Nāgārjuna's Logic

Aaron J. Cotnoir

8.1 Nāgārjuna and the Catuskoți

The traditional dominant view in Western philosophy is that *truth* and *falsity* are semantic categories that exhaust, without overlap, all (meaningful, declarative) sentences. A traditional Buddhist view is that, in addition to the Western categories, there may be (meaningful, declarative) sentences that are *neither* true *nor* false, and even some that are *both* true *and* false. So, the Buddhist might recognize four distinct truth values: t, f, b, n. These four truth values make up the "four corners of truth," or the *catuṣkoṭi*.

Jay Garfield and Graham Priest (2009) claim that Nāgārjuna (1995), in Mūlamadhyamakakārikā (MMK), uses the catuṣkoṭi positively and negatively. In the positive use, he argues that for a given proposition all four truth values hold. In the negative use, he argues that for a given proposition none of the four truth values hold.

Garfield and Priest then attempt to make sense of these apparently paradoxical uses of the *catuṣkoṭi* by presenting a series of lattices—orderings on truth values. The series of lattices are intended to model the process of *awakening*. Garfield and Priest are silent as to whether these lattices will suffice as forming the basis for models of valid inference—as the basis for modeling Nāgārjuna's *logic*. The question of finding adequate models for Nāgārjuna's *logic* is independently interesting. And since Garfield and Priest's lattices feature truth values of his claims, the question becomes more pressing. In what follows, I argue that Garfield and Priest's lattices cannot ground the logic at play in *MMK*. That is, Garfield and Priest's models are inadequate when applied to Nāgārjuna's arguments; their

semantic analysis of awakening cannot be an accurate semantic analysis of Nāgārjuna's arguments. In section 8.2, I present Garfield and Priest's view. In section 8.3, I argue that validity in the given logic is too weak to represent the arguments made by Nāgārjuna in *MMK*. In section 8.4, I argue for a new interpretation that places greater emphasis on a suggestion made by Deguchi, Garfield, and Priest in an earlier paper (2008). The suggestion is that the *positive catuṣkoṭi* requires disambiguation between the conventional and ultimate perspectives, whereas the negative *catuṣkoṭi* does not. This proposal, I think, solves the problem of applying Garfield and Priest's analysis to Nāgārjuna's logic. As a result, Garfield and Priest's interpretive analysis can be tweaked to apply more broadly than previously supposed.¹

8.2 The Garfield-Priest Interpretation

Nāgārjuna's positive uses of the *catuṣkoṭi* are primarily intended to undermine the appearances of "conventional reality," or *samvrti-satya*. Conventional reality is the world replete with properties, universals, causation, self, and the like—in other words, the world as it is usually seen. Nāgārjuna undermines this picture of reality by arguing that claims about this conventional reality have all four truth values.

Everything is real and is not real, both real and not real, neither unreal nor real. This is the Lord Buddha's teaching. (*MMK* 18.8)

According to Garfield and Priest, the positive *catuṣkoṭi* is best represented by a four-valued lattice, identical to the one proposed by Dunn for the four-valued semantics for the logic of First Degree Entailment (FDE: table 8.1).

We can then use the lattice to generate a semantics for the logic. We give truth values to all the sentences of our language via an interpretation function v such that $v(A) \cdot \{t,f,b,n\}$. We define the behavior of the standard connectives in the following way.

- The value of a conjunction, $v(A \land B)$, is the *greatest lower bound* of v(A) and v(B).
- The value of a disjunction, $v(A \vee B)$, is the least upper bound of v(A) and v(B).
- The value of a negation $v(\neg A) = v(A)$ whenever $v(A) \cdot \{b,n\}$, otherwise it toggles t and f.







AARON J. COTNOIR

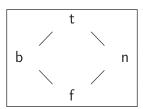


TABLE 8.1 Dunn's four-valued FDE

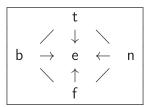


TABLE 8.2 The great death

Importantly, the material conditional ($A \supset B$) is defined in the standard way as $\neg A \lor B$.² This semantics, Garfield and Priest contend, is the correct depiction of the state of the "conventional situation."

