Submitted on August 8, 2024 to the Notre Dame Journal of Formal Logic
Volume ??, Number ??,

Iterating Both and Neither With Applications to the Paradoxes

Levin Hornischer

Abstract A common response to the paradoxes of vagueness and truth is to introduce the truth-values 'neither true nor false' or 'both true and false' (or both). However, this infamously runs into trouble with higher-order vagueness or the revenge paradox. This, and other considerations, suggest iterating 'both' and 'neither': as in 'neither true nor neither true nor false'. We present a novel explication of iterating 'both' and 'neither'. Unlike previous approaches, each iteration will change the logic, and the logic in the limit of iteration is an extension of paraconsistent quantum logic. Surprisingly, we obtain the same limit logic if we use (a) both and neither, (b) only neither, or (c) only neither applied to comparable truth-values. These results promise new and fruitful replies to the paradoxes of vagueness and truth. (The paper allows for modular reading: for example, half of it is an appendix studying involutive lattices to prove the results.)

1 Introduction

In addition to the classical truth-values *true* (1) and *false* (0), logicians also explored adding a third truth-value for *neither true nor false* (here written as $\{0,1\}_n$), and a fourth truth-value for *both true and false* (here written as $\{0,1\}_b$). The former was prompted by situations where sentences are neither true nor false: for example, in borderline cases of vague predicates ('A hundred grains of sand make a heap'); or in the case of the liar sentence ('This sentence is false') where having either classical truth-value leads to inconsistency. The latter was prompted by situations where sentences are both true and false: for example, when understood with respect to a database which happens to be inconsistent. (And some also argued that the former cases should actually be both true and false.)

But why stop here? If this motivation is conceded, it doesn't seem to be exhausted after just one iteration of 'both' and 'neither'. Consider the following examples.

2010 Mathematics Subject Classification: Primary 03G10, 06B20, 03B50; Secondary 18A35
Keywords: involutive lattice, first-degree entailment, paraconsistent quantum logic, Belnap computer, higher-order vagueness, liar paradox, revenge paradox

(1) Higher-order vagueness: The claim that a thousand grains of sand make a heap seems to be neither *true* nor *neither true nor false*, since it is 'in between' these two truth-values (here written as $\big\{\{1\},\{0,1\}_n\big\}_n$). (2) The revenge sentence 'This sentence is neither false nor neither true nor false': If it had one of the three truth-values suggested by the 'solution' to the liar sentence (*true*, *false*, or *neither*), we get inconsistency, so we could iterate the 'solution' and take it to be neither *true* nor *neither true nor false*. (3) Maybe a sentence can be both *neither true nor false* and *both true and false* according to the fusion of two databases, one claiming the former and one the latter (here written as $\big\{\{0,1\}_n,\{0,1\}_b\big\}_b$). (4) In Buddhist logic, Nagarjuna discusses cases where "Neither both nor neither should be asserted" [cited and discussed in 34, 35]. (5) Or, tongue-in-cheek, consider the last lines of the movie *Mowgli: Legend of the Jungle*: "Mowgli, man and wolf, both and neither".

Thus, even though this question of iterating 'both' and 'neither' may first sound like a mere curiosity, it is philosophically useful: especially for the (higher-order) paradoxes of vagueness and truth. In fact, the question is at least almost half a century old: it was already asked by Meyer [27, 19] and others.[1]

> [I]f we take seriously both true and false and neither true nor false separately, what is to prevent our taking them seriously conjunctively? As in "It is both true and false and neither true nor false that snow is white". That way, in the end, lies madness.

On one natural way of 'taking them conjunctively', this madness was subsequently investigated by Priest [31] and later, in greater generality, by Shramko and Wansing [37]. Both conclude that, actually, the result is quite coherent.

In this paper, we revisit that question. We consider another, to the best of our knowledge novel way of 'taking them conjunctively'. We generate in a different way new truth-values from old ones, using 'both' and 'neither'. This also seems natural, but has more intricate—albeit intriguing—structure. And it promises fruitful applications to the paradoxes.

In more detail, the structure and results of this paper are as follows. In section 2, we motivate why to iterate 'both' and 'neither' by discussing the mentioned applications to databases (Belnap computers), truth, and vagueness. In section 3, we describe the construction process in detail. And in section 4, we describe what lies at the end of iterating this process forever. In section 5, we investigate the resulting logic of these truth-values. Unlike the existing approaches, the logic fails to be distributive already after the second iteration. In fact, we prove that, after every iteration, there will be some logical law that fails which held previously. The tightest known upper bound is paraconsistent quantum logic (PQL): all its laws are valid for iterated both-and-neither, but whether it is exactly the logic of iterated both-and-neither is a curious open question. In section 6, we investigate the relationship between using (a) both and neither (suggestive for Belnap computers), (b) only neither (suggestive for truth), or (c) only neither applied to comparable truth-values (suggestive for vagueness). After one iteration, these yield different logics (e.g., strong Kleene vs. FDE). However, remarkably, in the limit of the iteration, their logics coincide. In section 7, we explore the philosophical consequences of these results. They promise new and fruitful replies to the paradoxes of vagueness and truth. We conclude in section 8 with some open questions.

The proof of the results are in appendices A–D. They constitute roughly half of the paper and are a mathematical study of involutive lattices developing the above philosophical ideas. Appendix A provides a more general but unified definition of both-and-neither like operations and the algebras that they form (this greater generality is needed afterward in the proofs). Appendix B describes the fixed point of iterating these operations forever. Appendix C investigates the congruences of the both-and-neither algebras to understand their logics (for section 5). Appendix D develops embeddability results between both-and-neither, neither-only, and comparable-neither-only algebras to show that they yield the same logic in the limit (for section 6).

2 Motivation

We discuss three applications motivating the usage and iteration of 'both' and 'neither'. First, Belnap computers provide an interpretation of a single use of 'both' and 'neither', and considering networks of Belnap computers motivates iterations of 'both' and 'neither'. Second, the liar paradox provides an interpretation of a single use of 'neither', and the revenge paradox motivates iterations of 'neither'. Third, vagueness provides an interpretation of a single use of 'neither', and higher-order vagueness motivates iterations of 'neither'.

These surely are not the only applications. As already mentioned, rather than taking vague or liar sentences to be neither true nor false (e.g., supervaluationism in vagueness [14] or paracomplete theories of truth [24]), some approaches take them to be both true and false (subvaluationism [6] and paraconsistent truth [30]). So we could develop the motivating applications equally well on those approaches; the choice for the former here is just for illustration. Concerning yet further applications, we already mentioned Buddhist logics, and others might possibly come, e.g., from the rich philosophical ideas behind Belnap–Dunn logic [28] or from the related logic of epistemic modals [18, 17].

2.1 Belnap computers & networks thereof A well-known connection between computers—specifically, partial recursive functions—and 'neither true nor false' is the (strong or weak) Kleene logic [23, ch. XII]. Since computer algorithms may fail to terminate, the 'neither' truth-value is interpreted along the lines of 'the computer neither establishes that the sentence is true nor that it is false'. Here, however, we consider another connection between computers and 'both' and 'neither'.

Belnap [3, 4] forcefully argued that the logic with which a computer should think is *first-degree entailment* FDE: A computer might receive various inputs about whether a given statement p is true. It might get no input, it might get some and only inputs saying that p is true, it might get some and only inputs saying that p is false, and it might get both some inputs saying that p is true and some inputs saying that p is false. This yields, respectively, the four truth-values $\emptyset, \{1\}, \{0\}, \{0,1\}$, i.e., the possible unions of the inputted classical truth-values.[2] So the computer has to reckon with the powerset $\mathbf{4} := \mathscr{P}\big(\{0,1\}\big)$ of the possible input truth-values $\mathbf{2} := \{0,1\}$.

The next question is how should the computer reason with these truth-values: i.e., how should it compute truth-values of complex sentences given the truth-values for the atomic statements? It turns out, the answer is this: While the subset order on the set of truth-values $\mathscr{P}\big(\{0,1\}\big)$ orders them by how much information they provide, we can also order them by how close they are to being true (and only true) as shown in figure 1. Then, if $v : \mathsf{P} \to \mathbf{4}$ is an assignment of truth-values to the set of atomic sentence $\mathsf{P} = \{p, q, \dots\}$, we can extend it to complex sentences as follows:

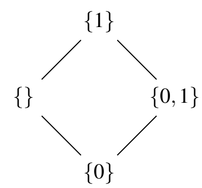

Figure 1 The truth-order on $\mathbf{4} = \mathscr{P}\big(\{0,1\}\big)$.

- $v(\varphi \vee \psi) =$ the least upper bound of $v(\varphi)$ and $v(\psi)$ in the truth-order on $\mathbf{4}$.
- $v(\varphi \wedge \psi) =$ the greatest lower bound of $v(\varphi)$ and $v(\psi)$ in the truth-order on $\mathbf{4}$.
- $v(\neg\varphi) =$ the set of classical negations of the classical truth-values in $v(\varphi)$.

The logic FDE then emerges as preservation of truth [32, ch. 8]: Writing $D := \big\{\{1\},\{0,1\}\big\}$ for the truth-values that contain truth, we say that a set Γ of sentences FDE-*entails* a sentence φ (written $\Gamma \vDash_{\mathsf{FDE}} \varphi$) iff, for any valuation $v : \mathsf{P} \to \mathbf{4}$, if $v(\psi) \in D$ for every $\psi \in \Gamma$, then $v(\varphi) \in D$.

Those computers are then also called *Belnap computers*. And, in the spirit of iterating this process, Shramko and Wansing [36] ask: what if we have a *network* of Belnap computers—how should *it* think? Specifically, if we have a server receiving input about the truth-value of an atomic sentence p from several Belnap computers, what truth-value should the server assign to p? Taking the same idea as above, if the server receives the truth-values $a_1, \dots, a_n$ (with $n \geq 0$, i.e., it might not receive anything), then it takes p to have the new truth-value $\{a_1, \dots, a_n\}$. Thus the set of new truth-values that the server should use is $\mathbf{16} := \mathscr{P}\big(\mathbf{4}\big)$. In determining the logic, there is a complication that, in addition to the truth-order, there also is a falsity-order. But when focusing on just the truth-order (or, equivalently, on just the falsity-order), then Shramko and Wansing [36] show that the resulting logic again is FDE. Hence, although the expressive power of the truth-valued increased, the logic remained the same.

But other approaches are possible, too. So far, the computer acted as an *accumulator*: it deterministically collected all the truth-values provided by the sources (modulo order and multiplicity). But, as we now explore, the computer could also be a *selector*: non-deterministically merging inputted truth-values.

Let's first consider only two input sources A and B saying that the sentence in question has a classical truth-value a and b, respectively. Now the computer (or reasoner) has to assess this information—maybe by taking into account the credibility and certainty of the source—and merge it into a truth-value. Its policy could be this. If the sources agree (i.e., $a = b$), it goes with their judgment and takes $\{a\}$ $(= \{b\})$. If the sources don't agree (i.e., $a \neq b$), then

- if the computer accepts A but not B, it takes $\{a\}$,
- if the computer accepts B but not A, it takes $\{b\}$,
- if the computer accepts neither A nor B, it takes $\{a,b\}_n$ representing *neither a nor b*,
- if the computer accepts both A and B, it takes $\{a,b\}_b$ representing *both a and b*.

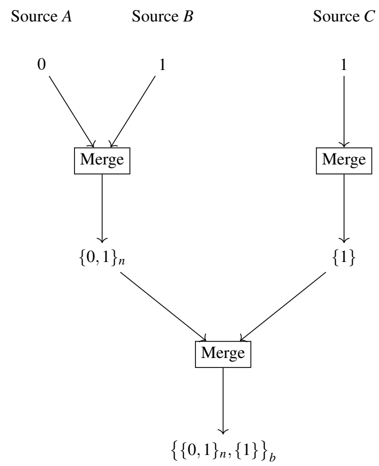

Figure 2 Example of merging the information of three sources.

Thus, for the possible choices of classical truth-values a and b, we again obtain the four truth-values $\{0\}, \{0,1\}_n, \{0,1\}_b, \{1\}$ and hence the logic FDE.

But what if a third source C becomes available and the computer now has to merge its decision with this new information? This could, for example, look like in figure 2: The original two sources A and B inputted 0 and 1, respectively. The computer took none of them to be convincing, hence merging their input into the new truth-value $\{0,1\}_n$. Now the computer first processes the new source C alone. Since it is a single source, there is no conflict (just like with two agreeing sources), and it takes over its truth-value $\{1\}$. Next, the computer has to merge the former result $\{0,1\}_n$ with the new $\{1\}$. Say it takes both of these (preprocessed) sources convincing, thus producing $\big\{\{0,1\}_n, \{1\}\big\}_b$.

Of course, many other versions of this example are possible. The sources could have different values. The computer could make different decisions in merging. There could be more sources, merged in different order (resulting in deeper and different nestings of $\{\cdot\}_n$ and $\{\cdot\}_b$). But, as before, the question is: If we consider *all* these possibilities—i.e., if we generate all these new truth-values—which laws are still valid that the computer can reason with? In short, what is the logic of these new truth-values? We answer this below (and it will be different from FDE).

2.2 The liar paradox & revenge Infamously, the liar sentence ('This sentence is false') poses a paradox in the context of the two classical truth-values: it is *true* iff it is *false* (i.e., not *true*). A popular response—in line with Kripke's theory of truth [24]—is to take the liar sentence to be *neither true nor false*. But this move faces the well-known revenge paradox: The *revenge sentence*

$$\text{This sentence is either } \textit{false} \text{ or } \textit{neither true nor false}. \tag{1}$$

again poses a contradiction in the context of the three values *true*, *false*, and *neither true nor false*: (1) is *true* iff it is not *true*.[3]

Now, it is suggestive to iterate the previous solution: We add the new truth-values *neither a nor b* where a and b are truth-values that we had so far. Then we regard the revenge sentence (1) as *neither* true *nor* neither true nor false. Then no contradiction arises, because (1) is not '*true*' and also not '*false* or *neither true nor false*'.

Of course, we can now build a new revenge sentence for those iterated truth-values. And we can respond again by further iterating 'neither'. Thus, a solution to the revenge paradox occurs in the limit of this process. We will explore this below.

Priest [31] arrived at the question of iterating 'both' in a related way. The starting point is that the liar sentence is *both true and false*, in line with a paraconsistent theory of truth [30]. The analogous revenge sentence then also leads to further truth-values like *both* true only *and* both true and false. Priest [31] formalizes this process and shows that, although the expressive power of the truth-values increases, the logic remains the same as after applying the process only once. Here we instead consider a broadly Kripkean theory of truth (that rather opts for paracompleteness than paraconsistency) and, on our explication of the iteration also the logic changes.

Moreover, one should also mention Field's theory of truth [12, 13]. To address the worry that the language of this theory is not expressive enough to state the 'defectiveness' of the liar sentence, the language contains a 'determinacy operator' D. So one can say that the liar sentence λ is neither determinately true ($D\lambda$) nor determinately false ($D\neg\lambda$). To deal with the revenge sentence λ_2 which, in this context, says of itself that it is not determinately true, the language has another determinacy operator D_2. So one can say that λ_2 is neither D_2-determinately true ($D_2\lambda_2$) nor D_2-determinately false ($D_2\neg\lambda_2$). For the next revenge sentence λ_3 which says of itself that it is not D_2-determinately true, the language has yet another determinacy operator D_3, and so on. So the language has a whole hierarchy of determinacy operators $D = D_1, D_2, \ldots, D_\alpha, \ldots$ (in fact, transfinitely many). And one might view the 'neither D_α-determinately true nor D_α-determinately false' as yet another way of iterating 'neither'.

2.3 Vagueness & higher-order vagueness Vagueness is another common reason for introducing *neither true nor false*. Consider a sentence p involving a vague predicate, like 'This is a heap of sand'. As a default, we could think that p can have the classical truth-values 0 and 1. But then there is an implausibly sharp cut-off between heaps and non-heaps. When p is about a borderline case of a heap, say 100 grains of sand, we neither want to call p true nor call p false.

It is more natural to speak of three kinds of cases: where p is definitely true (write $\{1\}$), where p is definitely false (write $\{0\}$), and where p is neither true nor false (write $\{0,1\}_n$).

However, this move faces the issue of *higher-order vagueness* (see, e.g., [14, 42]). Just as it was a blurry line where 1 ended and 0 started, it seems to be an equally blurry line where $\{1\}$ ends and $\{0,1\}_n$ starts (and similarly between $\{0,1\}_n$ and $\{0\}$).

Again, it now is suggestive to iterate the previous reply. So if p is said about the just mentioned border-borderline case of a heap, we take p to be *neither* definitely true *nor* neither true nor false ($\{\{1\},\{0,1\}_n\}_n$). And so on, for further iterations. We need a precise understanding of the neither operation to understand what 'logic of vagueness' this gives rise to.

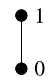

Figure 3 The classical truth-value algebra **2**.

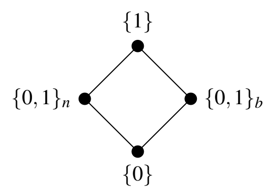

Figure 4 The truth-value algebra $\mathbf{4} = \mathsf{BN}(\mathbf{2})$.

3 The both and neither operations

We describe how to build from two 'old' truth-values a and b the 'new' truth-values $\{a,b\}_b$ (both a and b) and $\{a,b\}_n$ (neither a nor b). We do so by generalizing the intuitions that lead from classical logic with its two truth-values to FDE with its four truth-values. We find that the algebraic structure of the truth-values thus generated always is what is known as an involutive lattice. So our generalized both-and-neither operation takes an involutive lattice A and produces another involutive lattice $\mathsf{BN}(A)$ consisting of all the new values $\{a,b\}_b$ and $\{a,b\}_n$ with $a,b \in A$. Special cases are the neither-only involutive lattice $\mathsf{N}(A)$ and the comparable-neither-only involutive lattice $\underline{\mathsf{N}}(A)$.

3.1 The starting point: classical logic The starting point is classical logic with its two truth-values $\mathbf{2} := \{0,1\}$. They are naturally ordered by 'being more true than': i.e., $0 < 1$ as shown in figure 3. Conjunction and disjunction are simply interpreted as min and max in that ordering, and negation is interpreted by $\neg 0 = 1$ and $\neg 1 = 0$.

3.2 The first iteration: FDE The next, well-established step is to take copies of the truth-values in **2** (often written as $\{0\}$ and $\{1\}$) and add the truth-values *neither true nor false* and *both true and false* (often written as $\emptyset$ and $\{0,1\}$, respectively; here we write $\{0,1\}_n$ and $\{0,1\}_b$). This collection, which we denote **4**, is ordered by 'more true than' as shown in figure 4. (We have already seen this as figure 1.) We will now closely consider the three pieces of intuitions that yield this structure of truth-values to guide our general definition of 'both' and 'neither'.

Intuition 1: truth-values. So we obtain the four truth-values by taking a copy of the original truth-values 0 and 1 and by also applying 'both' and 'neither' to them yielding *neither* 0 *nor* 1 and *both* 0 *and* 1. These are all new ones because, whatever *neither a nor b* and *both a and b* are (for old truth-values a and b), we expect two constraints: First, *neither a nor b* arguably means the same as *neither b nor a*, and similarly for 'both'. So we don't need to additionally add *neither* 1 *nor* 0 or *both* 1 *and* 0. Second, saying *neither a nor b* is only felicitous when $a \neq b$, and similarly for 'both'. It sounds odd to say 'It neither rains nor rains'.[4]

Here is a way to define (or rationally reconstruct) *neither* 0 *nor* 1 and *both* 0 *and* 1 so as to meet these constraints: First, for two old truth-values a and b and $k \in \{n, b\}$, define

$$\{a,b\}_k := \begin{cases} \{a\} & \text{if } a = b \\ (\{a,b\},k) & \text{if } a \neq b. \end{cases} \tag{2}$$

Then define *neither a nor b* as $\{a,b\}_n$ and *both a and b* as $\{a,b\}_b$. This definition satisfies the first constraint (since $\{a,b\}_k = \{b,a\}_k$). It satisfies the second constraint since b (both) or n (neither) only occur if $a \neq b$. And the definition has the convenient side-effect that, if we have an old truth-value a, we can take the copy $\{a\}$ of it, and *stipulate* it to be $\{a,a\}_n = \{a,a\}_b$ purely for more efficient notation.

Intuition 2: order. The intuition behind the 'more true than' order is this: anything is more true than pure falsity $\{0\}$, and pure truth $\{1\}$ is more true than anything, so they are the least and greatest elements, respectively. Now, *neither* 0 *nor* 1 is in between: it is more true than pure falsity since it does not contain falsity, but it is less true than pure truth since it does not contain truth. Similarly, *both* 0 *and* 1 is in between: it is more true than pure falsity since it contains truth, but it is less true than pure truth since it contains falsity. Finally, since it is not clear which is a bigger defect—not containing truth or containing falsity—*neither* 0 *nor* 1 and *both* 0 *and* 1 are not comparable.

