) \in \{\{T\}, \{F\}\}$  for  $i = 1, \dots, n$ , delivers  $v(\varphi(p_1, \dots, p_n)) \in \{\{T\}, \{F\}\}$ , whereas the deployment of  $T$  and  $F$  conditions which are not complementary can destroy this feature. For an example of such behaviour, this time using  $\star$ , suppose  $v(p) = \{T\}$ . Then since  $T \notin v(p)$ , by  $[\star]_T$ ,  $T \notin v(\star p)$  and since  $F \notin v(p)$ , by  $[\star]_F$ ,  $F \notin v(\star p)$ , and thus  $v(\star p) = \emptyset$ . There may well be good reasons for considering a semantic treatment with effects like these, and indeed the account of Stephen Blamey in [4] famously does so with his interjunction and transplication connectives, which similarly give compounds whose components have defined values (corresponding to the present  $\{T\}$  and  $\{F\}$ ) an undefined value (corresponding to the present  $\emptyset$ , there being no analogue to  $\{T, F\}$  in [4]. The present point is that whatever their merits may be, they are not in the spirit of Dunn and Belnap’s discussions.<sup>15</sup>

Now, as is well known, we can re-package the semantic description provided by the conditions  $[\wedge]_T$ ,  $[\wedge]_F$ ,  $[\vee]_T$ , etc., as a four-valued matrix on an algebra whose elements are the subsets of  $\{T, F\}$ , conventionally relabelled – according to one among several conventions to be found in the literature – as **t** (“true only”) for  $\{T\}$ , **b** (“both”) for  $\{T, F\}$ , **f** (“false only”) for  $\{F\}$  and **n** (“neither”) for  $\emptyset$ , of which the first two are designated. (The assignments  $v$  are then matrix evaluations in the sense of Section 1.) In fact such a treatment could extend to non-standard truth and falsity clauses such as those suggested for  $\equiv$  by Dale, or the cases of  $\circ$  and  $\star$ . In fact, a principal aim of Priest [42] consists in looking at the logics determined by a certain five-element matrix and its submatrices, the fifth value, alongside the designated **t** and **b** and undesignated **f** and **n**, is a further undesignated element he writes as **e**, and which behaves not like the  $\emptyset$  of the Dunn–Belnap semantics (which is **n**’s job) but like the  $\emptyset$  of the power matrix semantics – satisfying Priest’s “gap in, gap out” description, where *gap* means “ $\emptyset$ ”. (Some motivational considerations appear in Priest [40]; **e** is what the title of Priest [41] refers to, “the above” being the Dunn–Belnap **t**, **f**, **b** and **n**.)

---

<sup>15</sup>Issues of expressive power and functional completeness as they arise for the Dunn–Belnap semantics have been discussed in Ruet [45], Pynko [43], Omori and Sano [35], De and Omori [14].

Our main interest is on aspects of the divergence between the two styles of semantics which among other things makes it necessary to distinguish these two roles played by the empty set in the first place – and to ask whether this divergence manifests itself when attention is confined to  $\mathbf{t}$ ,  $\mathbf{b}$ , and  $\mathbf{f}$ . (This question will be answered in Section 4.) But for the record, Figure 1 gives the full matrix. Priest [42] notes that every subset of  $\{\mathbf{t}, \mathbf{b}, \mathbf{n}, \mathbf{f}, \mathbf{e}\}$  which contains  $\mathbf{t}$  and  $\mathbf{f}$  is the universe of a submatrix of this matrix, and indicates how to provide proof systems codifying the consequence relations determined by the resulting 8 submatrices, some of which are independently familiar and some of which first appear in [40] or [42]. The  $\{\mathbf{t}, \mathbf{b}, \mathbf{f}\}$  and  $\{\mathbf{t}, \mathbf{n}, \mathbf{f}\}$  cases Priest calls the LP matrix and the strong Kleene matrix; they are referred to in Humberstone [30] as  $\mathbf{K}_1$  and  $\mathbf{K}_{1,2}$  respectively (“ $\mathbf{K}$ ” for “Kleene” in both cases, with a similar nomenclature in [27]); as Priest also notes, the  $\{\mathbf{t}, \mathbf{e}, \mathbf{f}\}$  case is that of Bochvar’s three-valued logic (sometimes called ‘weak Kleene’, and conveniently described at p. 29*ff.* of Rescher [44] – or see the original, [6]). Note that by the definition of *submatrix* given in Section 1, these elements retain the designated/undesignated status they have in Priest’s five-element matrix. Technical comprehensiveness suggests the exploration of an alternative five element matrix (and suitable submatrices) in which  $\mathbf{e}$  is designated instead of undesignated, though we do not enquire into that here; evidently  $\wedge$ -elimination style inferences would in this case stand in need of some restriction, just as Priest notes (e.g., in [40]) the need to restrict  $\vee$ -introduction style inferences in the version with  $\mathbf{e}$  undesignated.<sup>16</sup> Note also that the reason we can discard arbitrary subsets of  $\{\mathbf{b}, \mathbf{n}, \mathbf{e}\}$  is that all the elements of this set are irreducible in the algebra of the matrix of Figure 1, in the sense that none of these elements is the value of a fundamental operation as applied to a sequence of elements in which it does not figure – by contrast with  $\mathbf{t}$  and  $\mathbf{f}$  (with, for example,  $\mathbf{t} = \mathbf{b} \vee \mathbf{n}$ , and  $\mathbf{t} = \neg \mathbf{f}$ ).

A popular presentation of the  $\mathbf{e}$ -free reduct – i.e., the Dunn–Belnap four-valued matrix – of this five-element matrix uses a Hasse diagram to indicate the ‘diamond’ lattice ( $\mathbf{t}$  and  $\mathbf{f}$  as top and bottom elements,  $\mathbf{b}$  and  $\mathbf{n}$  as incomparable intermediate elements), so that  $\wedge$  and  $\vee$  are calculated by taking meets and joins, with a record of the effect of negation (interchanging top and bottom elements and leaving the intermediate elements fixed) superimposed on the diagram. The presence of  $\mathbf{e}$  spoils the lattice structure, since its own universally absorptive behaviour thwarts precisely the usual lattice

---

<sup>16</sup>What about having two absorbent elements,  $\mathbf{e}$  and  $\mathbf{e}^*$ , say, respectively undesignated and designated? That’s not an option because, for example,  $\mathbf{e} \wedge \mathbf{e}^*$  would have to equal both the undesignated  $\mathbf{e}$  and also the designated  $\mathbf{e}^*$ . (This is just a version of the standard argument for the uniqueness of zero elements in semigroups.)

$\wedge$	$t$	$b$	$n$	$f$	$e$	$\vee$	$t$	$b$	$n$	$f$	$e$	$\neg$	
$*t$	$t$	$b$	$n$	$f$	$e$	$t$	$t$	$t$	$t$	$t$	$e$	$t$	$f$
$*b$	$b$	$b$	$f$	$f$	$e$	$b$	$t$	$b$	$t$	$b$	$e$	$b$	$b$
$n$	$n$	$f$	$n$	$f$	$e$	$n$	$t$	$t$	$n$	$n$	$e$	$n$	$n$
$f$	$f$	$f$	$f$	$f$	$e$	$f$	$t$	$b$	$n$	$f$	$e$	$f$	$t$
$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$	$e$

Figure 1: Priest’s Five-Element Matrix

absorption laws.

While Priest is happy to have both  $\emptyset$  as it behaves in the Dunn–Belnap semantics, appearing here as  $n$ , and  $\emptyset$  as it behaves in the (non-negative) power matrix semantics, appearing here as  $e$ , both living under the same roof, we continue to ponder the relations between these two roles played by the empty set, as well as of the two semantic approaches, for which purpose we stick with a notation neutral between them:  $\{T\}$ ,  $\{T, F\}$  etc. Priest [42] continues the theme of Priest [38] but this time with respect to what he calls logics “in the FDE family”, meaning the consequence relations determined by the eight submatrices (already alluded to) of the above five-element matrix. More precisely, he details the effects of “going plurivalent,” as he puts it: sorting out the cases in which in which on passage from matrix in question to its power matrix – construed positively or construed non-negatively – gives rise to a strictly weaker consequence relation from the cases in which the change in matrix makes no difference to the consequence relation. But our interest here is in the mixture of power matrix and Dunn–Belnap considerations which get us from the 2-element Boolean matrix to this 5-element matrix in the first place, and in particular in the differences between  $e$  and  $n$ , the incarnations of  $\emptyset$  coming from these two sets of considerations, respectively.

