PROOFS WITH STAR AND PERP

Greg Restall*

Philosophy Department
The University of St Andrews
gr69@st-andrews.ac.uk

DECEMBER 23, 2021

Abstract: In this paper, I show how to incorporate insights from the model-theoretic semantics for negation (insights due the late J. Michael Dunn [4], among others [1, 12]), into a *proof-first* understanding of the semantics of negation. I then discuss the different ways a logical pluralist may understand the underlying accounts of proofs and their significance.

J. Michael Dunn's pioneering work in non-classical logic is profoundly important, covering many different topics. He has given us new tools and techniques, inspired new questions, and opened up us fresh perspectives on what we thought was familiar. In this paper I will pay tribute to Mike by transplanting some of his important insights about the semantics of negation [4] — insights grown in model-theoretic soil — to a new field, that of proof-theoretic semantics.

The key idea to be applied can be simply described. The behaviour of negation in a range of logical systems (classical logic, intuitionistic logic, orthologic, relevant logics, linear logic, etc.) can be understood in terms of an underlying relation of *incompatibility* on a model. Given some collection P of points, and an relation $\Vdash$ evaluating formulas at those points, then *whatever* we think about the details of the particular behaviour of negation, we can agree that if we have two points a and b where $a \Vdash \neg A$ (according to a, $\neg A$ holds) and $b \Vdash A$ (according to b, A holds), then in some sense, a and b are *incompatible*. If $\perp$ is the incompatibility relation, we have $a \perp b$. An incompatibility model (or a 'perp'-model, for short) takes this 'only if' to be an 'if and only if', and for it to be the defining feature of negation.

- $a \Vdash \neg A$ if and only if for each b, if $b \Vdash A$ then $a \perp b$.

*Thanks to Franz Berto, Aaron Cotnoir and colleagues in the Logic and Metaphysics research seminar in Arché at the University of St Andrews, for discussion on material that lead to these results. Thanks also to the European Network for the Philosophy of Logic for an opportunity to present on these results, and for useful questions and feedback.

This semantics for negation certainly sounds plausible enough, but it might look unfamiliar.[1] Despite its unfamiliarity, it is ubiquitous. If we impose different conditions on the incompatibility relation $\perp$, familiar logics arise. The truth-functional classical semantics of negation is recovered if $\perp$ is simply defined to be the nonidentity relation $\neq$. A point is incompatible with everything else, it is compatible with itself only. On a Kripke model for intuitionistic logic, we have $a \perp b$ iff a and b have no common descendents: that is, there is no c where $a \sqsubseteq c$ and $b \sqsubseteq c$.

However, if we allow incompatibility to range more widely, we can model paraconsistent logics. If a point a supports both A and its negation $\neg A$, then we have $a \perp a$. Contradictory points clash *with themselves*. Such points may clearly be defective in some sense or other, but that does not mean that a semantics should rule them out *tout court*. Ringing the changes with different kinds of incompatibility relations gives a wide range of logics. So wide, in fact, that if we allow incompatibility to fail to be *symmetric*, then we have models that provide counterexamples to the inference from A to $\neg\neg A$.[2]

One of Dunn's many insights is that the Routley star semantics [19] for a de Morgan negation, with the familar clause

- $a \Vdash \neg A$ if and only if $a^* \nVdash A$

can be naturally understood as a kind of incompatibility semantics. If we understand a^* as the *maximal* point compatible with a (so $a \not\perp b$ if and only if $b \sqsubseteq a^*$) then indeed the Routley semantics is a species of the broader genus of incompatibility models. Incompatibility models with a Routley star are particularly well-suited to modelling a non-classical *de Morgan* negation. In these models, the constructively invalid de Morgan law (from $\neg(A \wedge B)$ to $\neg A \vee \neg B$) is valid, and if our model satisfies the condition that $a^{**} = a$, then if $a \Vdash \neg\neg A$, we have $a \Vdash A$ too.

Incompatibility models (whether with *perp* primitive, or using *star*) provide a unifying framework for understanding the semantics of negation. We now have a map of the terrain, that takes in a great deal of the logical scenery indeed, including the placid and well-understood downtowns of classical and intuitionistic negation, and the rather less familiar wilds of paraconsistency.

One question that lurks in the background when we use this semantics for different logical systems is—what do such models *model*? What *are* these points that stand in a

[1] In the incompatibility models we will discuss, conjunction and disjunction have the standard 'local' truth conditions: $a \Vdash A \wedge B$ iff $a \Vdash A$ and $a \Vdash B$; $a \Vdash A \vee B$ iff $a \Vdash A$ or $a \Vdash B$, and so, the resulting logics extend distributive lattice logic with an order inverting 'negation' operator. To model logics without distribution we must complicate the semantics in some way or other, and to do so would take us beyond the scope of this paper.

[2] Here is a simple model that does the job. Take two points a, b where $b \perp a$, but no other incompatibilities hold. Let $a \Vdash p$ and $b \nVdash p$. Then we have $b \Vdash \neg p$, since a is the only p-point, and $b \perp a$. However, in this model we do not have $a \Vdash \neg\neg p$, since $b \Vdash \neg p$, but we do *not* have $a \perp b$.

relation of incompatibility with each other? Do they have any independent status, apart from their use in our semantic theorising? If so, where do we find them? And *how* do we come in contact with them? If they have no independent status, other than in our models, how do they do such a good job of modelling the behaviour of negation? How is it that negation could have the properties that are appropriately modelled by models like these?

I will not answer these questions in this essay, but I will provide for the friend of the incompatibility semantics a new perspective on the meaning of negation and how we might grasp it, in such a way as to help explain how the incompatibility model semantics turns out to be an appropriate class of models for the concept as we use it. For that, we don't start with models. We start with *proofs*.

1 PROOFS AND CONTEXTS

However we grasp concepts like conjunction, disjunction and negation, that grasp is intimately bound up with the way we use these concepts in our everyday reasoning practice. There are many things we could mean by terms like 'reasoning' and 'inference.'[3] Here, though, let us focus on a very simple connection between our pracitces of *proof*, and our grasp of logical concepts such as conjuction and negation.

If someone were to grant A and grant B and still take it (after this had been pointed out) that it remained an open question, in need of further justification, as to whether $A \wedge B$ were the case, this would seem to be good reason to doubt that this person understood that '$\wedge$' means *and*, in the usual logicians' sense of conjunction. Similarly, if you had just granted a conjunction, such as $A \wedge B$, a competent user of '$\wedge$' could see that if the issue of whether or not A is the case were to arise, the answer would again be 'yes,' and no further justification would be needed than to point back to $A \wedge B$, or whatever it was that justified $A \wedge B$. We can understand the inference rules of a natural deduction system (such as the rules in Figure 1) as encoding basic competencies with the logical concepts, at least under a specific idealisation.

One key feature of natural deduction rules like these is that they do not merely tell us how to grind out indubitable consequences of things that have been granted as true. The rules allowing for *discharging* of assumptions (here, $\vee E$ and $\neg I$), and the negation elimination rule $\neg E$ rely on our ability to engage in the question-and-answer practice of proving from assumptions that we don't necessarily take to be true. The

[3] See Gil Harman's influential *Change in View* [5] for a convincing defence of the idea that inference as understood as reasoned change in view is *not* to be identified with deductive validity, and see Bob Brandom's *Articulating Reasons* [2] for an introduction to *inferentialism*, according to which a primitive notion of *material inference* (not to be confused with formal deductive logical validity) lies at the root of our conceptual capacities. Then, for good measure, read Catarina Dutilh Novaes' superb *The Dialogical Roots of Deduction* [10] for an account of how the roots of our reasoning practice take nourishment in fertile soil of our dialogical norms for dispute resolution.

$$\frac{A \quad B}{A \wedge B}\wedge I \qquad \frac{A\wedge B}{A}\wedge E \qquad \frac{A\wedge B}{B}\wedge E$$

$$\frac{A}{A\vee B}\vee I \qquad \frac{B}{A\vee B}\vee I \qquad \frac{A\vee B \quad \begin{matrix}[A]^j\\ \Pi\\ C\end{matrix} \quad \begin{matrix}[B]^k\\ \Pi'\\ C\end{matrix}}{C}\vee E^{j,k}$$

$$\frac{\begin{matrix}[A]^i\\ \Pi\\ \bot\end{matrix}}{\neg A}\neg I^i \qquad \frac{\neg A \quad A}{\bot}\neg E \qquad \frac{\bot}{A}\bot E$$