With this order, **4** is a lattice: i.e., for any two elements x and y, the greatest lower bound $x \wedge y$ and the least upper bound $x \vee y$ exist. These lattice operations—called *meet* and *join*, respectively—then can interpret the conjunction and disjunction (of the language interpreted over these truth-values). Also, the lattice is bounded: it has a least and a greatest element, which interpret the logical symbols $\bot$ (falsum) and $\top$ (verum), respectively.

Intuition 3: negation. The idea of classical negation is that the sentence $\neg\varphi$ has truth-value 1 iff φ has truth-value 0. Hence $\neg 1 = 0$ and $\neg 0 = 1$. This extends to the new truth-values via the following (intuitive) equivalences:

$$\begin{aligned} & \neg\varphi \text{ has truth-value } \textit{neither } 0 \textit{ nor } 1 \\ \Leftrightarrow\ & \neg\varphi \text{ neither has truth-value } 0 \text{ nor truth-value } 1 \\ \Leftrightarrow\ & \varphi \text{ neither has truth-value } \neg 0 = 1 \text{ nor truth-value } \neg 1 = 0 \\ \Leftrightarrow\ & \varphi \text{ has truth-value } \textit{neither } 0 \textit{ nor } 1. \end{aligned}$$

So $\neg\{0,1\}_n = \{\neg 0, \neg 1\}_n = \{0,1\}_n$. Similarly, $\neg\{0,1\}_b = \{\neg 0, \neg 1\}_b = \{0,1\}_b$.

Summary: algebra. So, the set of four truth-values **4** still forms a bounded lattice (hence interprets conjunction, disjunction, falsum, verum) and it has a unary operator $\neg$ (hence interprets negation). Thus, we have a (universal) algebra with the signature $(2,2,1,0,0)$, i.e., a set with two binary operations $(\wedge, \vee)$, a unary operation $(\neg)$, and two constants $(0,1)$. Hence, in the spirit of universal algebra, we should be more precise about which kind of algebra we are dealing with. Arguably the most famous algebras of this signature are Boolean algebras; which **2** still is, but **4** is not anymore (e.g., $x \vee \neg x = 1$ fails). But **4** still is what is known as a bounded involutive lattice.[5]

Definition 3.1 A *bounded involutive lattice* is a structure $(A, \vee, \wedge, \neg, 0, 1)$ such that $(A, \vee, \wedge, 0, 1)$ is a bounded lattice (see, e.g., [5, 25] for the definition) and the unary operation $\neg$ satisfies the identities:

$$\neg\neg x = x \qquad\qquad \neg(x \wedge y) = \neg x \vee \neg y \qquad\qquad \neg 1 = 0.$$ [6]

For brevity, we drop the word 'bounded', so *involutive lattice* always means bounded involutive lattice (and *not* just a lattice with an involution). As usual, we write A both for the structure $(A, \vee, \wedge, \neg, 0, 1)$ and its underlying set.

In fact, **4** still is distributive, i.e., satisfies $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$. However, we will see that, at the next iteration of 'both' and 'neither', distributivity will fail, but we will always get involutive lattices.

3.3 The general case In the general case, assume we have a set A of truth-values with the structure of an involutive lattice—like **4**. How do we apply the 'both' and 'neither' operations to get an involutive lattice $A' = \mathsf{BN}(A)$? We describe it now by generalizing the preceding intuitions. Figure 5 below shows what the resulting $\mathsf{BN}(\mathbf{4})$ will look like.

Following the first intuition, we build, from the 'old' truth-values of A, the following set of 'new' truth-values

$$A' := \big\{\{a,b\}_k : a,b \in A, k \in \{n,b\}\big\}. \tag{3}$$

In particular, A' contains the copy $\{a\}$ of each $a \in A$.

The second intuition concerns the order of A'. It demanded that the new elements $\{a,b\}_n$ and $\{a,b\}_b$ are incomparable and in between the old elements $a \wedge b$ and $a \vee b$. Thus, we formally define the new order $\leq'$ on A' using the old order $\leq$ on A as follows: For $\{a,b\}_k$ and $\{c,d\}_l$ in A', set

$$\{a,b\}_k \leq' \{c,d\}_l \text{ iff } \{a,b\}_k = \{c,d\}_l \text{ or } a \vee b \leq c \wedge d. \tag{4}$$

The third intuition concerns negation. It immediately generalizes to define the new negation $\neg'$ on A' from the old negation $\neg$ on A: If $\{a,b\}_k$ is in A', then

$$\neg'\{a,b\}_k := \{\neg a, \neg b\}_k. \tag{5}$$

Now, the following theorem—which we prove in appendix A—shows that this way we indeed get again an involutive lattice.

Theorem 3.2 *Let $A = (A, \vee, \wedge, \neg, 0, 1)$ be an involutive lattice. Define A' is as in (3) with order $\leq'$ as in (4) and $\neg'$ as in (5). Set $0' := \{0\}$ and $1' := \{1\}$. Then $\mathsf{BN}(A) := (A', \vee', \wedge', \neg', 0', 1')$ is an involutive lattice where for any $x = \{a,b\}_k$ and $y = \{c,d\}_l$ in A',*

$$x \vee' y = \begin{cases} y & \text{if } x \leq' y \\ x & \text{if } y \leq' x \\ \{a \vee b \vee c \vee d\} & \text{if } x \text{ and } y \text{ are } \leq'\text{-incomparable} \end{cases}$$

and similarly for $\wedge$.

As mentioned, $\mathsf{BN}(\mathbf{4})$, the next step in iterating 'both' and 'neither', is depicted in figure 5. It is the first one to contain the truth-value of the Mowgli sentence: $\big\{\{0,1\}_b, \{0,1\}_n\big\}_b$. Further iterations become too big to draw.

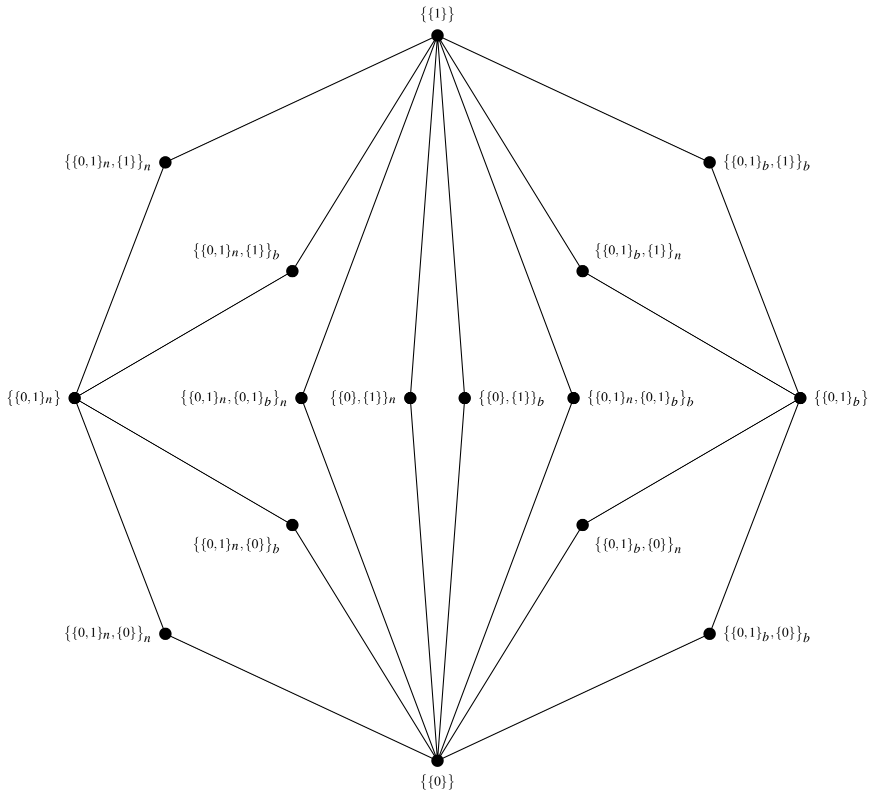

Figure 5 The truth-value algebra $\mathsf{BN}(\mathbf{4})$.

3.4 Neither-only, both-only, comparable-neither-only When constructing $\mathsf{BN}(A)$ from A, we used 'both' and 'neither', but in applications we may want to restrict their usage. For example, in the case of truth, we might only allow 'neither' (on a broadly Kripkean theory of truth) or allow only 'both' (on a broadly Priestian theory of truth). Or in the case of vagueness, where the intuition was that $\mathsf{N}(a,b)$ is the truth-value *in between* a and b, it is natural to only allow 'neither' *and* to require that a and b are comparable in A, i.e., either $a \leq b$ or $b \leq a$ (so it makes sense to speak of a truth-value lying in between the two).

We can recover these special cases as *subalgebras* of $\mathsf{BN}(A)$, i.e., subsets of the underlying set of $\mathsf{BN}(A)$ that are closed under the operations $\vee, \wedge, \neg, 0, 1$ of $\mathsf{BN}(A)$ (proven in appendix A).

- $\mathsf{BN}(A)$ the *both-and-neither algebra* of A
- $\mathsf{N}(A) := \big\{\{a,b\}_n : a,b \in A\big\}$ the *neither-only algebra* of A
- $\mathsf{B}(A) := \big\{\{a,b\}_b : a,b \in A\big\}$ the *both-only algebra* of A
- $\underline{\mathsf{N}}(A) := \big\{\{a,b\}_n : a,b \in A \text{ comparable}\big\}$ the *comparable-neither-only algebra* of A.

The neither-only and the both-only algebra are isomorphic (just replace n by b). And we have the inclusions $\underline{\mathsf{N}}(A) \subseteq \mathsf{N}(A) \subseteq \mathsf{BN}(A)$. In section 6, we discuss surprising embeddability results in the other direction (embedding $\mathsf{BN}(A)$ in iterated applications of $\underline{\mathsf{N}}$ to A). But let's first further investigate the iteration of 'both' and 'neither'.

4 Iterating forever

If we start with the classical truth-values $\mathbf{2}$ and keep on iterating 'both' and 'neither', what—as Meyer asked—lies in the end? Another way of asking this is: what is the smallest involutive lattice A that contains $\mathbf{2}$ and applying 'both' and 'neither' to it does not yield anything new, i.e., A is isomorphic to $\mathsf{BN}(A)$?

Computer scientists know the answer to such questions: A is the initial algebra for the construction BN. Or A is the least fixed point of the construction BN above $\mathbf{2}$.

Concretely, we construct A as follows (the details are in appendix B). First, we can naturally embed any involutive lattice A into $\mathsf{BN}(A)$ by mapping an old element a to its copy $\{a\}$: i.e., we have a function $e : A \to \mathsf{BN}(A)$ mapping a to $\{a\}$. And this is an involutive lattice embedding, i.e., an injective function that preserves $\vee, \wedge, \neg, 0, 1$. We call e the *natural embedding*. Next, whenever we have an embedding $f : A \to B$, we also have the embedding $\mathsf{BN}(f)$ mapping $\{a,b\}_k$ in $\mathsf{BN}(A)$ to $\{f(a), f(b)\}_k$ in $\mathsf{BN}(B)$. So we get the following chain of embeddings (with $\mathsf{BN}^2(\cdot) := \mathsf{BN}(\mathsf{BN}(\cdot))$, etc.)

$$\mathbf{2} \xrightarrow{\;e\;} \mathsf{BN}(\mathbf{2}) \xrightarrow{\mathsf{BN}(e)} \mathsf{BN}^2(\mathbf{2}) \xrightarrow{\mathsf{BN}^2(e)} \mathsf{BN}^3(\mathbf{2}) \longrightarrow \dots \tag{6}$$

For $n \leq m$, write $A_n := \mathsf{BN}^n(A)$ and $e_{nm} : A_n \to A_m$ for the result of chaining the above embeddings from A_n to A_m. Now, we take the 'limit' of this chain, i.e., collect all the truth-values built along the way. Formally, this is known as a direct limit (in model theory) or co-limit (in category theory).[7] The direct limit (A, e_n) is given as follows:

- A is the union $\bigcup_{n \geq 0} A_n$ modulo the equivalence relation $a \sim b$ iff there is $n \leq m$ with $e_{nm}(a) = b$ or $e_{nm}(b) = a$ (i.e., one is just the embedded version of the other). Let's write $[a]$ for the equivalence class of a.
- The operations between equivalence classes are given via their representatives: e.g., $[a] \vee [b] = [a \vee b]$ where, without loss of generality, the representatives a and b are chosen from the same A_n, so $a \vee b$ is defined in A_n. So they still satisfy the defining axioms of an involutive lattice.
- The embeddings $e_n : A_n \to A$ are given by mapping a to $[a]$.

We denote A also by $\mathsf{BN}^\infty(\mathbf{2})$ and call it the direct limit (leaving the embeddings e_n implicit).

Using the theory of initial algebras from theoretical computer science [25, 39, 1], we can show that $\mathsf{BN}^\infty(\mathbf{2})$ is indeed the desired limit:

Theorem 4.1 *The direct limit $\mathsf{BN}^\infty(\mathbf{2})$ of the chain (6) is the least fixed point under BN: i.e., $\mathsf{BN}^\infty(\mathbf{2})$ is isomorphic to $\mathsf{BN}(\mathsf{BN}^\infty(\mathbf{2}))$, and if B is another such nontrivial*[8] *involutive lattice, then $\mathsf{BN}^\infty(\mathbf{2})$ can be embedded into B. The analogous result holds for the other constructions $\mathsf{N}, \mathsf{B}, \underline{\mathsf{N}}$.*

5 The logic

By iterating 'both' and 'neither', we now have the algebra of truth-values $\mathsf{BN}^\infty(\mathbf{2})$. But we still need a logic telling us how to reason with these truth-values: we want to know when an argument that uses sentences having these truth-values is a good one. We also need this for the applications: If we use these truth-values to model our use of, say, vague expressions, we also want to know the logic of these expressions, i.e., their fundamental principles. And similarly for the other applications.

In subsection 5.1, we find this logic by generalizing (again) the intuitions from the four-valued case. The guiding idea is that sentence φ entails sentence ψ iff, for every valuation, the value of φ is less-or-equal-to the value of ψ—so entailment is, in a sense, truth-preservation.

In fact, we can consider the logic at each step of the iteration process:

- starting with classical logic: $\vDash_{\mathbf{2}}$,
- moving to FDE: $\vDash_{\mathsf{BN}(\mathbf{2})}$,
- continuing with the finitely iterated logics: $\vDash_{\mathsf{BN}^2(\mathbf{2})}, \vDash_{\mathsf{BN}^3(\mathbf{2})}, \dots$
- and eventually considering the limit logic: $\vDash_{\mathsf{BN}^\infty(\mathbf{2})}$.

This naturally poses the questions whether all these logics really are different or whether they collapse at some point. After all, in the mentioned previous investigations of iterating 'both' and 'neither', the logics collapsed at the second state: i.e., the logics of the third and fourth bullet points where identical to that of the second. In subsection 5.3, we show that here there is no collapse.

Finally, in subsection 5.4, we show that all these infinitely many both-and-neither logics lie in between the well-known logics *first-degree entailment* (FDE) and *paraconsistent quantum logic* (PQL). Whether the limit logic $\vDash_{\mathsf{BN}^\infty(\mathbf{2})}$ is identical to PQL is an intriguing open problem.

5.1 Defining logical consequence The intuition for defining logical consequence (or entailment) for four-valued logic is that $\varphi \vDash \psi$ means that, for every valuation v, $v(\varphi) \leq v(\psi)$ [3, 4]. This straightforwardly generalizes to any involutive lattice A other than $\mathsf{BN}(\mathbf{2})$: all that is needed are the operations $\vee, \wedge, \neg, 0, 1$ to provide the order $\leq$ and to interpret the logical connectives $\vee, \wedge, \neg, \bot, \top$. We can also extend this definition to a set Γ of sentences entailing a sentence φ by requiring there to be a finite subset of Γ whose conjunction entails φ. The formal definitions read as follows.

We use the *language* of propositional logic: We fix a countably infinite set At of propositional *atoms* $p_0, p_1, p_2, \dots$. The *sentences* are the atoms and, whenever φ and ψ are sentences, so are $\varphi \vee \psi, \varphi \wedge \psi, \neg\varphi, \top, \bot$. We use $p, q, r, \dots$ as variables ranging over atoms, and $\varphi, \psi, \chi, \dots$ as variables ranging over sentences.

If $A = (A, \vee, \wedge, \neg, 0, 1)$ is an involutive lattice, an *A-valuation* v is a function $v : \mathsf{At} \to A$. It is recursively extended to all sentences:

$$\begin{aligned} v(\varphi \vee \psi) &:= v(\varphi) \vee v(\psi) & v(\top) &:= 1 \\ v(\varphi \wedge \psi) &:= v(\varphi) \wedge v(\psi) & v(\bot) &:= 0 \\ v(\neg\varphi) &:= \neg v(\varphi). \end{aligned}$$

Definition 5.1 Let $A = (A, \vee, \wedge, \neg, 0, 1)$ be an involutive lattice. Let Γ be a set of sentences and φ a sentence. Define $\Gamma \vDash_A \varphi$ as: there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that, for all A-valuations v,

$$\bigwedge_{\psi \in \Gamma_0} v(\psi) \leq v(\varphi).$$ [9]

We call $\vDash_A$ the *consequence relation* or *logic* of A.

This definition of logical consequence is in the spirit of algebraic logic (see, e.g., [15, 9]). There also is another approach to define logical consequence in the context of many-valued logic by using designated values (see, e.g., [32, ch. 7]). It coincides with the above algebraic one for four-valued logic [15]. But generalizing it to iterated 'both' and 'neither' logics yields quite a wild logic: we discuss this in the next subsection (but it can also be skipped). That is why we don't take it as the guiding generalization.

5.2 Logical consequence via designated values The idea of the designated-value approach is to declare a subset D of the set of truth-values A to be 'designated' or 'desired'. Hence one requires logical consequence to preserve those: for all valuations, if all the premises are designated, so must be the conclusion. For four-valued logic $\mathsf{BN}(\mathbf{2})$, the designated values are 1 and $\{0,1\}_b$ because it is exactly those that 'contain truth'. Thus, logical consequence in the sense of preserving designated values also spells out a sense of truth-preservation. This intuition straightforwardly generalizes to $\mathsf{BN}^\infty(\mathbf{2})$ as follows:

- Starting point: Truth-values $\mathbf{2}$. Designated $D_0 := \{1\}$.
- First iteration: Truth-values $\mathsf{BN}(\mathbf{2})$. Designated D_1: those $\{a,b\}_k$ with $k = b$ and either a or b designated (in A_0).
- General case: Truth-values $\mathsf{BN}^{n+1}(\mathbf{2})$. Designated D_{n+1}: those $\{a,b\}_k$ with $k = b$ and either $a \in D_n$ or $b \in D_n$.
- Limit case: Truth-values $\mathsf{BN}^\infty(\mathbf{2})$. Designated $D_\infty := \bigcup_n e_n(D_n)$.

For example, $\{\{0,1\}_n, \{0,1\}_b\}_b$ is designated in $\mathsf{BN}^2(\mathbf{2})$.

However, the resulting logic is quite wild: Let's write $\vDash_{D_n}$ (with $n \in \{0,1,\dots,\infty\}$) for the corresponding designated value consequence relation. As mentioned, $\vDash_{D_0} = \vDash_{\mathbf{2}}$ and $\vDash_{D_1} = \vDash_{\mathsf{BN}(\mathbf{2})}$. But already at the second iteration, the logics become very different. On the algebraic definition, we have, for any involutive lattice A, the much desired conjunction introduction and elimination:

$$\varphi, \psi \vDash_A \varphi \wedge \psi \qquad\qquad \varphi \wedge \psi \vDash_A \varphi.$$

But this fails for $\vDash_{D_2}$.[10] Hence it also fails for $\vDash_{D_\infty}$, since $\vDash_{D_\infty} \subseteq \vDash_{D_n}$.[11] One might wonder: is this the madness that Meyer was referring to?

Of course, one can also investigate variations of the designated value approach. For example, one may consider the both-only construction B, so it becomes a generalization of the *logic of paradox* [30].[12] And one can choose other designated values: e.g., only the top element, so consequence becomes preserving 'nothing but the truth' [29]. Or one can choose designated values differently for premises and for conclusions, thus arriving, e.g., at a 'strict-tolerant' consequence [7]. This might not always yield desirable logics, but one beauty of generalizations is that they allow to tell apart intuitions that coincide in the original case. But, for now, we leave the designated value approach to future work and focus, for the rest of the paper, on the more well-behaved algebraic definition.

5.3 No collapse The logics that we get at each step of iterating 'both' and 'neither' are $\vDash_{\mathbf{2}}$, $\vDash_{\mathsf{BN}(\mathbf{2})}$, $\vDash_{\mathsf{BN}^2(\mathbf{2})}$, etc., with the limit logic $\vDash_{\mathsf{BN}^\infty(\mathbf{2})}$. In fact, these are increasingly weaker logics, because if we have an embedding $f : A \to B$, then $\vDash_A \supseteq \vDash_B$.[13] So we have the chain

$$\vDash_{\mathbf{2}} \ \supseteq \ \vDash_{\mathsf{BN}(\mathbf{2})} \ \supseteq \ \vDash_{\mathsf{BN}^2(\mathbf{2})} \ \supseteq \ \ldots \ \supseteq \ \vDash_{\mathsf{BN}^\infty(\mathbf{2})} . \tag{7}$$

As mentioned, a natural question is whether this chain eventually stabilizes. To get an intuition, let's check the first two cases.