A common view, indeed Garfield and Priest's view, is that Nāgārjuna uses the negative *catuṣkoṭi* in order to underscore the view that ultimate reality, or *paramārtha-satya*, is fundamentally ineffable. Consider the application of the negative *catuṣkoṭi* to the proposition that Buddha exists:

We do not assert "empty." We do not assert "non-empty." We neither assert both nor neither. They are asserted only for the purpose of designation. $(MMK\ 22.11)$

The proposition has none of the four truth values present in the positive *catuṣkoṭi*. So, Garfield and Priest, following Sylvan, suggest a fifth truth value, e, to represent the transition to the ultimate perspective—called "the great death" by Jōshu (table 8.2).

This ordering on values is not, strictly speaking, a lattice since e is meant to be incomparable to the other values. How then do sentences receive this value? We introduce a new valuation function, μ (indicated by the arrows in table 8.2), which maps all values in the set $\{t,f,b,n\}$ to e. Moreover, if $\mu(v(A)) = e$, then $\mu(v(\neg A)) = \mu(v(A \land B)) = \mu(v(A \lor B)) = e$. The negative *catuṣkoṭi* utilizes μ on the positive values to yield the denial of the entire conventional picture.







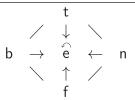


TABLE 8.3 The emptiness of emptiness

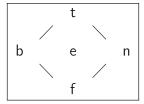


TABLE 8.4 The awakening

Since truth and falsity are themselves merely part of the conventional picture, the negative catuskoti is meant to explicitly say the unsayable, to truly say what cannot truly be said. So, there is a self-undermining paradoxical nature to the negative catuskoti as well. The next stage of awakening for Nāgārjuna, according to Garfield and Priest, is the realization that the distinction between conventional and ultimate reality is itself a convention. This has sometimes been referred to as "the emptiness of emptiness." But in table 8.2, μ (e) is undefined, and so a new ordering must be defined (table 8.3).

In the final stage, when one has appreciated the fact that distinction between the conventional and the ultimate is itself "empty," the usefulness of negative catuskoti, and so μ , becomes null. Garfield and Priest represent this as a lattice with the μ -arrows removed, but with the "empty" value still present (table 8.4).

Such is the extent of Garfield and Priest's presentation.

8.3 Reductio, Modus Ponens, and Validity

If we grant that Garfield and Priest's view is an accurate model of the stages of awakening, there is the additional question as to whether their view will serve as a model for valid inference. It is difficult to get a grasp on just what kind of logical arguments we should expect from the MMK. One must assume that, given the conventional perspective at least, the

acprof-9780190226879.indd 179



semantics are to be used in the same way as Dunn's FDE. That is, validity is defined as designation preservation, where the values t and b are designated. But the movement from lattice to lattice seems to alter the logic entirely. We must decide whether e is a designated value or not. One would expect not, since if it were designated, tables 8.2 and 8.3 would represent a kind of *trivialism*—the thesis that every sentence is designated—a view argued against by Priest (2006).³ Moreover, it appears that Nāgārjuna viewed the negative *catuṣkoṭi* as a kind of *denial*, and Garfield and Priest see μ as a kind of "external" negation.⁴ Hence, one would assume that e is undesignated.

Under these assumptions, it will be useful to consider what argument forms are valid and invalid in such a logic. In doing so, we will see that the lattices as presented cannot be extended to a charitable interpretation of the logic in *MMK*. In what follows, I shall assume that we should attribute Nāgārjuna as endorsing valid arguments whenever possible. In determining valid inferences for this logic, I will restrict my attention to tables 8.1 and 8.2. These tables represent the uses of the positive and negative *catuṣkoṭi*, and hence are central to the argument in *MMK*. Moreover, tables 8.3 and 8.4 are *transitional*; Priest and Garfield claim they are best viewed as *dynamic*. Hence I focus attention on two perspectives: the conventional, and the ultimate as seen in the "great death."