Figure 1: NATURAL DEDUCTION PROOF RULES

rule, $\neg E$ tells us that if we have ever proved $\neg A$ and also proved A, then we have landed in an inconsistency (marked here by writing $\bot$) in the conclusion point of the proof tree. The negation introduction rule gives us something to *do* when we land in such an inconsistency. We can take one of the assumptions we used to lead to the inconsistency, 'blame' it, and conclude on the basis of the *other* assumptions, that it is not true—that is, its negation is true. Here is a simple example of these rules in use in a proof from the assumption $\neg p \vee \neg q$ to the conclusion $\neg(p \wedge q)$.

$$\dfrac{\dfrac{\neg p\vee\neg q \qquad \dfrac{[\neg p]^1 \quad \dfrac{[p\wedge q]^3}{p}\wedge E}{\bot}\neg E \qquad \dfrac{[\neg q]^2 \quad \dfrac{[p\wedge q]^3}{q}\wedge E}{\bot}\neg E}{\bot}\vee E^{1,2}}{\neg(p\wedge q)}\neg I^3$$

This proof can be understood as not only reassuring us that the argument from $\neg p \vee \neg q$ to $\neg(p\wedge q)$ is valid, but that it gives us a way to *justify* the claim of $\neg(p\wedge q)$, appealing only to the assumption $\neg p \vee \neg q$ and using only the primitive inference rules. These rules, as we have indicated, are to be understood as giving us basic constraints on competent use of the logical concepts in a context where we are asking yes/no questions against a background in which certain assumptions may be taken for granted. The resulting justification is straightforward. Working backwards, we can conclude $\neg(p \wedge q)$ because the assumption of $p \wedge q$ together with $\neg p \vee \neg q$ is inconsistent. Why is that inconsistent? Because that assumption is inconsistent with either disjunct of our disjunctive assumption. Why is *that* the case? Because $p \wedge q$ leads to p, which is immediately inconsistent with the first disjunct $\neg p$, and $p \wedge q$ also leads to q and this is immediately inconsistent with the second disjunct. Each of the transitions in this piece of reasoning corresponds to one of the primitive inference rules, and these are the kinds of things we can *follow*.

This seems a very long way away from incompatibility models for non-classical logics of negation. At most, this gives us intuitionistic logic.[4] However, the gap is not as wide we might worry, once we note that the norms governing assertion and proof constrain not only assertions made under the scope of *additive* suppositions (where we add extra assumptions to draw out their consequences, at least locally), but also *alternative* suppositions. To illustrate what I have in mind, consider the following stretch of reasoning.

> Let's grant that either p is necessary or q is necessary. Let's show that the disjunction $p \lor q$ is necesarry. To do that, let's suppose things go some way or other, totally arbitrarily. On the option that it was p that was necessary, that means that in this case, we've still got p, and so, $p \lor q$, too. Alternatively, in the option that it was q was necessary, similarly, in that case we've got q, and so, $p \lor q$, too. So, on either option, we have $p \lor q$. So, it follows from all this that $p \lor q$ is necessary.

The wording here is clumsy, but one way to articulate the logical structure of this stretch of reasoning is following proof, in which the formulas are annotated with *context* markers.

$$\dfrac{\dfrac{\Box p \lor \Box q \cdot a \qquad \dfrac{\dfrac{[\Box p \cdot a]^1}{p \cdot b}\,\Box E}{p \lor q \cdot b}\,\lor I \qquad \dfrac{\dfrac{[\Box q \cdot a]^2}{q \cdot b}\,\Box E}{p \lor q \cdot b}\,\lor I}{p \lor q \cdot b}\,\lor E^{1,2}}{\Box(p \lor q) \cdot a}\,\Box I$$

Here, the formulas tagged with a are those asserted (or assumed) in our home context, not under the scope of the "suppose things go some way or other" shift in the dialogue. That shift opens up another *context* in which assertions or inferences can go. We label the formulas here with the b tag. The $\Box E$ inference is a context-shifting inference. Granting $\Box A$ (in some context of the discussion) can be used to justify A in another. That is the force of a claim of necessity: to grant that A not only holds but is necessary is to agree that even had things gone otherwise, we still would have had A.[5] How, then, do we show that something is necesary? The natural way to do so is to prove it *in an arbitrary context*. There is nothing special about a context that makes it arbitrary, other

[4] Or minimal logic, if we are prepared to do without $\bot E$ which *does* look rather out of place, it must be said [8]. But for more on $\bot E$, see the discussion in Section 3.

[5] This is not to say that all forms of "going otherwise" are the same. There is more than one notion of necessity, and more than one way that we can suppose that things go otherwise [3, 9, 16]. The details of these distictions, as interesting and fruitful as they are, are not important for the present discussion.

than the fact that at the step of the reasoning at which the inference us made (here, it is a $\Box I$ step), none of the active assumptions feature that context label. In our proof, we proved $p \vee q$ (in context b) from an assumption featuring context a. So, at least if we understand deductive practice with necessity in this way, our inference is legitimate.

Notice that in this reasoning we allowed the disjunction rules to operate on *tagged* formulas in a natural way. If the aim was to extend our propositional logic with modal operators, the natural choice for rules for the propositional connectives is relatively straightforward. Since we ordinarily do not think of context shifts as changing the meaning of the connectives, the rules remain almost unchanged, except for the addition of context labels. The results are compiled in Figure 2. In most of these rules, the

$$\frac{A \cdot x \qquad B \cdot x}{A \wedge B \cdot x}\wedge I \qquad\qquad \frac{A \wedge B \cdot x}{A \cdot x}\wedge E \qquad \frac{A \wedge B \cdot x}{B \cdot x}\wedge E$$

$$\frac{A \cdot x}{A \vee B \cdot x}\vee I \qquad \frac{B \cdot x}{A \vee B \cdot x}\vee I \qquad\qquad \frac{A \vee B \cdot x \qquad \begin{array}{c}[A \cdot x]^j \\ \Lambda \\ C\end{array} \qquad \begin{array}{c}[B \cdot x]^k \\ \Lambda' \\ C\end{array}}{C}\vee E^{j,k}$$

$$\frac{\begin{array}{c}[A \cdot x]^i \\ \Lambda \\ \bot\end{array}}{\neg A \cdot x}\neg I^i \qquad\qquad \frac{\neg A \cdot x \qquad A \cdot x}{\bot}\neg E \qquad\qquad \frac{\bot}{A \cdot x}\bot E$$

Figure 2: NATURAL DEDUCTION PROOF RULES WITH *CONTEXTS*

contexts are fixed from premise to conclusion. From A in a given context, I can conclude $A \vee B$ in *that* context. If I have A and B in the same context, I can conclude $A \wedge B$ in that context. And in particular, given our particular interest in negation, we reach a contradiction from A and $\neg A$ *in the same context*.

However, two small details deserve comment. The first is that the contradiction marker is not tagged with a context, since the contradiction marker is not, primarily, a statement which can be assumed or derived in a context, but is a sign that the assumptions leading to this point are contradictory. If we have concluded A and $\neg A$ in the same context, then the assumptions leading up to that point clash. As we have seen, the assumptions leading up to those contradictory conclusions may well have been granted in some other contexts, so there is no sense that we need locate the 'blame' for that clash in any one alternative context.

The second detail is the special role of the conclusion C in the $\vee E$ rule. Here, C could be anything that could occur as a conclusion of a proof: either a tagged formula or the contradiction marker. In the case of a tagged formula, there is no requirement that this tagged formula be tagged with the same context marker as the disjunctive assumption leading to that conclusion. We saw this in our proof from $\Box p \lor \Box q$ to $\Box(p \lor q)$. The premise was granted in context a, and so, we split into two cases in that context. In both of those two cases, we conclude $p \lor q$ in the fresh context b, and so we hold onto that conclusion on the basis of the disjunctive assumption. What drives the case-based reasoning of a disjunction elimination is that the same conclusion is derived in each case, and each case in question is opened up in the *same* context as the disjunctive conclusion being eliminated. The conclusion of each inference must be the same, but it need not be another claim in that original context. It could also be a contradiction marker, as seen in the first (untagged) natural deduction proof on page 4.