The first inclusion is proper: Famously, FDE discards many entailments of classical logic. For example, $\vDash_{\mathbf{2}} \varphi \vee \neg\varphi$ and $\varphi \wedge \neg\varphi \vDash_{\mathbf{2}} \psi$ but $\nvDash_{\mathsf{BN}(\mathbf{2})} \varphi \vee \neg\varphi$ (paracompleteness) and $\varphi \wedge \neg\varphi \nvDash_{\mathsf{BN}(\mathbf{2})} \psi$ (paraconsistency).

The second inclusion now is, on our approach, also proper: while $\vDash_{\mathsf{BN}(\mathbf{2})}$ still satisfies distributivity, $\vDash_{\mathsf{BN}^2(\mathbf{2})}$ doesn't anymore. An example involves the 'Mowgli sentence': Let v be a $\mathsf{BN}^2(\mathbf{2})$-valuation where p, q, r respectively get the following values from 'the middle' of $\mathsf{BN}^2(\mathbf{2})$ (see figure 5):

$$a := \big\{\{0,1\}_n\big\} \qquad b := \big\{\{0,1\}_n, \{0,1\}_b\big\}_b \qquad c := \big\{\{0,1\}_b\big\}.$$

Then distributivity fails, i.e., $p \wedge (q \vee r) \nvDash_{\mathsf{BN}^2(\mathbf{2})} (p \wedge q) \vee (p \wedge r)$, because

$$a \wedge (b \vee c) = a \not\leq \big\{\{0\}\big\} = (a \wedge b) \vee (a \wedge c).$$

However, for the other inclusions, it is getting increasingly harder to find counterexamples. After all, what should two sentences φ and ψ be such that φ entails ψ in $\mathsf{BN}^{999.999}(A)$ but not anymore in $\mathsf{BN}^{1.000.000}(A)$? Still, the main result of this section is that all these counterexamples must exist:

Theorem 5.2 *All inclusions in (7) are proper.*

The proof is in appendix C and uses a powerful tool from universal algebra: Jónsson's lemma.

5.4 Comparison to other logics Now that we have a new logic $\vDash_{\mathsf{BN}^\infty(\mathbf{2})}$, we would like to know how it compares to existing logics—to put it into perspective. Two tight lower and upper bounds are FDE and PQL: $\vDash_{\mathsf{FDE}} \supseteq \vDash_{\mathsf{BN}^\infty(\mathbf{2})} \supseteq \vDash_{\mathsf{PQL}}$. Let's explain them in turn.

One the one hand, FDE is characterized by $\mathbf{4} = \mathsf{BN}(\mathbf{2})$: $\Gamma \vDash_{\mathsf{FDE}} \varphi$ iff $\Gamma \vDash_{\mathbf{4}} \varphi$ [e.g. 15]. By the results of Kalman [21], FDE can equivalently also be characterized by the class of all distributive involutive lattices: i.e., $\Gamma \vDash_{\mathsf{FDE}} \varphi$ iff, for all distributive involutive lattices A, $\Gamma \vDash_A \varphi$.[14]

On the other hand, *paraconsistent quantum logic* (PQL) can be characterized by all—as opposed to all distributive—involutive lattices: $\Gamma \vDash_{\mathsf{PQL}} \varphi$ iff for all involutive lattices A, $\Gamma \vDash_A \varphi$ [9].[15] This logic generalizes—as the name indicates—the usual quantum logic. The usual quantum logic is based on ortho(modular) lattices capturing the algebraic structure of measurements (or observable properties) of a physical system. Paraconsistent quantum logic then considers a more general notion of measurement (abstracted from so-called effects) which need not satisfy the laws of non-contradiction anymore ($x \wedge \neg x = 0$).

Now, the logic of iterated 'both' and 'neither' sits in between these two logics FDE and PQL: We already saw $\vDash_{\mathsf{FDE}} \supsetneq \vDash_{\mathsf{BN}^\infty(\mathbf{2})}$. And, by definition, $\vDash_{\mathsf{BN}^\infty(\mathbf{2})} \supseteq \vDash_{\mathsf{PQL}}$. (Whether the connection to physics is just a coincidence has to be investigated.) We consider it a central open question whether the two logics actually are identical, i.e., whether $\vDash_{\mathsf{BN}^\infty(\mathbf{2})} = \vDash_{\mathsf{PQL}}$.[16]

For other logics in this ballpark, see, e.g., [35, 37, 18, 17].

6 Relationships between the constructions

So far we have investigated iterating both-and-neither (BN). However, as we saw in the applications, we also are interested in iterating neither-only (N), both-only (B), and comparable-neither-only ($\underline{\mathsf{N}}$). This section is about the relationships of these constructions. We will find that, although the constructions are philosophically different, they mathematically are closely related: in the limit, they all give rise to the very same logic.

Let's start with the differences. The philosophical motivations for the constructions BN, N, B, and $\underline{\mathsf{N}}$ differ: (1) they have rather dissimilar intended philosophical interpretations: e.g., databases, truth, and vagueness. (2) Some might object to an 'inconsistent' truth-value *both true and false*, while being friendly toward an 'incomplete' truth-value *neither true nor false*. (3) After one iteration, the corresponding logics are rather different: e.g., with just 'neither', one fails excluded middle but keeps the law of non-contradiction—unlike when also using 'both'. (4) After several iterations, the resulting truth-value algebras look quite different: e.g., figure 5 for BN vs. figure 6 for $\underline{\mathsf{N}}$.

So how can they be related? We know that $\underline{\mathsf{N}}$ is contained in N which is contained in BN, and N is isomorphic to B. (Because of this isomorphism, we focus on $\underline{\mathsf{N}}, \mathsf{N}, \mathsf{BN}$ in the remainder.[17]) This is expected: the constructions add more—as some would say, problematic—entities. The unexpected direction is if there would also be a sense in which BN is contained in N or $\underline{\mathsf{N}}$. Surprisingly, there is! That is the main result of this section: For a finite involutive lattice A, we have that $\mathsf{BN}(A)$ cannot be contained in $\mathsf{N}(A)$, but it is contained in $\mathsf{N}^2(A)$! The embedding maps

$$\{a,b\}_n \mapsto \{\{a,b\}_n\} \qquad\qquad \{a,b\}_b \mapsto \{\{a\},\{b\}\}_n.$$

With considerably more work—done in appendix D—one can also show that $\mathsf{N}^2(A)$, in turn, can be embedded into $\underline{\mathsf{N}}^m(A)$, for some m (which can be quite large). To summarize:

Theorem 6.1 *Let A be an involutive lattice. Then $\mathsf{BN}(A)$ can be embedded into $\mathsf{N}^2(A)$. And, if A was finite, this can, in turn, be embedded into $\underline{\mathsf{N}}^m(A)$ for some m.*

But we not only want to understand the relationships of the constructions at the level of truth-values. We also want to understand it at the level of logic: i.e., compare the consequence relations of the limits of these constructions (where the construction is 'completed'). The above embeddability result yields the other surprising theorem that these limit logics in fact are all identical!

Theorem 6.2 $\vDash_{\mathsf{BN}^\infty(\mathbf{2})} \;=\; \vDash_{\mathsf{N}^\infty(\mathbf{2})} \;=\; \vDash_{\underline{\mathsf{N}}^\infty(\mathbf{2})}$.

7 Applications

With the acquired theoretical understanding of 'both' and 'neither', we revisit the motivating applications from section 2. These were: $\mathsf{BN}^\infty(\mathbf{2})$ for (non-deterministic) Belnap computer networks, $\mathsf{N}^\infty(\mathbf{2})$ for the revenge paradox, and $\underline{\mathsf{N}}^\infty(\mathbf{2})$ for higher-order vagueness.

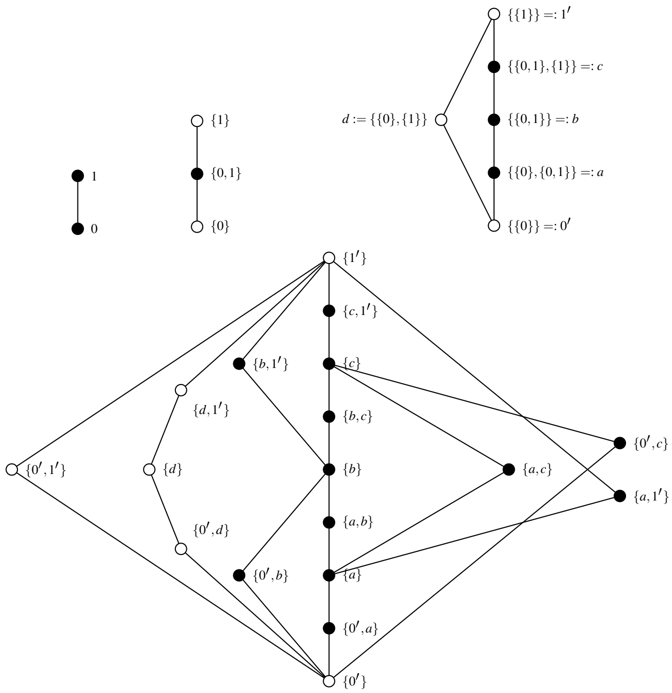

Figure 6 The first four iterations of comparable-neither-only $\underline{\mathsf{N}}$.

We already visualized the first iterations of BN (figures 3–5). Since we here are also deal with N and $\underline{\mathsf{N}}$, we visualize in figure 6 the first four iterations of $\underline{\mathsf{N}}$, where the first three iterations coincide with N. (The white dots are the image of the embeddings $\underline{\mathsf{N}}^k(e)$ from the preceding iteration; so the black dots are the 'new' elements.)

We saw that, remarkably, the logics of these applications coincide. So, somehow, these different applications are governed by the very same logical laws. Maybe this can be interpreted as further evidence of a deeper hidden connection between the prima facie different paradoxes of truth and vagueness [33]. But let's leave the philosophical interpretation of this coincidence to future work and focus first on the applications themselves.

7.1 Belnap computer networks In section 2.1, we saw that a single Belnap computer which receives classical sources should use the truth-values $\mathbf{4}$ and hence the logic FDE. When it comes to reasoning with input from (networks of) Belnap computers, there are two choices. If the input is processed deterministically by accumulation, we get more complex truth-values as described by [36], but the logic remains FDE. However, if we allow the input to be processed by judging the reliability of the sources on a case-by-case basis, we get the truth-values $\mathsf{BN}^\infty(\mathbf{2})$ and the logic $\models_{\mathsf{BN}^\infty(\mathbf{2})}$, which is properly contained in FDE. So, in short, their difference lies in the non-deterministic processing of the input sources.

7.2 Revenge paradox A common criticism of broadly Kripkean theories of truth is this: Yes, they consistently provide a language with its own truth-predicate, crucially by rendering the liar sentence neither true nor false. However, they cannot express this fact: otherwise they could formulate the revenge sentence, which leads to inconsistency [2, sec. 4.1.3]. As already suggested in section 2.2, iterating ‘neither’ might provide a reply. Let’s sketch it step by step.

The first step is the usual setting for theories of truth. We add to our base language a truth-predicate, i.e., the expressive means to say that a sentence has truth-value 1. So we can formulate in the language the liar sentence λ_1 (here the subscript just indicates that we are in step 1) which says that the truth-value of sentence λ_1 is 0 (or not 1). Even though there is no two-valued model, the Kripkean solution provides a three-valued model of the language: assigning truth-values from $\mathsf{N}(2)$ to sentences in a way that respects the connectives and the truth-predicate. In this model, λ_1 gets the value $a_1 := \{0,1\}_n$.

The second step is the setting of the revenge paradox. To counter the mentioned criticism, we add to our language the expressive resources to say ‘φ has truth-value a’ where φ is a sentence and a is in $\mathsf{N}(\mathbf{2})$. So we can formulate in this new language the revenge sentence λ_2 which says that the truth-value of sentence λ_2 is either $\{0\}$ or $\{0,1\}_n$. Now one would need to explore if an adjusted model construction yields a consistent model. The idea would be that, in such a model, λ_2 gets the value ‘neither true nor a_1’, i.e., $a_2 := \big\{\{1\},\{0,1\}_n\big\}_n$, because then λ_2 is not $\{1\}$ and also not $\{0\}$ or $\{0,1\}_n$.

Now we iterate these steps. In step $k+1$, the criticism still complains about the mismatch of language and truth-values: The language of step k was interpreted in $\mathsf{N}^k(\mathbf{2})$, but some of those truth-values cannot be expressed in the language (e.g., the truth-value a_k of the revenge sentence λ_k). We add to the language the required expressive resources. But then we can formulate the revenge sentence λ_{k+1} which says that the truth-value of sentence λ_{k+1} is in $\mathsf{N}^k(\mathbf{2}) \setminus \{1_k\}$, where 1_k is the top element of $\mathsf{N}^k(\mathbf{2})$. Again, one would need an adjusted model construction to find a consistent model of this extended language by moving to $\mathsf{N}^{k+1}(\mathbf{2})$ and giving λ_{k+1} the value $a_{k+1} := \big\{1_k, a_k\big\}_n$.[18]

In these steps, expressivity and revenge play a game of tag: once expressiveness caught up to express the latest solution to the revenge sentence, revenge produces a new sentence resulting in new truth-values to which expressiveness has to catch up again. However, or so the idea goes, by iterating ‘neither’ forever and moving to $\mathsf{N}^\infty(\mathbf{2})$, we reach a fixed point of this game where the two are in harmony: the language can express all those truth-values, and all revenge sentences have a truth-value.

So far, this is just a rough intuitive idea, and it needs to be formally developed and critically assessed.[19] Our point here merely is show that iterating ‘neither’ might fruitfully be applied to theories of truth.

7.3 Higher-order vagueness In section 2.3, we said that we get a semantics for higher-order vagueness by iterating ‘neither’ (or, more precisely, $\underline{\mathsf{N}}$): Whenever we hedge and want to say that a sentence φ actually is in between two truth-values a and b (which hence must be comparable), we can say that φ has truth-value $\{a,b\}_n$. Here we discuss if this can be extended to a logic of higher-order vagueness by using the logic $\vDash_{\underline{\mathsf{N}}^\infty(\mathbf{2})}$.

But what exactly should a logic of vagueness do? It should describe when sentences involving vague predicates entail each other. There are two common choices for the language in which these sentences are formulated. The first choice is a basic language where we can form (logically atomic) sentences using vague predicates and have standard logical connectives. Formally, we can take this to be the language of propositional logic that we have used above (section 5.1). The second choice is to extend this basic language with a 'definitely' or 'determinately' operator Δ. (Cf. Field's determinacy operator mentioned in section 2.2.) This is to be able to say, in the language, whether a sentence has borderline status: If φ is a sentence, the sentence $\Delta\varphi$ intuitively says that φ is definitely the case (and not borderline).

Providing a logic of vagueness for the basic language is considered by some to be the prime concern—extending it to Δ being a secondary step [e.g. 22, sec. 1.5]. Our framework delivers: The semantics of the basic language is given by valuations in $\underline{\mathsf{N}}^{\infty}(\mathbf{2})$, which can express higher-order vagueness by allowing to iterate 'neither' arbitrarily often. The corresponding logic of vagueness then is $\vDash_{\underline{\mathsf{N}}^{\infty}(\mathbf{2})}$. To better understand this suggestion, let's compare it in the remainder of this section to two other prominent approaches and see how the secondary step—interpreting Δ—could look like.

The first prominent approach is based on strong Kleene logic.[20] Here the semantics of the basic language is given by valuations in $\underline{\mathsf{N}}(\mathbf{2})$ (true, false, or neither) and the logic is $\vDash_{\underline{\mathsf{N}}(\mathbf{2})}$. Concerning the secondary step, the definitely operator is interpreted by the function $\Delta : \underline{\mathsf{N}}(\mathbf{2}) \to \underline{\mathsf{N}}(\mathbf{2})$ sending $\{1\}$ to $\{1\}$ and the other two values to $\{0\}$. A common criticism is that this does not allow for higher-order vagueness: if φ is vague, it seems that also $\Delta\varphi$ is vague, but now $\Delta\varphi$ always has a classical truth-value and hence no borderlines. Another way to say this is that $\Delta\Delta\varphi \vee \Delta\neg\Delta\varphi$ is a theorem.

The second, arguably most prominent logic of vagueness is (some form of) supervaluationism.[21] The semantics of the basic language is given by judging the truth of sentences relative to *precisifications* of the language. Here a precisification is a way of making the language precise (saying how many grains of sand are needed to form a heap, etc.). If φ is true at all relevant precisifications, it is *super-true*—which is what we typically mean by saying that a vague sentence is true. Logical consequence is preservation of super-truth. Concerning the secondary step, Δ expresses super-truth in the object language: so $\Delta\varphi$ is true at a precisification if φ is super-true. This gets many things right, but one criticism is that still $\Delta\Delta\varphi \vee \Delta\neg\Delta\varphi$ is a theorem.[22]

Thus, our approach generalizes the first approach, aiming to fix the issue of higher-order vagueness by iterating neither. But can it perform the secondary step of providing a new function interpreting Δ? The second approach offers an idea.

For the supervaluationist, $\Delta\varphi$ means that φ is determined in the sense that φ is true under all precisifications. In other words, even if other ways of evaluating φ become available, φ will still get the same truth-value. Reformulated in our setting, $\Delta\varphi$ says that even if further borderline truth-values become available, φ will keep the truth-value that it currently has. Indeed, this seems characteristic of vague predicates: Consider a borderline heap and the sentence $\varphi =$ 'This is a heap'. If only the classical truth-values 0 and 1 are available, we reluctantly give φ one of those classical truth-values, but once the borderline truth-value $\{0,1\}_n$ becomes available, we happily assign it to φ. So $\Delta\varphi$ indeed is not true, as expected.

Can we turn this intuition into a function $\Delta : \underline{\mathsf{N}}^\infty(\mathbf{2}) \to \underline{\mathsf{N}}^\infty(\mathbf{2})$? Many choices can be discussed, but a straightforward one maps $[a]$ to $[\{a\}]$, with the following explanation. Assume φ has been assigned to truth-value a in, say $\underline{\mathsf{N}}^k(\mathbf{2})$. This means that, after deliberation, we found enough truth-values at the k-th iteration of 'neither' to confidently give φ the truth-value a. Now we should give $\Delta\varphi$ a truth-value. The definitely operator makes us consider a further iteration of 'neither', qua expressive devise to talk about borderline status. So we should pick a truth-value for $\Delta\varphi$ in $\underline{\mathsf{N}}^{k+1}(\mathbf{2})$. But since φ received the definite truth-value a, we now choose its copy $\{a\}$ and need not resort to any new borderline value. So $\Delta\varphi$ has the truth-value $\{a\}$. Thus, in $\underline{\mathsf{N}}^\infty(\mathbf{2})$, we assign φ to $[a]$ and $\Delta\varphi$ to $[\{a\}]$.

It turns out that $\Delta : \underline{\mathsf{N}}^\infty(\mathbf{2}) \to \underline{\mathsf{N}}^\infty(\mathbf{2})$ mapping $[a]$ to $[\{a\}]$ is not just a well-defined function; it even is an involutive lattice embedding.[23] This has both desired consequences but also some in need of discussion. Among the desired ones are that 1_∞ (and only that) is mapped to 1_∞. So $\Delta\varphi$ is true iff φ is true (and nothing but true). Thus, Δ acts much like a truth-predicate, as one would expect if it expresses something like super-truth.[24] The same can be said for 0_∞ and super-falsity. Also, as desired, $\Delta\Delta\varphi \vee \Delta\neg\Delta\varphi$ is not a theorem.[25] And since Δ preserves $\wedge$, we have that $\Delta(\varphi \wedge \psi)$ is equivalent to $\Delta\varphi \wedge \Delta\psi$. The same holds for $\vee$, which is more debatable.[26] However, the most debatable consequence is that $\neg\Delta\varphi$ is equivalent to $\Delta\neg\varphi$. The obvious objection is that, when talking about a borderline heap, 'not definitely a heap' is *not* the same as 'definitely not a heap': the former seems true, while the latter seems false. Further work should explore this objection: Is it decisive, thus dismissing the above prima facie plausible interpretation of Δ? (This would also constitute a philosophical insight, albeit negative one.) Should the above straightforward choice for Δ be refined? Can the objection be explained away, since 'not' doesn't have a classical meaning anymore due to the presence of non-classical truth-values?[27] Or are we looking at an argument that there are (at least) two senses of 'definitely' (one on which $\neg\Delta\varphi$ and $\Delta\neg\varphi$ are not equivalent and one on which they are)?

To summarize, the iterated neither approach suggests a logic of higher-order vagueness for the basic language. It may be extended to also interpret the definitely operator, but more discussion is needed. Our goal here was to argue that this suggestion seems fruitful. Future works needs to assess it in detail.