Let us start with the four-valued lattice and corresponding semantics for FDE. I use $X \models_{\text{FDE}} A$ to abbreviate the claim that the argument from a set of premises X to a conclusion A is valid in FDE. It is well known that in many paraconsistent logics, like FDE, *disjunctive syllogism* fails.

$$A \vee B, \neg A v |_{E_{DE}} B \tag{8.1}$$

Let v(A) = b and v(B) = f. The *least upper bound* of b and f is b; so, we have it that $v(A \vee B) = b$. Hence, the first premise is designated. If v(A) = b then $v(\neg A) = b$, so the second premise is designated. But v(B) = f, and hence the conclusion is undesignated.

Since the material conditional is just a disguised disjunction, *modus ponens* will fail as well. Given the semantics for FDE, modus ponens is *not* valid in the framework.

$$A, A \supset Bv \models_{FDF} B \tag{8.2}$$





v(A) = b and v(B) = f yield a counterexample. Since v(A) = b, the first premise is designated. When v(A) = b, $v(\neg A) = b$. Since the *least upper bound* of b and f is b, we have it that $v(\neg A \lor B) = b$. By definition, $v(A \supset B) = v(\neg A \lor B)$, so the second premise— $v(A \supset B)$ —is designated. But v(B) = f, and hence the conclusion is undesignated.

The failure of modus ponens is a serious handicap. In fact, modus ponens makes a frequent appearance in *MMK*. Consider the following example:

When there is change, there is motion. Since there is change in the moving . . . motion is in that which is moving. (MMK 2.2)

Such instances of modus ponens in *MMK* should give us pause.⁶ Nāgārjuna's preferred logic ought to make such arguments *valid*.

We also have a failure of modus tollens in the framework.

$$A \supset B, \neg Bv \mid_{FDF} \neg A \tag{8.3}$$

Let v(A) = n and v(B) = b. So, $v(\neg A) = n$. The *least upper bound* of n and b is t. Hence, $v(\neg A \lor B) = t$; and thus $v(A \supset B) = t$. So the first premise is designated. Since v(B) = b, $v(\neg B) = b$ as well. Hence the second premise is designated. But notice that v(A) = n, and so $v(\neg A) = n$. This means that the conclusion is undesignated.

The failure of modus tollens from the conventional perspective is problematic because, again, Nāgārjuna explicitly reasons in accordance with it.

If apart from the cause of form, there were form, form would be without cause. But nowhere is there an effect without a cause. If apart from form there were a cause of form, it would be a cause without an effect. But there are no causes without effects. . . . Form itself without a cause is not possible or tenable. (MMK 4.2-5)

This passage is clearly made up of two instances of modus tollens, which again Nāgārjuna appears to take as valid.

Further, we also have the failure of hypothetical syllogism:

$$A \supset B, B \supset CV |_{EDE} A \supset C \tag{8.4}$$



1/28/2015 9:13:44 PM

Letting v(A) = t, v(B) = b and v(C) = f yields a counterexample. Since $v(\neg A) = f$, the *least upper bound* of f and b is b. Hence, $v(\neg A \lor B) = b$; and thus $v(A \supset B) = b$. So the first premise is designated. Since v(B) = b, $v(\neg B) = b$ as well. Since the *least upper bound* of b and f is b, $v(\neg B \lor C) = b$. Hence the second premise is designated as well. But notice that $v(\neg A) = f$ and v(C) = f, and so $v(\neg A \lor C) = c$. This means that the conclusion is undesignated.

MMK is rife with hypothetical syllogism. Indeed, Robinson (1957, 196) claims that "the hypothetical syllogism is Nāgārjuna's principal form of inference." For example, the following passage is an instance of hypothetical syllogism:⁷

Since this action is not arisen from a condition, nor arisen causelessly, it follows that there is no agent. If there is no action and agent, where could the fruit of action be? Without a fruit, where is there an experiencer? $(MMK_{17.29-30})$

It should be noted that the failure of modus ponens, modus tollens, and hypothetical syllogism are primarily due to the failures of the *material* conditional in this context. Given the truth functions defined by Garfield and Priest, the material conditional is the only candidate available. We could, however, use the FDE-lattice (with additional resources) as a basis for a relevant conditional. While none of this is explored in Garfield and Priest, the relevant conditional would solve some of the issues mentioned above. For example, a relevant conditional could be constructed to validate modus ponens, modus tollens, and hypothetical syllogism. A more detailed exploration of this option could prove productive. One particularly interesting problem with this approach is that it is difficult to correctly model restricted quantification in relevant logics due to the strength of the conditional. But Nāgārjuna makes frequent use of restricted quantification of the problematic sort. For example:

Whatever is deceptive is false. Compounded phenomena are all deceptive. Therefore they are all false. (*MMK* 13.1)

In any case, moving to a relevant logic is unnecessary; in section 8.4, I present a simple reinterpretation of Garfield and Priest's lattices which validates all these inferences using the *material* conditional.





Even on the relevant conditional semantics for FDE, some conditional-free inferences (primarily disjunctive syllogism) would still be problematic. Moreover, we still have the failure of *reductio ad absurdum*:

$$A \supset (B \land \neg B) v |_{FDE} \neg A \tag{8.5}$$

Let v(A) = t and v(B) = b. Then $v(\neg B) = b$, and since the *greatest lower bound* of b and b is b, we have it that $v(B \land \neg B) = b$. Further, it follows from our supposition that $v(\neg A) = f$. Since the *least upper bound* of b and f is b, we have it that $v(\neg A \lor (B \land \neg B)) = b$. But that's equivalent to $(A \supset (B \land \neg B))$, and so our premise is designated. But $v(\neg A) = f$, which is undesignated. So we have a counterexample.

It may also be objected that the reductio arguments are not intended to establish the *negation* of a proposition, but merely to provide reasons for *rejecting* that proposition.¹⁰ This response, however, would be successful only if one, in the conventional perspective, could distinguish between rejection and accepting a negation. I tentatively suggest that this distinction is not present in the conventional perspective, but only arises out of the transition to the ultimate perspective.¹¹ This seems to be the whole point of the positive *catuskoți*.

Are these arguments part of *upāya*—teachings which are, from the ultimate perspective, false but useful for a better understanding than the one currently held? If so, one may wish to endorse a part of the conventional perspective in order to allow it to lead one to a deeper understanding. Could this be the case with the invalidity of such argument forms?

Unfortunately, no. This interpretation of Nāgārjuna would make sense if these argument forms were valid from the conventional perspective, but invalid from the ultimate perspective. In that case, these argument forms could be considered $up\bar{a}ya$; acceptable for teaching but ultimately to be rejected. But as we have seen, according to Garfield and Priest's interpretation, the positive catuṣkoṭi makes these argument forms invalid from the conventional perspective. What is more, it turns out these forms of argument are valid from the ultimate perspective. So let us turn to that now.

We have seen that the negative *catuṣkoṭi* amounts to a rejection of all sentences using μ , our external negation, and a new value, e. Let us abbreviate the claim that an argument from a set of premises X to a conclusion A is valid according to these semantics thus: $X \vDash_{e} A$.





AARON J. COTNOIR

Remember that $\mu(v(A)) = e$ for every atomic A, and whenever $\mu(v(A)) = e$, $\mu(v(\neg A)) = \mu(v(A \land B)) = \mu(v(A \lor B)) = e$. Thus, any nonempty set of premises X can never be designated. It follows that every valuation according to which X is designated (there are none) will be one where a conclusion A is designated. Thus, it turns out that *every* argument from nonempty X to A is valid.

$$Xv|_{e}$$
 A for all A and nonempty X. (8.6)

Notice that this result does not turn on our earlier assumption that e is undesignated. For if e were designated, then every sentence would be designated, and hence every argument from *X* (whether empty or not) to *A* would be *non*-vacuously valid.

8.4 A Proposal

In light of these problems, in order to capture what Garfield and Priest claim Nāgārjuna is saying with respect to his *arguments*, we need to modify Garfield and Priest's analysis. Indeed, the following proposal is really just a natural extension of some suggestions that Garfield and Priest have made elsewhere. There are two thoughts that inspired to the proposed extension.

The first comes from the suggestion by Deguchi, Garfield, and Priest (2008) that the positive *catuṣkoṭi* is fundamentally a tool used by Nāgārjuna to *undermine* the conventional perspective; it is not the conventional perspective itself. If this is right, then the FDE lattice is not the correct account of conventional truth values.