The final thing to note about natural deduction proofs with context labels is that they at least open the way to understand how a possible world semantics might truly *model* the logic of necessity, without having to take possible worlds as the starting point of our semantic theorising. While it is understandably tempting to take the context label in a proof as naming a *world*, there is no need to do so. There is no need to take the world labels to be referring expressions of any kind, any more than the supersecript in the second variable in the ordered pair expression $\langle x, x' \rangle$ should be taken as referring to anything. The role of the context labels in proof exhausted by their identity and difference. $A \cdot a$ and $\neg A \cdot a$ clash, while $A \cdot a$ and $\neg A \cdot b$ need not, since there is some context shift between the claim of A and the claim of $\neg A$. It is that shift in context that is represented by the different labels. Of course we can—if we like—treat labels as referring expressions, and evaluate tagged formulas in a possible worlds model. We can think of assertions in some fixed context as evaluated with respect to some *possible world*. Perhaps this interpretation can do some explanatory work. However, it is a further question as to which direction the explanation goes. Do we justify the proofs on the terms that they fit with possible worlds models, or do we justify the models on the basis that they provide models that fit with the antecedently grasped rules for reasoning with modal expressions and the context shifts they involve.[6] Since any questions concering the explanatory power of possible worlds talk remain when we talk of *inconsistent* or *incomplete* 'worlds', and new worries are added to them, it seems worthwhile to see if such a story can be told for incompatibility models for negations, so let's turn our attention there.

[6]See my "Truth Values and Proof Theory" [15] for a development of the explanatory connection between proofs and models in this second direction, justifying models in terms of an antecedently motivated system of proof rules. Although the proof system discussed there is a hypersequent calculus and not a labelled natural deduction system, the broad lesson is the same.

2 PROOFS WITH PERP

If we start with the idea that proofs can combine claims made in different contexts, the issue immediately arises: what, if anything, follows if we conclude $\neg A$ one context and A in another? On the rules discussed in the previous section (see Figure 2), the answer is 'nothing, in general.' There is no rule telling us what follows. However, having learned the lessons Dunn taught us, we see that there *is* a way to describe what follows from $\neg A$ in context a and A in context b, and that is that the contexts a and b are incompatible. So, let's mark that in our reasoning writing '$a \perp b$' for the incompatibility claim.

$$\dfrac{\neg A \cdot x \qquad A \cdot y}{x \perp y}\,\neg E \qquad\qquad \dfrac{\begin{matrix}[A \cdot y]^i \\ \Lambda \\ x \perp y\end{matrix}}{\neg A \cdot x}\,\neg I^{i\dagger}$$

†(In $\neg I$, y appears nowhere else in the undischarged assumptions of the proof Λ.)

Figure 3: NEGATION RULES WITH PERP

With that as a new candidate for a negation *elimination* rule, what would a be matching *introduction* rule? Harmony considerations dictate that we are in position to *derive* $\neg A$ (in a given context x) whenever we can vouchsafe the transition from A in some context to the conclusion that x is incompatible with *that* context—whichever context that turns out to be. So, the natural negation introduction rule is no real surprise. The *incompatibility* rules for negation are given in Figure 3. If we have some labelled proof from $A \cdot y$ to the conclusion $x \perp y$, then provided we have made no other assumption in context y, we can conclude $\neg A$ in context x. That context, indeed, has what it takes to rule A out.

If we replace the negation rules from Figure 2 by these *incompatibility* rules, given in Figure 3, we have a simple, well-behaved natural deduction system for the logic of incompatibility models, and it is one can motivate the logic by general considerations of reasoning with negation *in contexts*. To illustrate, her is an example proof, from the assumption $\neg p \vee \neg q$, to the conclusion $\neg(p \wedge q)$, in the same context.

$$\dfrac{\dfrac{\neg p \vee \neg q \cdot a \qquad \dfrac{[\neg p \cdot a]^1 \qquad \dfrac{[p \wedge q \cdot b]^3}{p \cdot b}\,\wedge E}{a \perp b}\,\neg E \qquad \dfrac{[\neg q \cdot a]^2 \qquad \dfrac{[p \wedge q \cdot b]^3}{q \cdot b}\,\wedge E}{a \perp b}\,\neg E}{a \perp b}\,\vee E^{1,2}}{\neg(p \wedge q) \cdot a}\,\neg I^3$$

Notice that this proof has the same general shape as the unlabelled proof of $\neg(p \wedge q)$ from the premise $\neg p \vee \neg q$. The difference is the presence of the labels. We need to check that the negation introduction step at the conclusion of this proof satisfies the side-condition of that rule, that the context label b, occuring in the conclusion and in the assumptions discharged that that step (the two occurrences of $p \wedge q \cdot b$) occur nowhere else among the undischarged assumptions, and so, the b in this part of the proof is perfectly *general*. This is indeed the case, so the proof is well-formed.

There are a few notable features of labelled proofs: first, incompatibility facts $a \perp b$ arise only *inside* proofs, and never as assumptions occuring in the leaves. (The smallest proof is a labelled formula $A \cdot a$ on its own, in which the formula appears as both premise and conclusion.) Second, notice that the proof rules each have the same shape as the unlabelled proof rules in Figure 1. If we take any labelled proof and simply erase all the labels, the result is a proof in intuitionistic natural deduction. Third, as we had hoped as we introduced these proofs, they match Dunn's incompatibility models for negation, precisely. We have the following two facts. We start with *soundness*. To state the soundness fact, we need to be precise with how the premises and conclusions of a proof might be evaluated on a model. We will use the following, natural, definition:

DEFINITION 1 [HOLDING, VALIDITY] $A \cdot a$ HOLDS IN THE MODEL $\langle P, \perp, \Vdash \rangle$, relative to the assignment p of points to labels when $p(a) \Vdash A$. The the incompatibility claim $b \perp c$ HOLDS IN THE MODEL when $p(b) \perp p(c)$.

An argument from the labelled premises X to some conclusion C (either a labelled formula or an incompatibility claim) is VALID ON THE MODEL if and only if for any assignment p, if the premises hold in the model, so does the conclusion.

With that definition in place, we have the following fact:

FACT 1 [SOUNDNESS (FOR PERP PROOFS)] *If Λ is a labelled proof with premises X and conclusion $B \cdot c$ [or $b \perp c$], then corresponding argument is valid on every incompatibility model.*

Proof: This can be shown by a straightforward induction on the construction of the proof Λ, appealing to the match between the proof rules and the evaluation conditions for the connectives in incompatibility models. The base case consists of a proof with the single premise $A \cdot a$, with the very same conclusion, so the induction hypothesis holds trivially. The cases for the distributive lattice connectives are straightforward and need no extensive discussion, but let's consider $\vee E$ since it allows for context shifting. Let's fix on some incompatibility model $\langle P, \perp, \Vdash \rangle$ and some assigment p sending labels to points, and we suppose the hypothesis applies to the labelled proofs Λ, Λ', Λ'' leading up to this application of $\vee E$:

$$\dfrac{\begin{array}{c} X \\ \Lambda \\ A \vee B \cdot x \end{array} \quad \begin{array}{c} Y,\ [A \cdot x]^j \\ \Lambda' \\ C \end{array} \quad \begin{array}{c} Z,\ [B \cdot x]^k \\ \Lambda'' \\ C \end{array}}{C} \vee E^{j,k}$$

So, the arguments from X to $A \vee B \cdot x$, and from $Y, A \cdot x$ to C and $Z, B \cdot x$ to C are valid on our incompatibility model $\langle P, \perp, \Vdash \rangle$. Consider now our assignment p. We know that if each of the members of X hold in the model, so does $A \vee B \cdot x$, and that if the members of $Y, A \cdot x$ hold in the model, so does C, and if the members of $Z, B \cdot a$ hold in the model, C does as well. To show that the entire proof is valid in the model, suppose that each of the members of X, Y, Z hold in the model. Since the members of X hold, so does $A \vee B \cdot x$, and by the disjunction clause ($a \Vdash A \vee B$ iff $a \Vdash A$ or $a \Vdash B$), we know that either $A \cdot x$ holds in the model or $B \cdot x$ holds in the model. In the first case, the validity of the argument from $Y, A \cdot x$ to C assures us that C holds in the model, and in the second case, it is the validity of the second argument that does the job, and so, in either case, the desired conclusion indeed holds in the model. (Notice that this reasoning applies whether the conclusion C is a labelled formula, or an incompatibility fact. The only condition requirement is that it be derived in either case.)

For the negation rules, consider $\neg E$:

$$\dfrac{\begin{array}{c} X \\ \Lambda \\ \neg A \cdot x \end{array} \quad \begin{array}{c} Y \\ \Lambda' \\ A \cdot y \end{array}}{x \perp y} \neg E$$

By hypothesis, the arguments from X to $\neg A \cdot x$ and from Y to $A \cdot y$ are valid on our model. So, to show that the whole argument is valid, suppose that both X and Y hold in our model under assignment p. It follows that $p(x) \Vdash \neg A$, and $p(y) \Vdash A$. Given the evaluation clause for negation, this means that $p(x) \perp p(y)$, and hence, $x \perp y$ holds in our model, as desired.

