Negation on the Australian Plan

Francesco Berto[1,2] ⃝ · **Greg Restall**[3]

Abstract

We present and defend the Australian Plan semantics for negation. This is a comprehensive account, suitable for a variety of different logics. It is based on two ideas. The first is that negation is an exclusion-expressing device: we utter negations to express incompatibilities. The second is that, because incompat*ibility* is modal, negation is a modal operator as well. It can, then, be modelled as a quantifier over points in frames, restricted by accessibility relations representing compatibilities and incompatibilities between such points. We defuse a number of objections to this Plan, raised by supporters of the American Plan for negation, in which negation is handled via a many-valued semantics. We show that the Australian Plan has substantial advantages over the American Plan.

✉ Francesco Berto
   fb96@st-andrews.ac.uk; F.Berto@uva.nl

   Greg Restall
   restall@unimelb.edu.au

1   Department of Philosophy, University of St Andrews, St Andrews, UK

2   Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Amsterdam,
    The Netherlands

3   Department of Philosophy, University of Melbourne, Melbourne, Australia

> *It has occurred to several authors ...*
> *that we might refute $A \wedge \neg A \to B$*
> *by making $A \wedge \neg A$ true but B false.*
> *The trick is to make each of A, $\neg A$*
> *true. There is an American plan for*
> *doing this, on which A may be*
> *viewed as both true and false ... But*
> *we have come to praise the*
> *contrasting Australian Plan.*

R.K. Meyer & E.P. Martin, *Logic on*
the Australian Plan

1 Introduction

The last third of the Twentieth century saw a flowering of work in non-classical logics: the study of relevant logics, paraconsistent logics, orthologic, constructive logics, fuzzy logics, substructural logics and their cousins gave rise to a plethora of different kinds of models for such logics. One point of distinction in these models is the treatment of *negation*.[1] After the initial flurry of proposals for different ways to understand the logical connectives and, particularly, negation, the time came to survey the field. Robert K. Meyer and Errol P. Martin made a proposal for how to understand this landscape in their important, but relatively neglected paper, *Logic on the Australian Plan* [28].[2]

Given model-theoretic treatments that evaluate formulas at points (worlds, constructions, states, situations, setups, or whatever), Meyer and Martin distinguished those which take the relationship between formulas and points to be two-valued (for each point x and formula A, either $x \Vdash A$ or $x \nVdash A$), and those for which the relationship is fundamentally more complicated. In Meyer and Martin's sights were those models in which there are four possible semantic values: a formula can be *true*, *false*, *both* or *neither*. Two-valued semantics follow the *Australian Plan*. Four-valued semantics follow the *American Plan*.

Meyer and Martin's paper was directed towards understanding the costs and benefits of different semantic schemes for relevant logics. But the point is more general than this. For example, a traditional Kripke or Beth semantics for intuitionist logic [23] is also a kind of Australian Plan semantics, while Wansing's semantics for constructible negation [48] is a kind of American Plan semantics. The distinction applies very generally, to a sweep of different logics.

Here's why this is a salient distinction: in a range of non-classical logics, negation is not *Boolean*. In point semantics for relevant and paraconsistent logics, the argument from A, $\neg A$ to B is made to fail by allowing A and $\neg A$ to both hold at some

---

[1]For a general, but thorough introduction, see [47].

[2]A *Google Scholar* search finds only 39 citations for this paper as of late September, 2018. Thirteen of these citations are by the authors of *this* paper.

point. In paracomplete logics (allowing for 'gaps'), the argument from A to $B \vee \neg B$ is given a counterexample by allowing for B and $\neg B$ to both fail at some point. How, then, does the status of a negation $\neg C$ at a point depend on the status of C?

Following the Australian Plan requires the semantic value of a negation, $\neg C$, at a point x to depend on more than just the value of C at that point. If the semantics is to be compositional, it seems that negation must act like a modal: whether $\neg C$ holds *here* depends on whether C holds *elsewhere*, in the same way that whether $\Box C$ holds in *this* possible world depends on whether C holds in other (*relatively possible*) worlds. Simplicity for semantic values (the Boolean yes/no answer, at each point) comes at the cost of complexity for the evaluation clause for negation.