To conclude this section, let us confirm explicitly the claim quoted from Priest [38] in Section 1 that “Rather the point-wise definitions give the extension of the LP functors according to the rule: gap-in, gap-out,” in which, as is clear from the text of [38] the reference to pointwise definitions is to the power algebra construction.<sup>17</sup> Since the only departure from the Dunn–

<sup>17</sup>The ‘pointwise’ terminology carries the risk of suggesting, at the level of algebras, the direct product (rather than the power algebra) construction, which, taking the product of the two-element Boolean algebra with itself, gives us the algebra we get from the  $\{t, b, n, f\}$ -subalgebra of that depicted in Figure 1 by retaining the  $\wedge$  and  $\vee$  tables but changing that for negation so as to have not only  $t$  and  $f$  but now also  $b$  and  $n$  interchanged. (Traditionally the associated product matrix based on this algebra would have only  $t$  designated.) Now  $t, b, n$ , and  $f$  have become (not particularly suggestive) labels

Belnap semantics here described arises in the non-negative version of the power construction, the implicit suggestion is that if we use the positive version (excluding  $\emptyset$ , that is) then the power matrix semantics and the Dunn–Belnap semantics agree. We confirm this for the favoured primitives (“the LP functors”),  $\wedge, \vee, \neg$ , in Proposition 2.1 so that we can see (in the proof) exactly where excluding  $\emptyset$  is crucial. For ease of formulation, the following notational conventions will be useful: In the following, for a two-valued truth-function  $\#$ , its ‘power matrix’ lifting is denoted by  $\hat{\#}$ , while  $\#_{DB}$  denotes the corresponding four-valued Dunn–Belnap function. Illustrating in the case of  $\wedge$ , this means that for  $X, Y \subseteq \{T, F\}$ ,  $X \hat{\wedge} Y = \{x \wedge y \mid x \in X, y \in Y\}$  and  $X \wedge_{DB} Y$  is the set uniquely fixed by the stipulation that  $T \in X \wedge_{DB} Y$  iff  $T \in X$  and  $T \in Y$ , and  $F \in X \wedge_{DB} Y$  iff  $F \in X$  or  $F \in Y$ . To repeat, the following observation is all but explicit in Priest [38]:

**PROPOSITION 2.1** *For non-empty  $X, Y \subseteq \{T, F\}$ , we have  $X \hat{\wedge} Y = X \wedge_{DB} Y$ ; for non-empty  $X, Y \subseteq \{T, F\}$ , we have  $X \hat{\vee} Y = X \vee_{DB} Y$ ; for arbitrary  $X \subseteq \{T, F\}$ , we have  $\hat{\neg} X = \neg_{DB} X$ .*

*Proof.* The reasoning is simple but we go through it explicitly here, dividing the argument for the case of  $\wedge$  into four parts to make it very clear where the assumption that neither  $X$  nor  $Y$  is  $\emptyset$  is required: only for part (4). Parts (1) and (2) show that  $T \in X \hat{\wedge} Y$  if and only if  $T \in X$  and  $T \in Y$ ; parts (3) and (4) show that  $F \in X \hat{\wedge} Y$  if and only if  $F \in X$  or  $F \in Y$ .

(1) Suppose that  $T \in X \hat{\wedge} Y$ ; then  $T = x \wedge y$  for some  $x \in X, y \in Y$ ; but then, in virtue of the truth-function  $\wedge$ ,  $x = y = T$ , so  $T \in X$  and  $T \in Y$ , implying that  $T \in X \wedge_{DB} Y$  by the “if” direction of  $[\wedge]_T$ .

(2) Suppose that  $T \in X \wedge_{DB} Y$ . Then, by the “only if” direction of  $[\wedge]_T$ ,  $T \in X$  and  $T \in Y$ , so there exist  $x \in X, y \in Y$  with  $T = x \wedge y$ , since, in virtue of the truth-function  $\wedge$ , we may take  $x = y = T$ , establishing that  $T \in X \hat{\wedge} Y$ .

(3) Suppose that  $F \in X \hat{\wedge} Y$ ; then  $F = x \wedge y$  for some  $x \in X, y \in Y$ , so in virtue of the truth-function  $\wedge$ ,  $x = F$  or  $y = F$ , so  $F \in X$  or  $F \in Y$ , implying, by the “if” direction of  $[\wedge]_F$ , that  $F \in X \wedge_{DB} Y$ .

(4) Suppose that  $F \in X \wedge_{DB} Y$ . By the “only if” direction of  $[\wedge]_F$ , we have  $F \in X$  or  $F \in Y$ . Suppose first that  $F \in X$ . Then, since  $Y \neq \emptyset$ , we may take  $y \in Y$  and note that whether  $y$  is  $T$  or  $F$ ,  $F \wedge y = F$ , so there are  $x \in X$  (take  $x = F$ ) and  $y \in Y$  with  $x \wedge y = F$ , meaning that  $F \in X \hat{\wedge} Y$ .

---

for  $\langle T, T \rangle, \langle T, F \rangle, \langle F, T \rangle$  and  $\langle F, F \rangle$  and the four-valued truth-functions are indeed computed pointwise (or ‘coordinatewise’). For further details and some historical references, see pp. 96–102 of Rescher [44].

If we suppose instead that  $F \in Y$ , then similar reasoning applies, this time exploiting the assumption that  $X \neq \emptyset$ ; so, either way,  $F \in X \hat{\wedge} Y$ .

The case of  $\vee$  is similar, this time the need to have  $X$  and  $Y$  non-empty arising for the analogue of the above part (2), showing that if  $T \in X \vee_{DB} Y$ , then  $T \in X \hat{\vee} Y$ . The case of  $\neg$  is simpler still, since as the formulation of the result indicates, no non-emptiness assumption is required. ■

In the spirit of generalizing Proposition 2.1 one might consider the case of an arbitrary  $n$ -place two-valued truth-function  $\#$  – and note that, as in the preamble to Proposition 2.1, this does mean the truth-function and not a (primitive or derived) connective with which this truth-function is associated – with  $\#_{DB}$  and  $\hat{\#}$  understood as there; for non-empty  $X_1, \dots, X_n \subseteq \{T, F\}$ , we have:

$$\hat{\#}(X_1, \dots, X_n) = \#_{DB}(X_1, \dots, X_n).$$

To assess the truth – indeed the intelligibility – of this candidate generalization of Proposition 2.1 we need to recall a cautionary note sounded in the 1980s by A. J. Dale.

### 3 Dale's Point

We may introduce the objection of Dale [16]<sup>18</sup> to the work of Anderson and Belnap [1] (in particular Chapter 3 thereof) and Dunn [17] by means of a quotation from p. 151 of the latter:

Jeffrey's formalism includes sentence letters (...), and connectives for negation, conjunction and disjunction (...) as well as for the truth-functional conditional and biconditional. We shall ignore these last two since they are not primitive in the standard formulations of the system **E** (though they can of course be introduced as abbreviatory devices via their ordinary contextual definitions).