Finally, consider $\neg I$:

$$\dfrac{\begin{array}{c} X,\ [A \cdot y]^i \\ \Lambda \\ x \perp y \end{array}}{\neg A \cdot x} \neg I$$

The induction hypothsis is that the argument from $X, A \cdot y$ to $x \perp y$ is valid on our model. We wish to show that the argument from X to $\neg A \cdot x$ is *also* valid on our model. The one extra piece of information is that the label y is absent from X. So, let's suppose that p is some assignment for which X holds in our model. We wish to show that $\neg A \cdot x$

holds in our model, too. For that, we need to show that whenever $b \Vdash A$, then $p(x) \perp b$. So, suppose b is a point in the frame where $b \Vdash A$. Select a new assignment p' of points to labels such that $p'(y) = b$ while $p'(z) = p(z)$ for every *other* label z. (This includes the label x.) This includes the labels occuring in the premises X, since y is absent from X. The induction hypotheis applied to *this* assigment p' tells us that since every member of $X, A \cdot y$ hold under *this* assignment, so does the conclusion $x \perp y$. Since $p'(x) = p(x)$ and $p'(y) = b$, we have $p(x) \perp b$ as desired, and so, since b was arbitrary, we have shown that $\neg A \cdot x$ holds in our model. ∎

Now, since nothing in this reasoning assumed *any* features of the incompatibility relation $\perp$, this soundness theorem tells us that the logic arising out of these proof rules is *very* weak. The non-symmetric model given above[7] provides a counterexample to the argument from $p \cdot a$ to $\neg\neg p \cdot a$. In other words, there is no way to assign labels to the unlabelled proof

$$\dfrac{\dfrac{[\neg p]^1 \qquad p}{\bot}\,\neg E}{\neg\neg p}\,\neg I^1$$

—or to any other unlabelled proof from p to $\neg\neg p$—satisfying the conditions of our labelled calculus. Models where compatibility fails to be reflexive (and so, we allow $x \perp x$ for some points x) give us means to satisfy contradictory conjunctions $p \wedge \neg p$ at point, and so, to provide ways to fail to validate $\neg(p \wedge \neg p)$. This means that we also have no way to convert this

$$\dfrac{\dfrac{\dfrac{[p \wedge \neg p]^1}{\neg p}\,\wedge E \qquad \dfrac{[p \wedge \neg p]^1}{p}\,\wedge E}{\bot}\,\neg E}{\neg(p \wedge \neg p)}\,\neg I^1$$

—or any other proof to this conclusion from no premises—into a properly labelled proof. Of course, there are simple ways to extend the calculus with symmetry or reflexivity rules to allow for labellings of these proofs:

$$\dfrac{\dfrac{\dfrac{[\neg p \cdot b]^1 \qquad p \cdot a}{b \perp a}\,\neg E}{a \perp b}\,\mathit{Sym}}{\neg\neg p \cdot a}\,\neg I^1 \qquad\qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[p \wedge \neg p \cdot b]^1}{\neg p \cdot b}\,\wedge E \qquad \dfrac{[p \wedge \neg p \cdot b]^1}{p \cdot b}\,\wedge E}{b \perp b}\,\neg E}{\bot}\,\mathit{Refl}}{a \perp b}\,\bot E}{\neg(p \wedge \neg p) \cdot a}\,\neg I^1$$

[7] See footnote 2.

Rather than exploring the details of different rules such as *Sym* and *Refl* and the corresponding frame conditions (for space reasons only—the proving soundness for the extended proof systems with respect to the corresponding class of models is straightforward) let us end this section discussing the *Completeness* Theorem.

FACT 2 [COMPLETENESS (FOR PERP PROOFS)] *If an argument with premises X and conclusion $B \cdot c$ [or $b \perp c$] is valid on every incompatibility model, then there is some labelled proof with from X to $B \cdot c$ [or $b \perp c$].*

There are a number of different ways to show that our labelled proof rules are complete for validity on incompatibility models. We will sketch a direct proof of completeness of an extended proof system in Section 4. Here I will take a different approach, favouring theft over toil.

Proof [*Sketch*]: We borrow the completeness result for the *sequent* natural deduction system of for distributive lattice logic with split negation, given in *An Introduction to Substructural Logic* (*ISL*), Chapter 11 [13]. There, it is shown that incompatibility frames are sound and complete for a simple *unlabelled* sequent calculus for distributive lattice logic extended with two rules for *two* negations:

$$\frac{A \vdash \neg B \qquad X \vdash B}{X \vdash {\sim}A}\ {\neg E/{\sim}I} \qquad \frac{A \vdash {\sim}B \qquad X \vdash B}{X \vdash \neg A}\ {{\sim}E/\neg I}$$

Incompatibility models are extended to interpret the converse negation $\sim$ in the obvious way, using the *converse* of incompatibility.[8] The two negations are modelled together, like this:

- $a \Vdash \neg A$ if and only if for each b, if $b \Vdash A$ then $a \perp b$,
- $a \Vdash {\sim}A$ if and only if for each b, if $b \Vdash A$ then $b \perp a$.

We then can show that unabelled sequent proofs (which are complete for compatibility frames) can be translated into labelled proofs. A sequent consists of a multiset X of formulas on the left-hand side and a single formula on the right. For any label a, and any natural deduction proof π ending in a sequent $X \vdash A$, the a-translation of π will be some labelled proof π_a with undischarged assumptions selected from X, each labelled with a, and with conclusion $A \cdot a$. The axiomatic sequent $A \vdash A$ is a-translated into the atomic proof $A \cdot a$. The rules for conjunction and disjunction act exactly as one would expect with a homophonic translation from one system to the other. For sequent natural deduction proofs using the negation rules we translate as follows, using the expected

[8]To be precise, in *ISL*, the models are described using *compatibility* rather than an *incompatibility* relation, but the

converse I and E rules for the negation $\sim$.

$$\dfrac{\begin{matrix}\pi\\ A\vdash\neg B\end{matrix}\qquad\begin{matrix}\pi'\\ X\vdash B\end{matrix}}{X\vdash{\sim}A}\,\neg E/{\sim}I \quad\Rightarrow\quad \dfrac{\dfrac{\begin{matrix}[A\cdot b]^1\\ \pi_b\\ \neg B\cdot b\end{matrix}\qquad\begin{matrix}X\cdot a\\ \pi'_a\\ B\cdot a\end{matrix}}{b\perp a}\,\neg E}{{\sim}A\cdot a}\,{\sim}I^1$$

$$\dfrac{\begin{matrix}\pi\\ A\vdash{\sim}B\end{matrix}\qquad\begin{matrix}\pi'\\ X\vdash B\end{matrix}}{X\vdash\neg A}\,{\sim}E/\neg I \quad\Rightarrow\quad \dfrac{\dfrac{\begin{matrix}[A\cdot b]^1\\ \pi_b\\ {\sim}B\cdot b\end{matrix}\qquad\begin{matrix}X\cdot a\\ \pi'_a\\ B\cdot a\end{matrix}}{a\perp b}\,{\sim}E}{\neg A\cdot a}\,\neg I^1$$

The end of the translation process is some labelled proof from the premises to the conclusion, littered, alas, with the converse negation $\sim$ which was may be foreign matter to the premises and conclusion under discussion. This is no problem. We have been delivered a labelled natural deduction proof. It can be normalised, as usual (remember, if we delete the labels, this is simply an intuitionist propositional proof, satisfying extra conditions). The normalisation process (as we will see in the final section) preserves the special properties of the negation rules, and so, we can normalise our proof into a thorougly normal one, which will satisfy the subformula property. This will delete the extraneous converse negations from our proof and the result is a proof in the original calculus. ∎

What, then, of the de Morgan negation represented by Routley star models? On these models, the argument from $\neg\neg A$ to A comes out as *valid*, but no straightforward rules governing $\perp$ will give us a means to bridge that gap—at least, none will do so if the added rules remain intuitionistically acceptable when we erase labels. We'll have to do something else to bridge the gap. For that, we turn to natural deduction for *classical logic*.

3 ASSERTION AND DENIAL

There are many ways to modify the rules for Prawitz-style natural deduction to allow for properly classical proofs. The simplest thing to do is to add extra principles to some or other connective, beyond the harmonious introduction and elimination rules. This will *not* be our strategy here, since the harmonious introduction and elimination rules. Instead, we will pay attention to the kinds of speech acts we use in dialogical situations. Not only do we assert and suppose in different contexts. We also *deny*. Taking assertion and denial to be on equal footing has become an important theoretical perspective on proof-theoretical semantics for classical logic [6, 7, 11, 14, 18, 20, 21].

