Ways Things Can't Be

GREG RESTALL

Abstract   Paraconsistent logics are often semantically motivated by considering "impossible worlds." Lewis, in "Logic for equivocators," has shown how we can understand paraconsistent logics by attributing equivocation of meanings to inconsistent believers. In this paper I show that we can understand paraconsistent logics without attributing such equivocation. Impossible worlds are simply *sets* of possible worlds, and inconsistent believers (inconsistently) believe that things are like each of the worlds in the set. I show that this account gives a sound and complete semantics for Priest's paraconsistent logic **LP**, which uses materials any modal logician has at hand.

1 Introduction   Possible worlds semantics has been very useful in modeling not only the *intensionality* of necessity and possibility, future and past. It has also found its place in modeling the *intentionality* of propositional attitudes like belief and knowledge. There is something fruitful in analyzing a belief as a set of possible worlds. The belief is the set of possible worlds in which the belief is true. The belief is true if and only if the actual world is in the corresponding set of propositions. The possible worlds in the set corresponding to the belief represent *how the agent perceives the world to be*. If the belief is false, then the world isn't how the agent sees the world to be, and so the actual world isn't in the set of worlds corresponding to the belief (see Lewis [4] and Stalnaker [9]).

The same can be said of whole belief states just as much as it can be said of individual beliefs. My belief state is the set of worlds consistent with what I believe. This view has been very fruitful, not least because the set-theoretic structure of sets of possible worlds corresponds nicely with the logical structure of entailment relations among propositions and the behavior of propositional connectives like conjunction, disjunction, and negation.

However, the story does not deal well with inconsistent belief. Inconsistent beliefs are true in no possible worlds, so they are each modeled by the same set of worlds—the empty set. My beliefs are often inconsistent, and so are those of many

of you. For example, people have labored under the misconception that there are integers x, y, z and $n > 2$ such that $x^n + y^n = z^n$. As we now know, this is impossible. Frege thought for years that naïve set theory was consistent. Russell showed us that this theory is inconsistent. Lewis gives a more homely example.

> I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. (By 'roughly' I mean "to within 20°.") So each sentence in an inconsistent triple was true according to my beliefs, but not everything was true according to my beliefs. ([3], p. 436)

In each of these cases, there are *no* possible worlds consistent with the agent's explicit beliefs. Clearly, the agent's belief sets display a *degree* of logical coherence. There is a sense in which the agent's beliefs form a corpus of information, closed under *some* sort of consequence relation, even if they are inconsistent. Lewis notes this and posits four desiderata for an account of "truth according to a body of information."

> We can reasonably ask that [a conception of truth according to a corpus] satisfy the following desiderata. (1) Anything that is explicitly affirmed in the corpus is true according to it. (2) Truth according to the corpus is not limited to what is explicitly there, but is to some extent closed under implication . . . (3) Nevertheless, an inconsistency does not—or does not always—make everything true according to the corpus. Hence truth according to the corpus is closed not under classical implication generally, but under some sort of restricted implication capable of quarantining inconsistencies. Maybe not all inconsistencies can be quarantined successfully, but many can be. (4) A sentence is false according to the corpus if and only if its negation is true according to the corpus. ([3], p. 435)

Lewis goes on to develop an account which has these four properties. I'll call it the *chunking account* as it treats inconsistency by partitioning the agent's beliefs into consistent subsets [3], [4]. Take the case of Nassau Street and the railway. Lewis holds A, that Nassau street runs east-west; B, that the railway runs north-south; and C, that they run parallel. For Lewis, the way to understand this is to see the agent's beliefs as "chunked" into separate categories, to quarantine the inconsistency. The agent no longer has one real belief set, but several. One, perhaps, in which $A \land C$ is true, and one, perhaps, where $B \land C$ is true. But none in which $A \land B \land C$ is true. These consistent belief sets are treated as before, as giving rise to sets of possible worlds, which then are used to see the *content* of the beliefs. An agent's beliefs are those which appear in at least one of the corresponding belief sets. The agent here believes $A \land C$, and $B \land C$, but not $A \land B \land C$, as this appears in *neither* theory.

This "chunking" account clearly gets *something* right.[1] Not all beliefs are available to be compared and conjoined with all other beliefs at all times. However, it has problems. I will consider just a few.

First, Lewis doesn't give us a principled way of finding the relevant consistent subsets of the agent's explicit beliefs. There are many ways to generate such sets. How many of the agent's beliefs are we to add to $A \land C$? Second, the procedure of finding consistent subsets of a set of propositions is not in general even decidable, so even finding the *inputs* with which to perform our logical consequences (to find the final belief set) is not computable. Third, not all inconsistent beliefs are formed by conjoining two or more consistent beliefs. The straightforward ways of reading Fer-

mat's last theorem and of naïve set theory seem to be examples of this. There *may* be a way of analyzing these sorts of inconsistent beliefs under Lewis's analysis. However, to do so would probably take us very far afield from the "surface grammar" of the beliefs in question.

Finally, consider what the agent does when discovering that he or she holds inconsistent beliefs. They perform the deduction, see that they're committed to $A \wedge B \wedge C$ and see that they must reject something if they are to remain consistent. What can they do? One plausible way of describing the situation is as follows: they realize that *according to their beliefs* $A \wedge B \wedge C$ holds, but they know independently that $\sim(A \wedge B \wedge C)$. They are holding obviously inconsistent beliefs. They know this is impossible, and they see which one they ought to modify in order to resolve the inconsistency. At this stage the beliefs are *not* quarantined in separate compartments. They are mingling, and it is only a result of this mingling that the agent realizes that there is a problem. Lewis realizes this.

> I need not quarrel with anyone who wishes to put forward a fifth desideratum for a conception of truth according to a corpus: (5) the orthodox rules for $\wedge$ and $\vee$ must apply without exception ... What I doubt is that there is any useful and intuitive conception that satisfies all five desiderata. When asked to respect $(1)-(4)$, I come up with a conception that violates (5). ([3], p. 438)

In what follows I will construct a notion of "truth according to a corpus" which respects *all five* conditions. One obvious way to do this is to start with a different *logic*. A paraconsistent logic is one in which the inference $A, \sim A \vdash B$ fails. If we choose such a logic, we can construct "impossible worlds" on a par with possible worlds as sets of propositions satisfying some maximality condition and closed under consequence. A set can contain both A and $\sim A$ without being the total set of all propositions. This technique is helpful for those who understand the logics used, but it is not helpful for those who find the logic obscure.