The second thought comes from suggestions made in Deguchi, Garfield, and Priest (2008) and in Garfield and Priest (2002; 2009) to the effect that contradictions as applied to conventional reality are merely prima facie. It is suggested that these contradictions need to be disambiguated between the conventional and ultimate perspectives and their distinct notions of truth.

Thus, something may be true (conventionally), false (ultimately), true and false (conventionally and ultimately, respectively), and neither true nor false (ultimately and conventionally, respectively). (Garfield 2009, 2)



This suggests that the positive *catuṣkoṭi* is really an intermixing of perspectives, and not constitutive of the conventional perspective itself. Furthermore, Garfield and Priest (2002) suggest such an interpretation is supported by the fact that Nāgārjuna never explicitly endorses a contradiction at the level of conventional reality. Second, as we have already noted, Nāgārjuna takes reductio arguments to be decisive against his opponents. As such, it seems he cannot be committed to the possible truth of contradictions in the conventional picture.

Following these two lines of thought, I propose we revise the Garfield-Priest interpretation by starting with an alternative lattice. In this lattice, truth values reflect *both* perspectives; they are ordered pairs of the values 1,0, where 1 represents "yes" and 0 represents "no." The first member of the pair reflects the answer to the question "Is it conventionally true?"; the second member of the pair reflects the answer to the question "Is it ultimately false"? There are four possibilities: $\langle 1,1\rangle$, $\langle 1,0\rangle$, $\langle 0,1\rangle$, $\langle 0,0\rangle$. Intuitively, $\nu(A) = \langle 1,1\rangle$ when A is both conventionally true and ultimately false; $\nu(A) = \langle 0,1\rangle$ when A is conventionally true and ultimately not false; $\nu(A) = \langle 0,0\rangle$ when A is neither conventionally true nor ultimately false.

So, we can represent the order on these values by the following lattice (table 8.5).

We include semantics for the connectives as follows.

- The value of a conjunction, $v(A \land B)$, is the *greatest lower bound* of v(A) and v(B).
- The value of a disjunction, $v(A \vee B)$, is the *least upper bound* of v(A) and v(B).
- Where $v(A) = \langle x, y \rangle$, the value of the negation $v(\neg A) = \langle 1 x, 1 y \rangle$.

The only significant change from the Dunn semantics above is the treatment of negation. Here, negation toggles $\langle 1,0 \rangle$ and $\langle 0,1 \rangle$, similar to

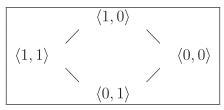


TABLE 8.5 B4





the Dunn semantics, but also toggles (1,1), (0,0). This reflects the suggestion emphasized by Garfield and Priest (2002; 2009) that Nāgārjuna never explicitly endorses contradictions from the conventional perspectives. According to our semantics for negation, a proposition and its negation may never both be conventionally true.

To fully explicate the logic, we need to specify designated values and consequence. Since the primary purpose of the positive *catuṣkoṭi* is to adopt the conventional perspective (if only to undermine it), we take conventional truth as designated. That means that any truth value which has 1 as its first member should be designated, otherwise not. So, $\langle 1,1 \rangle$ and $\langle 1,0 \rangle$ are designated values. On this semantics, we define *conventional* validity (\models_{B4}) in the usual designation-preserving way: an argument from X to A is *conventionally valid* ($X \models_{B4} A$) iff whenever v(x) is designated for all x in X, so too is v(A) designated.

It is important to notice that the B4 semantics for the connectives on this lattice generates a fully classical propositional logic, according to which disjunctive syllogism, modus ponens, modus tollens, hypothetical syllogism, and reductio are all valid. More specifically, these semantics are a four-valued *Boolean algebra*, and hence validate all the same inferences as is two-valued counterpart—the standard classical semantics. ¹² Thus, from the conventional perspective, Nāgārjuna has all the classical inference principles at his disposal, because they are indeed valid. The positive *catuṣkoṭi*, the four corners of truth, arise when we move between perspectives. The clash of perspectives generates the four values necessary for the positive *catuskoṭi*.