8 Conclusion

We conclude with a brief summary and open questions. Motivated by, among others, the higher-order paradoxes of truth and vagueness, we newly explicated iterating 'both' and 'neither' as a functor BN on involutive lattices. Special cases are iterating only neither N or only neither applied to comparable truth-values $\underline{\mathsf{N}}$. Each iteration of BN changes the logic, and the logic obtained in the limit of iteration lies between FDE and PQL. Surprisingly, the limit logics of BN, N, and $\underline{\mathsf{N}}$ are all identical even though differing at finite stages. These results promise new and fruitful applications to the paradoxes of vagueness and truth. The main technical open question is whether the logic of iterated both-and-neither is paraconsistent quantum logic. The main philosophical open question is to work out the sketched applications.

Appendix A Both and neither algebras

The objects that we will be dealing with are involutive lattices, which we always assume to be bounded (definition 3.1), and their morphisms are:

Definition A.1 Let A and B be involutive lattices. An *involutive lattice homomorphism* is a function $f : A \to B$ such that, for all $x, y \in A$,

$$f(x \vee y) = f(x) \vee f(y) \qquad f(\neg x) = \neg f(x) \qquad f(0) = 0.$$

These conditions already imply the other two expected preservation conditions: $f(x \wedge y) = f(x) \wedge f(y)$ and $f(1) = 1$. We say f is an *embedding* if it additionally is injective and an *isomorphism* if it additionally is bijective. If there is an embedding from A to B, we write $A \hookrightarrow B$. We say A is a *subalgebra* of B if $A \subseteq B$ and the inclusion function (mapping $a \in A$ to $a \in B$) is an embedding.

For $k \geq 0$, define

$$\{a,b\}_k := \begin{cases} \{a\} & \text{if } a = b \\ (\{a,b\},k) & \text{if } a \neq b. \end{cases}$$

So if $\{a,b\}_k = \{c,d\}_l$, we have $\{a,b\} = \{c,d\}$; a fact that we often use without mentioning.

Theorem A.2 *Let $A = (A, \vee, \wedge, \neg, 0, 1)$ be an involutive lattice. Let $n \geq 1$ and write $n = \{0, \dots, n-1\}$. Let $\star : n \to n$ be an involution (i.e., a function with $x^{\star\star} = x$). Define*

- $A' = \big\{\{a,b\}_k : a, b \in A, k \in n\big\}$
- $\{a,b\}_k \leq' \{c,d\}_l$ *iff* $\{a,b\}_k = \{c,d\}_l$ *or* $a \vee b \leq c \wedge d$
- $\neg'\{a,b\}_k = \{\neg a, \neg b\}_{k^\star}$
- $0' = \{0,0\}_0$ *and* $1' = \{1,1\}_0$

Then $\mathsf{N}_{n,\star}(A) := (A', \vee', \wedge', \neg', 0', 1')$ *is an involutive lattice where*

$$\{a,b\}_k \vee' \{c,d\}_l = \begin{cases} \{c,d\}_l & \textit{if } \{a,b\}_k \leq' \{c,d\}_l \\ \{a,b\}_k & \textit{if } \{a,b\}_k \geq' \{c,d\}_l \\ \{a \vee b \vee c \vee d\}_0 & \textit{if } x \textit{ and } y \textit{ are } \leq'\textit{-incomparable} \end{cases}$$

and similarly for $\wedge$.

So theorem 3.2 is the special case where $n = \{0,1\}$ and $\star = \mathrm{id}$ (the identity function). We will need the present more general formulation in section D.

Proof *Step 1*: Show that $\leq'$ is a partial order with least element $0'$ and greatest element $1'$. By construction, $\leq'$ is reflexive. For antisymmetry, given $x = \{a,b\}_k$ and $y = \{c,d\}_l$ with $x \leq' y$ and $y \leq' x$, we need to show $x = y$. If one of $x \leq' y$ or $y \leq' x$ holds due to the first clause, we immediately have $x = y$, so assume both hold due to the second clause. Then

$$a \vee b \leq c \wedge d \leq c \vee d \leq a \wedge b \leq a \vee b,$$

so all the terms are identical. In particular, $a \wedge b = a \vee b$, which implies $a = b$. And $c \wedge d = c \vee d$, which implies $c = d$. So $a = a \vee b = c \vee d = c$. Hence $a = b = c = d$, and $x = \{a,b\}_k = \{a\} = \{c\} = \{c,d\}_l = y$, as needed.

For transitivity, given $x = \{a,b\}_k$, $y = \{c,d\}_l$, and $z = \{e,f\}_m$ with $x \leq' y$ and $y \leq' z$, we need to show $x \leq' z$. If one of the inequalities is due to the first clause, this is trivial, so assume both are due to the second. Then

$$a \vee b \leq c \wedge d \leq c \vee d \leq e \wedge f,$$

so $x \leq' z$, as needed.

Finally, $0' = \{0,0\}_0$ is the $\leq'$-least element of A', because for any $\{a,b\}_k$, we have $0 \vee 0 = 0 \leq a \wedge b$. Similarly, $1' = \{1,1\}_0$ is the $\leq'$-greatest element of A', because for $\{a,b\}_k$, we have $a \vee b \leq 1 = 1 \wedge 1$.

Step 2: Show that $x \vee' y$ is the $\leq'$-least upper bound of $\{x,y\}$, and dually for $\wedge'$. Write $x = \{a,b\}_k$ and $y = \{c,d\}_l$. The claim is trivial if x and y are $\leq'$-comparable, so assume that they are incomparable. We need to show that $z := \{a \vee b \vee c \vee d\}_0$ is the $\leq'$-least upper bound of $\{x,y\}$.

First, it is an upper bound: We have $a \vee b \leq a \vee b \vee c \vee d$, so $x \leq' z$. Similarly, $y \leq' z$. To show z is the least upper bound, let $w = \{e,f\}_m$ be another upper bound, and show $z \leq' w$. Note that $x \neq w$, since otherwise $x = w \geq' y$ contradicts incomparability; and similarly $y \neq w$. So $x \leq' w$ must be because $a \vee b \leq e \wedge f$; and $y \leq' w$ must be because $c \vee d \leq e \wedge f$. Hence $a \vee b \vee c \vee d \leq e \wedge f$, which implies $z \leq w$.

For $\wedge$ we reason dually.

Step 3: Show that $\neg'$ satisfies the axioms for an involutive lattice. First, we show $\neg'\neg' x = x$. Write $x = \{a,b\}_k$. Then

$$\neg'\neg'\{a,b\}_k = \neg'\{\neg a, \neg b\}_{k^\star} = \{\neg\neg a, \neg\neg b\}_{k^{\star\star}} = \{a,b\}_k.$$

Second, we show that $x \leq' y$ implies $\neg' y \leq' \neg' x$. Write $x = \{a,b\}_k$ and $y = \{c,d\}_l$. If $x \leq' y$ because $x = y$, the claim is trivial, so assume $x \leq' y$ holds due to the second clause. So $a \vee b \leq c \wedge d$. Since A is an involutive lattice,

$$\neg c \vee \neg d = \neg(c \wedge d) \leq \neg(a \vee b) = \neg a \wedge \neg b.$$

Hence $\neg' y = \{\neg c, \neg d\}_{l^\star} \leq' \{\neg a, \neg b\}_{k^\star} = \neg' x$. $\square$

The following two propositions provide important subalgebras of $\mathsf{N}_{n,\star}(A)$.

Proposition A.3 *In the setting of theorem A.2, define*

$$\mathsf{N}^{\leq}_{n,\star}(A) := \big\{\{a,b\}_k \in \mathsf{N}_{n,\star}(A) : a \text{ and } b \text{ are } \leq\text{-comparable}\big\}$$

Then $\mathsf{N}^{\leq}_{n,\star}(A)$ is a subalgebra of $\mathsf{N}_{n,\star}(A)$ (with the inherited operations).

Proof We need to show that $\mathsf{N}^{\leq}_{n,\star}(A)$ is closed under $0', \neg', \vee'$. Note that, if $a \in A$, then $\{a\} = \{a,a\}_0$ is in $\mathsf{N}^{\leq}_{n,\star}(A)$, because $a \leq a$. In particular, $0' \in \mathsf{N}^{\leq}_{n,\star}(A)$. Moreover, if $x = \{a,b\}_k$ is in $\mathsf{N}^{\leq}_{n,\star}(A)$, then either $a \leq b$ or $a \geq b$, so either $\neg a \geq \neg b$ or $\neg a \leq \neg b$, so $\neg' x = \{\neg a, \neg b\}_k$ is in $\mathsf{N}^{\leq}_{n,\star}(A)$. So assume $x = \{a,b\}_k$ and $y = \{c,d\}_l$ are in $\mathsf{N}^{\leq}_{n,\star}(A)$, and show that $x \vee' y$ is, too. If x and y are $\leq'$-comparable, this is trivial. If they are incomparable, then $x \vee' y = \{a \vee b \vee c \vee d\}_0$ which is in $\mathsf{N}^{\leq}_{n,\star}(A)$, as noted. $\square$

Proposition A.4 *In the setting of theorem A.2, $\mathsf{N}_{n,\mathsf{id}}(A)$ is a subalgebra of $\mathsf{N}_{n+1,\mathsf{id}}(A)$.*

Proof It is a subset: If $\{a,b\}_k \in \mathsf{N}_{n,\mathsf{id}}(A)$, then $a,b \in A$ and $0 \leq k < n < n+1$, so $\{a,b\}_k \in \mathsf{N}_{n+1,\mathsf{id}}(A)$. And the inclusion is an embedding: first applying the algebraic operation of $\mathsf{N}_{n,\mathsf{id}}(A)$ and then mapping is the same as first mapping and then applying the algebraic operation of $\mathsf{N}_{n+1,\mathsf{id}}(A)$. $\square$

The both and neither constructions that we are particularly interested in (as discussed in section 3.4) are the following special cases.

Definition A.5 Given an involutive lattice A and $n \geq 1$, write $\mathsf{N}_n(A) := \mathsf{N}_{n,\mathsf{id}}(A)$ and $\mathsf{N}^{\leq}_n(A) := \mathsf{N}^{\leq}_{n,\mathsf{id}}(A)$. We call:

- $\mathsf{BN}(A) := \mathsf{N}_2(A)$ the *both-and-neither algebra* of A
- $\mathsf{N}(A) := \mathsf{N}_1(A)$ the *neither-only algebra* of A
- $\underline{\mathsf{N}}(A) := \mathsf{N}^{\leq}_1(A)$ the *comparable-neither-only algebra* of A.

So $\underline{\mathsf{N}}(A)$ is a subalgebra of $\mathsf{N}(A)$, which is a subalgebra of $\mathsf{BN}(A)$.

Appendix B Fixed points

For this section, let $n \geq 1$ and let $\star$ be an involution on $\{0,\dots,n-1\}$. Let F be $\mathsf{N}_{n,\star}$ or $\mathsf{N}^{\leq}_{n,\star}$. We want to understand the fixed points of the construction F.

To do so, the following category-theoretic language will be useful (but not necessary: it just highlights the more general ideas behind the proofs). Let C be the category whose objects are nontrivial involutive lattices (nontrivial means that the least element 0 is distinct from the greatest element 1) and whose morphisms are involutive lattice embeddings. (Taking all involutive lattice homomorphisms as morphisms does not work as intended here.)

We first observe that the construction F is 'compositional': F is a functor from C to C (i.e., an endofunctor on C).

Proposition B.1 *If $f : A \to B$ is an involutive lattice embedding, then*

$$\begin{aligned}\mathsf{N}_{n,\star}(f) : \mathsf{N}_{n,\star}(A) &\to \mathsf{N}_{n,\star}(B)\\ \{a,b\}_k &\mapsto \{f(a),f(b)\}_k\end{aligned}$$

is an involutive lattice embedding. This remains true when adding $\leq$.

Proof Write $g := \mathsf{N}_{n,\star}(f)$. This is injective: If $\{f(a),f(b)\}_k = \{f(c),f(d)\}_l$, then $\{f(a),f(b)\} = \{f(c),f(d)\}$, so, by injectivity of f, $\{a,b\} = \{c,d\}$. If $f(a) \neq f(b)$, then $k = l$, so $\{a,b\}_k = \{c,d\}_l$. And if $f(a) = f(b)$, then $a = b$ and hence $c = d$, so $\{a,b\}_k = \{a\} = \{c\} = \{c,d\}_l$.

Preserving the top-element: $g(\{1,1\}_0) = \{f(1),f(1)\}_0 = \{1,1\}_0$.

Preserving negation: We have

$$\begin{aligned}g(\neg\{a,b\}_k) = g(\{\neg a,\neg b\}_{k^\star}) &= \{f(\neg a),f(\neg b)\}_{k^\star}\\ &= \{\neg f(a),\neg f(b)\}_{k^\star} = \neg\{f(a),f(b)\}_k = \neg g(\{a,b\}_k).\end{aligned}$$

Preserving meet: First note that the following equivalences

$\{a,b\}_k \leq \{c,d\}_l$
iff $\{a,b\}_k = \{c,d\}_l$ or $a \vee b \leq c \wedge d$
iff $g(\{a,b\}_k) = g(\{c,d\}_l)$ or $f(a) \vee f(b) \leq f(c) \wedge f(d)$ (using the injectivity of g and the meet preservation and injectivity of f)
iff $g(\{a,b\}_k) \leq g(\{c,d\}_l$.

We need to show $g(\{a,b\}_k \wedge \{c,d\}_l) = g(\{a,b\}_k) \wedge g(\{c,d\}_l)$. This is immediate if $\{a,b\}_k$ and $\{c,d\}_l$ are comparable, so assume they are not, hence also $g(\{a,b\}_k)$ and $g(\{c,d\}_l)$ are incomparable. So

$$\begin{aligned} g(\{a,b\}_k \wedge \{c,d\}_l) &= g(\{a \wedge b \wedge c \wedge d\}_0) = \{f(a \wedge b \wedge c \wedge d)\}_0 \\ &= \{f(a) \wedge f(b) \wedge f(c) \wedge f(d)\}_0 = g(\{a,b\}_k) \wedge g(\{c,d\}_l). \end{aligned}$$

Finally, g restricts to a map $\mathsf{N}^{\leq}_{n,\star}(A) \to \mathsf{N}^{\leq}_{n,\star}(B)$, because if a and b are comparable, so are $f(a)$ and $f(b)$. $\square$

Next, we can naturally view A to be contained in $\mathsf{N}^{\leq}_{n,\star}(A)$, and hence also in $\mathsf{N}_{n,\star}(A)$. (So we might call the endofunctor F inflationary.)

Proposition B.2 *Let A be an involutive lattice. Then*

$$\begin{aligned} e : A &\to \mathsf{N}^{\leq}_{n,\star}(A) \subseteq \mathsf{N}_{n,\star}(A) \\ a &\mapsto \{a\} = \{a,a\}_0 \end{aligned}$$

is an involutive lattice embedding, which we call the natural embedding.

Proof By construction, $e(1) = \{1,1\}_0 = 1'$. Moreover,

$$e(\neg a) = \{\neg a\} = \{\neg a, \neg a\}_{0^\star} = \neg'\{a,a\}_0 = \neg' e(a).$$

Finally, for $a,b \in A$, we need to show $e(a \vee b) = e(a) \vee' e(b)$. First note that $a \leq b$ iff $\{a\} \leq' \{b\}$. So, if a and b are $\leq$-comparable, the claim follows, and it they are not, also $e(a)$ and $e(b)$ are not $\leq'$-comparable, so $e(a \vee b) = \{a \vee b\}_0 = e(a) \vee e(b)$. $\square$

To find the least fixed point for F, we use the idea that it is the initial algebra for the endofunctor F on the category C. In good cases, the initial algebra can be obtained as the colimit of applying the endofunctor to the initial object of the category. Conveniently, our C has an initial object, namely $\mathbf{2}$: i.e., for every nontrivial involutive lattice A, there is exactly one embedding $h : \mathbf{2} \to A$ (it maps 0 to 0_A and 1 to 1_A).

Theorem B.3 *Write F for $\mathsf{N}_{n,\star}$ or $\mathsf{N}^{\leq}_{n,\star}$. The direct limit A of*

$$\mathbf{2} \xrightarrow{e} \mathsf{F}(\mathbf{2}) \xrightarrow{\mathsf{F}(e)} \mathsf{F}^2(\mathbf{2}) \xrightarrow{\mathsf{F}^2(e)} \mathsf{F}^3(\mathbf{2}) \xrightarrow{\mathsf{F}^3(e)} \dots$$

is the least nontrivial fixed point of F.

Proof For $n \leq m$, write $A_n := \mathsf{F}^n(\mathbf{2})$ and $e_{nm} := \mathsf{F}^{m-1} \circ \ldots \circ \mathsf{F}^n(e)$ (and $e_{nn} := \mathrm{id}$). So (A_n, e_{nm}) is a direct system. The direct limit A is constructed as usual for universal algebras (described in section 4), with embeddings $e_n : A_n \to A, a \mapsto [a]$.

Toward building a cocone $(\mathsf{F}(A), f_n)$, consider the following diagram:

$$\begin{tikzcd} & & & \mathsf{F}(A) \\ \mathbf{2} \arrow[r, "e"'] \arrow[urrr, "\mathsf{F}(e_0)\circ e"] & \mathsf{F}(\mathbf{2}) \arrow[r, "\mathsf{F}(e)"'] \arrow[urr, "\mathsf{F}(e_0)"] & \mathsf{F}^2(\mathbf{2}) \arrow[r, "\mathsf{F}(e_1)"'] \arrow[ur, "\mathsf{F}(e_1)"] & \dots \end{tikzcd}$$

The first triangle on the left commutes by construction, and the others since $\mathsf{F}(e_0) = \mathsf{F}(e_1 \circ e_{01}) = \mathsf{F}(e_1) \circ e_{12}$, etc. So, writing f_n for the arrow from A_n to $\mathsf{F}(A)$ in the above diagram, we have $f_n = f_m \circ e_{nm}$, so $(\mathsf{F}(A), f_n)$ indeed forms a cocone.

So there is a mediating morphism $u : A \to \mathsf{F}(A)$, mapping $[a]$ to $f_n(a)$ for n such that $a \in A_n$. We show that the embedding u actually is an isomorphism, i.e., we show that it is surjective. Indeed, if $\{[a],[b]\}_k$ is in $\mathsf{F}(A)$, let, without loss of generality, n be such that $a, b \in A_n$. Then $\{a,b\}_k \in \mathsf{F}(A_n)$. So $[\{a,b\}_k] \in A$ and

$$u\big([\{a,b\}_k]\big) = f_{n+1}\big(\{a,b\}_k\big) = \big\{e_n(a), e_n(b)\big\}_k = \big\{[a],[b]\big\}_k.$$

So A is indeed a fixed point. To show it is least, let B be another non-trivial one, with isomorphism $i : \mathsf{F}(B) \to B$. We build a cocone (B, f_n) as follows. Consider

$$\begin{tikzcd}
B & \mathsf{F}(B) \arrow[l, "i"'] & \mathsf{F}^2(B) \arrow[l, "\mathsf{F}(i)"'] & \mathsf{F}^3(B) \arrow[l, "\mathsf{F}^2(i)"'] & \dots \arrow[l, "\mathsf{F}^3(i)"'] \\
\mathbf{2} \arrow[u, "h"] \arrow[r, "e"] & \mathsf{F}(\mathbf{2}) \arrow[u, "\mathsf{F}(h)"] \arrow[r, "\mathsf{F}(e)"] & \mathsf{F}^2(\mathbf{2}) \arrow[u, "\mathsf{F}^2(h)"] \arrow[r, "\mathsf{F}^2(i)"] & \mathsf{F}^3(\mathbf{2}) \arrow[u, "\mathsf{F}^3(h)"] \arrow[r, "\mathsf{F}^3(e)"] & \dots
\end{tikzcd}$$

Since h is the only embedding from $\mathbf{2} \to B$, the first square commutes. Hence, since F is a functor, also the other squares commute. So, defining $f_n : A_n \to B$ as the shortest path in the diagram from A_n to B, we get that (B, f_n) is indeed a cocone ($f_n = f_m \circ e_{nm}$). The mediating morphism $u : A \to B$ is the desired embedding. $\square$

Appendix C Congruences, simplicity, and varieties

In this section, we prove theorem 5.2 from section 5.3: that, for all n, $\vDash_{\mathsf{BN}^n(\mathbf{2})}$ is a proper superset of $\vDash_{\mathsf{BN}^{n+1}(\mathbf{2})}$.

To do so, we need three central concepts from universal algebra: congruence, simplicity, and variety. (For a textbook, see, e.g., [5].) Let $A = (A, \vee, \wedge, \neg, 0, 1)$ be an involutive lattice. A *congruence* θ on A is an equivalence relation such that, for all $a, b, a', b' \in A$,

If $a\theta a'$ and $b\theta b'$, then $a \vee b \theta a' \vee b'$, $a \wedge b \theta a' \wedge b'$, and $\neg a \theta \neg a'$.