In the American Plan, negation can have a relatively simple interpretation: $\neg A$ is true (that is, *true only* or *both true and false*) if and only if A is false (that is, *false only* or *both true and false*), and $\neg A$ is false if and only if A is true. Simplicity in the clauses for negation is bought at the cost of making semantic values more complex. Having four values is only a little more than the two Boolean truth values. But the shift to four values ramifies throughout the entire semantics: each time one introduces a new concept or operator (necessity, a relevant or counterfactual conditional, a non-standard quantifier), this must now be given independent truth as well as falsity conditions. Things get cumbersome for the four-valued approach, for instance, when a relevant conditional is introduced: see e.g. [31], [27, p. 89]. Under the Australian plan, no such complexity arises. Once the truth conditions for a concept are given on the set of points, this automatically determines the interaction between that concept and negation. All of this has been well understood since Meyer and Martin's original mapping of the terrain. However, much has changed in the decades since. The time has come to revisit some issues.

Starting in the 1990s, the Australian Plan has been generalized into a comprehensive approach to negation. Attention has shifted from Meyer and Martin's treatment of a de Morgan negation, modelled by the distinctive semantic device of the Routley star [42]. Negation is understood as a modal operator, whose semantics is given in terms of relations of *compatibility* and *incompatibility* between points [15]. It has been connected to broader themes in philosophical logic, such as logical pluralism [4, 5, 8] and the interpretation and applications of substructural logics [33, 34]. Furthermore, claims to the fundamental role of compatibility and incompatibility relations in giving an account of negation, also independently of direct appeal to the model theory of the Australian Plan, have been made in Aristotelian metaphysics [6, 7, 44] and in normative pragmatic accounts of semantics [10, 11, 22, 29, 36].

However, the Australian Plan has recently come under attack by proponents of a version of the American Plan, De and Omori [12]. In this paper we will update the Australian plan in the light of recent developments in logic. We will examine De and Omori's criticisms,[3] show that the Australian Plan—not only the Routley star

---

[3]Henceforth we use 'D&O' for De and Omori, 'FB' and 'GR' for the authors of this paper when we need to be referred to independently.

semantics for negation, but the entire approach of treating negation as a point shift operator—survives unscathed, that its advantages over the American Plan remain intact, and substantial.[4]

2 Grounding Negation

Critics of the Australian Plan tend to run together different matters. So it falls on us to distinguish them. There are two key ideas in the (generalized) Australian Plan's (in)compatibility semantics. **Idea 1**: We utter claims of the form '¬A' in order to *rule out* something. That is: to express a certain *exclusion*, or *incompatibility*.

This may be further specified in different ways. Incompatibility may be understood (*a*) in a normatively pragmatic fashion, as signaling that the utterer takes a certain stance and commits to it. This reading is favored by authors who have entertained a characterization of negation in terms of incompatible pragmatic and inferential commitments, such as Robert Brandom [10, 11], or Huw Price, to whose work [29] we will come back soon. Such a normative pragmatic account can also be understood in terms of the primitive incompatibility of the speech acts of assertion and denial, combined with an account of negation as a means to make explicit what is implicit in the incompatibility of assertion and denial [36–38].

Incompatibility may also be specified (*b*) in more realist, metaphysically committing ways, as expressing that some *thing* is the case (obtains, is instantiated, realized, or whatnot), which rules out something else in the world. Here 'ruling out' is, thus, understood as a metaphysical relation between worldly items: properties, states of affairs, circumstances, or whatever else. This is the reading favoured by FB, but also by other authors, such as [21]. Being an Australian, GR is also comfortable with such robust metaphysical vocabulary; but in this paper we will be largely agnostic between (*a*) the normatively pragmatic and (*b*) the metaphysical specifications of incompatibility.

Either way, negation is understood as an *exclusion-expressing device*: its existence in the language (indeed, in any natural language we know of) is explained by grounding it in notions involving compatibility and incompatibility or exclusion. A first, legitimate question semanticists following this path are liable to be asked, therefore, is: What does 'grounding' mean here?

Some remarks by D&O indicate that they take the salient sense of grounding to be reduction via definition. They are troubled, then, by the fact that negative particles or prefixes show up in the names of various concepts proposed to do the grounding work, as this would make the attempted definition circular:

---

[4]A Referee of this journal kindly reminded us to clarify the narrower and wider reading of 'Australian plan'. We focus in this paper on the *more general* approach to negation; not only because we take modal approaches to negation to be important and interesting, but also because D&O's arguments apply not just to the Routley star semantics in particular, but also to compatibility semantics in general. As we show below and one of us, GR, has shown in such works as [33], the Routley star semantics is a special case of the general framework.