However, instead of complicating our natural deduction system to incorporate positively and negatively signed judgements at every step (which results in a *fully bilateral* natural deduction system), it suffices for our purposes to be only very *mildly* bilateral, by allowing for negative *premises* in our proofs, and leaving the of the rest of the proof system unchanged.[9] So, for our original unlabelled natural deduction system, we allow not only formulas to occur in the leaves of a proof, we also allow *slashed* formulas, $\cancel{A}$, indicating the (primitive) *denial* of the claim A. If to assert A is to answer *yes* to the polar question A?, then to deny A is to answer the qustion A? with a *no*.

$$\dfrac{\cancel{A} \qquad A}{\bot}\uparrow (\textit{Store}) \qquad \dfrac{\begin{matrix}[\cancel{A}]^i \\ \Pi \\ \bot\end{matrix}}{A}\downarrow^i (\textit{Retrieve}) \qquad\qquad \dfrac{\cancel{A\cdot x} \qquad A\cdot x}{\bot}\uparrow \qquad \dfrac{\begin{matrix}[\cancel{A\cdot x}]^j \\ \Lambda \\ \bot\end{matrix}}{A\cdot x}\downarrow^j$$

Figure 4: *DENIAL* RULES, UNLABELLED AND LABELLED

The rules governing this primitive denial are straightforward, and are presented in Figure 4 (focus on the first pair, of *unlabelled* rules, for the moment). Asserting and denying the same thing is, of course, inconsistent. On the other hand, if we reach an inconsistency, one way out of that dead end is to say 'yes' to something we had previously ruled out. (These rules are named '*Store*' and '*Retrieve*' because we can think of the negatively tagged formulas as temporarily stored conclusions, which can be retrieved from storage when needed.) With the addition of these purely structural rules, the existing logical rules for our unlabelled system now suffice for all of classical logic. Here is a proof from $\neg\neg p$ to p:

$$\dfrac{\dfrac{\neg\neg p \qquad \dfrac{\dfrac{[\cancel{p}]^2 \qquad [p]^1}{\bot}\uparrow}{\neg p}\neg I^1}{\bot}\neg E}{p}\downarrow^2$$

At the first step we have assumed p and denied it at the same time, to reach a contradiction. We immediately blame that contradiction on the assumption of p, discharging that to deduce $\neg p$. So, at this stage of the proof we have inferred $\neg p$ in a context where p is denied. This, of course, clashes with the assumption $\neg\neg p$, and so, in the presence of

[9] For a much more extensive treatment of this mildly bilateralist natural deduction, its connection to Michel Parigot's $\lambda\mu$-calculus for classical logic, and its many pleasant properties, see "Speech Acts & the Quest for a Natural Account of Classical Proofs" [17].

this contradiction we retrieve the denied p, to conclude p, as desired. We recover classical reasoning by paying the coin of two structural rules involving denials.

Adding these two rules to the unlabelled natural deduction system has an added benefit. The awkwardly justified $\bot E$ rule is now redundant, and can be replaced by the *Retrieve* rule, in the special (*explicitly* irrelevant) case where we retrieve *zero* instances of a stored A.

$$\dfrac{p}{q \to p}\,{\to}I \qquad \dfrac{\dfrac{\neg p \qquad p}{\bot}\,\neg E}{q}\,\downarrow$$

In this case, in just the same way that the classically valid inference from p to $q \to p$ arises from assuming p, and discharging *zero* occurrences of the assumption of q to conclude $q \to p$, the classically valid inference from $p, \neg p$ to q arises in the same way. The p and $\neg p$ lead to a contradiction, and we retrieve *zero* instances of the denied q to conclude q. The two kinds of irrelevance present in classical logic come down to the same underlying phenomenon, vacuous discharge of positive or negative assumptions, and if we wish to be properly *relevant*, one way to do this is to ban vacuous discharge, whether positive or negative.

Instead of exploring the finer details of this classical natural deduction system, let's immediately see how we can use ideas from it in the setting of a *labelled* natural deduction system. The generalisation we use, as seen in the second pair of rules in Figure 4, is the simplest possible. Just as in the unlabelled system we allowed for leaves of the tree to be slashed to represent denial, we do the same thing here, so the denial of A in the context x is represented by slashing the entire judgement, like so: $\cancel{A \cdot x}$. This one change will allow us to construct natural deduction proofs for the Routley star semantics, as we will see in the next section.

4 PROOFS WITH STAR

Recall the truth conditions for negation in Routley star models:

- $x \Vdash \neg A$ if and only if $x^* \nVdash A$.

If we use this modelling condition in a frame with a map $^* : P \to P$ of period two (so $x^{**} = x$), then the appropraite proof rules for negation to match these models will be as given in Figure 5.

Despite the presence of the Routley star with period two, the $\neg I/\neg E$ rules, together with the label-manipulating rules $^{**}I$ and $^{**}E$ do not suffice to derive constructively invalid inferences from $\neg\neg p$ to p, or the constructively invalid de Morgan's law. These rules, as usual, remain constructively valid if we erase the labels, so on their own, they would not be enough to generate the non-constructive de Morgan validities. In the

$$
\dfrac{\neg A\cdot x \qquad A\cdot x^*}{\bot}\,\neg E
\qquad\qquad
\dfrac{\begin{matrix}[A\cdot x^*]^i\\ \Lambda\\ \bot\end{matrix}}{\neg A\cdot x}\,\neg I^i
\qquad\qquad
\dfrac{A\cdot x}{A\cdot x^{**}}\,{**}I
\qquad\qquad
\dfrac{A\cdot x^{**}}{A\cdot x}\,{**}E
$$

Figure 5: NEGATION RULES WITH STAR

presence of the denial rules, however, they suffice. For example, we have this proof for double negation elimination:

$$
\dfrac{\dfrac{\dfrac{\neg\neg p\cdot a \qquad \dfrac{\dfrac{[\cancel{p\cdot a^{**}}]^2 \qquad [p\cdot a^{**}]^1}{\bot}\,\uparrow}{\neg p\cdot a^*}\,\neg I^1}{\bot}\,\neg E}{p\cdot a^{**}}\,\downarrow^2}{p\cdot a}\,{**}E
$$

This longer proof suffices for the constructively invalid de Morgan's Law:

$$
\dfrac{\dfrac{[\cancel{\neg p\vee\neg q\cdot a}]^4 \qquad \dfrac{\dfrac{\dfrac{[\cancel{\neg p\vee\neg q\cdot a}]^4 \qquad \dfrac{\dfrac{\dfrac{[\cancel{\neg p\cdot a}]^3 \qquad \dfrac{\dfrac{\neg(p\wedge q)\cdot a \qquad \dfrac{[p\cdot a^*]^1 \qquad [q\cdot a^*]^2}{p\wedge q\cdot a^*}\,\wedge I}{\bot}\,\neg E}{\neg p\cdot a}\,\neg I^1}{\bot}\,\uparrow}{\neg q\cdot a}\,\neg I^2}{\neg p\vee\neg q\cdot a}\,\vee I}{\bot}\,\uparrow}{\neg p\cdot a}\,\downarrow^3}{\neg p\vee\neg q\cdot a}\,\vee I}{\bot}\,\uparrow}{\neg p\vee\neg q\cdot a}\,\downarrow^4
$$

We can combine the *denial* rules with the incompatibility rules for negation, too. With either the star rules or the incompatibility rules in place for negation, the resulting proofs are still sound for the logic modelled by our star or perp models. To show this, we need to expand our definition of validity and holding, to allow for slashed premises in our arguments.

DEFINITION 2 [HOLDING AND VALIDITY, AGAIN] For any model with an evaluation $\Vdash$, the slashed labelled formula $\cancel{A \cdot a}$ HOLDS in the model when $p(a) \nVdash A$. The inconsistency sign $\perp$ never holds in a model.

In any *star* model, a map p from labels to points must respect stars on labels: that is, $p(a^*) = p(a)^*$ for each label a. Then, as before, an argument from X to C is VALID ON A MODEL, if and only if, under any assginment p, if the members of X hold in the model, so does the C.

With this definition, soundness is straightforward to state and prove.