The set of congruences on A is denoted $\mathsf{Con}(A)$. The identity relation Δ (i.e., $a\Delta b$ iff $a = b$) and the trivial relation (i.e., $a\nabla b$ for all $a, b \in A$) are always in $\mathsf{Con}(A)$. If they are the only congruences on A (i.e., $\mathsf{Con}(A) = \{\Delta, \nabla\}$), then A is called *simple*. For example, $\mathbf{2}$ is simple. Finally, a *variety* is a nonempty class of algebras of the same type (e.g., involutive lattices) that is closed under taking products, subalgebras, and homomorphic images. The smallest variety containing a class of algebras K is denoted $\mathsf{Var}(K)$.

The plan of the proof is as follows: We first show a key lemma (lemma C.1) from which we deduce that BN preserves simplicity: if A is simple, so is $\mathsf{BN}(A)$ (proposition C.2). Then we derive theorem 5.2 using standard tools from universal algebra. (And we end this section with a lemma on the interplay of varieties and logic that is useful in the main text.)

Visually, the key lemma says: In a 'diamond' in $\mathsf{BN}(A)$ as in figure 7, if one of the depicted lines is a θ-relation, then all of the lines are θ-relations. We write k' for the opposite of k (i.e., if $k = 0$, then $k' = 1$, and if $k = 1$, then $k' = 0$).

Lemma C.1 *Let A be an involutive lattice and $\theta \in \mathsf{Con}(\mathsf{BN}(A))$.*

1. *If $\{a,b\}_k \theta \{a \vee b\}$, then all of $\{a \wedge b\}, \{a,b\}_k, \{a,b\}_{k'}, \{a \vee b\}$ are θ-related.*
2. *The same follows if $\{a \wedge b\} \theta \{a,b\}_k$.*

Proof We show (1). (Then (2) will be immediate.) If A is the trivial involutive lattice, then also $\mathsf{BN}(A)$ is trivial, so all these elements are identical and the claim

$$
\begin{tikzcd}[column sep=small]
 & \{a \vee b\} \arrow[dl, dash] \arrow[dr, dash] & \\
\{a,b\}_0 \arrow[dr, dash] & & \{a,b\}_1 \arrow[dl, dash] \\
 & \{a \wedge b\} &
\end{tikzcd}
$$

Figure 7 A diamond in $\mathsf{BN}(A)$.

follows. So let A be nontrivial (i.e., $0 \neq 1$). Similarly, if $a = b$, then all these elements are identical, and the claim follows. So let $a \neq b$. The crucial trick is to consider the elements $\{0, a \vee b\}_k$ and $\{0, a \vee b\}_{k'}$. Though, we need to consider three cases.

Case 1: $\{a,b\} \neq \{0, a \vee b\}$. Write $x := \{a,b\}_k$ and $y_k := \{0, a \vee b\}_k$ and $y_{k'} := \{0, a \vee b\}_{k'}$.

Let $l \in \{k, k'\}$. Then x and y_l are incomparable: By the case assumption, they are not identical, and $a \vee b \not\leq 0 \wedge (a \vee b)$ (otherwise $a \vee b = 0$, so $a = 0 = b$) and $0 \vee (a \vee b) \not\leq a \wedge b$ (otherwise $a \wedge b = a \vee b$, which implies $a = b$). Moreover, $y_l \leq \{a \vee b\}$ (since $0 \vee (a \vee b) \leq a \vee b$). So

$$y_l = y_l \wedge \{a \vee b\} \; \theta \; y_l \wedge \{a,b\}_k = \{0 \wedge (a \vee b) \wedge a \wedge b\} = \{0\}.$$

Hence $y_k \; \theta \; \{0\} \; \theta \; y_{k'}$. Moreover, y_k and $y_{k'}$ are incomparable: We have $a \vee b \neq 0$ (otherwise $a = 0 = b$), so y_k and $y_{k'}$ cannot be identical (since $k \neq k'$) and $0 \vee (a \vee b) \not\leq 0 \wedge (a \vee b)$ (since $a \vee b \not\leq 0$). So

$$\{0\} = \{0\} \vee \{0\} \; \theta \; y_k \vee y_{k'} = \{0 \vee (a \vee b) \vee 0 \vee (a \vee b)\} = \{a \vee b\}.$$

Now all the elements of the diamond are $\leq$-between $\{0\}$ and $\{a \vee b\}$, so all of them are θ-related.

Case 2: $\{a,b\} \neq \{1, a \wedge b\}$. We can play a similar trick: Write $x := \{a,b\}_k$ and $y_k := \{1, a \wedge b\}_k$ and $y_{k'} := \{1, a \wedge b\}_{k'}$. Similarly to the first case, $y_k \; \theta \; \{1\} \; \theta \; y_{k'}$ and hence $\{1\} \; \theta \; \{a \wedge b\}$. Since the elements of the diamond are $\leq$-between $\{a \wedge b\}$ and $\{1\}$, they are all θ-related.

Case 3: If neither case 1 nor case 2 obtains, then $\{0, a \vee b\} = \{a,b\} = \{1, a \wedge b\}$. Since $0 \neq 1$ and $a \vee b \neq a \wedge b$ (otherwise $a = b$), we must have $0 = a \wedge b$ and $1 = a \vee b$. Now the trick is to consider $\{0,1\}_{k'}$. It is incomparable with $\{a,b\}_k$: they cannot be identical (since $k \neq k'$) and $1 = a \vee b \not\leq 0 \wedge 1$ and $0 \vee 1 \not\leq a \wedge b = 0$. So

$$\{0,1\}_{k'} = \{0,1\}_{k'} \wedge \{1\} \; \theta \; \{0,1\}_{k'} \wedge \{a,b\}_k = \{0 \wedge 1 \wedge a \wedge b\} = \{0\}.$$

Hence also $\neg\{0,1\}_{k'} \; \theta \; \neg\{0\} = \{1\}$. Since $\neg\{0,1\}_{k'} = \{0,1\}_{k'}$, we have $\{0\} \; \theta \; \{1\}$, hence all elements are θ-related (since they are $\leq$-between $\{0\}$ and $\{1\}$).

It remains to show (2): If $\{a \wedge b\} \theta \{a,b\}_k$, then

$$\{a,b\}_{k'} = \{a,b\}_{k'} \vee \{a \wedge b\} \; \theta \; \{a,b\}_{k'} \vee \{a,b\}_k = \{a \vee b\},$$

so (1) implies the desired conclusion. $\square$

Proposition C.2 *Let A be an involutive lattice. If A is simple, then* $\mathsf{BN}(A)$ *is simple.*

Note, though, that this is neither true for N nor for $\underline{\mathsf{N}}$: for example, $A := \mathsf{N}(\mathbf{2}) = \underline{\mathsf{N}}(\mathbf{2})$ is simple, but $\mathsf{N}(A) = \underline{\mathsf{N}}(A)$ is not (it is depicted in the top-right of figure 6, and we can identify the three black dots).

Proof Using the natural embedding $e : A \to \mathsf{BN}(A)$, define a function

$$\bar{\cdot} : \mathsf{Con}(\mathsf{BN}(A)) \to \mathsf{Con}(A) \qquad\qquad \theta \mapsto \overline{\theta}$$

where $\overline{\theta}$ is defined by: $a\overline{\theta}b$ iff $e(a)\theta e(b)$. Note that, since e is a homomorphism, $\overline{\theta}$ is indeed a congruence. It suffices to show that this function is injective: then also $\mathsf{Con}(\mathsf{BN}(A))$ can have at most two elements.

Let $\theta, \vartheta \in \mathsf{Con}(\mathsf{BN}(A)))$ with $\overline{\theta} \subseteq \overline{\vartheta}$ and show $\theta \subseteq \vartheta$. So assume $x := \{a,b\}_k$ and $y := \{c,d\}_l$ are θ-related, and show $x\vartheta y$.

If x and y are incomparable,

$$\{a \wedge b \wedge c \wedge d\} = x \wedge y \;\theta\; x \;\theta\; x \vee y = \{a \vee b \vee c \vee d\}$$

so $a \wedge b \wedge c \wedge d$ and $a \vee b \vee c \vee d$ are $\overline{\theta}$-related, hence also $\overline{\vartheta}$-related, so $\{a \wedge b \wedge c \wedge d\}$ and $\{a \vee b \vee c \vee d\}$ are ϑ-related, so the $\leq$-between elements x and y also are ϑ-related, as needed.

So assume x and y are comparable, say $x \leq y$. If $x = y$, we have $x\vartheta y$, so assume $x \neq y$. Hence $a \vee b \leq c \wedge d$. Since $\{a,b\}_k \leq \{a \vee b\} \leq \{c \wedge d\} \leq \{c,d\}_l$ and $x\theta y$, they are all θ-related. So lemma C.1 implies $\{a \wedge b\}\theta\{a \vee b\}\theta\{c \wedge d\}\theta\{c \vee d\}$. Hence $a \wedge b$ and $c \vee d$ are $\overline{\theta}$-related, hence $\overline{\vartheta}$-related, so $\{a \wedge b\}\vartheta\{c \vee d\}$. So the $\leq$-between elements x and y also are ϑ-related, as needed. □

Now we can prove theorem 5.2 with standard methods from universal algebra.

Proof of theorem 5.2 Given n, we show $\models_{\mathsf{BN}^n(\mathbf{2})} \neq \models_{\mathsf{BN}^{n+1}(\mathbf{2})}$. Since $\mathbf{2}$ is simple and BN preserves this, we know that all $\mathsf{BN}^n(\mathbf{2})$ are simple.

Now we apply (a corollary of) Jónsson's lemma [e.g. 5, 149]. It states the following (definitions afterward): If K is a finite set of finite algebras and $\mathsf{Var}(K)$ is congruence-distributive, then the subdirectly irreducible algebras of $\mathsf{Var}(K)$ are in $\mathsf{HS}(K)$. We choose $K := \{\mathsf{BN}^n(\mathbf{2})\}$. Then $\mathsf{Var}(K)$ is congruence-distributive, because it is 'lattice-based'. Finally, being simple implies being subdirectly irreducible [5, 59]. So $\mathsf{BN}^{n+1}(\mathbf{2})$ cannot be in $\mathsf{Var}(K)$: Otherwise, Jónsson's lemma implies that it is in $\mathsf{HS}(K)$, which is the set of homomorphic images of subalgebras of K; but all of these have cardinality $\leq |\mathsf{BN}^n(\mathbf{2})|$ and $|\mathsf{BN}^{n+1}(\mathbf{2})| > |\mathsf{BN}^n(\mathbf{2})|$.

Next, by Birkhoff's theorem [5, 79], varieties are precisely equational classes, which means here that there is a set Σ of identities in the language of involutive lattices such that $\mathsf{Var}(K)$ is precisely the set of involutive lattices satisfying these identities. Since $\mathsf{BN}^{n+1}(\mathbf{2}) \notin \mathsf{Var}(K)$, there must be an identity $\varphi(p_1, \ldots, p_n) = \psi(p_1, \ldots, p_n)$ that holds in $\mathsf{BN}^n(\mathbf{2})$ (and in all other members of $\mathsf{Var}(K)$) but not in $\mathsf{BN}^{n+1}(\mathbf{2})$. But this means that the sentences φ and ψ are $\models_{\mathsf{BN}^n(\mathbf{2})}$-equivalent (i.e., $\varphi \models_{\mathsf{BN}^n(\mathbf{2})} \psi$ and $\psi \models_{\mathsf{BN}^n(\mathbf{2})} \varphi$) but not $\models_{\mathsf{BN}^{n+1}(\mathbf{2})}$-equivalent (i.e., either $\varphi \not\models_{\mathsf{BN}^{n+1}(\mathbf{2})} \psi$ or $\psi \models_{\mathsf{BN}^{n+1}(\mathbf{2})} \varphi$). Hence $\models_{\mathsf{BN}^{n+1}(\mathbf{2})} \neq \models_{\mathsf{BN}^n(\mathbf{2})}$. □

We also state a basic lemma on the interplay between varieties and logics for reference in the main text.

Lemma C.3 *Let A be an involutive lattice and V a variety of involutive lattices. Write $\Gamma \models_V \varphi$ iff, for all $A \in V$, $\Gamma \models_A \varphi$. Then*

1. $\Gamma \vDash_A \varphi$ *iff* $\Gamma \vDash_{\mathsf{Var}(A)} \varphi$.
2. $\mathsf{Var}(A) = V$ *iff* $\vDash_V = \vDash_A$.

Proof Concerning (1), since $A \in \mathsf{Var}(A)$, the right-to-left direction is trivial. For the other direction, assume $\Gamma \vDash_A \varphi$, let $B \in \mathsf{Var}(A)$, and show $\Gamma \vDash_B \varphi$. By the assumption, there is a finite Γ_0 such that, for all A-valuations v, $\bigwedge_{\psi \in \Gamma_0} v(\psi) \leq v(\varphi)$. Let $\overline{p} = (p_1, \ldots, p_n)$ be the atoms occurring in the formulas of Γ_0 and in φ. Then $\bigwedge_{\psi \in \Gamma_0} \psi(\overline{p}) \vee \varphi(\overline{p}) = \varphi(\overline{p})$ is an identity satisfied in A. A basic lemma of universal algebra says that $\mathsf{Var}(A)$ satisfies the same identities as A [5, 72]. In particular, B satisfies the above identity, which means that, for any B-valuation w, $\bigwedge_{\psi \in \Gamma_0} w(\psi) \leq w(\varphi)$, as needed.

Concerning (2), we have, by (1), that $\vDash_A = \vDash_{\mathsf{Var}(A)}$. So we need to show $\mathsf{Var}(A) = V$ iff $\vDash_V = \vDash_{\mathsf{Var}(A)}$. If $\mathsf{Var}(A) = V$, then trivially $\vDash_V = \vDash_{\mathsf{Var}(A)}$. And if $\vDash_V = \vDash_{\mathsf{Var}(A)}$, then, for every identity $\varphi = \psi$, it is satisfied in each algebra of V (i.e., $\varphi \vDash_V \psi$ and $\psi \vDash_V \varphi$) iff it is satisfied in each algebra of $\mathsf{Var}(A)$ (i.e., $\varphi \vDash_{\mathsf{Var}(A)} \psi$ and $\psi \vDash_{\mathsf{Var}(A)} \varphi$). Since varieties are equational classes, this means $V = \mathsf{Var}(A)$. □

Appendix D Embeddings

In this section, we prove theorem 6.1 (saying: if A is an involutive lattice, then $\mathsf{BN}(A) \hookrightarrow \mathsf{N}^2(A)$ and, if A is finite, $\mathsf{N}^2(A) \hookrightarrow \underline{\mathsf{N}}^m(A)$ for some m), and theorem 6.2 (saying: $\vDash_{\mathsf{BN}^\infty(\mathbf{2})} = \vDash_{\mathsf{N}^\infty(\mathbf{2})} = \vDash_{\underline{\mathsf{N}}^\infty(\mathbf{2})}$).

The proof of theorem 6.1 is in three steps (D.1–D.3) before putting everything together in step D.4, where we also prove theorem 6.2 as a corollary. The theorems established in the first three steps are also independently interesting as they go beyond just showing theorem 6.1.

D.1 Embedding BN into N: finding more space

Theorem D.1 *Let A be an involutive lattice and $n \geq 1$. The following defines an embedding*

$$e : \mathsf{N}_{n+1}(A) \to (\mathsf{N}_n)^2(A)$$

$$\{a,b\}_k \mapsto \begin{cases} \{\{a,b\}_k\}_0 & \text{if } 0 \leq k < n \\ \{\{a\}_0, \{b\}_0\}_0 & \text{if } k = n \end{cases}$$

The embedding restricts to an embedding $\mathsf{N}^{\leq}_{n+1}(A) \to (\mathsf{N}^{\leq}_n)^2(A)$.

Proof Note that e is well-defined: The definition is independent of the order of a and b, and if $a = b$, the two cases agree. And e preserves the top element: $e(\{1,1\}_0) = \{\{1,1\}_0\}_0$. Also, e preserves negation: Given $\{a,b\}_k$, we have, if $k < n$ that

$$e(\neg\{a,b\}_k) = \{\{\neg a, \neg b\}_k\}_0 = \neg\{\{a,b\}_k\}_0 = \neg e(\{a,b\}_k),$$

and if $k = n$, then

$$e(\neg\{a,b\}_k) = \big\{\{\neg a\}_0, \{\neg b\}_0\big\}_0 = \neg\big\{\{a\}_0, \{b\}_0\big\}_0 = \neg e(\{a,b\}_k).$$

What requires work is to show that e preserves $\wedge$. As a first step, we show that e is an order-embedding:

$$\{a,b\}_k \leq \{c,d\}_l \text{ iff } e(\{a,b\}_k) \leq e(\{c,d\}_l).$$

(In particular, this shows that e is injective.)

($\Rightarrow$). If $\{a,b\}_k \leq \{c,d\}_l$ holds because they are identical, the claim is immediate, so assume $a \vee b \leq c \wedge d$. Note that $e(\{a,b\}_k) \leq \{\{a \vee b\}_0\}_0$: If $k < n$, we have $e(\{a,b\}_k) = \{\{a,b\}_k\}_0 \leq \{\{a \vee b\}_0\}_0$ (using the natural embedding). If $k = n$, then $e(\{a,b\}_k) = \{\{a\}_0, \{b\}_0\}_0 \leq \{\{a \vee b\}_0\}_0$, where the inequality holds since $\{a\}_0 \vee \{b\}_0 = \{a \vee b\}_0$ (by the natural embedding). Similarly, $\{\{c \wedge d\}_0\}_0 \leq e(\{c,d\}_l)$. So the claim follows since $\{\{a \vee b\}_0\}_0 \leq \{\{c \wedge d\}_0\}_0$.

($\Leftarrow$). We analyze the assumption $e(\{a,b\}_k) \leq e(\{c,d\}_l)$ in four cases.

Case 1: $k < n$ and $l < n$. Then $\{\{a,b\}_k\}_0 \leq \{\{c,d\}_l\}_0$, which implies $\{a,b\}_k \leq \{c,d\}_l$.

Case 2: $k < n$ and $l = n$. Then $\{\{a,b\}_k\}_0 \leq \{\{c\}_0, \{d\}_0\}_0$. If this holds because they are identical, then $\{c\} = \{a,b\} = \{d\}$, so $a = b = c = d$, so $\{a,b\}_k = \{a\} = \{c\} = \{c,d\}_l$, as needed. If they are not identical, then

$$\{a,b\}_k \leq \{c\}_0 \wedge \{d\}_0 = \{c \wedge d\}_0 \leq \{c,d\}_l.$$

Case 3: $k = n$ and $l < n$. Similar.

Case 4: $k = n$ and $l = n$. Then $\{\{a\}_0, \{b\}_0\}_0 \leq \{\{c\}_0, \{d\}_0\}_0$. If this holds because they are identical, then $\{a,b\} = \{c,d\}$, so $\{a,b\}_k = \{c,d\}_l$, as needed. If they are not identical, then

$$\{a,b\}_k \leq \{a \vee b\}_0 = \{a\}_0 \vee \{b\}_0 \leq \{c\}_0 \wedge \{d\}_0 = \{c \wedge d\}_0 \leq \{c,d\}_l.$$

Now, it remains to show, for any $\{a,b\}_k$ and $\{c,d\}_l$ that

$$e\big(\{a,b\}_k \wedge \{c,d\}_l\big) = e\big(\{a,b\}_k\big) \wedge e\big(\{c,d\}_l\big).$$

If $\{a,b\}_k$ and $\{c,d\}_l$ are comparable, this follows from monotonicity. So assume they are incomparable. So

$$e\big(\{a,b\}_k \wedge \{c,d\}_l\big) = e\big(\{a \wedge b \wedge c \wedge d\}_0\big) = \{\{a \wedge b \wedge c \wedge d\}_0\}_0.$$

By the intermediate claim, also $e\big(\{a,b\}_k\big)$ and $e\big(\{c,d\}_l\big)$ are incomparable. To compute $e\big(\{a,b\}_k\big) \wedge e\big(\{c,d\}_l\big)$ and show that it equals $\{\{a \wedge b \wedge c \wedge d\}_0\}_0$ we consider the different cases.

Case 1: $k < n$ and $l < n$. Then

$$e\big(\{a,b\}_k\big) \wedge e\big(\{c,d\}_l\big) = \big\{\{a,b\}_k \wedge \{c,d\}_l\big\}_0 = \big\{\{a \wedge b \wedge c \wedge d\}_0\big\}_0,$$

as needed.

Case 2: $k < n$ and $l = n$. Then

$$e\big(\{a,b\}_k\big) \wedge e\big(\{c,d\}_l\big) = \big\{\{a,b\}_k \wedge \{c\}_0 \wedge \{d\}_0\big\}_0 = \big\{\{a,b\}_k \wedge \{c \wedge d\}_0\big\}_0.$$

We cannot have $\{a,b\}_k \leq \{c \wedge d\}_0$: otherwise, if they are identical, $a = c \wedge d = b$, so $a \vee b \leq c \wedge d$, so $\{a,b\}_k \leq \{c,d\}_l$ are comparable; and if they are not identical, $a \vee b \leq c \wedge d$, and $\{a,b\}_k \leq \{c,d\}_l$ again are comparable.