In explaining incompatibility, using expressions like 'exclude', 'preclude', and 'rule out' [...] does not suggest that incompatibility is primitive. It is after all possible, for instance, that the prefixes 'ex-' and 'pre-' here signal the use of negation, as does the 'out' in 'ruling out', or that the expressions in any case have meanings or truth conditions that depend on negation whether or not those expressions contain subexpressions signaling the use of negation. [...] It is important to notice that if incompatibility is defined from compatibility and negation, (S¬) [scil. the targeted semantic clause for negation] becomes circular since the definiendum occurs in the definiens. The version of (S¬) given in terms of *in*compatibility (got by contraposing and removing double negations)[5] would remain circular on the assumption that incompatibility is understood in terms of negation and compatibility. [12, p. 5 and fn, notation adjusted for consistency with ours.]

Now we agree, of course, that it would be bad if one were to define negation by using notions which are themselves defined using negation.

However, the Australian Plan is no attempt to define negation away by reducing it to some other notion which makes no mention of negation. One of us, FB, has expressed, in print and in the very paper [8] targeted by D&O, skepticism (on which GR agrees) on any attempted definitional reduction of fundamental notions like *reference, identity, necessity*, or *negation*. Any elucidation of such notions is likely to make essential use of those very notions somewhere, in such a way that the explanation as a whole cannot count as a reductive definition. A few examples of the pervasiveness of the phenomenon (of course, we don't claim that these settle the respective issues):

- In *Naming and Necessity*, Kripke claimed (or, went very close to claiming) that the notion of *reference* is primitive: 'Philosophical analyses of some concepts like reference, in completely different terms which make no mention of reference, are very apt to fail.' [24, p. 94].
- Wiggins [49] famously argued that the concept of *identity* is primitive and co-originary with predication: a is F iff a is some F, that is, Fa iff $\exists x(x = a \And Fx)$.[6]
- Kit Fine stated of the notion of *reality* that, while 'we seem to have a good intuitive grasp of the concept', he does 'not see any way to define the concept of reality in essentially different terms' [18, p. 175].
- Many take the concept of *set* as a candidate primitive. We give examples, and we elucidate it by saying that a set is a collection or an aggregate of objects, but that is no definition of the notion *set* in set-free vocabulary.

---

[5]We will come back below to the version of (S¬) using compatibility, and its contraposed, using incompatibility, after we have presented a formal semantics to serve as the target for some of D&O's objections.

[6]'No reduction of the identity relation has ever succeeded. [...] Nor yet is it called for, once we realize how much can be achieved in philosophy by means of elucidations that put concepts to use without attempting to reduce it but, in using the concept, exhibit its connections with other concepts that are established, genuinely coeval or collateral, and independently intelligible.' [49, p. 5].

- Examine the clauses for *conjunction*, *disjunction*, or the *quantifiers* in your favorite model-theoretic or proof-theoretic semantics. You will find that these use the very notions that are being explicated. Nonetheless, the clauses are informative.

Australian Plan semanticists have a similar attitude towards the notion of negation and its relationship to incompatibility. Negation is as basic to our grasp of language as these other concepts are to our grasp of mathematics, logic, language, and the world at large. We should not expect explications of such notions to be reductions.

How, then, is negation grounded in compatibility and incompatibility, since 'grounded' is not to mean 'reduced by definition to'? The question making for the title of [29], which inspired FB, was: 'Why "Not"?'. That is: *Why* do we possess the concept of negation? (Thus, the question wasn't, 'What "Not"?', understood as a request for a reductive definition.) Price's opening words:

> This paper addresses some questions about negation. What is negation good for? What is its linguistic function? How might it plausibly have developed in natural language, and what if anything does this tell us about its properties? The project is thus to *explain* the existence and nature of negation in ordinary language. [29, p. 221]

And the answer given by Price was a psychological-genetic account of how creatures capable of perceiving the world around themselves, and willing to pass on to their peers information about such shared world, may benefit by having an exclusion-expressing device in their communicative toolkit. D&O claim to be 'not much concerned with the psychological account' [12, p. 5]. But this is essential to the whole story, so *we* are concerned with it.

Price asks us to imagine a community, the Ideological Positivists, whose language is devoid of negation. Here's how a conversation between me and you as members of such community may go:

> *Me*: 'Fred is in the kitchen.' (Sets off for kitchen.)
>
> *You*: 'Wait! Fred is in the garden.'
>
> *Me*: 'I see. But he is in the kitchen, so I'll go there.' (Sets off.)
>
> *You*: 'You lack understanding. The kitchen is Fred-free.'
>
> *Me*: 'Is it really? But Fred's in it, and that's the important thing. (Leaves for kitchen.).

> Your problem is to get me to appreciate that your claims are incompatible with mine. [29, p. 224]

What would make things easier is a 'Lo, Fred is *not* in the kitchen'. That is: 'Fred is somewhere else – in the garden – and his being there *rules out* his being in the kitchen'.