In this paper I will proceed in the other direction. I will use the techniques readily available to theorists like Lewis and Stalnaker—possible worlds and a first-order language, related in the usual classical way—and out of this machinery we will see how "impossible worlds" can be constructed in a well-motivated way. Then finally, the logic is read off the semantics. It will turn out (in the end) to be a familiar paraconsistent logic. However, it will be motivated by our semantic considerations, not imposed from the outset.

2 Propositional logic   Let's start with propositional logic to see how the technique works in a simple context.

Definition 2.1   A *propositional model* is a set W of possible worlds, a set P of atomic propositions, and an interpretation $\Vdash$ relating atomic propositions to worlds. The relation $\Vdash$ is then extended to the entire vocabulary in the usual way.

1. $x \Vdash A \wedge B$ iff $x \Vdash A$ and $x \Vdash B$.
2. $x \Vdash A \vee B$ iff $x \Vdash A$ or $x \Vdash B$.
3. $x \Vdash \sim A$ iff $x \nVdash A$.

These models are standard fare. Given such a model, a proposition A is possible if and only if there is some world w where $w \Vdash A$. It is impossible otherwise. Clearly, contradictions of the form $A \wedge {\sim}A$ are impossible. Other propositions might be impossible too. It might be that in the model, there is no possible world w such that $w \Vdash p$, for some atomic proposition p. This is allowed by the constraints of the model.

Now consider the case where there are three propositions, A, B, and C, which are each pairwise consistent but their conjunction $A \wedge B \wedge C$ is inconsistent. So, there is no possible world w such that $w \Vdash A \wedge B \wedge C$. What are we to say about a belief state which includes all three beliefs? We have already seen the Lewis/Stalnaker account, which is to "chunk" the belief set into consistent compartments. There are no worlds at which the three-way conjunction is true, but there are worlds at which any pair is true. Our belief set is quarantined into consistent subsets. Say X is the set of all worlds in which $A \wedge B$; Y is the set of all worlds in which $A \wedge C$ (we could add Z, the set of worlds in which $B \wedge C$ if we wished, but X and Y are enough to illustrate the point). X and Y are disjoint. They together somehow represent the belief state of the inconsistent believer. How can we understand the belief set in a way which doesn't require it to be divided up?

Consider X. Any world in X is consistent with part of the belief set, but not the whole. It is consistent with $A \wedge B$, but not with $A \wedge C$. If there were a world *both* in X and in Y, this would be enough to model $A \wedge B \wedge C$. Of course, there are no such worlds, as $A \wedge B \wedge C$ is inconsistent. However, we can construct something to do the job. If $x \in X$ and $y \in Y$, consider "pasting together" x and y, like superimposing together two maps which inconsistently describe the landscape, or concatenating two stories which inconsistently describe the situation. How do *x and y together* describe the world to be? There are many ways you could define this notion. I will choose a general framework for defining the notion of 'truth at a *set* of possible worlds' which agrees with the notion of truth at a possible world in the case where the set of possible worlds is a singleton. The idea is that something is true in the set of possible worlds just when it's true in some member of that set. It is false in the set of possible worlds just when it is false in some member in that set. However, instead of leaving the condition at that (which would simply be the "chunking" idea all over again) we will allow overlap. A conjunction is true at a set just when both conjuncts are true at that set. A conjunction is false at a set just when either conjunct is false. This leads to the following generalization.

Definition 2.2   A *world* in a propositional model is a nonempty set of possible worlds in that model. Truth and falsity ($\Vdash^{+}$ and $\Vdash^{-}$) are defined for worlds in propositional models as follows.

1. $X \Vdash^{+} p$ iff $x \Vdash p$ for some $x \in X$.
2. $X \Vdash^{-} p$ iff $x \nVdash p$ for some $x \in X$.
3. $X \Vdash^{+} {\sim}A$ iff $X \Vdash^{-} A$.
4. $X \Vdash^{-} {\sim}A$ iff $X \Vdash^{+} A$.
5. $X \Vdash^{+} A \wedge B$ iff $X \Vdash^{+} A$ and $X \Vdash^{+} B$.
6. $X \Vdash^{-} A \wedge B$ iff $X \Vdash^{-} A$ or $X \Vdash^{-} B$.
7. $X \Vdash^{+} A \vee B$ iff $X \Vdash^{+} A$ or $X \Vdash^{+} B$.
8. $X \Vdash^{-} A \vee B$ iff $X \Vdash^{-} A$ and $X \Vdash^{-} B$.

So, we have an answer to our question "what is a way the world can't be?" It is a *set* of ways the world can be. The world could be like *x* and it could be like *y* for two different possible worlds *x* and *y*. However, it can't be *both* like *x* and like *y*.

The definition of $\Vdash^+$ is a simple way of defining "truth in a model" which (it will turn out) satisfies Lewis's five desiderata. The following results show that the notion is well behaved.

Lemma 2.3   *In any propositional model, $\{x\} \Vdash^+ A$ if and only if $x \Vdash A$, and $\{x\} \Vdash^- A$ if and only if $x \not\Vdash A$.*

Proof: The proof is a simple induction on the complexity of *A*.   □

Lemma 2.4   *If $X \Vdash^+ A$ and $X \subseteq Y$ then $Y \Vdash^+ A$. Similarly, if $X \Vdash^- A$ and $X \subseteq Y$ then $Y \Vdash^- A$.*

Proof: This too is a simple induction on the complexity of *A*.   □

Theorem 2.5   *In any propositional model, if $x \Vdash A$ then $X \Vdash^+ A$ for all $x \in X$. However, the converse doesn't hold.*

Proof: The first part is yet another simple induction on the complexity of *A*. The second part is also simple. If $x \Vdash A$ and $y \Vdash {\sim}A$, then $\{x, y\} \Vdash^+ A \wedge {\sim}A$, but as both *x* and *y* are consistent, neither $x \Vdash A \wedge {\sim}A$ nor $y \Vdash A \wedge {\sim}A$.   □

These results go some way toward showing that the notion of truth and falsity at a world (in our sense) is reasonably well behaved. Another way to understand the 'logic' of the semantics is to examine how propositions behave. As is well known, on the 'standard account' propositions can be modeled as sets of possible worlds (the set of worlds at which the proposition is true). Then the logical operations of conjunction, disjunction, and negation are modeled by the set theoretic operations of intersection, union, and negation.