If $\{a,b\}_k \geq \{c \wedge d\}_0$, then, since they cannot be identical, $c \wedge d \leq a \wedge b$, so $\big\{\{a,b\}_k \wedge \{c \wedge d\}_0\big\}_0 = \big\{\{c \wedge d\}_0\big\}_0 = \big\{\{a \wedge b \wedge c \wedge d\}_0\big\}_0$, as needed.

If $\{a,b\}_k$ and $\{c \wedge d\}_0$ are incomparable, then

$$\big\{\{a,b\}_k \wedge \{c \wedge d\}_0\big\}_0 = \big\{\{a \wedge b \wedge c \wedge d\}_0\big\}_0,$$

as needed.

Case 3: $k = n$ and $l < n$. Similar.

Case 4: $k = n$ and $l = n$. Then

$$e\big(\{a,b\}_k\big) \wedge e\big(\{c,d\}_l\big) = \big\{\{a\}_0 \wedge \{b\}_0 \wedge \{c\}_0 \wedge \{d\}_0\big\}_0 = \big\{\{a \wedge b \wedge c \wedge d\}_0\big\}_0,$$

as needed.

Finally, to show that e restricts to $\mathsf{N}^{\leq}_{n+1}(A) \to (\mathsf{N}^{\leq}_{n})^2(A)$ we need to show, for $\{a,b\}_k$ in $\mathsf{N}^{\leq}_{n+1}(A)$, that $e(\{a,b\}_k)$ is in $(\mathsf{N}^{\leq}_{n})^2(A)$ (it then will still be an embedding). By definition, a and b are comparable in A. First consider the case $k < n$. Then $e(\{a,b\}_k) = \{\{a,b\}_k\}_0$. Since a and b are comparable, $\{a,b\}_k \in \mathsf{N}^{\leq}_{n}(A)$. Hence, by the properties of the natural embedding, $\{\{a,b\}_k\}_0 \in (\mathsf{N}^{\leq}_{n})^2(A)$, as needed. Now consider the case $k = n$. Then, by the properties of the natural embedding, also $\{a\}_0$ and $\{b\}_0$ are comparable in $\mathsf{N}^{\leq}_{n}(A)$, so $e(\{a,b\}_k) = \{\{a\}_0, \{b\}_0\}_0 \in (\mathsf{N}^{\leq}_{n})^2(A)$, as needed. □

D.2 Embedding N into $\mathsf{N}^{\leq}_{n,\star}$: squeezing into intervals

Theorem D.2 *Let A be a finite involutive lattice. Let*

$$i : \big\{\{a,b\} : a,b \in A\big\} \to \{0,\ldots,n-1\} =: n$$

be a bijection. Define the involution $\star : n \to n$ by mapping k to $i(\{\neg a, \neg b\})$ where $k = i(\{a,b\})$. Then the following defines an embedding

$$\begin{aligned} e : \mathsf{N}_1(A) &\to \mathsf{N}^{\leq}_{n,\star}(A) \\ \{a,b\}_0 &\mapsto \big\{a \wedge b, a \vee b\big\}_{i(\{a,b\})}. \end{aligned}$$

Proof Well-defined: First, $\star$ is an involution since, for k with $k = i(\{a,b\})$, we have $k^{\star\star} = i(\{\neg a, \neg b\})^\star = i(\neg\neg a, \neg\neg b\}) = i(\{a,b\}) = k$. And e is well-defined: Since $a \wedge b \leq a \vee b$ in A, $e(\{a,b\})$ is indeed an element of $N^{\leq}_{n,\star}(A)$, and the definition of e doesn't depend on the order of a and b.

Injective: Assume $\{a \wedge b, a \vee b\}_{i(\{a,b\})} = \{c \wedge d, c \vee d\}_{i(\{c,d\})}$, and show that $\{a,b\}_0 = \{c,d\}_0$. If $\{a \wedge b, a \vee b\}_{i(\{a,b\})}$ is a singleton, then $a = b$ and also $\{c \wedge d, c \vee d\}_{i(\{c,d\})}$ is a singleton, so $c = d$. Hence $a = b = c = d$, and the claim follows. If $\{a \wedge b, a \vee b\}_{i(\{a,b\})}$ is not a singleton, then $a \neq b$ and also $\{c \wedge d, c \vee d\}_{i(\{c,d\})}$ is not a singleton and $c \neq d$. So $i(\{a,b\}) = i(\{c,d\})$, and the injectivity of i implies $\{a,b\}_0 = \{c,d\}_0$, as needed.

Preserving the top element: $e(\{1,1\}_0) = \{1 \wedge 1, 1 \vee 1\}_{i(\{1,1\})} = \{1\}$.

Preserving negation: We have

$$\begin{aligned} e(\neg\{a,b\}_0) &= \{\neg a \wedge \neg b, \neg a \vee \neg b\}_{i(\{\neg a, \neg b\})} \\ &= \{\neg(a \vee b), \neg(a \wedge b)\}_{i(\{a,b\})^*} \\ &= \neg\big\{a \wedge b, a \vee b\big\}_{i(\{a,b\})} = \neg e(\{a,b\}_0). \end{aligned}$$

Preserving meet: We first show that $\{a,b\}_0 \leq \{c,d\}_0$ iff $e(\{a,b\}_0) \leq e(\{c,d\}_0)$.

($\Rightarrow$) If $\{a,b\}_0 \leq \{c,d\}_0$ holds because of identity, the claim is trivial, so assume this holds because $a \vee b \leq c \wedge d$. Then

$$(a \wedge b) \vee (a \vee b) = a \vee b \leq c \wedge d = (c \wedge d) \wedge (c \vee d),$$

hence $e(\{a,b\}_0) \leq e(\{c,d\}_0)$.

($\Leftarrow$) If $e(\{a,b\}_0) \leq e(\{c,d\}_0)$ holds because they are identical, injectivity implies $\{a,b\}_0 = \{c,d\}_0$. So assume this holds due to the second clause. Then

$$a \vee b = (a \wedge b) \vee (a \vee b) \leq (c \wedge d) \wedge (c \vee d) = c \wedge d,$$

hence $\{a,b\}_0 \leq \{c,d\}_0$.

Now, we show $e(\{a,b\}_0 \wedge \{c,d\}_0) = e(\{a,b\}_0) \wedge e(\{c,d\}_0)$. If $\{a,b\}_0$ and $\{c,d\}_0$ are comparable, this now is immediate, so assume they are not comparable, hence also $e(\{a,b\}_0)$ and $e(\{c,d\}_0)$ are not comparable. Then

$$\begin{aligned} e(\{a,b\}_0 \wedge \{c,d\}_0) &= e(\{a \wedge b \wedge c \wedge d\}_0) \\ &= \big\{a \wedge b \wedge c \wedge d\big\} \\ &= \big\{(a \wedge b) \wedge (a \vee b) \wedge (c \wedge d) \wedge (c \vee d)\big\}_0 \\ &= e(\{a,b\}_0) \wedge e(\{c,d\}_0), \end{aligned}$$

as needed. □

D.3 Embedding $\mathsf{N}^{\leq}_{n,\star}$ into $\mathsf{N}^{\leq}_{n,\mathbf{id}}$: removing the swaps In this subsection, we use the shorthand $\{^3a,b\}^3_k = \{\{\{a,b\}_k\}\}$.

Theorem D.3 *Let A be an involutive lattice, let $n = \{0,\ldots,n-1\}$, and let $\star : n \to n$ be an involution with $i^\star = j$ for $i \neq j$. Let $\bar{\star} : n \to n$ be like $\star$ except that $\bar{\star}(i) = i$ and $\bar{\star}(j) = j$ (i.e., $\bar{\star}$ removes the swap of i and j that $\star$ has). Then the following defines an embedding*

$$\begin{aligned} e : \mathsf{N}^{\leq}_{n,\star}(A) &\to \big(\mathsf{N}^{\leq}_{n,\bar{\star}}\big)^3(A) \\ \{a,b\}_k &\mapsto \begin{cases} \varphi(a,b) & \text{if } k = i \\ \psi(a,b) & \text{if } k = j \\ \{^3a,b\}^3_k & \text{otherwise} \end{cases} \end{aligned}$$

where $\varphi,\psi : A \times A \to \big(\mathsf{N}^{\leq}_{n,\bar{\star}}\big)^3(A)$ are defined by

$$\begin{aligned} \varphi(a,b) &= \Big\{\big\{\{a \wedge b\}_i\big\}_i\,,\ \big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i\Big\}_i \\ \psi(a,b) &= \Big\{\big\{\{a \wedge b\}_j, \{a \wedge b, a \vee b\}_j\big\}_j\,,\ \big\{\{a \vee b\}_j\big\}_j\Big\}_j. \end{aligned}$$

Lemma D.4 *In the setting of theorem D.3, we have*

1. *e is well-defined: in particular, $\varphi(a,b), \psi(a,b) \in \big(\mathsf{N}^{\leq}_{n,\bar{\star}}\big)^3(A)$.*
2. *$\{^3a \wedge b\}^3_0 \leq \varphi(a,b), \psi(a,b) \leq \{^3a \vee b\}^3_0$.*
3. *$\neg\varphi(a,b) = \psi(\neg a, \neg b)$ and $\neg\psi(a,b) = \varphi(\neg a, \neg b)$.*
4. *If $\varphi(a,b)$ or $\psi(a,b)$ are a singleton, then $a = b$.*

Proof Ad (1). Recall that singletons of elements of A are always in $\mathsf{N}^{\leq}_{n,\bar{\star}}(A)$. We have $a \wedge b \leq a \vee b$, so $\{a \wedge b, a \vee b\}_i \in \big(\mathsf{N}^{\leq}_{n,\bar{\star}}\big)^1(A)$. And $\{a \wedge b, a \vee b\}_i \leq \{a \vee b\}_i$ so $\big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i$ is in $\big(\mathsf{N}^{\leq}_{n,\bar{\star}}\big)^2(A)$. Finally,

$$\big\{\{a \wedge b\}_i\big\}_i \leq \big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i,$$

so $\varphi(a,b)$ is in $\big(\mathsf{N}^{\leq}_{n,\bar{\star}}\big)^3(A)$. Similarly for ψ. The function e is well-defined since φ and ψ do not depend on the order of a and b, and all cases agree if $a = b$.

Ad (2). For φ the claim follows since

$$\big\{\{a \wedge b\}_0\big\}_0 \leq \big\{\{a \wedge b\}_i\big\}_i = \big\{\{a \wedge b\}_i\big\}_i \wedge \big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i$$

and

$$\big\{\{a\wedge b\}_i\big\}_i \vee \big\{\{a\wedge b, a\vee b\}_i, \{a\vee b\}_i\big\}_i$$
$$= \big\{\{a\wedge b, a\vee b\}_i, \{a\vee b\}_i\big\}_i \leq \big\{\{a\vee b\}_0\big\}_0.$$

Similarly for ψ.

Ad (3). We have

$$\begin{aligned}
&\neg\varphi(a,b)\\
=&\neg\Big\{\big\{\{a\wedge b\}_i\big\}_i\,,\ \big\{\{a\wedge b, a\vee b\}_i, \{a\vee b\}_i\big\}_i\Big\}_i\\
=&\Big\{\big\{\{\neg(a\wedge b)\}_{i^\star}\big\}_{i^\star}\,,\ \big\{\{\neg(a\wedge b), \neg(a\vee b)\}_{i^\star}, \{\neg(a\vee b)\}_{i^\star}\big\}_{i^\star}\Big\}_{i^\star}\\
=&\Big\{\big\{\{\neg a\vee\neg b)\}_j\big\}_j\,,\ \big\{\{\neg a\vee\neg b, \neg a\wedge\neg b\}_j, \{\neg a\wedge\neg b\}_j\big\}_j\Big\}_j\\
=&\psi(\neg a,\neg b).
\end{aligned}$$

Hence also $\neg\psi(a,b)=\neg\psi(\neg\neg a,\neg\neg b)=\neg\neg\varphi(\neg a,\neg b)=\varphi(\neg a,\neg b)$.

Ad (4). Assume $\varphi(a,b)=\{x\}$. Then

$$\big\{\{a\wedge b\}_i\big\}_i = x = \big\{\{a\wedge b, a\vee b\}_i, \{a\vee b\}_i\big\}_i,$$

so $\{a\wedge b\}_i=\{a\wedge b, a\vee b\}_i$, so $a\wedge b=a\vee b$, which implies $a=b$. Similarly, if $\psi(a,b)$ is a singleton. □

Lemma D.5 *In the setting of theorem D.3, we have: $\{a,b\}_k\leq\{c,d\}_l$ iff $e(\{a,b\}_k)\leq e(\{c,d\}_l)$.*

Proof ($\Rightarrow$) If $\{a,b\}_k\leq\{c,d\}_l$ holds due to identity, the claim is trivial, so assume it is because $a\vee b\leq c\wedge d$. Note that $e(\{a,b\}_k)$ is one of $\varphi(a,b)$, $\psi(a,b)$ and $\{^3a,b\}^3_k$. They are all $\leq\{^3a\vee b\}^3_0$ (lemma D.4). This, in turn, is $\leq\{^3c\wedge d\}^3_0$. Which, in turn, is $\leq$ all of $\varphi(c,d)$, $\psi(c,d)$ and $\{^3c,d\}^3_k$ (lemma D.4). And $e(\{c,d\}_l)$ is one of them.

($\Leftarrow$) Case 1: $e(\{a,b\}_k)=e(\{c,d\}_l)$. Show $\{a,b\}_k\leq\{c,d\}_l$. If $a=b$, then $e(\{a,b\}_k)=\{^3a\}^3$. We claim that then $c=d$; then $e(\{a,b\}_k)=\{^3a\}^3$ and $e(\{c,d\}_l)=\{^3c\}^3$, so $\{a,b\}_k=\{a\}=\{c\}=\{c,d\}_l$, as needed. Indeed, if $l\notin\{i,j\}$, then $\{^3a\}^3=e(\{c,d\}_l)=\{^3c,d\}^3_k$ implies $c=a=d$. If $l=i$, then $\varphi(c,d)=\{^3a\}^3$ is a singleton, so, by lemma D.4 (4), $c=d$. Similarly, if $l=j$. Analogously, the claim follows if $c=d$.

So assume $a\neq b$ and $c\neq d$. We consider the possible cases of k and l being neither i nor j, or one of them.

If $k,j\notin\{i,j\}$, then $\{\{\{a,b\}_k\}\}=e(\{a,b\}_k)=e(\{c,d\}_l)=\{\{\{c,d\}_l\}\}$, so $\{a,b\}_k=\{c,d\}_l$ follows.

If $k=i$ and $l\notin\{i,j\}$, then $\varphi(a,b)=\{\{\{c,d\}_l\}\}$ is a singleton, which implies, by lemma D.4 (4), that $a=b$, which we excluded.

If $k=i$ and $l=i$, then $\varphi(a,b)=\varphi(c,d)$. So

$$\Big\{\big\{\{a\wedge b\}_i\big\}_i\,,\ \big\{\{a\wedge b, a\vee b\}_i, \{a\vee b\}_i\big\}_i\Big\}$$
$$=\Big\{\big\{\{c\wedge d\}_i\big\}_i\,,\ \big\{\{c\wedge d, c\vee d\}_i, \{c\vee d\}_i\big\}_i\Big\}$$

Since $a\wedge b\neq a\vee b$ (otherwise $a=b$) and $c\wedge d\neq c\vee d$ (otherwise $c=d$), $\big\{\{a\wedge b\}_i\big\}_i$ cannot be identical to $\big\{\{c\wedge d, c\vee d\}_i, \{c\vee d\}_i\big\}_i$, so $\big\{\{a\wedge b\}_i\big\}_i=\big\{\{c\wedge d\}_i\big\}_i$, hence

$a \wedge b = c \wedge d$. Similarly, $\big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i$ cannot be $\big\{\{c \wedge d\}_i\big\}_i$, so it is $\big\{\{c \wedge d, c \vee d\}_i, \{c \vee d\}_i\big\}_i$, which implies $\{a \vee b\}_i = \{c \vee d\}_i$, so $a \vee b = c \vee d$. Since a and b are comparable and c and d are comparable, this implies $\{a,b\} = \{c,d\}$. Hence $\{a,b\}_k = \{a,b\}_i = \{c,d\}_i = \{c,d\}_l$.

If $k = i$ and $l = j$, then $\varphi(a,b) = \psi(c,d)$. So

$$\Big\{\big\{\{a \wedge b\}_i\big\}_i\,,\ \big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i\Big\}_i$$
$$= \Big\{\big\{\{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j\big\}_j\,,\ \big\{\{c \vee d\}_j\big\}_j\Big\}_j$$

Since $a \wedge b \neq a \vee b$ (otherwise $a = b$) and $c \wedge d \neq c \vee d$ (otherwise $c = d$), we must have, similar to above,

$$\big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i = \big\{\{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j\big\}_j,$$

so $a \vee b = c \wedge d$. Hence $\{a,b\}_k \leq \{c,d\}_l$, as needed.

The other cases are similar.

Case 2: $e(\{a,b\}_k) \neq e(\{c,d\}_l)$, i.e., $e(\{a,b\}_k) < e(\{c,d\}_l)$. We consider the possible cases of k and l being neither i nor j or one of them.

If $k, l \notin \{i, j\}$, then $\{^3a,b\}^3_k < \{^3c,d\}^3_l$, so, by the properties of the natural embedding, $\{a,b\}_k < \{c,d\}_l$, as needed.

If $k = i$ and $l \notin \{i, j\}$, then $\varphi(a,b) < \{^3c,d\}^3_l$, so

$$\big\{\{a \wedge b\}_i\big\}_i \vee \big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i \leq \{^2c,d\}^2_l.$$

Hence $\big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i \leq \{^2c,d\}^2_l$. Either the two are identical and $\{a,b\}_k \leq \{a \vee b\}_i = \{c,d\}_l$ as needed, or they are not identical and

$$\{a,b\}_k \leq \{a \vee b\}_i = \{a \wedge b, a \vee b\}_i \vee \{a \vee b\}_i \leq \{c,d\}_l.$$

If $k = i$ and $l = i$, then $\varphi(a,b) < \varphi(c,d)$. So

$$\begin{aligned}
&\big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i \\
=&\big\{\{a \wedge b\}_i\big\}_i \vee \big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i \\
\leq&\big\{\{c \wedge d\}_i\big\}_i \wedge \big\{\{c \wedge d, c \vee d\}_i, \{c \vee d\}_i\big\}_i \\
=&\big\{\{c \wedge d\}_i\big\}_i.
\end{aligned}$$

If they are identical, $\{a,b\}_k \leq \{a \vee b\}_i = \{c \wedge d\}_i \leq \{c,d\}_l$, and if they are not identical,

$$\{a,b\}_k \leq \{a \vee b\}_i = \{a \wedge b, a \vee b\}_i \vee \{a \vee b\}_i \leq \{c \wedge d\}_i \leq \{c,d\}_l.$$

If $k = i$ and $l = j$, then $\varphi(a,b) < \psi(c,d)$. So

$$\begin{aligned}
&\big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i \\
=&\big\{\{a \wedge b\}_i\big\}_i \vee \big\{\{a \wedge b, a \vee b\}_i, \{a \vee b\}_i\big\}_i \\
\leq&\big\{\{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j\big\}_j \wedge \big\{\{c \vee d\}_j\big\}_j \\
=&\big\{\{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j\big\}_j.
\end{aligned}$$

If they are identical, then either $\{a \vee b\}_i = \{c \wedge d\}_j$ and hence $\{a,b\}_k \leq \{c,d\}_l$, or $\{a \vee b\}_i = \{c \wedge d, c \vee d\}_j$, so $a \vee b = c \wedge d$, hence $\{a,b\}_k \leq \{c,d\}_l$. If they are not identical,

$$\{a,b\}_k \le \{a\vee b\}_i = \{a\wedge b, a\vee b\}_i \vee \{a\vee b\}_i \\ \le \{c\wedge d\}_j \wedge \{c\wedge d, c\vee d\}_j = \{c\wedge d\}_j \le \{c,d\}_l.$$

The other cases are similar. □

Proof of theorem D.3 Preserving the top-element: We have $e(\{1,1\}_0) = \{^3 1,1\}^3_0$, which is the top element of $\left(\mathsf{N}^{\le}_{n,\bar{\star}}\right)^3(A)$.

Preserving negation: For $\{a,b\}_k$, if $k \notin \{i,j\}$, then also $k^\star \notin \{i,j\}$ (since $i^\star = j$), and

$$e(\neg\{a,b\}_k) = e(\{\neg a,\neg b\}_{k^\star}) = \{\{\{\neg a,\neg b\}_{k^\star}\}\} \\ = \{\{\neg\{a,b\}_k\}\} = \neg\{^3\{a,b\}^3_k = \neg e(\{a,b\}_k).$$

If $k = i$, then, by lemma D.4 (3),

$$e(\neg\{a,b\}_i) = e(\{\neg a,\neg b\}_j) = \psi(\neg a,\neg b) = \neg\varphi(a,b) = \neg e(\{a,b\}_i).$$

Similarly if $k = j$.