Price then asks, 'Where might a sense of incompatibility first arise?' [29, p. 226]. Answer: it comes from our experience of the world as agents that face choices

between performing some action or other – something we think animals as well deal with every day. To face a choice is to experience an incompatibility: one cannot have it both ways.

One could further specify **Idea 1** in normatively pragmatic terms as per option (*a*), flagged above. The clash of incompatibility is first expressed when we rule options *in* or *out*, whether to ourselves or in dialog. When I consider options in planning by myself, or in arguing with someone else, I rule an option in by making an assertion, and I rule it out by denying it. To say 'yes' to the claim that Fred is in the kitchen is to say 'no' to the claim that he is in the garden. We take *pro* and *con* attitudes to options before we have an embeddable, composable negation. The incompatibility between acts of assertion and denial – that we take them to clash – is prior to any incompatibility expressed in assertions of negations. Assertions of negations then make explicit what is implicit in the clash between assertion and denial. Such is the direction of explanation between incompatibility and negation favoured in GR's works.

One could go for a more metaphysically committing route, as per option (*b*) above, favoured by FB. One may track our experience of incompatibility to our basic capacity of locating objects in space-time (*this* – located entirely here, with such and such boundaries – cannot be *that* – which is just down there); or of appreciating perceptual incompatibilities concerning colours and dimensions (this is blue all over, which rules out its being red all over; this is about one inch long, which rules its being six inches long).[7]

Either way, here's how incompatibility grounds negation: it *explains* why we have the concept of negation. It allows one to answer questions like: Why is negation *ubiquitous*? Why do we have negations in any natural language we know of? We have them, the answer goes, because things ruling out things make for a universal feature of experience, whether this boils down to the need to take exclusionary commitments, or it involves more metaphysical incompatibilities between worldly features. We need a linguistic device to express and report instances of such ubiquity.

It's no use, then, to complain that expressions like 'ex-' and 'in-' show up in the names of the relevant concepts. If the story is right, we have negation in our language in order to express exclusions. So it's no surprise that negation shows up in such expressions as 'incompatibility' and 'ruling out', which we use to talk about exclusions. On the contrary, that's precisely what one should expect.

So when D&O 'agree with Price [and, both of us] that there is a sense of incompatibility that is more primitive than sentential negation; surely very young children and animals can see when two states are incompatible *in some sense* before they grasp anything like sentential negation' [12, p. 5], they have already granted a good deal of what an Australian Plan semanticist asks for. And when they add that 'where we disagree with is that it is this very notion of incompatibility that grounds the truth

---

[7]Such examples come from the literature on the metaphysics of exclusion: [45] mentions phenomenological colour incompatibilities, concepts that express our categorization of physical objects in space and time, such as x being here right now and x being way over there right now, for a suitably small x. Other cases are provided by Grim [21] and Tahko [44].

conditions of negated sentences' (Ibid), that may be because they take the Australian Plan to aim at something it never aimed at to begin with.

Regardless, D&O express various other reasons for dissatisfaction with the Australian Plan. To deal with them, we need to make the Plan more precise. We begin in the next section.

3 Algebras, Frames, Points, Worlds

One who claims that negation is grounded, in the aforementioned sense, in compatibility and incompatibility, is liable to be asked to make one's views more precise by providing a formal treatment of these relations and their relata. Now 'Incompati*bility*' and 'compati*bility*' sound modal: to claim that x is compatible with y is to claim that x and y *can* coexist. (How to fine-tune 'coexist' depends, of course, on what x and y are – a point to which we shall return.) The key **Idea 2** of the Australian Plan is that, therefore, one can expect an account of negation grounding it in such notions to be a modal one.

This is no very specific commitment yet. In particular, it does not mandate resorting to (what we nowadays call) Kripke-style frame semantics, using points in frames and accessibility relations between them. The origins of incompatibility semantics can be traced back to the Birkoff–von Neumann–Goldblatt characterization of *ortho-negation* in quantum logic [9, 20]. This was based on frames comprising points and relations between them, but the points were narrowly taken as possible outcomes of experimental measurements, of the kind performed by quantum physicists. One such relation was labelled as 'orthogonality', also called 'perp' (say, '$\perp$'), expressing the idea that two outcomes are incompatible with one another: if a and b are possible outcomes, '$a \perp b$' means that outcome a rules out outcome b.