Of the two non-primitives mentioned here,  $\supset$  and  $\equiv$ , there may seem little room for choice as to what might be meant by their 'ordinary' definitions in

---

<sup>18</sup>[16] also discusses ideas from Casimir Lewy, Michael Clarke and Timothy Smiley, but this material is not germane to the present topic. (Most of the remarks on Dale [16] in the current section were made in a note privately circulated by the author in 1983, under the title 'Memorandum on an Objection by A. J. Dale'; one to whom it was sent – Mike Dunn – remarked that his own reaction to [16] had been along similar lines.)

the case of the former, but not so, Dale argues, in the case of the latter: surely if I want to define  $\equiv$  in terms of  $\wedge$ ,  $\vee$  and  $\neg$ , it would be equally natural to offer the conjunctive definition (3.1) or the disjunctive definition (3.2):

$$\varphi \equiv \psi = (\neg\varphi \vee \psi) \wedge (\varphi \vee \neg\psi) \quad (3.1)$$

$$\varphi \equiv \psi = (\varphi \wedge \psi) \vee (\neg\varphi \wedge \neg\psi) \quad (3.2)$$

Dale's point, illustrated with the example of " $\equiv$ ", is that the right-hand sides of (3.1) and (3.2) are not equivalent in the logic of first degree entailment (FDE) under discussion in [1] and [17], so no unique candidate for the appropriate analogue of classical material equivalence has been provided. (Concerning the present pair of candidates, we have the r.h.s. of (3.1) as an  $\vdash_{\text{fde}}$ -consequence of the r.h.s. of (3.2) but not conversely.) The situation is superficially somewhat different from the case of  $\supset$ , where the first two options that perhaps come to mind for  $\varphi \supset \psi$ , namely  $\neg\varphi \vee \psi$  and  $\neg(\varphi \wedge \neg\psi)$ , are by contrast FDE-equivalent, and one might envisage a weak plea for (3.1) coming from the consideration that  $\varphi \equiv \psi$  is strongly associated, when  $\equiv$  is not taken as primitive, with  $(\varphi \supset \psi) \wedge (\psi \supset \varphi)$ .<sup>19</sup> But of course the point could equally well be made – and Dale does make it – with reference to exclusive disjunction, which similarly comes in two matching flavours, conjunctive and disjunctive;<sup>20</sup> here we use  $\underline{\vee}$  for this connective; again the disjunctive version is strictly stronger than the conjunctive version, from an FDE perspective:

$$\varphi \underline{\vee} \psi = (\varphi \vee \psi) \wedge (\neg\varphi \vee \neg\psi) \quad (3.3)$$

$$\varphi \underline{\vee} \psi = (\varphi \wedge \neg\psi) \vee (\neg\varphi \wedge \psi) \quad (3.4)$$

The general point has in any case nothing to do with the naturalness of conventional frequency of this or that representation of Boolean connectives other than  $\wedge$ ,  $\vee$  and  $\neg$  in terms of these three, so much as the fact that there is in general no unique representation to within FDE-equivalence of any given Boolean compound. Thus Dale [16] correctly chides, for example,

---

<sup>19</sup>Indeed this association is built into the notation if the variants  $\rightarrow$  and  $\leftrightarrow$  of  $\supset$  and  $\equiv$  are employed – something that would lead to confusion in the discussion of relevant logic where they are reserved for use as corresponding intensional connectives. Note that in the line of argument in support of (3.1) here envisaged, no fuss is made of the distinction between  $\psi \supset \varphi$ , alias  $\neg\psi \vee \varphi$ , and the commuted form  $\varphi \vee \neg\psi$  actually appearing in (3.1), since this is a distinction without an FDE-difference.

<sup>20</sup>Dale [16], p. 437, alludes in this connection to conjunctive and disjunctive normal form, which is potentially misleading, as  $\varphi$  and  $\psi$  in (3.1, 2) are not restricted to being sentence letters.



Anderson and Belnap, for temporarily overlooking this in calling Chapter 3 of [1] ‘Entailment Between Truth-Functions’.<sup>21</sup> His point is in the first instance an illustration of the widespread though not universal fact that the weaker the logic one is considering, the more distinctions the logic registers; see [29] for a discussion of this generalization and its limits.<sup>22</sup> But secondly, in one form or another the point has always been a familiar fact in the specifically relevant-logical community: consider for example, for formulas  $\varphi, \psi$ , the compounds  $\varphi \wedge \psi$  and  $\varphi \wedge (\neg\varphi \vee \psi)$ , both of which are classically unexceptionable representations of the conjunction truth-function but which fare very differently by relevant lights, the first but not the second having  $\psi$  as a relevant consequence – Disjunctive Syllogism and all that. We could throw in as a third candidate  $(\varphi \vee \neg\psi) \wedge \psi$ , and still not be finished, even subject to the further constraint that no more than the given schematic letters  $\varphi, \psi$ , are to be employed. And, setting aside that constraint – or rather, its analogue in the 1-ary case – no-one in the Anderson–Belnap relevance tradition would expect the classically equivalent  $\varphi$  and  $\varphi \wedge (\psi \vee \neg\psi)$  to be FDE-equivalent in general since, instantiating the schematic letters  $\varphi, \psi$ , to distinct sentence letters  $p, q$ , this would make  $q \vee \neg q$  an  $\vdash_{\text{fde}}$ -consequence of a formula (namely  $p$ ) with which it shares no non-logical vocabulary.

Returning specifically to Dale’s examples (3.1)-vs.-(3.2) and (3.3)-vs.-(3.4), it should be remarked that in referring to the first and second of each pair as respectively the conjunctive and disjunctive forms of material equivalence and exclusive disjunction is just a convenience and should not be taken too seriously. Every  $\{\wedge, \vee, \neg\}$ -formula has, after all, both an FDE-equivalent in conjunctive normal form and one in disjunctive normal form. Taking the (3.1) ‘conjunctive’ case (“ $p \equiv^{\wedge} q$ ” in the more explicit notation of used below) with  $\varphi, \psi$ , atomic, we have  $(\neg p \vee q) \wedge (p \vee \neg q)$ , with an FDE-equivalent in disjunctive normal form as follows:

$$(p \wedge \neg p) \vee (\neg p \wedge \neg q) \vee (p \wedge q) \vee (q \wedge \neg q).$$

It’s just that this disjunction is not FDE-equivalent to the what we called the disjunctive form of material equivalence (“ $p \equiv^{\vee} q$ ”) meaning the shorter disjunction which drops the initial and final disjuncts (instantiating the r.h.s.

<sup>21</sup>The present objection aside, there is no reason to describe  $\neg$  as truth-functional in the present setting; this connective is not even extensional according to  $\vdash_{\text{fde}}$ , in the sense of §3.2 of [30], which is a necessary (though not sufficient) condition for being amenable to a semantic treatment invoking an associated truth-function.

<sup>22</sup>Several illustrations of this phenomenon as it arises in connection with intuitionistic (rather than relevant) logic as a weakening of classical logic are given under Examples 4.38.5 of [30], p. 618; the case of exclusive disjunction is treated for these two logics Exercise 6.12.6 there (p. 785) and in the discussion following that.

of (3.2)). Similarly, this shorter disjunction itself has an FDE-equivalent conjunctive normal form representation, which is not  $(\neg p \vee q) \wedge (p \vee \neg q)$  but an extended form with some classically redundant but relevant-logically essential ‘excluded middle’ conjuncts.

**Digression.** The contrast between (3.1) and (3.2) will be familiar to anyone who has thought about the simplest kind of definition by cases – of, say a monadic predicate letter – in quantified FDE (though the point could be made at the purely propositional level). Classically if we want to introduce a predicate symbol  $E$ , say, in terms of primitive (also monadic) predicates  $F, G, H$ , by saying that something is  $E$  just in case it is  $G$  or  $H$ , according as it is  $F$  or not  $F$  respectively, one can do so either by setting  $Ex$  equivalent to (3.5), in which  $\varphi \supset \psi$  means  $\neg\varphi \vee \psi$ , or instead setting  $Ex$  equivalent to (3.6):

$$(Fx \supset Gx) \wedge (\neg Fx \supset Hx) \quad (3.5)$$

$$(Fx \wedge Gx) \vee (\neg Fx \wedge Hx) \quad (3.6)$$

and it makes no difference which form one chooses, since (3.5) and (3.6) are equivalent. But when  $\wedge, \vee$  and  $\neg$  are governed instead by FDE, the two formulas are not equivalent and it is necessary to decide between the conjunctive  $E^\wedge x$  (say), defined by (3.5) and the disjunctive  $E^\vee x$  defined by (3.6). As with the earlier examples, in the absence of further non-logical assumptions, the disjunctive form is the stronger of the two by FDE lights, in the sense that (extending our turnstile notation to apply to predicate logic and using  $a$  as an individual parameter)  $E^\vee a \vdash_{\text{fde}} E^\wedge a$  while  $E^\wedge a \not\vdash_{\text{fde}} E^\vee a$ . One instance of the choice between forms like (3.5) and (3.6) is touched on in the Appendix below, where we will see Dunn drawing attention to such contrasts as this as they bear on Goodman’s *grue* example. **End of Digression.**