How are propositions modeled once we consider impossible worlds as well as possible worlds? Take a particular simple possible worlds model, and the relations $\Vdash^+$ and $\Vdash^-$ which arise from considering impossible worlds as well as possible ones. Consider the set of worlds (that is, nonempty sets of possible worlds) at which a proposition is true. What must this set be like? Well, Lemma 2 tells us that if *X* is in this set, so is any superset *Y* of *X*. That is, the set of worlds at which a proposition is true must be *closed upwards*. If *X* is in the set, so are all supersets of *X*.

A proposition, however, is not uniquely defined by the set of worlds at which it is true. Consider the following simple model with two possible worlds *a* and *b*, at which $a \Vdash p \wedge {\sim}q$ and $b \Vdash {\sim}p \wedge {\sim}q$. Clearly $X \Vdash^+ p \vee {\sim}p$ and $X \Vdash^+ {\sim}q$ for any nonempty set *X* of worlds. However, the propositions $p \vee {\sim}p$ and ${\sim}q$ are importantly different. The negation of $p \vee {\sim}p$ is true at one impossible world, $\{a, b\}$, whereas the negation of *q* is true nowhere—not even at $\{a, b\}$. So, propositions ought not be modeled just at the set of worlds at which they are true, but also at the set of worlds at which they are *false*. An appropriate model for propositions is as follows.

Definition 2.6   A *proposition* in a simple possible worlds model is a *pair* $\langle \mathcal{X}^+, \mathcal{X}^- \rangle$ of sets of worlds, satsifying the following conditions.

1. Both $\mathcal{X}^+$ and $\mathcal{X}^-$ are closed upwards.

   2. For any $w \in W$, $\{w\}$ is in exactly one of $\mathcal{X}^+$ and $\mathcal{X}^-$.

In any proposition $\langle \mathcal{X}^+, \mathcal{X}^- \rangle$, $\mathcal{X}^+$ is the set of worlds at which the proposition is true, and $\mathcal{X}^-$ is the set of worlds at which it is false. As propositions are either true or false (and not both) at *possible* worlds, for each $w \in W$, $\{w\}$ must be in exactly one of $\mathcal{X}^+$ and $\mathcal{X}^-$. It is simple to show that conjunction, disjunction, and negation work as you would expect on these 'propositions'.

Theorem 2.7   *The conjunction of* $\langle \mathcal{X}^+, \mathcal{X}^- \rangle$ *and* $\langle \mathcal{Y}^+, \mathcal{Y}^- \rangle$ *is* $\langle \mathcal{X}^+ \cap Y^+, \mathcal{X}^- \cup Y^- \rangle$. *The disjunction of* $\langle \mathcal{X}^+, \mathcal{X}^- \rangle$ *and* $\langle \mathcal{Y}^+, \mathcal{Y}^- \rangle$ *is* $\langle \mathcal{X}^+ \cup Y^+, \mathcal{X}^- \cap Y^- \rangle$. *The negation of* $\langle \mathcal{X}^+, \mathcal{X}^- \rangle$ *is* $\langle \mathcal{X}^-, \mathcal{X}^+ \rangle$.

Proof: This is a simple matter of taking the definitions of $\Vdash^+$ and $\Vdash^-$ and the connectives and translating them into our talk of propositions.   ☐

This construction is not new—Dunn used this as a semantics for first-degree entailment (see Anderson and Belnap [1] and Dunn [2]).

   So propositions, in these models, have a richer structure than that in simple possible worlds models. However, the connectives maintain their simple set theoretic interpretation—complicated only by the need to keep track of, not only where a proposition is true, but also where it is also false. You might protest, and rightly so, that it seems that negation is not interpreted by set theoretic negation or complement but by a "swapping" operation. However, all is not lost—the two sets in the pair of sets are already related by a negation. For any possible world x, $\{x\}$ is in $\mathcal{X}^+$ or $\mathcal{X}^-$ but *not* both. So, the set theoretic analogue of negation is still recognizably "negative."

   In fact, if you modeled old-style propositions, not as sets of possible worlds but as "cuts" of the set of all possible worlds into two sets—one in which the proposition is true and the other in which it is false—this would be a merely notational difference. The logic would be the same. However, the definitions of conjunction, disjunction, and negation in such a modeling would be *exactly the same* as what we have for our propositions. So, the definition is not "nonclassical" in the sense of changing the meanings of the connectives.

   The resulting "logic" is rather simple. Let's say that $\Sigma \vdash_P A$ ("P" for paraconsistent) just when in each model, for each world X, if $X \Vdash^+ B$ for each $B \in \Sigma$ then $X \Vdash^+ A$ too. This is an extension of the standard definition of consequence to our case in which we admit impossible worlds. It turns out that the consequence relation $\vdash_P$ so defined is the consequence relation of Priest's paraconsistent logic **LP** [5]. An **LP**-evaluation is a function v from the language to the set $\big\{\{1\}, \{0\}, \{0, 1\}\big\}$ of truth values 'true', 'false', and 'both', satisfying the following conditions.

   1. $1 \in v(\sim A)$ iff $0 \in v(A)$.
   2. $0 \in v(\sim A)$ iff $1 \in v(A)$.
   3. $1 \in v(A \wedge B)$ iff $1 \in v(A)$ and $1 \in v(B)$.
   4. $0 \in v(A \wedge B)$ iff $0 \in v(A)$ or $0 \in v(B)$.
   5. $1 \in v(A \vee B)$ iff $1 \in v(A)$ or $1 \in v(B)$.
   6. $0 \in v(A \vee B)$ iff $0 \in v(A)$ and $0 \in v(B)$.