It is worth noting that we have not yet discussed the relation between "true" and "not false," nor the relation between "false" and "not true." It is plausible to assume that, from the conventional perspective, being false is equivalent to being not true. The distinction between "false" and "not true" is, in part, what is supposed to result from attending carefully to the positive uses of the *catuṣkoṭi*. So, in short, a 1 in the first argument place of a truth value denotes conventional truth and a 0 in the first argument place denotes conventional falsity (which is just to say conventional untruth). Of course, from the conventional perspective, the above semantics dictates that a proposition and its negation are never both conventionally not true. But this is to be expected, if falsity is conventionally identical to untruth. The result is that a proposition and its negation may never both be conventionally untrue because they may never both be conventionally false. This is simply because conventionally (i.e., *classically*) the falsity of A implies





the truth of $\neg A$, and the falsity of $\neg A$ implies the truth of A. In other words, if we allowed conventionally false contradictions, we would be forced to accept conventionally true ones.

On the other hand, from the ultimate perspective, the negative *catuṣkoṭi* teaches us that being *not false* is not the same as being *true*. What the positive *catuṣkoṭi* adds to the conventional perspective is the possibility of *rejecting* or *denying* propositions and their negations. So, in an important way, our truth values build an external negation into the ultimate perspective. Since a 1 in the second place denotes "not ultimately false," this need not imply that anything is ultimately *true*. Taking this external negation seriously moves us from positive *catuṣkoṭi* to negative *catuṣkoṭi*, as understood by Garfield and Priest, since the attitude there is *rejection* all the way down. It may be objected that the above semantics appears to imply that no proposition and its negation may both be ultimately not false.¹³ Here, I think, it is important to see that the above semantics is useful only for *positive* uses of the *catuṣkoṭi*. *Negative* uses are best understood via Garfield and Priest's μ valuation function and the value e.

In summary, Garfield and Priest's series of lattices are best understood as models of the stages of awakening, but do not give adequate models for valid inference. In order to extend their account to valid inference, one must see that pure conventional reality includes only two truth values, t and f, and that Nāgārjuna's logic is entirely classical. The positive *catuṣkoṭi* utilizes ambiguity between the conventional and ultimate perspectives, as outlined using the B4 semantics. The logic here is still fully classical. Once the transition is made to negative uses of the *catuṣkoṭi*, we employ a new valuation function μ and a new semantic value e corresponding to "emptiness." This leads us directly to the full ultimate perspective.¹⁴

Notes

- 1. I will not here be engaged in textual analysis of MMK, aside from providing some example passages. I am interested in showing that Garfield and Priest's interpretive analysis cannot be applied to coherently across the board. Likewise, my appeal to Deguchi, Garfield, and Priest (2008) in section 8.4 should be viewed as a friendly suggestion on their behalf, not an assessment or endorsement of its being the correct interpretation of MMK.
- 2. Given the truth functions defined by Garfield and Priest, the material conditional is the only candidate available. With additional resources (e.g.,







AARON J. COTNOIR

non-normal frame semantics), a basic relevant conditional could be modeled. For difficulties regarding this approach, see the discussion in section 8.3. In any case, conditionals are clearly beyond the scope of Garfield and Priest's aims in their paper.

- 3. See, in particular, pp. 200-202.
- 4. See the suggestions in Garfield and Priest (2009), §5.
- 5. This assumption may be denied. See the discussion of *upāya* below.
- 6. For another example, see MMK 5.4-5.
- 7. The conclusion is implied in what follows.
- 8. Thanks to an anonymous referee for suggesting this response.
- 9. For details, and some progress toward a solution see Beall et al. (2006).
- 10. Thanks to an anonymous referee for suggesting this response.
- 11. For more discussion of this difference, see section 8.4.
- 12. For proof of this fact, see Beall and Van Fraassen's (2003, 169–172) discussion of the liberal B4 matrix.
- 13. Thanks to an anonymous referee for pointing this out.
- 14. Thanks to Joel Kupperman, Graham Priest, Jay Garfield, and two anonymous referees for suggestions and discussions that led to many improvements in this chapter.