Transposed into the setting of Dunn–Belnap semantics, Dale’s observation becomes the point that there is no such thing as *the* Dunn–Belnap condition induced by a given bivalent truth-function; to drive the point home, the pair  $([\equiv^\wedge]_T, [\equiv^\wedge]_F)$  below, following the lead of (3.1), and the pair  $([\equiv^\vee]_T, [\equiv^\vee]_F)$ , following the lead of (3.2), are pairs of internally complementary Dunn–Belnap conditions each of which describes the bivalent truth-function concerned when only the values  $\{T\}$  and  $\{F\}$  are involved:

$$T \in v(\varphi \equiv^\wedge \psi) \text{ iff } (F \in v(\varphi) \text{ or } T \in v(\psi)) \text{ and } (T \in v(\varphi) \text{ and } F \in v(\psi)); \quad [\equiv^\wedge]_T$$

$$F \in v(\varphi \equiv^\wedge \psi) \text{ iff } (T \in v(\varphi) \text{ and } F \in v(\psi)) \text{ or } (F \in v(\varphi) \text{ or } T \in v(\psi)). \quad [\equiv^\wedge]_F$$

$$T \in v(\varphi \equiv^\vee \psi) \text{ iff } (T \in v(\varphi) \text{ and } T \in v(\psi)) \text{ or } (F \in v(\varphi) \text{ and } F \in v(\psi)); \quad [\equiv^\vee]_T$$

$$F \in v(\varphi \equiv^\vee \psi) \text{ iff } (F \in v(\varphi) \text{ or } F \in v(\psi)) \text{ and } (T \in v(\varphi) \text{ or } T \in v(\psi)). \quad [\equiv^\vee]_F$$

The upshot of these conditions is summarised in the first two tables of Figure 2. (We retain the neutral  $\{T\}, \{T, F\}, \emptyset, \{F\}$  notation – though in tables write these more compactly as  $\top, \top F, \emptyset,$  and  $F$  respectively – since this allows us to write simply  $\emptyset$  rather than choosing between  $\mathbf{n}$  and  $\mathbf{e}$  à la Figure 1, precisely so as to be able to compare the behaviour of  $\emptyset$  in the Dunn–Belnap semantics and in the power matrix semantics.) The third table gives the power-algebraic case for comparison, with  $X \hat{=} Y$  being  $\{x \equiv y \mid x \in X, y \in Y\}$ . It of course differs from them in several cases where  $\emptyset$  is an argument, but the present point – Dale’s point – is that these two tables differ from each other. The moral for the mooted generalization of Proposition 2.1 from Section 2 about the equality  $\hat{\#}(X_1, \dots, X_n) = \#_{DB}(X_1, \dots, X_n)$  for a given  $n$ -ary (two-valued) truth-function  $\#$  is of course that the notation on the right, “ $\#_{DB}(X_1, \dots, X_n)$ ” does not make sense: there is no such thing as *the* Dunn–Belnap (four-valued) function corresponding to a given  $\#$ . If  $\#$  were taken as the binary  $\equiv$ , which – if either – of  $\hat{\equiv}_{DB}, \equiv_{DB}^{\vee}$ , would  $\equiv_{DB}$  be, for instance?

$\hat{\equiv}_{DB}$	T	TF	$\emptyset$	F	$\equiv_{DB}^{\vee}$	T	TF	$\emptyset$	F	$\hat{=}$	T	TF	$\emptyset$	F
*T	T	TF	$\emptyset$	F	T	T	TF	$\emptyset$	F	T	T	TF	$\emptyset$	F
*TF	TF	TF	F	TF	TF	TF	TF	T	TF	TF	TF	TF	$\emptyset$	TF
$\emptyset$	$\emptyset$	F	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	T	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
F	F	TF	$\emptyset$	T	F	F	TF	$\emptyset$	T	F	F	TF	$\emptyset$	T

Figure 2: Three Biconditional Tables

Looking at the tables of Figure 2, and with the condition – taken from Proposition 2.1 – on the envisaged generalization

$$\hat{\#}(X_1, \dots, X_n) = \#_{DB}(X_1, \dots, X_n)$$

in mind, to the effect that this should hold “For *non-empty*  $X_1, \dots, X_n \subseteq \{T, F\}$ ”, the reader may naturally wonder if the failure of uniqueness just observed for such notations as  $\equiv_{DB}$  has been pressed too hard. There may, for all that has been said so far, be no such failure once  $\emptyset$  is set aside. If we discard the rows and columns of the three tables in Figure 2 – noting that  $\emptyset$  arises only in the body of such a table in one of these rows or columns – then the first two tables coincide with each other, so the uniqueness problem has vanished, and, further, they now match the  $\{\top, \top F, F\}$ -reduct of the third table, so not only the conceptual coherence but also the correctness of the envisaged generalization of Proposition 2.1 remains intact despite Dale’s point, at least insofar as the case of  $\equiv$  is concerned. The same goes for the

case of  $\underline{\vee}$ , as the reader wanting to construct the analogue of Figure 2 for this case, with tables for  $\underline{\vee}_{DB}^{\wedge}$  and  $\underline{\vee}_{DB}^{\vee}$ , summarising the Dunn–Belnap conditions corresponding to (3.1) and (3.2) respectively, is invited to verify.

## 4 Another Example – and the Gautam Effect Revisited

What the examples just reviewed do not themselves provide is made available by minor tinkering, however, as we illustrate with the case of  $\equiv^{\wedge}$ , i.e., with  $\equiv$  as defined in (3.1). Our tinkering consists in discarding the second disjunct,  $\neg\psi$ , of the second conjunct; we abbreviate the result with “ $\wedge$ ” – to be thought of as an unstructured symbol – though in fact we write the two conjuncts in reverse order for convenience:

$$\varphi \wedge' \psi = \varphi \wedge (\neg\varphi \vee \psi) \quad (4.1)$$

This is only one of many examples that could be chosen to make the present point, selected because its r.h.s. was mentioned already apropos of the Disjunctive Syllogism inference in the previous section. In terms of Dunn–Belnap conditions, we have:

$$\begin{aligned} T \in v(\varphi \wedge' \psi) &\text{ iff } T \in v(\varphi) \text{ and } (F \in v(\varphi) \text{ or } T \in v(\psi)); & [\wedge']_T \\ F \in v(\varphi \wedge' \psi) &\text{ iff } F \in v(\varphi) \text{ or } (T \in v(\varphi) \text{ and } F \in v(\psi)). & [\wedge']_F \end{aligned}$$

We compare the tabular representation of these conditions, on the left of Figure 3, using the notation  $\wedge'_{DB}$ , with that of the power operation on the corresponding bivalent truth-function on the right, which we could write either as  $\widehat{\wedge}'$  or simply as  $\widehat{\wedge}$ , since the (bivalent) truth-function is just the conjunction function; for familiarity, we use the latter notation.

$\wedge'_{DB}$	T	TF	$\emptyset$	F	$\widehat{\wedge}$	T	TF	$\emptyset$	F
*T	T	TF	$\emptyset$	F	T	T	TF	$\emptyset$	F
*TF	TF	TF	TF	TF	TF	TF	TF	$\emptyset$	F
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
F	F	F	F	F	F	F	F	$\emptyset$	F

Figure 3: Two Conjunction Tables

There are three discrepancies between the tables in Figure 3, one at  $\langle \text{TF}, \emptyset \rangle$ , where we have F in the case of  $\wedge'_{DB}$  but  $\emptyset$  in the case of  $\widehat{\wedge}$ , a second at  $\langle \text{F}, \emptyset \rangle$ , giving values F and  $\emptyset$  respectively, and the third at  $\langle \text{TF}, \text{F} \rangle$ ,

yielding respectively TF and F. The first two are nothing new: like those in Figure 2, they illustrate the universally absorptive nature of  $\emptyset$  in the power matrix semantics – “gap in, gap out,” in Priest’s formulation. But the third discrepancy is what we have been looking for: an indication that the power matrix treatment and the Dunn–Belnap treatment diverge even in the case of compounds all of whose (here, both of whose) components are assigned values in the range  $\{T, TF, F\}$ . Thus if one took Priest’s remark (quoted in Section 1) that logic of  $\hat{\neg}$ ,  $\hat{\wedge}$  and  $\hat{\vee}$  “is not, what it might at first be thought to be, first degree entailment” to which the “gap in, gap out” comment is appended, to suggest that it was only when  $\emptyset$  was involved that there was a difference according as one calculated by the Dunn–Belnap matrix or the power matrix, one would be dealing with a misleading suggestion. Note that the third discrepancy just observed in the case of Figure 3 involves not just a simple numerical distinction between values – TF vs. F – but a difference in respect of designation status. (Thus it immediately gives rise to a failure of the “power conjunction” of  $p$  and  $q$  to be a consequence of  $p \wedge' q$  in a language supporting both connectives: assign TF to  $p$  and F to  $q$  for an invalidating evaluation. That would leave open the possibility that the only logical difference emerges when the two connectives are both present, as in Łukasiewicz’s ‘twins’ analogy,<sup>23</sup> but we can contrast them in isolation from each other too:  $q$  is a consequence of what for emphasis we may write as  $p \hat{\wedge} q$  but not – this being the Disjunctive Syllogism point – of  $p \wedge' q$ .)