FACT 3 [SOUNDNESS FOR LABELLED PROOFS WITH DENIAL] *If Λ is a labelled incompatibility* [*star*] *proof with premises X and conclusion $B \cdot c$* [*or $b \perp c$, or $\perp$*]*, then corresponding argument is valid on every incompatibility* [*star*] *model.*

Proof: The proof takes the same shape as our previous soundness proof, an induction on the construction of the labelled proof. The new cases to consider are the *Store* and *Retrieve* rules, as well as the star rules. Consider the *Store* and *Retrieve* rules first. Suppose the argument from X to $A \cdot a$ is valid on our model, and we extend some proof from X to $A \cdot a$ with the *Store* rule, adding the premise $\cancel{A \cdot a}$, leading to the conclusion $\perp$. Since on any choice of values for the labels in which the members of X hold, $A \cdot a$ holds, there is no assignment of values in which the members of X and $\cancel{A \cdot a}$ holds, so the argument from X, $\cancel{A \cdot a}$ to $\perp$ is indeed valid on our model.

Conversely, if we have a proof Λ from X, $\cancel{A \cdot a}$ to $\perp$ and this argument is valid on our model, this means there is no assignment of points to labels that such that every member of X, $\cancel{A \cdot a}$ holds. So, for any assignment of points to labels where every member of X holds, $A \cdot a$ must hold, too, and the argument from X to $A \cdot a$ is valid on our model, too.

For the star rules, soundness is straightforward for the $^{**}I$ and $^{**}E$ rules, since any assignment of points to labels respects star, and in any Routley star model, $a^{**} = a$ for every point a. Soundness for $\neg I$ and $\neg E$ is also straightforward, given the truth conditions for negations on Routley star models. There is *no* model in which $\neg A \cdot x$ and $A \cdot x^*$ can both hold, on the same assignment of values to points. On the other hand, if the argument from X and $A \cdot x^*$ to $\perp$ is valid on a model, that means that there is no assignment of points to labels where X and $A \cdot x^*$ holds. That means on any assignment where which each member of X holds is one whre $A \cdot x^*$ doesn't, but that means $\neg A \cdot x$ holds, and so, the argument from X to $\neg A \cdot x$ is valid, as desired. ∎

The *completeness* fact for proofs with denial is straightforward to state and prove.

FACT 4 [COMPLETENESS FOR LABELLED PROOFS WITH DENIAL] *An argument from X and C valid in every incompatibility* [*star*] *model, also has some labelled incompatibility* [*star*] *proof.*

Proof [*Sketch*]: This proof is a standard canonical model construction, and as usual, we prove the contrapositive. If there is no proof from X to the conclusion C, we construct a model and assignment of points to labels in which each member of X holds, under that assignment and C fails to hold, again, under that assignment. Instead of constructing the model out of maximal sets of formulas (or maximal prime/ideal pairs, as is appropriate for distribuitve lattice logics [13]), the construction is rather straightforward. We start with our the pair $\langle X, \{C\}\rangle$, which is *available*, in the sense that there is no proof from any subset of elements on the left side to any member of the right side. Our aim is to fill this out, constructing a partition $\langle Yes, No\rangle$ of the whole language, where the statements on the left side are ruled *in* while those on the right are ruled *out*. The constraint, as we fill out our pair of sets is that the pair *remains* available as we fill it out. We must be careful, since the constraint of availability is stricter than the condition that is no proof from members of *Yes* to any members of *No*, since the pair $\langle\{p \vee q \cdot a\}, \{p \cdot a, q \cdot a\}\rangle$ is indeed available in that weak sense (there is no proof from $p \vee q \cdot a$ to $p \cdot a$, and nor is there a proof to $q \cdot a$) but there will be no model in which $p \vee q \cdot a$ holds and both $p \cdot a$ and $q \cdot a$ doesn't. No, this pair shouldn't be available, and the appropriate definition of availability of a pair $\langle Y, N\rangle$ is that there is no proof from members of Y together with *denied* (slashed) members of N to a member of N. There is a proof from $p \vee q \cdot a$, $\cancel{p \cdot a}$ to $q \cdot a$, and so the pair $\langle\{p \vee q \cdot a\}, \{p \cdot a, q \cdot a\}\rangle$ is not available, as desired.

So, the construction of our partition $\langle Yes, No\rangle$ can proceed in the way familiar from Henkin completeness proofs. We take for ourselves a countable supply of labels, and enumerate our collection of labelled formulas and compatibility claims. Then, given any finite available pair $\langle Y, N\rangle$, we select the next object in our enumeration C (either a labelled formula or an incompatibility statement). We consider whether we have a proof from $Y, \cancel{N}$ to C. If we do, then there is no proof from $Y, \cancel{N}, C$ to $\perp$. If we *did* have such a proof, we could compose this with the proof from $Y, \cancel{N}$ to C, to give us a proof which leads us from $Y, \cancel{N}$ to $\perp$, in which case $\langle Y, N\rangle$ would not be available. So, either $\langle Y \cup \{C\}, N\rangle$ is available or $\langle Y, \{C\} \cup N\rangle$ is. Choose one to be $\langle Y', N'\rangle$, and continue the process until the limit, where every labelled formula and compatibility statement has been chosen. The result is a partition of the language, and it is available.

There is one complication to the process. If the labelled formula C is a *negation* $\neg A \cdot a$, we add it to the *right* of the partition, we must add a *witness*, too. In this case, we know that $\langle Y, \{\neg A \cdot a\} \cup N\rangle$ is available. So, there is no proof from $Y, \cancel{N}$ to $\neg A \cdot a$. So, choose a fresh label b. We know that there is no proof $Y, \cancel{N}, \cancel{\neg A \cdot a}, A \cdot b$ to $a \perp b$, since if there *were* such a proof, we could extend it by $\neg E$ (and some store/retrieve steps) to give a proof of $\neg A \cdot a$ from $Y, \cancel{N}$, and we know there is no such proof. So, it follows that $\langle Y \cup \{A \cdot b\}, \{\neg A \cdot a, a \perp b\} \cup N\rangle$ is also available, so whenever our construction asks us to make a negation *not* true, we add a fresh witness label and witnessing facts in this way.

Then, the limit $\langle Yes, No\rangle$ of this process will be a description of a compatibility model. The class of points is the set of all labels used in the process, the incompatibility relation is read off the set of $a \perp b$ claims added to the *No* side (these are the incompatibility claims that *don't* hold in the canonical model), we say $a \Vdash A$ when $A \cdot a \in \mathit{Yes}$ and $a \nVdash A$ when $A \cdot a \in \mathit{No}$. The standard truth conditions can be seen to hold by an easy inspection. For example, if $a \Vdash A$ and $a \Vdash B$ then since $A \cdot a, B \cdot a \in \mathit{Yes}$, we must have $A \wedge B \cdot a \in \mathit{Yes}$ since $\langle \mathit{Yes}, \mathit{No} \rangle$ is a partition, and we cannot have $A \wedge B \cdot a \in \mathit{No}$, since there is a one-step proof from $A \cdot a, B \cdot a$ to $A \wedge B \cdot a$, so $a \Vdash A \wedge B$ as desired. For the converse, if $a \Vdash A \wedge B$, then $A \wedge B \cdot a \in \mathit{Yes}$, and since there are one-step proofs from the conjunction to each conjunct, we must have $A \cdot a$ and $B \cdot a \in \mathit{Yes}$ too.

The most difficult cases to check are negation. If $a \Vdash \neg A$, then we wish to show that for any b where $b \Vdash A$, we have $a \perp b$ in the model. That is, we want $a \perp b \notin \mathit{No}$. This is immediate, since if $\neg A \cdot a \in \mathit{Yes}$ and $A \cdot b \in \mathit{Yes}$, we cannot have $a \perp b \notin \mathit{No}$, since we can infer $a \perp b$ by $\neg$E. For the converse, we show that if $a \nVdash \neg A$ then there is some b where $b \Vdash A$ and $a \perp b$ *fails* in our model. But this is *exactly* the condition on partitions that is ensured by the special condition on adding negations to the *No* side of the partition, covered above.

That sketches the completeness construction for *incompatibility* frames. The case for *star* frames is much simpler, since we do not need to throw in fresh labels for witnesses for negations in the construction. The details are simpler, and are left to the reader as an exercise. ∎

Let's leave the formal reflections on this proof system here. There are many more questions to be answered, of course, and there is much more to explore, but I hope that I have at least sketched a way forward for developing proof-first understanding of how we might acquire the concept of negation, and *why* such a concept might appropriately be modelled with an incompatibility semantics. If we use the concepts of conjunction, disjunction, negation like *this*, by making suppositions, assertions, denials, and inferences *in context*; if context shifts allow for us to at least potentially take *some* suppositional contexts as self-undermining (in the sense of being incompatible with themselves), then we have given some sense to a paraconsistent negation. And we have done so without having to start with a worlds semantics and its ontology. But the result is not moving away from the logic of incompatibility frames. On the contrary, we have shown how just such a logic can arise.