It is clear that **LP**-evaluations mirror closely the definition of truth in a world. We say that an argument from Σ to A is **LP**-valid (written '$\Sigma \vdash_{\textbf{LP}} A$') if and only if for

every **LP**-evaluation v such that $1 \in v(B)$ for each $B \in \Sigma$, we also have $1 \in v(A)$. It should come as no great surprise that the following result holds.

Theorem 2.8 $\Sigma \vdash_P A$ *if and only if* $\Sigma \vdash_{\mathbf{LP}} A$.

Proof: To show that if $\Sigma \vdash_{\mathbf{LP}} A$ then $\Sigma \vdash_P A$, consider any model and any world X in any model. Define an **LP**-valuation v as follows. $1 \in v(p)$ if and only if $X \Vdash^+ p$, and $0 \in v(p)$ if and only if $X \Vdash^- p$. It is simple to show that $1 \in v(A)$ if and only if $X \Vdash^+ A$ and $0 \in v(A)$ if and only if $X \Vdash^- A$. So, by the definition of **LP**-validity, if $\Sigma \vdash_{\mathbf{LP}} A$, then if $X \Vdash^+ B$ for each $B \in \Sigma$, then since $1 \in v(B)$ for each $B \in \Sigma$, we have $1 \in v(A)$ and hence $X \Vdash^+ A$ as desired.

The converse is slightly more difficult to prove. Suppose that $\Sigma \nvdash_{\mathbf{LP}} A$. We wish to construct a model containing a world X at which Σ is true but A is not. To do this, consider an **LP**-valuation v for which each element of Σ is (at least) true, and for which A is not. Σ', the set of formulas which are true (at least) according to v, is a superset of Σ, such that $\Sigma' \nvdash_{\mathbf{LP}} A$. We wish to find a world which makes each element of Σ' true but not A. This is reasonably straightforward. The map v sends each atomic proposition to one of $\{0\}$, $\{1\}$, and $\{0, 1\}$. Given such a map v, we will call a map v' assigning either 0 or 1 to each atomic proposition a *consistent subevaluation* of v just when for each p, if $v'(p) = i$ then $i \in v(p)$.

Now we construct our model. What are the possible worlds in our model? Any set of worlds will do, provided that we have one world for each consistent subevaluation v' of v. Let the world corresponding to v' be $w(v')$. We set the evaluation $\Vdash$ on worlds to satisfy

$$w(v') \Vdash p \text{ iff } v'(p) = 1$$

let the usual inductive conditions do the work for the connectives. We don't care how $\Vdash$ evaluates propositions at other worlds (if there are any) for they are irrelevant for our purposes. We will take our world X to be the set of all the worlds $w(v')$, where v' is a consistent subevaluation of v. We need just show that $X \Vdash^+ B$ for each $B \in \Sigma'$ and that $X \nVdash^+ A$. We will do more. We will prove Lemma 2.9 by induction on B.

Lemma 2.9   *For each B, $X \Vdash^+ B$ if and only if $B \in \Sigma'$ and $X \Vdash^- B$ if and only if $\sim B \in \Sigma'$.*

Proof: The proof is direct. For the base cases, consider $p \in \Sigma'$. If $p \in \Sigma'$ then $1 \in v(p)$. That means there is some consistent subevaluation v' of v for which $v'(p) = 1$, and so, that $w(v') \Vdash p$ and hence, that $X \Vdash^+ p$. Conversely, if $X \Vdash^+ p$ then $w(v') \Vdash p$ for some v' and so, $1 \in v(p)$, giving $p \in \Sigma'$. Similarly, if $\sim p \in \Sigma'$ then $0 \in v(p)$. That means there is some consistent subevaluation v' of v for which $v'(p) = 0$, and so, $w(v') \nVdash p$ and hence, $X \Vdash^- p$. Conversely, if $X \Vdash^- p$ then $w(v') \nVdash p$ for some v' and so, $0 \in v(p)$, giving $\sim p \in \Sigma'$.

The connective rules are trivial. Consider conjunction: $A \wedge B \in \Sigma'$ is equivalent to $A \in \Sigma'$ and $B \in \Sigma'$. This happens if and only if $X \Vdash^+ A$ and $X \Vdash^+ B$, which is in turn equivalent to $X \Vdash^+ A \wedge B$ as desired. On the other hand, $\sim(A \wedge B) \in \Sigma'$ if and only if $\sim A \in \Sigma'$ or $\sim B \in \Sigma'$, if and only if $X \Vdash^- A$ or $X \Vdash^- B$ if and only if $X \Vdash^- A \wedge B$. For disjunction, swap 'and' and 'or' in this reasoning.

Finally, for negation, $\sim A \in \Sigma'$ if and only if $X \Vdash^- A$, if and only if $X \Vdash^+ \sim A$ as desired, and $\sim\sim A \in \Sigma'$ if and only if $A \in \Sigma'$ if and only if $X \Vdash^+ A$ if and only if

$X \Vdash^{-} {\sim}A$, completing the proof of the lemma.   □

So, if $\Sigma \nvdash_{\mathbf{LP}} A$, we can construct a simple possible worlds model giving us $\Sigma \nvdash_{P} A$ as follows. Take the **LP**-evaluation v which makes all of Σ true but not A. Consider this evaluation and look at all of its consistent subevaluations. Take these to be the worlds which are inconsistently combined to model Σ. A is not true at this inconsistent combination. So, you have a simple possible worlds model which invalidates $\Sigma \vdash_{P} A$, as desired.   □

What has this result shown? We can see that our definition of truth in a world (for both possible and impossible worlds) agrees with a prior definition of a paraconsistent logic. As a result, we can have *two* notions of truth according to a body of information. First, you can chunk your belief set into consistent parts and consider those. Second, you can allow interaction between those parts to get an inconsistent theory that still has a degree of logical closure. The notion of logical consequence relevant to these inconsistent theories is a preexisting paraconsistent logic **LP**. We did not start out to motivate **LP**. We started out to consider a notion of 'truth in theory' which respected Lewis's five desiderata. It turns out that this kind of inconsistency tolerance gives us a known logic. The result is important because it shows that **LP** is adequate not only as a logic of inconsistency, but as the logic of inconsistency *generated by preexisting consistency*. For in our models the 'ground' for truth is still the possible world. The impossible worlds are 'epiphenomenal'. Truth at an impossible world is determined completely by truth at the possible worlds in which it consists. There is no more to inconsistency in these propositional models.