While the suggestion thus extracted from Priest’s discussion – no divergence except where  $\emptyset$  is involved – is correct for  $\wedge$ ,  $\neg$  and  $\vee$ , and since this is what in the passage quoted in Section 1 Priest means by the phrase “the LP functors” in that passage, the text of [38] is not itself misleading and our recent discussion may be summed up by saying that what holds for  $\wedge$ ,  $\neg$  and  $\vee$  in this regard does not hold generally for connectives defined in terms of them, as the case of  $\wedge'$  illustrates. This answers affirmatively the question raised in Section 2 as to whether the difference between the Belnap–Dunn and the power-algebraic semantics manifests itself even when attention is confined to  $\mathbf{t}$ ,  $\mathbf{b}$ , and  $\mathbf{f}$ .

One possible misinterpretation of the example of  $\wedge'$  should be addressed, concerning the following reaction to (4.1): shouldn’t the application of the power algebraic construction in this case be as dictated by the r.h.s. of (4.1), with the interpretation of  $X \hat{\wedge}' Y$  ( $X, Y \subseteq \{T, F\}$ ) being:  $X \hat{\wedge} (\hat{\neg} X \hat{\vee} Y)$ ? One could have a similar reaction in the case of the earlier examples ( $\hat{\equiv}$  etc.) but let us address it here in any case. The answer to the question is that we are understanding a function  $f$  to induce a power operation in

<sup>23</sup>See pp. 471, 484, of [30] for discussion and references

accordance with (1.1), which is written in a notation making the algebras concerned explicit, and so strictly for that introduction of the idea to be in play we should envisage the two-element Boolean algebra as being expanded to support as fundamental operations whichever bivalent truth-function is envisaged as subject to the power algebra construction. In particular  $\wedge'$  gives rise to the function  $\widehat{\wedge}'$  defined by:

$$\begin{aligned} X \widehat{\wedge}' Y &= \{x \wedge' y \mid x \in X, y \in Y\} \\ &= \{x \wedge (\neg x \vee y) \mid x \in X, y \in Y\} \\ &= \{x \wedge y \mid x \in X, y \in Y\} \quad [= X \widehat{\wedge} Y] \end{aligned}$$

This is a different matter from  $X \widehat{\wedge} (\widehat{\neg} X \widehat{\vee} Y)$ , which is like  $\wedge'_{DB}$  from Figure 3, except for having all entries in the  $\emptyset$ -column be  $\emptyset$ . (Proposition 2.1 already tells us that the only differences that could arise would be cases with  $\emptyset$  among the inputs.) We are concerned to apply the power construction to the function itself, rather than reading it off, seriatim, this or that polynomial (Boolean term) for the function. Indeed the difference here is a manifestation of what we called in Section 1 the Gautam effect: the fact that one does not in general expect identities which are (even if regular) not linear to be preserved on the passage from algebra to power algebra (whether the latter is construed positively or non-negatively). The identity at issue in the present case is  $x \wedge y \approx x \wedge (\neg x \vee y)$ , with a conspicuously non-linear term on the right.

Now, in Section 1 we had occasion to notice that in passing from the two-element Boolean algebra to its (non-negative) power algebra the non-linearity in the idempotence identities for  $\wedge$  and  $\vee$  did not prevent them from surviving the passage. Now although the identities holding in the two-element Boolean algebra are precisely those holding in every Boolean algebra, we continue to use this formulation with its emphasis on identities inherited by from the two-element algebra to its power algebra. Some other cases were also mentioned in Section 1: the absorption identities and the distribution laws, in particular. The absorption laws ( $x \approx (x \wedge y) \vee x$  and its dual) have no hope of surviving to the (non-negative) power algebra because they are not regular. We conclude our discussion by looking at the distribution case and then, as a stimulus for further work, formulating two conjectures consistent with our findings. Let us consider distribution in the following form rather than its dual:  $x \wedge (y \vee z) \approx (x \wedge y) \vee (x \wedge z)$ . We need to check that for all  $X, Y, Z \subseteq \{T, F\}$ , we have:

$$X \widehat{\wedge} (Y \widehat{\vee} Z) = (X \widehat{\wedge} Y) \widehat{\vee} (X \widehat{\wedge} Z).$$

The  $\subseteq$ -direction is easy, and does not require the hypothesis that we are in the power algebra of a 2-element algebra, or any details of how  $\wedge$  and  $\vee$

behave on that algebra (other than that they satisfy the distributive law): if  $T \in X \widehat{\wedge} (Y \widehat{\vee} Z)$ , this means that  $T = x \wedge (y \vee z)$  for some  $x \in X, y \in Y, z \in Z$ . By distributivity in the underlying algebra we have  $T = (x \wedge y) \vee (x \wedge z)$ , so  $T \in (X \widehat{\wedge} Y) \widehat{\vee} (X \widehat{\wedge} Z)$ . Since the same reasoning would go through for  $F$ , we have  $(Y \widehat{\vee} Z) \subseteq (X \widehat{\wedge} Y) \widehat{\vee} (X \widehat{\wedge} Z)$ . For the converse inclusion, begin by showing that  $T \in (X \widehat{\wedge} Y) \widehat{\vee} (X \widehat{\wedge} Z)$  implies that  $T \in X \widehat{\wedge} (Y \widehat{\vee} Z)$ , after which we will check that the same holds with “ $F$ ” in place of “ $T$ ”. Thus suppose  $T \in (X \widehat{\wedge} Y) \widehat{\vee} (X \widehat{\wedge} Z)$ , i.e.,  $T = (x_1 \wedge y) \vee (x_2 \wedge z)$  for some  $x_1, x_2 \in X, y \in Y, z \in Z$ . Accordingly, either  $x_1 = y = T$  or  $x_2 = z = T$ . In the first case,  $x_1 = T$  and  $y \vee z = T$ , so  $T \in X \widehat{\wedge} (Y \widehat{\vee} Z)$ , and in the second case  $x_2 = T$  and  $y \vee z = T$  so again  $T \in X \widehat{\wedge} (Y \widehat{\vee} Z)$ . Now for  $F$ : suppose that  $F \in (X \widehat{\wedge} Y) \widehat{\vee} (X \widehat{\wedge} Z)$ . Thus  $F = (x_1 \wedge y) \vee (x_2 \wedge z)$  for some  $x_1, x_2 \in X, y \in Y, z \in Z$ . So  $F = x_1 \wedge y$  and  $F = x_2 \wedge z$ ; thus (1)  $x_1 = F$  or  $y = F$ , and also (2)  $x_2 = F$  and  $z = F$ . If  $x_1 = F$  then  $x_1 \wedge (y \vee z) = F$  so  $F \in X \widehat{\wedge} (Y \widehat{\vee} Z)$ . If  $x_2 = F$  then  $x_2 \wedge (y \vee z) = F$  so again  $F \in X \widehat{\wedge} (Y \widehat{\vee} Z)$ . If we have neither  $x_1$  nor  $x_2 = F$ , then the only way for (1) and (2) both to obtain is to have  $y = z = F$  in which case  $x \wedge (y \vee z) = F$ , where  $x$  is taken indifferently either as  $x_1$  or as  $x_2$ , and in this case again we get  $F \in X \widehat{\wedge} (Y \widehat{\vee} Z)$ .