5 CLASSICAL AND NON-CLASSICAL PROOF

We could stop there, but this raises one further question which I will begin to address. We have a plurality of logics of negation. There is the logic of the *unlabelled* proof system, and the logic with the labelled proof system.[10] Must we be forced to choose, or

[10] There is also the question of whether to use denial or not, but I will set that aside and help myself to denial in what follows.

can we find a way for those to live alongside one another, as two parts of a larger picture, as two kinds of logical consequence, as two different kinds of *proof*?

In this last section of the paper, I will sketch an approach to harmonising a classical and a non-classical account of *proof*, which can be seen as extending the approach to harmonising classical and relevant consequence in "Negation in Relevant Logics" [12]. In that paper, I argued that the incompatibility semantics can be used to defend *both* a relevant paraconsistent consequence relation, and a classical consequence relation. In any incompatibility model $\langle P, \perp, \Vdash \langle$ we can isolate a *subset* of the points, consisting of all those points that act like *worlds*. These are points that are consistent and complete with respect to negation. If w is a world, then we do not have $w \perp w$ (it is *consistent*), and furthermore, if $w \nVdash A$ and we have $y \Vdash A$, then we must have $w \perp y$, since y verifies A and w, per *force* must somehow rule A out, since it is *complete*. If the model is a *star* model, the condition for w being a world is simpler to state: it is simply the condition that $w^* = w$.

Once we have worlds at hand, we can recover classical logic by attending to whether the model furnishes a *world* that renders the premises true and the conclusion false. Our model has both a coarse-grained consequence relation (looking only at worlds) and a finer-grained one (casting its eyes over all points). We have two validity notions for the price of one.

What does this look like if we start with proofs? If we start with *star* proofs, we can split the consistency and completeness conditions on points in two. A point is x *consistent* if whenever $x \Vdash A$, we also have $x^* \Vdash A$. It is *complete* if the converse holds. We can capture this notion in a pair of proof rules, like so:

$$\frac{\mathit{Cons}\,x \qquad A \cdot x}{A \cdot x^*}\,\mathit{ConsE} \qquad\qquad \frac{\mathit{Comp}\,x \qquad A \cdot x^*}{A \cdot x}\,\mathit{CompE}$$

Figure 6: CONSISTENCY AND COMPLETENESS RULES WITH STAR

These rules use a *flag*, entered into the proof tree as an assumption, which records that we are treating the context x as *consistent*, in the first case, and *complete*, in the second. (These need not be taken to be formulas, and I will not do so here.) Once we add the consistency and completeness rules to the proof system, we can recover classical principles. Here is a proof of the law of the excluded middle, showing why $p \vee \neg p$ holds in a complete context.

$$
\dfrac{\dfrac{[\cancel{p \vee \neg p \cdot x}]^2 \qquad \dfrac{\dfrac{\dfrac{\dfrac{[\cancel{p \vee \neg p \cdot x}]^2 \qquad \dfrac{\dfrac{\mathit{Comp}\,x \qquad [p \cdot x]^1}{\neg p \cdot x}\,\mathit{CompE}}{p \vee \neg p \cdot x}\,\vee I}{\bot}\,\uparrow}{\neg p \cdot x}\,\neg I^1}{p \vee \neg p \cdot x}\,\vee I}{}}{\bot}\,\uparrow}{p \vee \neg p \cdot x}\,\downarrow^2
$$

Proving a couple of results like this (showing how to prove the prinicples that suffice to lift us from the weak paraconsistent logic to classical logic) would be enough to show that what is *provable* classically can be *proved* in the paraconsistent proof system, whether with *star* or with *perp*. This is enough if all we are concerned with is *validity*. However, we could ask for more. We could wonder whether any classical (unlabelled) *proof* can be rewritten as a paraconsistent (labelled) proof, enhanced only with judicious applications of the assumption that the context is consistent and complete. In this section I will explain why this is indeed the case, there is a simple embedding of classical proof into labelled proofs, an embedding that is so intimate that the steps we use for *normalising* classical proofs correspond to steps used for normalising labelled proofs. This is strong evidence for the view that classical unlabelled proofs can be further analysed into labelled incompatibility proofs, which make explicit where the assumptions of consistency and completeness play a role in the reasoning, assumptions that are left completely *implicit* in classical proof.

To start presenting this translation from classical proofs into labelled proofs, let's see the consistency and completeness rules appropriate for *incompatibility* proofs. The rules are presented in Figure 7. The *completeness* rule is a little more complex than the

$$
\dfrac{\mathit{Cons}\,x \qquad x \perp x}{\bot}\,\mathit{ConsE}
\qquad\qquad
\dfrac{\mathit{Comp}\,x \qquad \begin{array}{c}[A \cdot x]^i \\ \Lambda \\ \bot\end{array} \qquad A \cdot y}{x \perp y}\,\mathit{CompE}^i
$$

Figure 7: CONSISTENCY AND COMPLETENESS RULES WITH PERP

completeness rule for star proofs, but it encodes the idea that if a context x fails to include A and a context y includes it, then, provided that x is complete, then x is indeed incompatible with y. In the presence of the denial rules, this condition is not too strong, though it might be thought that the second premise of the rule, which requires a reduction of A (at x) to absurdity, which seems stronger than the idea that A simply fails to hold at x. This worry is unfounded, because the labelled proof Λ may have other premises, including simply the premise that A fails at x.

$$\dfrac{Comp\,x \qquad \dfrac{\cancel{A \cdot x} \qquad [A \cdot x]^i}{\bot}\,\uparrow \qquad A \cdot y}{x \perp y}\,CompE^i$$

With the consistency and completeness rules at hand, we can encode the classical unlabelled proof rules in to the incompatibility proofs with labels. The translation will systematically send an unlabelled proof Π from X to C [or $\bot$] to a *labelled* proof Π_w from the premises in X, all labelled with the single context label w, (we abbreviate this set $X \cdot x$, for obvious reasons), to the conclusion $C \cdot w$ [or $\bot$], with the addition of premises $Comp\,w$ and $Cons\,w$ if necessary.

Most of the rules in the labelled and unlabelled proof system differ merely in the presence or absence of labels, and the conjunction, disjunction, and store and retrieve rules may be homophonically translated by adding w labels everywhere. The work must be done with the *negation* rules, which differ subsantially from one system to the other. In particular, the $\neg I$ rule in the labelled system has the side condition involving the freshness of the context label used in the discharged assumption. There is no such condition in the classical system, and the translation of classical proofs to labelled proofs annotates all the premises in the proof with the same context label, so we must find a way to insert a fresh label *somewhere*. But first, let's see the translation for $\neg E$.

$$\dfrac{\begin{matrix}\Pi \\ \neg A\end{matrix} \qquad \begin{matrix}\Pi' \\ A\end{matrix}}{\bot}\,\neg E \quad \Rightarrow \quad \dfrac{\dfrac{\begin{matrix}\Pi_w \\ \neg A \cdot w\end{matrix} \qquad \begin{matrix}\Pi'_w \\ A \cdot w\end{matrix}}{w \perp w}\,\neg E \qquad Cons\,w}{\bot}\,ConsE$$

Given a proof that ends in a $\neg E$ step, we translate it by translating the subproofs first, and then the work of the unlabelled $\neg E$ step is achieved first by a labelled $\neg E$ step, which gives us the conclusion that the world context w is inconsistent. We then appeal to consistency to reduce that conclusion to absurdity.

The translation of $\neg I$ has the same form, except in reverse. The subproof of our $\neg I$ proof is a proof of absurdity. To transform this into a proof of *incompatibility*, we appeal to the *completeness* of our context w, and this assumption provides *just* what we need. We take y to be a fresh label, which can be immediately discharged in a $\neg I$ step, which gives us the negation at w as desired.

$$
\dfrac{\begin{matrix}[A]^i\\ \Pi\\ \perp\end{matrix}}{\neg A}{\scriptstyle\neg I^i}
\quad\Rightarrow\quad
\dfrac{\dfrac{Comp\ w \qquad \begin{matrix}[A\cdot w]^i\\ \Pi_w\\ \perp\end{matrix} \qquad [A\cdot y]^j}{w\perp y}{\scriptstyle CompE^i}}{\neg A\cdot w}{\scriptstyle\neg I^j}
$$

With these translations, and homophonic translations of all the other inference rules, we have a systematic translation of classical unlabelled proofs into labelled incompatibility proofs. We can see the distinctions that classical *proofs* draw, faithfully represented inside the wilder and more varied world of labelled incompatibility proofs. We have two logics of *proofs*, for the price of one.