Preserving meet: We need to show $e(\{a,b\}_k \wedge \{c,d\}_l) = e(\{a,b\}_k) \wedge e(\{c,d\}_l)$. If $\{a,b\}_k$ and $\{c,d\}_l$ are comparable, then, since e is an order-embedding by lemma D.5, this is immediate; so assume they are incomparable, hence also $e(\{a,b\}_k)$ and $e(\{c,d\}_l)$ are incomparable. Then, on the one hand

$$e(\{a,b\}_k \wedge \{c,d\}_l) = e(\{a\wedge b\wedge c\wedge d\}_0) = \{\{\{a\wedge b\wedge c\wedge d\}_0\}_0\}_0.$$

To compute $e(\{a,b\}_k) \wedge e(\{c,d\}_l)$ and show that it equals $\{\{\{a\wedge b\wedge c\wedge d\}_0\}_0\}_0$, we consider the different cases.

If $k,l \notin \{i,j\}$, then, by the natural embedding,

$$e(\{a,b\}_k) \wedge e(\{c,d\}_l) = \{\{\{a,b\}_k\}_0\}_0 \wedge \{\{\{c,d\}_l\}_0\}_0 \\ = \{\{\{a,b\}_k \wedge \{c,d\}_l\}_0\}_0 = \{\{\{a\wedge b\wedge c\wedge d\}_0\}_0\}_0.$$

If $k = i$ and $l \notin \{i,j\}$, then

$$\begin{aligned} &e(\{a,b\}_k) \wedge e(\{c,d\}_l) = \varphi(a,b) \wedge \{\{\{c,d\}_l\}_0\}_0 \\ =&\Big\{\big\{\{a\wedge b\}_i\big\}_i \wedge \big\{\{a\wedge b, a\vee b\}_i, \{a\vee b\}_i\big\}_i \wedge \big\{\{c,d\}_l\big\}_l\Big\}_0 \\ =&\Big\{\big\{\{a\wedge b\}_0\big\}_0 \wedge \big\{\{c,d\}_l\big\}_0\Big\}_0 \\ =&\Big\{\big\{\{a\wedge b\}_0 \wedge \{c,d\}_l\big\}_0\Big\}_0 \end{aligned}$$

We cannot have $\{a\wedge b\}_0 \ge \{c,d\}_l$: otherwise $\{c,d\}_l \le \{a\wedge b\}_0 \le \{a,b\}_k$ are comparable. If $\{a\wedge b\}_0 \le \{c,d\}_l$, then, since they are not identical, $a\wedge b \le c\wedge d$, so we continue

$$= \Big\{\big\{\{a\wedge b\}_0\big\}_0\Big\}_0 = \Big\{\big\{\{a\wedge b\wedge c\wedge d\}_0\big\}_0\Big\}_0.$$

And if $\{a\wedge b\}_0$ and $\{c,d\}_l$ are incomparable, then we continue

$$= \Big\{\big\{\{a\wedge b\wedge c\wedge d\}_0\big\}_0\Big\}_0.$$

If $k = i$ and $l = i$, then

$$\begin{aligned}
&e(\{a,b\}_k) \wedge e(\{c,d\}_l) = \varphi(a,b) \wedge \varphi(c,d) \\
=&\Big\{ \big\{ \{a \wedge b\}_i \big\}_i \wedge \big\{ \{a \wedge b, a \vee b\}_i, \{a \vee b\}_i \big\}_i \\
&\qquad\qquad \wedge \big\{ \{c \wedge d\}_i \big\}_i \wedge \big\{ \{c \wedge d, c \vee d\}_i, \{c \vee d\}_i \big\}_i \Big\}_0 \\
=&\Big\{ \big\{ \{a \wedge b\}_i \big\}_i \wedge \big\{ \{c \wedge d\}_i \big\}_i \Big\}_0 \\
=&\Big\{ \big\{ \{a \wedge b \wedge c \wedge d\}_0 \big\}_0 \Big\}_0.
\end{aligned}$$

If $k = i$ and $l = j$, then

$$\begin{aligned}
&e(\{a,b\}_k) \wedge e(\{c,d\}_l) = \varphi(a,b) \wedge \psi(c,d) \\
=&\Big\{ \big\{ \{a \wedge b\}_i \big\}_i \wedge \big\{ \{a \wedge b, a \vee b\}_i, \{a \vee b\}_i \big\}_i \\
&\qquad\qquad \wedge \big\{ \{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j \big\}_j \wedge \big\{ \{c \vee d\}_j \big\}_j \Big\}_0 \\
=&\Big\{ \big\{ \{a \wedge b\}_i \big\}_i \wedge \big\{ \{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j \big\}_j \Big\}_0
\end{aligned}$$

We cannot have $\big\{ \{a \wedge b\}_i \big\}_i \geq \big\{ \{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j \big\}_j$: otherwise, if they are identical, $\{a \wedge b\}_i = \{c \wedge d, c \vee d\}_j$, so $a \wedge b = c \vee d$, so $\{a,b\}_k$ and $\{c,d\}_l$ are comparable; and if they are not identical, $\{c \wedge d\}_j \vee \{c \wedge d, c \vee d\}_j \leq \{a \wedge b\}_i$, so $\{c \wedge d, c \vee d\}_j \leq \{a \wedge b\}_i$, which again cannot be identical, hence

$$c \vee d = (c \wedge d) \vee (c \vee d) \leq a \wedge b,$$

which again implies that $\{a,b\}_k$ and $\{c,d\}_l$ are comparable.

If $\big\{ \{a \wedge b\}_i \big\}_i \leq \big\{ \{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j \big\}_j$, then, since they are not identical,

$$\{a \wedge b\}_i \leq \{c \wedge d\}_j \wedge \{c \wedge d, c \vee d\}_j = \{c \wedge d\}_j,$$

so $a \wedge b \leq c \wedge d$ and the earlier equation continues with

$$= \Big\{ \big\{ \{a \wedge b\}_i \big\}_i \Big\}_0 = \Big\{ \big\{ \{a \wedge b \wedge c \wedge d\}_0 \big\}_0 \Big\}_0.$$

If $\big\{ \{a \wedge b\}_i \big\}_i$ and $\big\{ \{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j \big\}_j$ are incomparable, then the earlier equation continues with

$$\begin{aligned}
=&\Big\{ \big\{ \{a \wedge b\}_i \wedge \{c \wedge d\}_j \wedge \{c \wedge d, c \vee d\}_j \big\}_0 \Big\}_0 \\
=&\Big\{ \big\{ \{a \wedge b\}_i \wedge \{c \wedge d\}_j \big\}_0 \Big\}_0 \\
=&\Big\{ \big\{ \{a \wedge b \wedge c \wedge d\}_0 \big\}_0 \Big\}_0.
\end{aligned}$$

If $k \notin \{i, j\}$ and $l = j$ (unlike before, this case is importantly different and hence done in detail),

$$\begin{aligned}
&e(\{a,b\}_k) \wedge e(\{c,d\}_l) = \{^{3}a,b\}^{3}_{k}\}_k\}_k \wedge \psi(c,d) \\
=&\Big\{ \big\{ \{a,b\}_k \big\}_k \wedge \big\{ \{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j \big\}_j \wedge \big\{ \{c \vee d\}_j \big\}_j \Big\}_0 \\
=&\Big\{ \big\{ \{a,b\}_k \big\}_k \wedge \big\{ \{c \wedge d\}_j, \{c \wedge d, c \vee d\}_j \big\}_j \Big\}_0
\end{aligned}$$

We cannot have $\big\{\{a,b\}_k\big\}_k \le \big\{\{c\wedge d\}_j,\{c\wedge d,c\vee d\}_j\big\}_j$: otherwise, if they are identical, $\{a,b\}_k = \{c\wedge d\}_j \le \{c,d\}_l$ are comparable, and if they are not identical, $\{a,b\}_k \le \{c\wedge d\}_j \wedge \{c\wedge d,c\vee d\}_j = \{c\wedge d\}_j \le \{c,d\}_l$ are comparable.

We also cannot have $\big\{\{a,b\}_k\big\}_k \ge \big\{\{c\wedge d\}_j,\{c\wedge d,c\vee d\}_j\big\}_j$: Otherwise, since they are not identical,

$$\{c\wedge d,c\vee d\}_j = \{c\wedge d\}_j \vee \{c\wedge d,c\vee d\}_j \le \{a,b\}_k.$$

However, we cannot have $\{c\wedge d,c\vee d\}_j = \{a,b\}_k$: Otherwise, if $a\neq b$, we get $k=j$, and if $a=b$, then $c\wedge d = a = c\vee d$, so $c=d$, so $\{a,b\}_k = \{a\} = \{c,d\}_l$ are comparable. Hence we must have $c\vee d = (c\wedge d)\vee(c\vee d) \le a\wedge b$. So $\{a,b\}_k \ge \{c,d\}_l$ are comparable.

So $\big\{\{a,b\}_k\big\}_k$ and $\big\{\{c\wedge d\}_j,\{c\wedge d,c\vee d\}_j\big\}_j$ must be incomparable, and the equation continues:

$$\begin{aligned}&=\Big\{\big\{\{a,b\}_k \wedge \{c\wedge d\}_j \wedge \{c\wedge d,c\vee d\}_j\big\}_0\Big\}_0\\ &=\Big\{\big\{\{a,b\}_k \wedge \{c\wedge d\}_j\big\}_0\Big\}_0.\end{aligned}$$

Now, $\{a,b\}_k \not\le \{c\wedge d\}_j$: Otherwise, if they are identical, $\{a,b\}_k = \{c\wedge d\}_j \le \{c,d\}_l$ are comparable; and if they are not identical, then

$$\{a,b\}_k \le \{a\vee b\}_0 \le \{c\wedge d\}_0 \le \{c,d\}_l$$

are comparable. If $\{a,b\}_k \ge \{c\wedge d\}_j$, then, since they are not identical, $c\wedge d \le a\wedge b$, so

$$\Big\{\big\{\{a,b\}_k \wedge \{c\wedge d\}_j\big\}_0\Big\}_0 = \Big\{\big\{\{c\wedge d\}_j\big\}_0\Big\}_0 = \Big\{\big\{\{a\wedge b\wedge c\wedge d\}_0\big\}_0\Big\}_0.$$

And if $\{a,b\}_k$ and $\{c\wedge d\}_j$ are incomparable,

$$\Big\{\big\{\{a,b\}_k \wedge \{c\wedge d\}_j\big\}_0\Big\}_0 = \Big\{\big\{\{a\wedge b\wedge c\wedge d\}_0\big\}_0\Big\}_0.$$

The other cases are similar. □

D.4 Putting everything together We now can prove theorems 6.1 and 6.2. We derive them as a consequence of a sequence of independently interesting corollaries. For this, we use the following pattern of reasoning a lot:

Lemma D.6 *Let $\mathsf{F} \in \{\mathsf{N}_{n,\star}, \mathsf{N}^{\le}_{n,\star}\}$ and $\mathsf{G} \in \{\mathsf{N}_{n',\star'}, \mathsf{N}^{\le}_{n',\star'}\}$. Then the following are equivalent for $j \ge 0$:*

1. *For all (finite) involutive lattices A, $\mathsf{F}(A) \hookrightarrow \mathsf{G}^j(A)$.*
2. *For all $i \ge 1$ and (finite) involutive lattices A, $\mathsf{F}^i(A) \hookrightarrow \mathsf{G}^{ij}(A)$.*

Proof The implication (2)⇒(1) follows by taking $i=1$. For the other direction, assume (1) and show, by induction on $i\ge 1$, that, for any (finite) A, $\mathsf{F}^i(A) \hookrightarrow \mathsf{G}^{ij}(A)$. If $i=1$, this is just the assumption. So assume that, for all $k\le i$ and (finite) A, $\mathsf{F}^k(A) \hookrightarrow \mathsf{G}^{kj}(A)$, and show, for a given (finite) A, $\mathsf{F}^{i+1}(A) \hookrightarrow \mathsf{G}^{(i+1)j}(A)$. Since F is a functor (proposition B.1), the assumption with $k=i$ yields

$$\mathsf{F}\big(\mathsf{F}^i(A)\big) \hookrightarrow \mathsf{F}\big(\mathsf{G}^{ij}(A)\big),$$

so, applying the assumption with $k=1$ to $B := \mathsf{G}^{ij}(A)$ (which is finite if A was finite),

$$\mathsf{F}\big(\mathsf{G}^{ij}(A)\big) \hookrightarrow \mathsf{G}^j\big(\mathsf{G}^{ij}(A)\big)$$

hence $\mathsf{F}^{i+1}(A) \hookrightarrow \mathsf{G}^{(i+1)j}(A)\big)$, as needed. $\square$

Corollary D.7 *For $n \geq 1$ and an involutive lattice A, we have $\mathsf{N}_n(A) \hookrightarrow \mathsf{N}^{2^n}(A)$ and $\mathsf{N}_n^{\leq}(A) \hookrightarrow \underline{\mathsf{N}}^{2^n}(A)$.*

Proof We only prove the first claim; the second follows by adding '$\leq$' to the proof. By induction on n. If $n = 1$, then $\mathsf{N}_n(A) = \mathsf{N}(A) \hookrightarrow \mathsf{N}^{2^1}(A)$ via the natural embedding. So assume we have, for any involutive lattice B, $\mathsf{N}_n(B) \hookrightarrow \mathsf{N}^{2^n}(B)$. Let A be an involutive lattice and show $\mathsf{N}_{n+1}(A) \hookrightarrow \mathsf{N}^{2^{n+1}}(A)$. From theorem D.1, we get $\mathsf{N}_{n+1}(A) \hookrightarrow \mathsf{N}_n^2(A)$. From lemma D.6 with $\mathsf{F} = \mathsf{N}_n$, $\mathsf{G} = \mathsf{N}$, and $j = 2^n$, we get that the induction hypothesis implies $\mathsf{N}_n^2(A) \hookrightarrow \mathsf{N}^{22^n}(A)$, as needed. $\square$

For an involution $\star : n \to n$ with $n = \{0, \dots, n-1\}$, call

$$\big\{(i,j) : 0 \leq i < j \leq n, \star(i) = j\big\}$$

the set of *swaps* of $\star$.[28] The cardinality of this set is the *number of swaps* of $\star$.

Corollary D.8 *For $n = \{0, \dots, n-1\}$, an involution $\star : n \to n$, and a finite involutive lattice A, $\mathsf{N}_{n,\star}^{\leq}(A) \hookrightarrow (\mathsf{N}_{n,\mathsf{id}}^{\leq})^{3^k}(A)$, where k is the number of swaps of $\star$.*

Proof We show by induction on k, if k is the number of swaps of an involution $\star : n \to n$ and A is a finite involutive lattice, then $\mathsf{N}_{n,\star}^{\leq}(A) \hookrightarrow (\mathsf{N}_{n,\mathsf{id}}^{\leq})^{3^k}(A)$. If $k = 0$, then $\star$ is the identity, so $\mathsf{N}_{n,\star}^{\leq}(A)$ and $(\mathsf{N}_{n,\mathsf{id}}^{\leq})^{3^k}(A)$ are even identical. So assume the claim holds for k and show it for $k+1$.

So let $\star : n \to n$ be an involution with $k+1$ many swaps, and let A be a finite involutive lattice. We need to show $\mathsf{N}_{n,\star}^{\leq}(A) \hookrightarrow (\mathsf{N}_{n,\mathsf{id}}^{\leq})^{3^{k+1}}(A)$. Let (i,j) be a swap of $\star$. Let $\bar\star : n \to n$ be the involution that is like $\star$ except that $\bar\star(i) = i$ and $\bar\star(j) = j$. By theorem D.3, $\mathsf{N}_{n,\star}^{\leq}(A) \hookrightarrow (\mathsf{N}_{n,\bar\star}^{\leq})^3(A)$. From lemma D.6 with $\mathsf{F} = \mathsf{N}_{n,\bar\star}^{\leq}$, $\mathsf{G} = \mathsf{N}_{n,\mathsf{id}}^{\leq}$, and $j = 3^k$, we get that the induction hypothesis implies $(\mathsf{N}_{n,\bar\star}^{\leq})^3(A) \hookrightarrow (\mathsf{N}_{n,\mathsf{id}}^{\leq})^{33^k}(A)$, as needed. $\square$

Proof of theorem 6.1 The first part—i.e., that for an involutive lattice A, we have $\mathsf{BN}(A) \hookrightarrow \mathsf{N}^2(A)$— is just theorem D.1 with $n := 2$. So assume A is finite, and show $\mathsf{N}^2(A) \hookrightarrow \underline{\mathsf{N}}^m(A)$ for some m.

By theorem D.2, there is n and an involution $\star : n \to n$ with

$$\mathsf{N}(A) = \mathsf{N}_1(A) \hookrightarrow \mathsf{N}_{n,\star}^{\leq}(A).$$

By corollary D.8, with k being the number of swaps of $\star$,

$$\mathsf{N}_{n,\star}^{\leq}(A) \hookrightarrow (\mathsf{N}_{n,\mathsf{id}}^{\leq})^{3^k}(A).$$

By corollary D.7, we have, for any A, that $\mathsf{N}_{n,\mathsf{id}}^{\leq}(A) \hookrightarrow (\mathsf{N}_{1,\mathsf{id}}^{\leq})^{2^n}(A)$. So, using lemma D.6,

$$(\mathsf{N}_{n,\mathsf{id}}^{\leq})^{3^k}(A) \hookrightarrow (\mathsf{N}_{1,\mathsf{id}}^{\leq})^{3^k 2^n}(A).$$

Hence, for any finite A, there is i_A such that $\mathsf{N}(A) \hookrightarrow \underline{\mathsf{N}}^{i_A}(A)$.

So, for our given A, we have $\mathsf{N}(A) \hookrightarrow \underline{\mathsf{N}}^{i_A}(A)$ and, for $B := \underline{\mathsf{N}}^{i_A}(A)$, we have $\mathsf{N}(B) \hookrightarrow \underline{\mathsf{N}}^{i_B}(B)$. Since N is a functor (proposition B.1),

$$\mathsf{N}\big(\mathsf{N}(A)\big) \hookrightarrow \mathsf{N}\big(\underline{\mathsf{N}}^{i_A}(A)\big) \hookrightarrow \underline{\mathsf{N}}^{i_B}\big(\underline{\mathsf{N}}^{i_A}(A)\big)$$

hence $\mathsf{N}^2(A) \hookrightarrow \underline{\mathsf{N}}^m(A)\big)$ with $m = i_B + i_A$, as needed. $\square$

Proof of theorem 6.2 We need to show $\models_{\mathsf{BN}^\infty(\mathbf{2})} = \models_{\mathsf{N}^\infty(\mathbf{2})} = \models_{\underline{\mathsf{N}}^\infty(\mathbf{2})}$. By lemma C.3, it is enough to show $\mathsf{Var}(\mathsf{BN}^\infty(\mathbf{2})) = \mathsf{Var}(\mathsf{N}^\infty(\mathbf{2})) = \mathsf{Var}(\underline{\mathsf{N}}^\infty(\mathbf{2}))$. We have $\mathsf{Var}(\mathsf{BN}^\infty(\mathbf{2})) = \mathsf{Var}(\mathsf{BN}^n(\mathbf{2}) : n \geq 0)$, and similarly for N^∞ and $\underline{\mathsf{N}}^\infty$.[29] So it suffices to show

$$\mathsf{Var}(\underline{\mathsf{N}}^n(\mathbf{2}) : n \geq 0) \subseteq \mathsf{Var}(\mathsf{N}^n(\mathbf{2}) : n \geq 0) \\ \subseteq \mathsf{Var}(\mathsf{BN}^n(\mathbf{2}) : n \geq 0) \subseteq \mathsf{Var}(\underline{\mathsf{N}}^n(\mathbf{2}) : n \geq 0).$$

The first inclusion holds since each $\underline{\mathsf{N}}^n(\mathbf{2})$ is a subalgebra of $\mathsf{N}^n(\mathbf{2})$. Similarly for the second inclusion: each $\mathsf{N}^n(\mathbf{2})$ is a subalgebra of $\mathsf{BN}^n(\mathbf{2})$.

The point is that, with theorem 6.1, we can also show the third inclusion: By induction on n, we show that $\mathsf{BN}^n(\mathbf{2})$ can be embedded into some $\underline{\mathsf{N}}^m(\mathbf{2})$. If $n = 0$, choose $m := 0$, and if $\mathsf{BN}^n(\mathbf{2}) \hookrightarrow \underline{\mathsf{N}}^{m_n}(\mathbf{2})$ for some m_n, then, by theorem 6.1, $\mathsf{BN}\big(\mathsf{BN}^n(\mathbf{2})\big) \hookrightarrow \underline{\mathsf{N}}^l\big(\mathsf{BN}^n(\mathbf{2})\big)$ for some l, and, since $\underline{\mathsf{N}}^l$ is a functor (proposition B.1), $\underline{\mathsf{N}}^l\big(\mathsf{BN}^n(\mathbf{2})\big) \hookrightarrow \underline{\mathsf{N}}^l\big(\underline{\mathsf{N}}^{m_n}(\mathbf{2})\big)$, hence $\mathsf{BN}^{n+1}(\mathbf{2}) \hookrightarrow \underline{\mathsf{N}}^{m_{n+1}}(\mathbf{2})$ for $m_{n+1} := l + m_n$. □

Acknowledgments

For inspiring discussions and helpful comments, I am grateful to Franz Berto, Frederik Möllerström Lauridsen, Hannes Leitgeb, an anonymous referee, and the audience of the *Workshop on Logic and Philosophy of Mathematics* at *LMU Munich* (April 22, 2023). Part of this work was done within the project 'Foundations of Analogical Thinking' (Project No. 322-20-017) of the research program 'PhDs in the Humanities', financed by the *Dutch Research Council* (NWO).