3 Quantifiers   We have seen how to model inconsistent belief and other inconsistent propositions in propositional languages. The result agrees with Priest's logic of paradox. However, beliefs have more logical structure than is exhibited by propositional logic. The predicate calculus is another degree of logical structure useful in examining content of beliefs. The addition of predicates and quantifiers to the picture makes things more complex for our project. Again we start with our classical models.

Definition 3.1   A *predicate model* for a language with n-ary predicates F^n and m-ary function symbols f^m is a set W of possible worlds, and a set D of objects with the following interpretations. Each n-ary predicate F^n is interpreted by a function $\|F^n\| : W \mapsto \mathcal{P}D^n$ which maps worlds to sets of n-tuples of D elements, and each m-ary function symbol f^m is interpreted by an m-ary function $\|f^m\| : D^m \mapsto D$ on D. The relation $\Vdash$ is defined between worlds and propositions as you would expect. First you define for each assignment a of variables the denotation of terms.

   1. $\|x\|_a = a(x)$.
   2. $\|f^m t_1 \cdots t_m\|_a = \|f^m\| \langle \|t_1\|_a , \dots , \|t_m\|_a \rangle$.

Then you define truth of propositions at worlds.

   1. $a, w \Vdash F^n t_1 \cdots t_n$ iff $\langle \|t_1\|_a , \dots , \|t_m\|_a \rangle \in \|F^n\| (w)$.
   2. $a, w \Vdash A \wedge B$ iff $a, w \Vdash A$ and $a, w \Vdash B$.
   3. $a, w \Vdash A \vee B$ iff $a, w \Vdash A$ or $a, w \Vdash B$.
   4. $a, w \Vdash {\sim}A$ iff $a, w \nVdash A$.

5. $a, w \Vdash \forall x A$ iff $a', w \Vdash A$ for each x-variant a' of a.[2]

If identity is among the predicates, we fix the interpretation $\|=\|$ of identity to be the function that returns the identity relation on D for each world w. Note that we have fixed our attention on constant domain models. This is a simplification but one chiefly to simplify presentation. Note too that predicates can vary in extension from world to world but functions cannot. The generalization to nonrigid functions is simple too.

Now, we could define truth and falsity at worlds (sets of possible worlds) just as before. The generalization to predicates and quantifiers is immediate. However, some inconsistencies will remain true at no worlds (possible or impossible worlds). Consider two constants (zero-place functions) c and d, in a model in which c and d have different denotations. The proposition $c = d$ is true in no possible world w, as $\|c\| \neq \|d\|$—the denotations differ in all worlds. This is simply the necessity of true identity statements and the impossibility of false identity statements when the terms involved are rigid designators. Simply "identifying worlds" is not going to introduce the inconsistency $c = d$.

It is clear that another sort of identification will do it. We can identify c and d. If we do that, then we can hold that c and d are identical after all. We are motivated to extend our definition of ways things can't be to allow the identification of objects as well as worlds. Not just any equivalence relation will do. If f is a function, and if c and d are identified, then it makes sense to identify $f(c)$ and $f(d)$ too, at least if f is to act as a function in the impossible world under construction. We will assume that our equivalence relations are *congruence relations* under the functions in the original model, so that the functions remain to act as functions on the impossible worlds. Of course, this is a choice. We need not require this if we don't wish to—however considering the alternate possibility here is one we will leave for another time.

Definition 3.2   A *world* in a predicate model is a set X of possible worlds together with a congruence relation $\equiv$ on the domain D.

Not only is it impossible that w and v are both actual, but if c and d are distinct objects, then it cannot be that c and d are identical. As before, we define the relations $\Vdash^+$ and $\Vdash^-$ on worlds.

Definition 3.3   Truth and falsity are defined for worlds (with assignments a as before) as follows:

1. $a, \equiv, X \Vdash^+ F^n t_1, \ldots, t_n$ iff for some objects $d_i \in D$ where $d_i \equiv \|t_i\|_a$, $\langle d_1, \ldots, d_n \rangle \in \|F^n\|(w)$ for some $w \in X$;
2. $a, \equiv, X \Vdash^- F^n t_1, \ldots, t_n$ iff for some objects $d_i \in D$ where $d_i \equiv \|t_i\|_a$, $\langle d_1, \ldots, d_n \rangle \notin \|F^n\|(w)$ for some $w \in X$;
3. $a, \equiv, X \Vdash^+ \sim A$ iff $a, \equiv, X \Vdash^- A$;
4. $a, \equiv, X \Vdash^- \sim A$ iff $a, \equiv, X \Vdash^+ A$;
5. $a, \equiv, X \Vdash^+ A \wedge B$ iff $a, \equiv, X \Vdash^+ A$ and $a, \equiv, X \Vdash^+ B$;
6. $a, \equiv, X \Vdash^- A \wedge B$ iff $a, \equiv, X \Vdash^- A$ or $a, \equiv, X \Vdash^- B$;
7. $a, \equiv, X \Vdash^+ A \vee B$ iff $a, \equiv, X \Vdash^+ A$ or $a, \equiv, X \Vdash^+ B$;
8. $a, \equiv, X \Vdash^- A \vee B$ iff $a, \equiv, X \Vdash^- A$ and $a, \equiv, X \Vdash^- B$;
9. $a, \equiv, X \Vdash^+ \forall x A$ iff $a', \equiv, X \Vdash^+ A$ for each x-variant a' of a;
10. $a, \equiv, X \Vdash^- \forall x A$ iff $a', \equiv, X \Vdash^- A$ for some x-variant a' of a;

This construction is inspired by the "collapsing" construction of Priest, which constructs inconsistent **LP** models from consistent models [6].