Thus, like idempotence, distributivity survives into the power algebra of the two-element Boolean algebra. We already know, however, from the case of  $\wedge'$ , that this is not so for arbitrary regular identities. There – or transposing the example from sentential to equational logic, at least – we denoted by  $x \wedge' y$  the term  $x \wedge (\neg x \vee y)$  and found that the regular identity  $x \wedge y \approx x \wedge (\neg x \vee y)$  did not similarly survive. But this, as a Boolean identity, relies on certain interactions between what we will call negative and positive occurrences of variables: in particular  $x$  occurs both positively and negatively in it, in a sense we now make precise in order to articulate an appropriately restricted version of the conjectures we shall finish with, concerning preservation of regular identities in a rather special case.<sup>24</sup>

An occurrence of a variable is *positive* in a term  $t$  if it lies in the scope of (altogether) an even number (including 0) of occurrences of  $\neg$  in  $t$  and is

---

<sup>24</sup>The general issue of when such identities are preserved has long been a topic of interest. Early ventures include Płonka [37], John [32]; the characterization provided in the second theorem on p. 4 of Taylor’s survey [51] is a variation on the latter; note that in the 2-element ‘sup-algebras’ described there, the absorbent element is the element labelled “1” rather than that labelled “0”. (The third theorem on that same page deals with definability by linear equations and gives as necessary and sufficient conditions for this that the class of algebras is closed not only under direct products, subalgebras and homomorphic images – the basic conditions for equational definability in general – but also under ‘complex algebras,’ by which Taylor here means passing to the power algebra, positively construed.)

a *negative* occurrence otherwise. (For example, if  $t$  is  $\neg\neg x \wedge \neg(y \vee x)$ , the first occurrence of  $x$  is positive, and its second occurrence, as well as the sole occurrence of  $y$ , is negative.) A variable  $v$  *occurs positively* in an identity  $t \approx u$  if  $v$  has a positive occurrence in  $t$  or  $v$  has a positive occurrence in  $u$ , and  $v$  *occurs negatively* in  $t \approx u$  if  $v$  has a negative occurrence in  $t$  or  $v$  has a negative occurrence in  $u$ .

With all this talk of positive and negative occurrences, it would be confusing to include in the same sentence the term *non-negative* to single out the power algebra construction which includes the empty set, so we just say “power algebra” to mean this in the formulation of the conjectures below. (Its correctness would imply the correctness of the corresponding claim for power algebras positively construed, of course.)

FIRST CONJECTURE. If no variable occurring in a regular Boolean identity occurs both positively and negatively in that identity, then the identity also holds in the power algebra of the 2-element Boolean algebra.

This conjecture would imply a similar transfer for some identities in which a variable does occur both positively and negatively, such as  $x \vee \neg x \approx \neg x \vee x$ , when one of these arises by substitution from another identity (in the present case  $x \vee y \approx y \vee x$ ) which does not exhibit such mixed occurrences. Though we are not intending to return from equational logic to sentential logic here, note also the resemblance to an aspect of  $\vdash_{\text{fde}}$  emphasized in §5.4 of Burgess [13]: negative and positive occurrences of a sentence letter in formulas involved in a correct  $\vdash_{\text{fde}}$ -claim can be replaced by corresponding occurrences of unrelated sentence letters without rendering that claim incorrect.

The correctness of this first conjecture would not subsume all cases in which identities are inherited by the power algebra mentioned in it. We have already seen examples showing that even if the conjectured condition is sufficient for such inheritance, it is not necessary. Consider the case of  $\equiv$  (the bivalent truth-function, rather than the connective). Even though there is no linear  $\{\wedge, \vee, \neg\}$ -term  $t(x, y)$  for which  $t(x, y) \approx x \equiv y$  is a Boolean identity,<sup>25</sup> for the two non-linear terms corresponding to Dale’s conjunctive and disjunctive incarnations of  $\equiv$  in FDE, we find that equating them:

$$(x \wedge y) \vee (\neg x \wedge \neg y) \approx (\neg x \vee y) \wedge (x \vee \neg y),$$

gives an identity which continues to hold in the power algebra (of the two-element Boolean algebra). Yes this identity is not obtainable by substitution from one satisfying the condition in the first conjecture above – one in which

---

<sup>25</sup>Potentially confusing here is that fact that the truth-function  $\equiv$  counts in Emil Post’s classification of Boolean functions (Urquhart [52], p. 448), as a linear function.



no variable occurs both positively and negatively. This suggests a bolder conjecture, with which we close:

SECOND CONJECTURE. If every variable occurring positively (negatively) on either side of a regular Boolean identity occurs positively (resp. negatively) on the other side, then the identity also holds in the power algebra of the 2-element Boolean algebra.

The word “regular” can be omitted without affecting the content here, and is retained only for parity of formulation with the first conjecture.

## Appendix: Note on FDE and *Grue*

One philosophically famous – though, on first encounter, apparently strained – instance of this kind of case-based definition illustrated by (3.5) and (3.6) in the Digression in Section 3 repays closer inspection. It is that of Nelson Goodman’s (see [24]) in which the  $Ex$  of that discussion is taken as “ $x$  is grue” and  $Fx$ ,  $Gx$ , and  $Hx$  are “ $x$  is first examined before (some fixed time)  $t$ ”, “ $x$  is green” and “ $x$  is blue”. Goodman’s motivation was to pose a problem for the presumed objectivity of the idea of as yet unsampled cases resembling already cases, which the request for a justification of (empirical) induction took for granted, since if the sampled cases in question consist of green emeralds examined before the chosen  $t$ , it can be equally well described as consisting of grue emeralds. And while for later emeralds to resemble the sampled cases in respect of being green they will have to be green, to resemble those sampled in respect of being grue they will have to be blue. So there is no objective sense to be made of the unsampled cases resembling *tout court* the sampled cases. In the course of his discussion (for a reason alluded to below) Goodman also considers the introduction of another predicate by a definition along the lines of (3.5)/(3.6), namely “bleen” (as  $E$ ), for which  $F$  is as before, while  $G$  and  $H$  change places, now representing the properties of being blue and being green, respectively. We will simplify the situation below, discarding the blue and bleen elements but have had to introduce them here to make intelligible a remark of Goodman’s quoted in the following paragraph, before that simplification is effected. Those wanting more information on the debate over *grue* will find Stalker [49] to be a useful compendium of contributions; for an interesting more recent contribution, see Godfrey-Smith [22].

Apropos of what is essentially the contrast between (3.5) and (3.6), Dunn ([18] p. 472) makes the key point about discriminatory strength: “To begin with, there are several different ways of defining ‘ $x$  is grue,’ equivalent in classical logic, but nonequivalent in relevance logic”. He goes on to consider

these alternatives with a view to seeing whether, given that *blue* and *green* pick out what Dunn calls *relevant properties*, the same will have to be said for *grue* (and Goodman’s *bleen*), noting that this does not follow unless the same is also said – implausibly, according to Dunn – for the property of being examined before a certain time. Protesting against any proposed dichotomy of predicates into *qualitative* and *non-qualitative*, or (we may add in anticipation of David Lewis) *natural* and *non-natural*, or (this time anticipating Dunn) *relevant* and *irrelevant*, Goodman had written ([24], p. 79) to forestall the objection that *grue* is defined in terms of *green*, *blue*, and this “examined before *t*” predicate (which Goodman calls a ‘temporal term’): “But equally truly, if we start with ‘grue’ and ‘bleen’, then ‘blue’ and ‘green’ will be explained in terms of ‘grue’ and ‘bleen’ and a temporal term.” Quoting this ([18], p. 476) Dunn concedes that he “cannot argue with the purely formal point that Goodman has made” and emphasizes that his theory of relevant predication only tells us that if certain predicates are deemed to be relevant (or to express relevant properties) than the theory tells us that certain other predicates must also be counted as relevant – somewhat as the theory of probability does not tell us the probability of particular outcomes (extreme cases of probability 1 and 0 aside), but only what the consequences of a assigning such and such probabilities to such and such outcomes are for the probabilities of other outcomes.<sup>26</sup>

In fact, however, one can argue a little over the ‘purely formal point that Goodman has made’, since it is no longer evident that the interdefinabilities claimed by Goodman still hold when the underlying propositional logic is weakened to FDE. Because so many predicate letters are involved in the full example, we discuss a simpler case here, for which we replace the “blue” in the definition of “grue” by “not green” and so can put the definability point more simply (without reference to “bleen”). That is, we take the grue things, for present purposes, to be, among the examined (before *t*) things, those which are green, and among the unexamined things, those which are not green. Thus we have the following simplified version of the situation recorded by (3.5) and (3.6), with  $Hx$  appearing, instead, as  $\neg Gx$ :

$$(Fx \supset Gx) \wedge (\neg Fx \supset \neg Gx) \tag{3.7}$$

$$(Fx \wedge Gx) \vee (\neg Fx \wedge \neg Gx) \tag{3.8}$$

---

<sup>26</sup>Dunn actually chooses to discuss a variant of *grue* as defined by Goodman which, as he says, “often appears in the literature as a seemingly unconscious simplification of Goodman’s original definition,” though the details do not matter here. (The opening passages of Frank Jackson’s paper in the collection [49] discuss this and other distortions of the original example.)