Let me end this section (and this paper) by showing how this representation is so faithful that normalisation steps operating on classical proofs correspond to normalisation steps operating on the underlying labelled proofs. The key normalisation steps for our purposes, again, are the negation rules, since these differ most, so I will explain these and leave the rest to the reader. Recall, the principle case of normalisation in a proof is the elimination of a *detour* where a formula is *introduced* and then *eliminated*. In the unlabelled proof system, a negation introduction followed by an elimination goes like the proof on the left, and the detour formula $\neg A$ is highlighted.

$$
\dfrac{\dfrac{\begin{matrix}[A]^i\\ \Pi\\ \perp\end{matrix}}{\neg A}{\scriptstyle\neg I^i} \qquad \begin{matrix}\Pi'\\ A\end{matrix}}{\perp}{\scriptstyle\neg E}
\quad\rightsquigarrow\quad
\begin{matrix}\Pi'\\ A\\ \Pi\\ \perp\end{matrix}
$$

We can eliminate this detour by replacing this proof section by the proof on the right, which stacks the subproof Π' some number of times into the proof Π. (How many times? Exactly as many times as the discharged assumption A was used in Π. In general, this is any finite number, including zero.)

In the case of a *labelled* proof, we can normalise in exactly the same way, though with the wrinkle of incorporating labels:

$$
\dfrac{\dfrac{\begin{matrix}[A\cdot y]^i\\ \Lambda\\ x\perp y\end{matrix}}{\neg A\cdot x}{\scriptstyle\neg I^i} \qquad \begin{matrix}\Lambda'\\ A\cdot z\end{matrix}}{x\perp z}{\scriptstyle\neg E}
\quad\rightsquigarrow\quad
\begin{matrix}\Lambda'\\ A\cdot z\\ \Lambda[y/z]\\ x\perp z\end{matrix}
$$

Here, we take the subproof Λ' of $A\cdot z$ and use this in the place of the discharged assumptions $A\cdot y$. To do this, we replace each y in Λ by z, reassured that this does not change any *other* premise in Λ, and the result replaces the conclusion of the proof by $x \perp z$, the required conclusion.

Another normalising pair to consider in our labelled proof system involves the consistency and completeness rules. Consider how these could be iterated. If I stack a *CompE* before a *ConsE* step in some proof, we have a subproof which ends in this pair of steps:

$$\dfrac{\mathit{Cons}\,x \qquad \dfrac{\mathit{Comp}\,x \qquad \begin{array}{c}[A\cdot x]^i\\ \Lambda\\ \bot\end{array} \qquad \begin{array}{c}\Lambda'\\ A\cdot x\end{array}}{x\perp x}\,\mathit{CompE}^i}{\bot}\,\mathit{ConsE}$$

It seems quite natural to see the $x \perp x$ is a detour formula, since the detour through this formula (and the concominant appeals to completeness and consistency) is thoroughly redundant. We could replace this subproof by this subproof

$$\begin{array}{c}\Lambda'\\ A\cdot x\\ \Lambda\\ \bot\end{array}$$

which makes no such appeal to consistency and completeness, and arrives at the same conclusion from the same remaining premises. With this pair of normalisation steps at hand, consider the labelled *translation* of an unlabelled proof which invovles a $\neg I/\neg E$ detour. The translation of such a classical detour has this shape:

$$\dfrac{\dfrac{\dfrac{\dfrac{\mathit{Comp}\,w \qquad \begin{array}{c}[A\cdot w]^i\\ \Pi_w\\ \bot\end{array} \qquad [A\cdot y]^j}{w\perp y}\,\mathit{CompE}^i}{\neg A\cdot w}\,\neg I^j \qquad \begin{array}{c}\Pi'_w\\ A\cdot w\end{array}}{w\perp w}\,\neg E \qquad \mathit{Cons}\,w}{\bot}\,\mathit{ConsE}$$

In this labelled proof, the $\neg A \cdot w$ is indeed a detour formula, and in this translated proof, it is introduced by a $\neg I$ and immediately eliminated by a $\neg E$. So, we can normalise it away. Once we do, we get exactly the *Completeness/Consistency* pair we saw

$$\dfrac{\dfrac{\mathit{Comp}\,w \qquad \begin{array}{c}[A\cdot w]^i\\ \Pi_w\\ \bot\end{array} \qquad \begin{array}{c}\Pi'_w\\ A\cdot w\end{array}}{w\perp w}\,\mathit{CompE}^i \qquad \mathit{Cons}\,w}{\bot}\,\mathit{ConsE} \qquad \rightsquigarrow \qquad \begin{array}{c}\Pi'_w\\ A\cdot w\\ \Pi_w\\ \bot\end{array}$$

and so, it normalises away to *exactly* what we would expect, no more and no less than the translation of the classical proof that eliminated the unlabelled detour. I take it that this is *some* evidence that we have a faithful representation of classical proof inside the more flexible, more general and more detailed world of labelled incompatibility proofs.

REFERENCES

[1] FRANCESCO BERTO AND GREG RESTALL. "Negation on the Australian Plan". *Journal of Philosophical Logic*, 48(6):1119–1144, 2019.

[2] ROBERT B. BRANDOM. *Articulating Reasons: an introduction to inferentialism*. Harvard University Press, 2000.

[3] MARTIN DAVIES AND LLOYD HUMBERSTONE. "Two notions of necessity". *Philosophical Studies*, 38(1):1–30, 1980.

[4] J. MICHAEL DUNN. "Star and Perp: Two Treatments of Negation". In JAMES E. TOMBERLIN, editor, *Philosophical Perspectives*, volume 7, pages 331–357. Ridgeview Publishing Company, Atascadero, California, 1994.

[5] GILBERT HARMAN. *Change In View: Principles of Reasoning*. Bradford Books. MIT Press, 1986.

[6] LUCA INCURVATI AND JULIAN J. SCHLÖDER. "Weak Rejection". *Australasian Journal of Philosophy*, 95(4):741–760, 2017.

[7] LUCA INCURVATI AND PETER SMITH. "Rejection and valuations". *Analysis*, 70(1):3–10, 2010.

[8] NILS KÜRBIS. *Proof and Falsity*. Cambridge University Press, Apr 2019.

[9] MARK LANCE AND W. HEATH WHITE. "Stereoscopic Vision: Persons, Freedom, and Two Spaces of Material Inference". *Philosophers' Imprint*, 7(4):1–21, 2007.

[10] CATARINA DUTILH NOVAES. *The Dialogical Roots of Deduction*. Cambridge University Press, 2021.

[11] HUW PRICE. "Why 'Not'?". *Mind*, 99(394):222–238, 1990.

[12] GREG RESTALL. "Negation in Relevant Logics: How I Stopped Worrying and Learned to Love the Routley Star". In DOV GABBAY AND HEINRICH WANSING, editors, *What is Negation?*, volume 13 of *Applied Logic Series*, pages 53–76. Kluwer Academic Publishers, 1999.

[13] GREG RESTALL. *An Introduction to Substructural Logics*. Routledge, 2000.

[14] GREG RESTALL. "Multiple Conclusions". In PETR HÁJEK, LUIS VALDÉS-VILLANUEVA, AND DAG WESTERSTÅHL, editors, *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*, pages 189–205. KCL Publications, 2005.

[15] GREG RESTALL. "Truth Values and Proof Theory". *Studia Logica*, 92(2):241–264, 2009.

[16] GREG RESTALL. "A Cut-Free Sequent System for Two-Dimensional Modal Logic, and why it matters". *Annals of Pure and Applied Logic*, 163(11):1611–1623, 2012.

[17] GREG RESTALL. "Speech Acts & the Quest for a Natural Account of Classcial Proof". Article in progress. Draft available at https://consequently.org/writing/speech-acts-for-classical-proofs/, 2021.

[18] DAVID RIPLEY. "Bilateralism, coherence, warrant". In FRIEDERIKE MOLTMANN AND MARK TEXTOR, editors, *Act-Based Conceptions of Propositional Content*, pages 307–324. Oxford University Press, 2017.

[19] RICHARD ROUTLEY AND VALERIE ROUTLEY. "Semantics of First Degree Entailment". *Noûs*, 6(4):335–359, 1972.

[20] IAN RUMFITT. ""Yes" and "No"". *Mind*, 109(436):781–823, 2000.

[21] TIMOTHY SMILEY. "Rejection". *Analysis*, 56:1–9, 1996.