The only clauses in this definition which need comment are the base clauses. In our worlds, the domain is the same as the domain in the original predicate model. However, a claim such as Fxy is true, not only when the pair $\langle a, b \rangle$ of objects denoted by x and y is in the extension $\|F\|$ of F, but also when a and b are "identified" by $\equiv$ with other objects c and d which are in the extension of F.

As before, there is a corresponding notion of an **LP**-evaluation in predicate calculus. Now, each predicate F^n has an extension $\|F^n\|^+$ and an anti-extension $\|F^n\|^-$. These extensions together cover (perhaps with some overlap) the product D^n of n-tuples from the domain. The constraint on identity is that its extension is the identity relation on the domain. Its anti-extension may overlap with the extension. (So all objects are self-identical and some might also be non-self-identical.) Valuations are relative to assignments of variables, and we have the following straightforward conditions to add to the usual clauses for the propositional connectives

1. $1 \in v_a(F^n t_1, \ldots, t_n)$ iff $\left\langle \|t_1\|_a, \ldots, \|t_n\|_a \right\rangle \in \|F^n\|^+$;
2. $0 \in v_a(F^n t_1, \ldots, t_n)$ iff $\left\langle \|t_1\|_a, \ldots, \|t_n\|_a \right\rangle \in \|F^n\|^-$;
3. $1 \in v_a(\forall x A)$ iff $1 \in v_{a'}(A)$ for each x-variant a' of a;
4. $0 \in v_a(\forall x A)$ iff $0 \in v_{a'}(A)$ for some x-variant a' of a.

LP-validity of arguments is defined just as before. We have the same result as before, on the predicate level.

Theorem 3.4 $\Sigma \vdash_P A$ *if and only if* $\Sigma \vdash_{\mathbf{LP}} A$.

Proof overview: The proof is both tedious and (notationally) complex. However it has the same structure as the propositional case. First, you can show that any possible worlds model gives rise to an **LP**-model, which makes true exactly the same propositions. This ensures that if $\Sigma \vdash_{\mathbf{LP}} A$ then $\Sigma \vdash_P A$. This is rather straightforward. You simply identify the objects 'identified' by the congruence relation $\equiv$, and proceed in the natural way to get an **LP**-model.

For the other direction, we need to "unwrap" an **LP**-model into a consistent family of possible worlds, with the one domain, with the one interpretation of function symbols, and with extensions of predicates varying from world to world. For this, the technique is surprisingly simple in concept but rather technical in detail. The idea is to take the domain D' of the extended model to be the objects in the domain D of the **LP**-model plus a few more to deal with objects which are non-self-identical. For any object taken to be non-self-identical by the **LP**-model, add *one* new object (its *shadow*) to which it is related by $\equiv$. That is enough for $\equiv$. Functions are interpreted just as they were in the **LP**-model, except for shadows. If a shadow is in the input of a function, the function treats it as the object of which it is a shadow. In other words, functions ignore the difference between shadow and substance, and they output only substantial objects. Functions so defined respect $\equiv$-equivalence, so this definition is permissible. For predicates, we let an extension of a predicate *approximate* an **LP**-extension/anti-extension pair, just when anything the classical extension takes to be true, so does the **LP**-extension, and anything the classical extension takes to be false, so does the anti-extension (taking shadows into account, of course, as they are in our

domain). We have a classical possible world for any family of approximating extensions. The whole set of these families of extensions, together with the equivalence relation, is our impossible world. The beautiful thing is: this definition works! The impossible world takes to be true exactly those propositions true in the **LP**-model. The proof is an induction on the complexity of propositions.   □

For those who prefer a more detailed proof, the detail is in the appendix to this paper.

4 Interpretation   What have we achieved? Let's start by considering what we *have not* achieved. The construction of this paper is not the final answer for a semantics of belief. All worlds, whether possible or impossible, contain all logical truths, and people certainly need not believe all logical truths. Perhaps this might be remedied by considering incompleteness dually with inconsistency, in the manner of my [8]. Even with this possible modification, more needs to be said about the nature of belief modeled in this way. It is integral to this account that beliefs are closed under *some kind* of logical consequence. This means that we are not considering explicit belief, but something akin to the notion of commitment. Much more must be done to make this a truly useful model.

Granting this weakness, we have achieved a number of things. First, we have models for **LP** in which the primitive notions are consistent possible worlds. We used the same raw machinery as is available to those who use classical possible worlds. We have used this machinery to construct a notion of 'truth in a world' which allows inconsistent worlds just as much as consistent ones. However, inconsistent worlds are *not* mysterious entities nor merely theoretical devices constructed using some unintelligible logic. Impossible worlds are just more than one possible world taken together. If w and v are two different ways the world can be, then the world *can't* be like w and v taken together. That's inconsistent. All we have done is constructed a way to understand how two worlds like w and v clash, in such a way as to show how the clash between w and v is *not* the same as the clash between two different worlds, w' and v'. Both impossibilities exhibit contradictions but they exhibit *different* contradictions. They are different impossible worlds.

However, we noted that this kind of impossibility is but one kind of impossibility. Another is found by not superimposing different worlds. Another notion can be found by superimposing different objects. In this way, other impossibilities can be modeled.

Second, we have shown that these two different kinds of identification are *sufficient* to model the pre-existing logic **LP**. We have some grip, therefore, in what *kinds* of inconsistencies are allowed by the paraconsistent logic **LP**. In the event of adding new connectives to the language of **LP**, we might find that different sorts of 'identification' or other techniques again are required to model the arising inconsistency using a classical base. One which comes to mind is the addition of intensional operators. It *may* be that we might have to consider the identification of different accessibility relations in order to treat inconsistent modal claims.

Third, Lewis, in his "Logic for equivocators" [3], shows how to interpret paraconsistent logics like **LP** as logics for equivocation. On his intepretation, we achieve inconsistency by allowing the extensions of our predicates to vary. In some part of

our speech, the expression might mean one thing, in another, something else. It's clear how we might appear to contradict ourselves when we do this.