In other words, we have the contrast between what was notated in Section 3 as  $Fx \equiv^\wedge Gx$ , on the one hand, and  $Fx \equiv^\vee Gx$ , on the other; since the point concerns the propositional logic of the situation, we now drop the “ $x$ ” and treat  $F$  and  $G$  simply as sentence letters (propositional variables). While classically the superscripts play no discriminatory role and we can simply write “ $\equiv$ ” here. Still speaking classically, we can use the equivalence between  $(F \equiv G) \equiv F$  and  $G$  to define *grue* in terms of *grue* (here  $F \equiv G$ ) and the property of being examined before  $t$  (here  $F$ ), since taking  $\vdash$  as the consequence relation of classical logic, we have  $G \dashv\vdash (F \equiv G) \equiv F$ . If  $\vdash$  is taken instead as  $\vdash_{\text{fde}}$ , however, then there is no definable binary connective  $\#$  for which we have  $G \dashv\vdash (F \equiv G) \# F$ , whether  $\equiv$  is understood as  $\equiv^\vee$  or as  $\equiv^\wedge$ .<sup>27</sup> To see this, consider the Dunn–Belnap evaluations  $v, v'$ , with  $v(F) = v(G) = \mathbf{b}$  and  $v'(F) = \mathbf{b}, v'(G) = \mathbf{f}$ . (Since the power matrix semantics is not in play here, we write  $\mathbf{b}$  rather than  $\{T, F\}$  or  $\text{TF}$ , etc.) Calculating, we find that  $v(F \equiv^\vee G) = v'(F \equiv^\vee G) = \mathbf{b}$ . So while  $v$  and  $v'$  agree on  $F$  and on  $F \equiv^\vee G$ , they do not agree on  $G$ , and thus no four-valued binary operation (to interpret a candidate  $\#$  connective) exists for determining the value of  $G$  from those of  $F$  and  $F \equiv^\vee G$ . Since  $\mathbf{n}$  is not involved in these calculations, by the observations in Section 3 (concerning Figure 2), these considerations show that the same holds for  $F \equiv^\wedge G$ . In principle the fact that the value of  $F \equiv G$  for either interpretation of  $\equiv$ , together with that of  $F$  do not fix the value of  $G$  shows only that we cannot interpret  $\#$  in such a way that our hypothetical compound  $(F \equiv^\wedge G) \# F$  and  $G$  always have the same value, this leaves open the possibility that we should still have  $(F \equiv^\wedge G) \# F \dashv\vdash_{\text{fde}} G$  because all we need is that the values of the left and right formulas should be equi-designated. However, a brief examination of cases shows whenever there a four-valued evaluations assigns different values to formulas  $\varphi$  and  $\psi$ , these formulas cannot be FDE-equivalent (mutual  $\vdash_{\text{fde}}$ -consequences, that is); for example, if we have  $\mathbf{t} = v(\varphi) \neq v(\psi) = \mathbf{b}$  – a “both values designated” case – we have  $v(\neg\varphi) = \mathbf{f}$  and  $v(\neg\psi) = \mathbf{b}$ , showing that  $\neg\psi \vdash_{\text{fde}} \neg\varphi$ , and thus, by the considerations about contraposition in Section 2,  $\varphi \vdash_{\text{fde}} \psi$ , giving us the desired failure of FDE-equivalence.

A treatment of these issues subsuming more candidates (than just  $\equiv^\wedge$  and  $\equiv^\vee$ ) would no doubt be welcome, and some may feel little light is thrown by such considerations on the issues Goodman had in mind with his *grue* example – though a sensitivity to unwanted language-dependence (‘*grue* as an illegitimate primitive predicate’ type responses) was certainly a

<sup>27</sup>Similarly, if  $\vdash$  is taken as the consequence relation of intuitionistic logic and  $\leftrightarrow$  is the usual intuitionistic biconditional, there is no binary  $\#$  definable within (or even conservative addable to) intuitionistic logic for which we have  $\psi \dashv\vdash (\varphi \leftrightarrow \psi) \# \varphi$ , for arbitrary formulas  $\varphi, \psi$ . See Exercise 7.31.14(ii) and surrounding discussion in [30].

Goodman trademark, so it is of interest to note that the intertranslatability or interdefinability claims made in the course of pressing such language-dependence objections themselves exhibit a conspicuous *logic-dependence*.<sup>28</sup> In any case, the title of this Appendix was chosen more as an attention-grabbing headline and with a view to acknowledging the Dunn’s comments in [18]; the point of logical interest is rather the question of when a logic provides us with some means of recovering (to within equivalence) a component  $\varphi_i$  from an  $n$ -ary compound  $\star(\varphi_1, \dots, \varphi_n)$  and the remaining components (i.e.,  $\varphi_1, \dots, \varphi_{i-1}, \varphi_{i+1}, \dots, \varphi_n$ ) with the aid of a further  $n$ -ary connective  $\blacksquare$ , in the sense of making  $\blacksquare(\varphi_1, \dots, \varphi_{i-1}, \star(\varphi_1, \dots, \varphi_n), \varphi_{i+1}, \dots, \varphi_n)$  equivalent to  $\varphi_i$ . In our case  $n$  was 2,  $\star$  was  $\equiv$  (with “ $\wedge$ ” or “ $\vee$ ” superscripted), and we denoted by  $\#$  a candidate ‘reciprocal’  $\blacksquare$  connective.

**Acknowledgments,** I am most grateful to Graham Priest for providing me with successive drafts of [42] and other resources, after my interest was piqued as an audience member hearing [41].

## References

- [1] A. R. Anderson and N. D. Belnap, *Entailment: the Logic of Relevance and Necessity, Vol. I*, Princeton University Press, Princeton, NJ 1975.
- [2] Eric Barnes, ‘Beyond Verisimilitude: a Linguistically Invariant Basis for Scientific Progress’, *Synthese* **88** (1991), 309–339.
- [3] Nuel D. Belnap, ‘A Useful Four-Valued Logic’, pp. 8–37 in J. M. Dunn and G. Epstein (eds.), *Modern Uses of Multiple-Valued Logic*, Reidel, Dordrecht 1977.
- [4] Stephen Blamey, ‘Partial Logic’, pp. 261–354 in D. Gabbay and F. Guenther (eds.), *Handbook of Philosophical Logic, Second Edition, Volume 5*, Kluwer, Dordrecht 2002. (Original version: pp. 1–70 in Gabbay and Guenther (eds.), *Handbook of Philosophical Logic, Volume III*, Reidel, Dordrecht 1986.)

---

<sup>28</sup>Mortensen [34], *q.v.* for references, discussed the extent to which David Miller’s and Pavel Tichý’s objections to Popper’s theory verisimilitude. The above discussion of not being able to ‘undo’ a biconditional and recover one of its components bears more strongly on Miller’s later language-dependence objection (with the aid of the “Minnesotan–Arizonan” language) to Tichý’s own positive proposal. For sources, discussion, and related issues, see the index entries in Humberstone [30] under ‘language-dependence objections’, as well as [28], esp. note 28; and also, for a discussion of the verisimilitude case in particular, Barnes [2].