Notes

1. Shramko and Wansing [36, endnote 6] note that they were informed by J.M. Dunn that this question was first asked by Manfred von Thun in 1975 at a lecture of Dunn. On page 125, they also cite Dunn and Hardegree [10, 277] mentioning information states that are both inconsistent and incomplete. The question is also brought up in an interview with Prof. Nuel D. Belnap [28, 109].

2. If, e.g., the computer received inputs $0,0,1,0,1$ concerning p, it would give it the truth-value $\{0,0,1,0,1\} = \{0,1\}$.

3. If (1) is *true*, then what it says is the case, so it is either *false* or *neither true nor false*, hence it is not *true*. If (1) is not *true*, then it is either *false* or *neither true nor false* (the other two possible truth-values), so what it says is the case, hence it is *true*.

4. An indication for this is that, upon hearing this sentence, we arguably would not interpret it literally, but maybe rather take it to pragmatically convey that there are two senses of 'raining'.

5. Involutive lattices and related structures are well-known in universal algebra. Involutive lattices were studied under the name *i-lattices* in 1958 by Kalman [21]. If the underlying lattice is distributive, one speaks of *DeMorgan lattices*, which were independently introduced by Moisil in 1935. (In fact, just as **2** generates the variety of Boolean algebras, **4** generates the variety of DeMorgan lattices [21].) If one also adds the bounds 0 and 1, one speaks of *DeMorgan algebras*, which are also known as *quasi-Boolean algebras*. (For references, see, e.g., [15, 3].) Thus, bounded involutive lattices are also called

generalized DeMorgan algebras [41, 86]. There are many related notions of negation: for an overview, see, e.g., [17, sec. 3] and [19]. For example, ortholattices are bounded involutive lattices with $x \wedge \neg x = 0$ and $x \vee \neg x = 1$.

6. We could equivalently replace the last two conditions by '$x \leq y$ implies $\neg y \leq \neg x$'.

7. Not to be confused with a limit in category-theory.

8. Meaning $0 \neq 1$. The trivial involutive lattice has just one element and is a fixed point of BN, albeit an uninteresting one (hence it is excluded here).

9. Here $\bigwedge_{\psi \in \Gamma_0} v(\psi)$ is 1 if Γ_0 is empty, and if $\Gamma_0 = \{\psi_1, \dots, \psi_n\}$ is nonempty, it is $v(\psi_1) \wedge \dots \wedge v(\psi_n)$.

10. Proof: Consider $\mathsf{BN}^2(\mathbf{2})$ from figure 5 and the elements $a := \big\{\{0,1\}_n, \{0,1\}_b\big\}_b$, $b := \big\{\{0,1\}_b\big\}$, $c := \big\{\{0,1\}_b, \{1\}\big\}_n$, and $d := \big\{\{0,1\}_b, \{1\}\big\}_b$. Then conjunction introduction fails since $a, b \in D_2$ but $a \wedge b = \big\{\{0\}\big\} \notin D_2$. And conjunction elimination fails since $c \wedge d = \big\{\{0,1\}_b\big\} \in D_2$ but $c \notin D_2$.

11. Proof sketch: First show, by induction on n, that $x \in D_n$ iff $e_{nn+1}(x) \in D_{n+1}$. So $x \in D_n$ iff $e_{nm}(x) \in D_m$. Conclude $x \in D_n$ iff $e_n(x) \in D_\infty$. Hence, if v is an $\mathsf{BN}^n(\mathbf{2})$-valuation with $v(\psi) \in D_n$ for each $\psi \in \Gamma$, then $w := e_n \circ v$ is an $\mathsf{BN}^\infty(\mathbf{2})$-valuation with $v(\psi) \in D_\infty$ for each $\psi \in \Gamma$, so $w(\varphi) \in D_\infty$, so $v(\varphi) \in D_n$.

12. If we use B and define the D_n analogously, then an easy induction argument shows that D_n is all of $\mathsf{B}^n(\mathbf{2})$ except the least element. Then conjunction elimination works: if $v(\varphi \wedge \psi) \in D_n$, then, since $v(\varphi \wedge \psi) \leq v(\varphi)$, also $v(\varphi) \in D_n$. But conjunction introduction can still fail: Consider $\mathsf{B}^2(\mathbf{2})$: here $a := \{\{0\}, \{1\}\}_b$ and $b := \{\{0,1\}_b\}$ are designated, but $a \wedge b = \{\{0\}\}$ is not.

13. Proof: Assume $\Gamma \vDash_B \varphi$ and show $\Gamma \vDash_A \varphi$. So there is a finite $\Gamma_0 \subseteq \Gamma$ such that, for all B-valuations v, $\bigwedge_{\psi \in \Gamma_0} v(\psi) \leq v(\varphi)$. In particular, if we have an A-valuation w, consider the B-valuation $v := f \circ w$, so, since f is a homomorphism,
$$f\Big(\bigwedge_{\psi \in \Gamma_0} w(\psi)\Big) = \bigwedge_{\psi \in \Gamma_0} v(\psi) \leq v(\varphi) = f\big(w(\varphi)\big),$$
hence, since f is an embedding, $\bigwedge_{\psi \in \Gamma_0} w(\psi) \leq w(\varphi)$.

14. Since $\mathbf{4}$ generates the variety of all distributive (bounded) involutive lattices, this claim follows with lemma C.3. There is a subtlety: If $\perp$ and $\top$ are in the language (as we chose here), the interpreting algebraic structures should have 0 and 1. So for FDE, we are looking at DeMorgan algebras. Otherwise, we work with DeMorgan lattices.

15. They are using a slightly different formulation [8, 431]: $\Gamma \vDash_* \varphi$ iff for all involutive lattices A, for all A-valuations v, for all $a \in A$, if $v(\psi) \geq a$ for all $\psi \in \Gamma$, then $v(\varphi) \geq a$. If Γ is finite, this is equivalent to $\bigwedge_{\psi \in \Gamma} v(\psi) \leq v(\varphi)$. So, for finite Γ, $\Gamma \vDash_* \varphi$ iff, for all involutive lattices A, $\Gamma \vDash_A \varphi$. Extending the equivalence to all Γ then is a matter of the compactness theorem.

16. Algebraically, the question is, using lemma C.3, whether the variety generated by $\mathsf{BN}^\infty(\mathbf{2})$ is the variety of all involutive lattices.

17. Also, for B the arguably more natural definition of logical consequence would be in terms of designated values as discussed in section 5.2, while here we look at the algebraic definition.

18. An anonymous referee helpfully remarks: Shouldn't the truth-values of the revenge sentences be fixed points under negation, just like for the original liar sentence? We can arrange for this, e.g., as follows. As usual, the liar λ_1 gets the value $a_1 = \{0,1\}_n = \{0,\neg 0\}_n$; and we continue this pattern by choosing for the revenge sentence λ_{k+1} the value $a_{k+1} := \{a_k, \neg a_k\}_n$. Then all a_k's are fixed points under negation (in fact, $a_{k+1} = \{a_k\}$) and a_{k+1} is different from all the ($\mathsf{N}^k(e)$-embeddings of) the truth-values in $\mathsf{N}^k(\mathbf{2})$. So the neither algebras allow for either, but should the truth-values of the revenge sentences be fixed points under negation? Intuitively, the defining property of the revenge sentence leaves this open: Just like for the usual liar, the defining property of λ_{k+1} arguably implies that its truth-value $a_{k+1} \in \mathsf{N}^{k+1}(\mathbf{2})$ cannot be an $\mathsf{N}^k(e)$-embedding of the previous truth-values $\mathsf{N}^k(\mathbf{2})$; otherwise, it is (the embedding of) 1_k iff it is (the embedding of something) in $\mathrm{N}^k(2) \setminus \{1_k\}$. So a_{k+1} is in $\mathsf{N}^{k+1}(\mathbf{2}) \setminus \mathsf{N}^k(e)(\mathsf{N}^k(\mathbf{2}))$, and hence also $\neg a_{k+1}$ is in there (if $\neg a_{k+1}$ is in $\mathsf{N}^k(e)(\mathsf{N}^k(\mathbf{2}))$, then, since this is a subalgebra, also $\neg\neg a_{k+1} = a_{k+1}$ is in there). In the case of the liar ($k = 0$), we have $\mathsf{N}^1(\mathbf{2}) \setminus \mathsf{N}^0(e)(\mathsf{N}^0(\mathbf{2})) = \big\{\{0,1\}\big\}$, which entails $a_1 = \{0,1\} = \neg a_1$. But for the revenge liar ($k > 0$), the set $\mathsf{N}^{k+1}(\mathbf{2}) \setminus \mathsf{N}^k(e)(\mathsf{N}^k(\mathbf{2}))$ can have many elements, so this argument cannot be used. Concretely, the elements of this set are the black dots in figure 6. For $k = 1$, i.e., the first revenge sentence, these are the dots labeled a, b, c (in the top right). Dot c was used in the main text as the truth-value for λ_2 (its negation is dot a, hence not a fixed point under negation), and dot b was used as the alternative choice above (which is a fixed point under negation). But, as noted, should the truth-values of the revenge sentences be fixed points under negation for some other reason, the neither algebras can afford this.

19. It involves finding fixed points in two 'dimensions': on the one hand, the usual fixed point of the 'jump' operation involved in building truth models, and, on the other hand, a fixed point of the 'neither' operation. To do so, the theory of Leitgeb [26] will be very helpful as it describes how to build truth models over De Morgan algebras (which are our involutive lattices plus distributivity). A challenge is that both operations might have to be iterated more than ω-many times to find a fixed point. For the 'jump' operation this is typically the case to deal with the quantifiers that are usually assumed to be in the language (e.g., for Kripke's construction, one needs to go to the first non-recursive ordinal). But also for the 'neither' operation this might be necessary for reasons similar to those discussed by Field [12]: to avoid sentences that say of themselves that they have some of the non-true values occurring in the neither-iteration.

20. See, e.g., [38, sec. 3.4] for a textbook discussion.

21. See, e.g., [14, 42, 40].

22. At least in the standard version; and even if it is not, there still are issues of higher-order vagueness, as Fara [11] points out.

23. Proof: Recall from theorem 4.1 that $\underline{\mathsf{N}}^\infty(\mathbf{2})$ with embeddings $[\cdot]_k$ is the colimit to the direct system built from the $\underline{\mathsf{N}}^k(\mathbf{2})$ and embeddings $\underline{\mathsf{N}}^k(e)$ (with $\underline{\mathsf{N}}^0(e) = e : \mathbf{2} \to \underline{\mathsf{N}}(\mathbf{2})$). Write $\varepsilon_k : \underline{\mathsf{N}}^k(\mathbf{2}) \to \underline{\mathsf{N}}^{k+1}(\mathbf{2})$ for the natural embedding, and set $\psi_k := [\cdot]_{k+1} \circ \varepsilon_k$. Then $(\underline{\mathsf{N}}^\infty(\mathbf{2}), \psi_k)$ is a cocone to the direct system: since $\underline{\mathsf{N}}^{k+1}(e)\big(\{a\}\big) = \{\underline{\mathsf{N}}^k(e)(a)\}$, we

have $\psi_k(a) = [\{a\}]_{k+1} = [\{\underline{\mathsf{N}}^k(e)(a)\}]_{k+2} = \psi_{k+1} \circ \underline{\mathsf{N}}^k(e)(a)$. So there is connecting embedding $\underline{\mathsf{N}}^\infty(\mathbf{2}) \to \underline{\mathsf{N}}^\infty(\mathbf{2})$, which maps $[a]$, with $a \in \underline{\mathsf{N}}^k(\mathbf{2})$, to $\psi_k(a) = [\{a\}]_{k+1}$; so this precisely is Δ.

24. For this idea, see, e.g., [22, 27], [11, 199-200], or [20, 1562].

25. If φ has value $[\{0,1\}_n]$, this sentence gets value $\Delta\Delta([\{0,1\}_n]) \vee \Delta\neg\Delta([\{0,1\}_n])$, which computes to $[\{\{\{0,1\}_n\}\}] \vee [\{\{\neg\{0,1\}_n\}\}]$, which is $[\{\{\{0,1\}_n\}\}] \vee [\{\{\{0,1\}_n\}\}]$, or simply $[\{\{\{0,1\}_n\}\}]$, which is not the top element 1_∞.

26. This makes Δ look like a provability predicate. This may be plausible if definiteness is linked with assertability and if this, in turn, is linked with provability in a broadly Dummettian way (summarized, e.g., in [32, sec. 6.5 and 6.9]).

27. The above intuition would explain the equivalence as follows. At a high-enough level of iterating 'neither', we confidently give φ a truth-value, say a. So $\Delta\varphi$ gets value $\{a\}$ at the next level, since we need not resort to new borderline values. Negation is just a matter of applying the corresponding operator, so $\neg\Delta\varphi$ gets value $\neg\{a\}$. Similarly, $\neg\varphi$ gets value $\neg a$ and $\Delta\neg\varphi$ gets value $\{\neg a\}$. But, by how negation works on these truth-values, this is the same as $\neg\{a\}$. Hence $\neg\Delta\varphi$ and $\Delta\neg\varphi$ always get the same value.

28. We require $i < j$ to avoid counting (j,i) as another swap: after all, we also have $j^\star = i^{\star\star} = i$.

29. Proof: ($\supseteq$) Each $\mathsf{BN}^n(\mathbf{2})$ is in $\mathsf{Var}(\mathsf{BN}^\infty(\mathbf{2}))$ qua subalgebra of $\mathsf{BN}^\infty(\mathbf{2})$. ($\subseteq$) Varieties are closed under taking direct limits. (The direct limit satisfies an equation if all algebras of the direct system satisfy it [16, p. 156, exc. 34].)

References

[1] J. Adámek, S. Milius, and L. S. Moss. Initial algebras without iteration. In F. Gadducci and A. Silva, editors, *9th Conference on Algebra and Coalgebra in Computer Science (CALCO 2021)*, volume 211, pages 5:1–5:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. . URL https://drops.dagstuhl.de/opus/volltexte/2021/15360. 11

[2] J. Beall, M. Glanzberg, and D. Ripley. Liar Paradox. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2020 edition, 2020. 17

[3] N. Belnap. How a computer should think. In G. Ryle, editor, *Contemporary aspects of philosophy*, pages 30–55. Oriel Press, 1977. Reproduced in [28, 35-53]. 3, 12

[4] N. Belnap. A useful four-valued logic. In J. Dunn and G. Epstein, editors, *Modern Uses of Multiple-Valued Logic*, pages 8–37. D. Reidel Publishing Co., 1977. Reproduced in [28, 55-76]. 3, 12

[5] S. Burris and H. P. Sankappanavar. *A Course in Universal Algebra*. Springer, New York, 1981. 8, 24, 26, 27

[6] P. Cobreros. Vagueness: Subvaluationism. *Philosophy Compass*, 8(5):472–485, 2013. . 3

[7] P. Cobreros, P. Egré, D. Ripley, and R. van Rooij. Tolerant, classical, strict. *Journal of Philosophical Logic*, 41(2):347–385, 2012. 13

[8] M. L. Dalla Chiara. Quantum logic. In D. Gabbay and F. Guenthner, editors, *Handbook of philosophical logic*, volume III of *Synthese library 166*, pages 427–469. D. Reidel Publishing Company, Dordrecht, 1986. 38

[9] M. L. Dalla Chiara and R. Giuntini. Paraconsistent quantum logics. *Foundations of Physics*, 19:891–904, 1989. 13, 14

[10] M. J. Dunn and G. M. Hardegree. *Algebraic Methods in Philosophical Logic*. Number 41 in Oxford Logic Guides. Oxford University Press, Oxford, 2001. 37

[11] D. G. Fara. Gap principles, penumbral consequence and infinitely higher-ordervagueness. In J. Beall, editor, *Liars and Heaps*, pages 195–221. Oxford University Press, Oxford, 2003. Originally published under the name 'Delia Graff'. 39, 40

[12] H. Field. Solving the paradoxes, escaping revenge. In J. Beall, editor, *Revenge of the Liar*, pages 78–144. Oxford University Press, Oxford, 2007. 6, 39

[13] H. Field. *Saving truth from paradox*. Oxford University Press, Oxford, 2008. 6

[14] K. Fine. Vagueness, truth and logic. *Synthese*, 30:265–300, 1975. . 3, 6, 39

[15] J. M. Font. Belnap's four-valued logic and de morgan lattices. *Logic Journal of the IGPL*, 5(3):1–29, 1997. . URL `https://doi.org/10.1093/jigpal/5.3.1-e`. 13, 14, 37

[16] G. Grätzer. *Universal Algebra*. Springer, New York, NY, 2 edition, 2008. 40

[17] W. H. Holliday. A fundamental non-classical logic, 2022. URL `https://arxiv.org/abs/2207.06993`. 3, 15, 38

[18] W. H. Holliday and M. Mandelkern. The orthologic of epistemic modals, 2022. URL `https://arxiv.org/abs/2203.02872`. 3, 15

[19] L. R. Horn and H. Wansing. Negation. In E. N. Zalta and U. Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2022 edition, 2022. 38

[20] L. Incurvati and J. J. Schlöder. Meta-inferences and supervaluationism. *Journal of Philosophical Logic*, 2021. . 40

[21] J. A. Kalman. Lattices with involution. *Transactions of the American Mathematical Society*, 87:485–491, 1958. 14, 37

[22] R. Keefe. *Theories of Vagueness*. Cambridge University Press, Cambridge, 2000. 18, 40

[23] S. C. Kleene. *Introduction to Metamathematics*. North-Holland, Amsterdam, 1952. 3

[24] S. Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72(19):690–716, 1975. 3, 5

[25] J. Lambek. A fixpoint theorem for complete categories. *Mathematische Zeitschrift*, 103: 151–161, 1968. . 11

[26] H. Leitgeb. Truth and the liar in de morgan-valued models. *Notre Dame Journal of Formal Logic*, 40(4):496–514, 1999. . 39

[27] R. K. Meyer. Why I am not a relevantist, 1978. 2

[28] H. Omori and H. Wansing, editors. *New Essays on Belnap-Dunn Logic*. Synthese Library 418. Springer, Cham, 2019. 3, 37, 40

[29] A. Pietz and U. Rivieccio. Nothing but the truth. *Journal of Philosophical Logic*, 42(1): 125–135, 2013. . 13

[30] G. Priest. The logic of paradox. *Journal of Philosophical Logic*, 8:219–241, 1979. 3, 6, 13

[31] G. Priest. Hyper-contradictions. *Logique et Analyse*, 27(107):237–243, 1984. 2, 6

[32] G. Priest. *An Introduction to Non-classical Logic*. Cambridge University Press, Cambridge, 2 edition, 2008. 4, 13, 40

[33] G. Priest. The logic of the catuskoti. *Comparative Philosophy*, 1(2):24–54, 2010. 16

[34] G. Priest. Inclosures, vagueness, and self-reference. *Notre Dame Journal of Formal Logic*, 51(1):69–84, 2010. . 2

[35] G. Restall and F. Paoli. The geometry of non-distributive logics. *The Journal of Symbolic Logic*, 70(4):1108–1126, 2005. 15

[36] Y. Shramko and H. Wansing. Some useful 16-valued logics: How a computer network should think. *Journal of Philosophical Logic*, 34:121–153, 2005. . 4, 16, 37

[37] Y. Shramko and H. Wansing. *Truth and Falsehood: An Inquiry into Generalized Logical Values*. Trends in Logic 36. Springer, Dordrecht, 2011. 2, 15

[38] T. Sider. *Logic for Philosophy*. Oxford University Press, Oxford, 2010. 39

[39] V. Trnková, J. Adámek, V. Koubek, and J. Reiterman. Free algebras, input processes and free monads. *Commentationes Mathematicae Universitatis Carolinae*, 16(2):339–351,, 1975. URL `http://dml.cz/dmlcz/105628`. 11

[40] A. C. Varzi. Supervaluationism and Its Logics. *Mind*, 116(463):633–676, 2007. . URL `https://doi.org/10.1093/mind/fzm633`. 39

[41] D. Ševčovič. Free non-distributive morgan-stone algebras. *New Zealand Journal of Mathematics*, 25:85–94, 1996. 38

[42] T. Williamson. On the structure of higher-order vagueness. *Mind*, 108(429):127–143, 1999. 6, 39