In our models for **LP** we get by without changing the *meanings* of our predicates at all. They are just as fixed as they are in classical possible worlds models. True, the extensions of predicates vary from world to world, but that's simply the trivial fact that objects have properties contingently. There is no *equivocation* on this analysis. The extensions of predicates in our new models remain just as they were in the original classical possible worlds models. The inconsistency arises by taking the world to be in two inconsistent ways, or taking objects to be the same which are, as a matter of fact, different.

Finally, we have seen a way we enjoy the fruits of both paraconsistent and classical logic. In our classical possible worlds models, classical first-order consequence is available as it always was. The addition of impossible worlds—not as extra ontology but rather as a new way of looking at what was always there—gives us a new, paraconsistent notion of logical consequence. If this helps people understand the formalism of paraconsistent logic and some of the ideas underlying it, then this paper will have served its purpose.

5 *Appendix*   Here is the detailed proof of Theorem 3.4.

Proof: To show that if $\Sigma \vdash_{\mathbf{LP}} A$ then $\Sigma \vdash_P A$, consider any model and any world $\equiv, X$ in any model. Define an **LP**-valuation v as follows. Its domain is the set of $\equiv$ congruence classes from the domain D of the predicate model. Take the extension of the function symbols to be given by such an identification (this is possible, as the extension of each function symbol respects $\equiv$-congruence). The extension of a predicate $||F||^+$ is just the union of the extensions $||F||(w)$, and $1 \in v(p)$ if and only if $X \Vdash^+ p$, and $0 \in v(p)$ if and only if $X \Vdash^- p$. It is simple to show that $1 \in v(A)$ if and only if $X \Vdash^+ A$ and $0 \in v(A)$ if and only if $X \Vdash^- A$. So, by the definition of **LP**-validity, if $\Sigma \vdash_{\mathbf{LP}} A$, then if $X \Vdash^+ B$ for each $B \in \Sigma$, then since $1 \in v(B)$ for each $B \in \Sigma$, we have $1 \in v(A)$ and hence $X \Vdash^+ A$ as desired.

Conversely, suppose $\Sigma \nvdash_{\mathbf{LP}} A$. We wish to construct a model containing a world X at which Σ is true but A not. Consider an **LP**-model which makes Σ true and A not. It will have a domain D, an extension and an anti-extension for each predicate (including identity) and an interpretation for each function symbol.

How do we then construct a world X in a possible worlds model to do the work of the **LP**-model? First of all, we need to find the domain of the possible worlds model. This cannot be simply the domain D of the **LP**-model, as we need some objects to identify to get the inconsistent identities (if there are any). We do this by 'unraveling the domain'. We expand the domain D such that for each object $d \in D$ such that $\langle d, d \rangle \in || = ||^-$ (that is, if the **LP**-model takes d to be nonidentical to itself) we add a new object d' (d's "shadow") to the domain of our new model. So the extended domain D' is the set $D \cup \{d' : d \in D$ where $\langle d, d \rangle \in || = ||^-\}$. The congruence relation $\equiv$ relates each $d \in D'$ to itself, and furthermore, relates each d to its shadow (if d has a shadow) and the shadow of d to d itself. This is a congruence relation on D'. For any $e \in D'$, we let $\overline{e} \in D$ be the object in D to which e is $\equiv$-related. For an assignment a of variables to objects in D' the corresponding D-assignment $\overline{a}$ assigns x the object $\overline{e}$ wherever a assigns x the object e.

We then take the extension of each function symbol f to be defined as follows: $||f||(e_1, \ldots, e_n)$ (where $e_i \in D'$) is simply $||f||(\overline{e_1}, \ldots, \overline{e_n})$, where the interpretation $||f||$ here is that of the **LP**-model. In other words, the shadow objects are never outputs of functions, and they function as inputs in just the same way as the objects which they shadow. It is simple to see that for any term t in the language $\overline{||t||} = ||t||$, where the left hand side is the interpretation in our classical model and the right hand side is the interpretation from the **LP**-model.

That will do for the domain and the interpretations of function symbols. The extensions of predicates are to vary from world to world, so we need to understand what the worlds are to be. Here, a world will simply be an index for different possible classical extensions of predicates which somehow approximate the inconsistent extensions of predicates from the **LP**-model. We will say that the extension $||F||$ of F on the domain D' *approximates* the extension $||F||^+$ and the anti-extension $||F||^-$ on the domain D just when if $\langle e_1, \ldots, e_n \rangle \in ||F||$ then $\langle \overline{e_1}, \ldots, \overline{e_n} \rangle \in ||F||^+$, and if $\langle e_1, \ldots, e_n \rangle \notin ||F||$ then $\langle \overline{e_1}, \ldots, \overline{e_n} \rangle \in ||F||^-$. In other words, any judgment that the approximating extension makes is also made by the **LP**-extension or anti-extension (keeping in mind the unraveling of the domain). An *approximating extension family* is an extension of D' for each predicate in the language (other than identity) such that each extension approximates the extension and anti-extension of the **LP**-model. The worlds in our classical model are simply the approximating extension families. So, adapting our notation somewhat, we have a set X of approximating extension families, such that for any $w \in X$, and any predicate F, $||F||(w)$ is an extension on D' approximating the extension and anti-extension of F in D'. We will show that taking this as our model, for any assignment a of variables on D' and for the congruence relation $\equiv$ defined above, that for every formula A, $a, \equiv, X \Vdash^+ A$ if and only if $1 \in v_{\overline{a}}(A)$ and $a, \equiv, X \Vdash^- A$ if and only if $0 \in v_{\overline{a}}(A)$. The proof is an induction on the complexity of A as in the propositional case. The conjunction, disjunction, and negation clauses are as before, so we need attend only to the atomic and the quantifier cases.