- [5] M. N. Bleicher, H. Schneider and R. L. Wilson, ‘Permanence of Identities on Algebras’, *Algebra Universalis* **3** (1973), 72–93.
- [6] D. A. Bochvar, ‘On a Three-valued Logical Calculus and Its Application to the Analysis of the Paradoxes of the Classical Extended Functional Calculus’ (with an introduction by the translator, Merrie Bergman), *History and Philosophy of Logic* **2** (1981), 87–112. (Originally published in Russian in 1937.)
- [7] I. Bošnjak and R. Madarász, ‘On Power Structures’, *Algebra and Discrete Mathematics* **2** (2003), 14–35.
- [8] Chris Brink, ‘Power Structures and Logic’, *Quaestiones Mathematicae* **9** (1986), 69–94.
- [9] Chris Brink, ‘ $R^-$ -Algebras and  $R^-$ -Model Structures as Power Constructs’, *Studia Logica* **48** (1989), 85–109.
- [10] Chris Brink, ‘A Comment on Relevant Truth Table Logic’, *Journal of Applied Non-Classical Logics* **2**(1992), 243–246.
- [11] Chris Brink, ‘Power Structures’, *Algebra Universalis* **30** (1993), 177–216.
- [12] Chris Brink, ‘From Two- to Four-Valued Logic’, pp. 9–16 in C. Rauszer (ed.), *Algebraic Methods in Logic and Computer Science* (= Vol. 28 of Polish Academy of Sciences, Institute of Mathematics, Banach center Publications), Warsaw 1993.
- [13] John P. Burgess, *Philosophical Logic*, Princeton University Press, Princeton, NJ 2009.
- [14] Michael De and Hitoshi Omori, ‘Dialetheism and Expressibility’, conference presentation at the Priestfest (Conference in Honour of Graham Priest), Melbourne University, June 12–14, 2013.
- [15] A. J. Dale, ‘Smiley’s Matrices and Dunn’s Semantics for Tautological Entailment’, *Logique et Analyse* **23** (1980), 323–325.
- [16] A. J. Dale, ‘Material Equivalence and Tautological Entailment’, *Notre Dame Journal of Formal Logic* **23** (1982), 435–442.
- [17] J. Michael Dunn, ‘Intuitive Semantics for First-Degree Entailments and “Coupled Trees”’, *Philosophical Studies* **29** (1976), 149–168.

- [18] J. Michael Dunn, ‘A Logical Framework for the Notion of *Natural Property*’, pp. 458–497 in J. Earman and J. D. Norton (eds.), *The Cosmos of Science: Essays of Exploration*, University of Pittsburgh Press, PA 1997.
- [19] J. Michael Dunn, ‘Partiality and its Dual’, *Studia Logica* **65** (2000), 5–40.
- [20] Anthony Everett ‘A Note on Priest’s “Hypercontradictions”’, *Logique et Analyse* **36** (1993), 39–43.
- [21] N. D. Gautam, ‘The Validity of Equations of Complex Algebras’, *Archiv. Math. Logik Grundlagenforschung* **3** (1957), 117–124.
- [22] Peter Godfrey-Smith, ‘Goodman’s Problem and Scientific Methodology’, *Journal of Philosophy* **100** (2003), 573–590.
- [23] Robert Goldblatt, ‘Varieties of Complex Algebras’, *Annals of Pure and Applied Logic* **44** (1989), 173–242.
- [24] Nelson Goodman, *Fact, Fiction, and Forecast* (second edn.), Bobbs-Merrill, New York 1965.
- [25] G. Grätzer and H. Lakser, ‘Identities for Globals (Complex Algebras) of Algebras’, *Colloquium Mathematicum* **56** (1988), 19–29.
- [26] G. Grätzer and S. Whitney, ‘Infinitary Varieties of Structures Closed Under the Formation of Complex Structures’, *Colloquium Mathematicum* **48** (1984), 1–5.
- [27] Lloyd Humberstone, ‘Many-Valued Logics, Philosophical Issues in’, pp. 84–91 in Vol. 6 of E. Craig (ed.), *Routledge Encyclopedia of Philosophy*, Routledge, London 1998.
- [28] Lloyd Humberstone, ‘Contra-Classical Logics’, *Australasian Journal of Philosophy* (2000) **78**, 437–474.
- [29] Lloyd Humberstone, ‘Logical Discrimination’, pp. 207–228 in Jean-Yves Béziau (ed.), *Logica Universalis: Towards a General Theory of Logic*, Birkhäuser, Basel 2005. (Second Edn. 2007.)
- [30] Lloyd Humberstone, *The Connectives*, MIT Press, Cambridge MA 2011.
- [31] Lloyd Humberstone, ‘Aggregation and Idempotence’, *Review of Symbolic Logic* **6** (2013), 680–708.

- [32] Reinhard John, ‘On Classes of Algebras Definable by Regular Equations’, *Colloquium Mathematicum* **36** (1976), 17–21.
- [33] E. M. Mogiljanskaja, ‘Non-Isomorphic Semigroups with Isomorphic Semigroups of Subsets’, *Semigroup Form* **6** (1973), 330–333.
- [34] Chris Mortensen, ‘Relevance and Verisimilitude’, *Synthese* **55** (1983), 353–364.
- [35] H. Omori and K. Sano, ‘Generalizing Functional Completeness in Belnap–Dunn’s Four-Valued System’, pp. 117–123 in K. Terui and N. Preining (eds.), *Logic, Algebra and Truth Degrees 2012: Book of Abstracts*, Japan Advanced Institute of Science and Technology, Kanazawa, Japan 2012.
- [36] A. Pietz and U. Rivieccio, ‘Nothing but the Truth’, *Journal of Philosophical Logic* **42** (2013) 125–135.
- [37] J. Płonka, ‘On a Methods of Construction of Abstract Algebras’, *Fundamenta Mathematicae* **61** (1967), 183–189.
- [38] Graham Priest, ‘Hyper-Contradictions’, *Logique et Analyse* **27** (1984), 237–243.
- [39] Graham Priest, ‘Everett’s Trilogy’, *Mind* **105** (1996), 631–647.
- [40] Graham Priest, ‘The Logic of the *Catuskoti*’, *Comparative Philosophy* **1** (2010), 24–54.
- [41] Graham Priest, ‘None of the Above’, talk presented to the Summer Reflectorium of the St Andrews Philosophy Departments (of Moral Philosophy and of Logic & Metaphysics), June 18, 2012.
- [42] Graham Priest, ‘Plurivalent Logics’, *Australasian Journal of Logic*, this issue.
- [43] A. P. Pynko, ‘Functional Completeness and Axiomatizability within Belnap’s Four-Valued Logic and its Expansions’, *Journal of Applied Non-Classical Logics* **9** (1999), 61–105.
- [44] Nicholas Rescher, *Many-Valued Logic*, McGraw-Hill, New York 1969.
- [45] Paul Ruet, ‘Complete Sets of Connectives and Complete Sequent Calculus for Belnap’s Logic’, Technical report, LIENS-96-28, Département de Mathématiques et Informatique (CNRS URA 1327), Ecole Normale Supérieure, Paris 1996.

- [46] A. Shafaat, ‘On Varieties Closed Under the Construction of Power Algebras’, *Bulletin of the Australian Math. Soc.* **11** (1974), 213–218.
- [47] Yaroslav Shramko and Heinrich Wansing, *Truth and Falsehood: An Inquiry into Generalized Logical Values*, Springer, Dordrecht 2011.
- [48] Timothy Smiley, ‘Sense Without Denotation’, *Analysis* **20** (1960), 125–135.
- [49] Douglas Stalker (ed.), *Grue! The New Riddle of Induction*, Open Court Publ., Chicago 1994.
- [50] T. Tamura and J. Shafer, ‘Power Semigroups’, *Math. Japonicae* **12** (1967), 25–32.
- [51] Walter Taylor, ‘Equational Logic’, *Houston Journal of Mathematics* **5** (Supplement, 1979), 1–83.
- [52] Alasdair Urquhart, ‘Emil Post’, pp.429–478 in Dov M. Gabbay and John Woods (eds.), *Handbook of the History of Logic, Vol. 5: Logic from Russell to Church*, Elsevier, Amsterdam 2008.