First, $a, \equiv, X \Vdash^+ F^n t_1, \ldots, t_n$ if and only if for some objects $d_i \in D$, where $d_i \equiv ||t_i||_a$, $\langle d_1, \ldots, d_n \rangle \in ||F^n||(w)$ for some $w \in X$. Now since $||F^n||(w)$ approximates the **LP**-extension and anti-extension of F, so $\langle \overline{d_1}, \ldots, \overline{d_n} \rangle \in ||F^n||^+$, and so, $\langle \overline{||t_1||_a}, \ldots, \overline{||t_n||_a} \rangle \in ||F^n||^+$, by the behavior of $\equiv$. Now $\overline{||t_i||_a} = ||t_i||_{\overline{a}}$, so $\langle ||t_1||_{\overline{a}}, \ldots, ||t_n||_{\overline{a}} \rangle \in ||F^n||^+$, giving $1 \in v_{\overline{a}}(F^n t_1, \ldots, t_n)$ as desired. Conversely, if $1 \in v_{\overline{a}}(F^n t_1, \ldots, t_n)$, we have $\langle ||t_1||_{\overline{a}}, \ldots, ||t_n||_{\overline{a}} \rangle \in ||F^n||^+$. There is some extension $||F^n||(w)$ approximating the **LP**-extension and anti-extension for which $\langle ||t_1||_{\overline{a}}, \ldots, ||t_n||_{\overline{a}} \rangle \in ||F^n||(w)$. Select such an extension (i.e., such a w). It follows that $a, \equiv, X \Vdash^+ F^n t_1, \ldots, t_n$, as $w \in X$. The case for the anti-extension is identical.

We need to treat identity separately, as it is interpreted as identity on the domain D' of the classical model. The only wrinkle is the congruence relation $\equiv$. So $a, \equiv, X \Vdash^+ t_1 = t_2$ if and only if there is an object $e \in D'$ such that $||t_1||_a \equiv e$ and $||t_2||_a \equiv e$, that is, if and only if $||t_1||_a \equiv ||t_2||_a$. This obtains if and only if $||t_1||_{\overline{a}} = ||t_2||_{\overline{a}}$, which in turn happens if and only if $1 \in v_a(t_1 = t_2)$. For the negative case, $a, \equiv, X \Vdash^- t_1 = t_2$ if and only if there are *distinct* objects e_1 and e_2 in D' where $||t_1||_a \equiv e_1$ and $||t_2||_a \equiv e_2$. This obtains if and only if $||t_1||_{\overline{a}} = \overline{e_1}$ and $||t_2||_{\overline{a}} = \overline{e_2}$. Now e_1 and e_2 are different if and only if $\langle \overline{e_1}, \overline{e_2} \rangle \in ||=||^-$ by the construction of the extended domain. Why is this? Well, if $\overline{e_1} \neq \overline{e_2}$, then $\langle \overline{e_1}, \overline{e_2} \rangle \in ||=||^-$, and we must have $e_1 \neq e_2$ (as

$^{-}$ is a function). If $\overline{e_1} = \overline{e_2}$, then e_1 and e_2 can differ only if one is the shadow of the other, which happens if and only if $\langle \overline{e_1}, \overline{e_2} \rangle \in || = ||^-$. So, $a, \equiv, X \Vdash^- t_1 = t_2$ if and only if $0 \in v_{\overline{a}}(t_1 = t_2)$, as desired.

Finally for quantifiers, $a, \equiv, X \Vdash^+ \forall xA$ if and only if $a', \equiv, X, \Vdash^+ A$ for each x-variant a' of a. This holds if and only if $1 \in v_{\overline{a'}}(A)$ for each x-variant a' of a. These x-variants are x-variants in D'. However, this doesn't matter. If b is an x-variant (in D) of $\overline{a}$, then $b = \overline{a'}$ for some x-variant (in D') a' of a. Conversely, if a' is an x-variant (in D') of a, then $\overline{a'}$ is an x-variant (in D) of $\overline{a}$. So, $1 \in v_{\overline{a'}}(A)$ for each x-variant a' of a if and only if $1 \in v_b(A)$ for each x-variant (in D) b of $\overline{a}$, if and only if $1 \in v_{\overline{a}}(\forall xA)$, as desired. Similarly, $a, \equiv, X \Vdash^- \forall xA$ if and only if $a', \equiv, X, \Vdash^- A$ for some x-variant a' of a. This holds if and only if $0 \in v_{\overline{a'}}(A)$ for some x-variant (in D') a' of a, if and only if $0 \in v_b(A)$ for each x-variant (in D) b of $\overline{a}$, if and only if $0 \in v_{\overline{a}}(\forall xA)$, as desired. The case for the existential quantifier is dual, and this completes the proof. $\square$

NOTES

1. Priest reminds me that the chunking account is not original with Lewis, but that Rescher and Manor [7] used the account in their paper published in 1970.

2. An x-variant of an assignment a is any assignment a' which agrees with a for every variable in the language except possibly x.

REFERENCES

[1] Anderson, A. R., and N. D. Belnap, *Entailment: The Logic of Relevance and Necessity*, vol. 1, Princeton University Press, Princeton, 1975. Zbl 0323.02030 MR 53:10542 2

[2] Dunn, J. M., "Intuitive semantics for first-degree entailments and 'coupled trees'," *Philosophical Studies*, vol. 29 (1976), p. 149–68. MR 58:10311 2

[3] Lewis, D., "Logic for equivocators," *Noûs*, vol. 16 (1982), p. 431–41. MR 85d:03034 1, 1, 1, 1, 4

[4] Lewis, D. K., *On the Plurality of Worlds*, Blackwell, Oxford, 1986. 1, 1

[5] Priest, G., "The logic of paradox," *Journal of Philosophical Logic*, vol. 8 (1979), pp. 219–41. Zbl 0402.03012 MR 80g:03007 2

[6] Priest, G., "Minimally inconsistent **LP**," *Studia Logica*, vol. 50 (1991), pp. 321–31. Zbl 0748.03017 MR 93e:03037 3

[7] Rescher, N., and R. Manor, "On inference from inconsistent premises," *Theory and Decision*, vol. 1 (1970), pp. 179–217. Zbl 0212.31103 5

[8] Restall, G., "Truthmakers, entailment and necessity," *Australasian Journal of Philosophy*, vol. 74 (1996), pp. 331–40. 4

[9] Stalnaker, R., *Inquiry*, The MIT Press, Cambridge, 1984. 1

Department of Philosophy
Macquarie University
Sydney NSW 2109
AUSTRALIA
email: Greg.Restall@mq.edu.au