One could generalize the insight beyond quantum experiments in an algebraic (as opposed to Kripke-frame) setting. To get a feel of how this may go, take a quadruple $\langle S, \leq, \perp, \bigvee \rangle$, where S is a nonempty set of points (read: states); $\leq$ and $\perp$ are binary relations on S: $s \leq t$ (read: 's entails t') is a pre-order (reflexive and transitive); $s \perp t$ (read: 's rules out t') is our incompatibility relation between states; $\bigvee$ is a join operation defined for all subsets of S (so $\langle S, \bigvee \rangle$ is a complete join semilattice): given a set of states $T \subseteq S$, $\bigvee T$ is the (possibly infinitary) join of all items in T.

A state may have one or more incompatible peers, thus, e.g., Grim [21] talks about 'exclusionary classes': for a given s, let $E_s = \{t \mid s \perp t\}$ be the set of states s is incompatible with. We could then characterize negation (not-s: 'It is not the case that s') in our algebra just as $\bigvee E_s$. If the exclusionary class for s is finite, not-s is like a plain join of states $s_1 \vee \cdots \vee s_n \in E_s$. If on the other hand we have an infinity of incompatibles, not-s may be recaptured by quantifiying over states in S:

- not-$s = \exists t\,(t\ \&\ t \perp s)$

'There's some state t, such that: t, and t rules s out'. Either way, we make sense of not-s as the weakest incompatible via the following equivalence:

- $x \leq$ not-s iff $x \perp s$

The left-to-right direction says that whatever entails not-s is incompatible with s. The right-to-left direction says that not-s is the weakest incompatible, i.e., that which is entailed by any incompatible x. Plugging in not-s for x and detaching, we get:

- not-$s \curlywedge s$

that is, not-s rules s out. Talk of 'weakest incompatible' connects to how asserting that one exclusionary state is the case ('Our new sofa is red', ruling out that it's blue) generally sounds more informative than asserting the entailed negation ('Our new sofa is not blue', i.e., it has some color or other incompatible with blue: yellow, white, orange...).

We shall add no more to this sketchy presentation of the algebraic approach because, as a matter of historical fact, negation on the Australian Plan has not been phrased algebraically, but rather in terms of frame semantics generalising the standard Kripke semantics for modal logics. Semantics of this kind typically consist of structures including points at which formulas are evaluated, and a relation between these points, understood as compatibility. Negation is a quantifier over such points, restricted via that relation from the perspective of a given point. Thus, negation is a modal in the following, plain sense: the status of a negated sentence at a point may depend on the status of the sentence at other points. It is a framework of this kind (see, for example, [26, 27, 33, 34, 39, 40]) that is targeted by D&O when they claim that 'the modal account of negation is implausible for providing an explanation as to when a negated sentence is true.' [12, p. 3].

The points of evaluation in such frame semantics, as a matter of fact, have oftentimes been labeled as 'worlds'. This may be misleading, but it is so, we submit, in no more dangerous a way than when points in the frame semantics of various nonclassical logics are so labeled. One of us (GR) has talked of 'cases' in various works, e.g., [5]. When we speak of worlds, or cases, in the Kripke semantics for intuitionistic logic, or in the Routley-Meyer semantics for relevant logics, or in the frame semantics for various substructural logics (linear logics, Lambek calculus, or else), the crucial question is how those points ought to be interpreted, that is, what the mathematical formalisms are to represent. We will get back to this. Let us first see how a simple Kripke-style frame semantics in the tradition of the Australian Plan could go.

We have a sentential language $\mathcal{L}$ with a set $\mathcal{L}_{\mathcal{AT}}$ of atoms p, q, r ($p_1, p_2, \ldots$), the binary connectives $\wedge$ and $\vee$, the unary connectives $\square$ and $\neg$, the 0-ary connectives $\top$ and $\bot$, round brackets as auxiliary symbols. We use A, B, C, ($A_1, A_2, \ldots$) as metavariables for formulas. The well-formed formulas are the items in $\mathcal{L}_{\mathcal{AT}}$, $\top$ and $\bot$ and, if A and B are formulas, $(A \wedge B)$, $(A \vee B)$, $\square A$, $\neg A$ (outermost brackets are normally omitted).

A *frame* for $\mathcal{L}$ is a quadruple $\mathfrak{F} = \langle W, P, C, \sqsubseteq \rangle$, where W is a nonempty set, $P, C \subseteq W \times W$, $\sqsubseteq$ is a partial ordering on W. We use a, b, c, x, y, z ($x_1, x_2, \ldots$) in the metalanguage as variables ranging on items in W, as well as the set-theoretic notation and the symbols $\forall, \exists, \Rightarrow, \Leftrightarrow, \&, or$, with the usual reading.

The official reading of the frame is: W is a set of worlds (we will not speak of *possible* worlds, for reasons that will become clear soon). P and C are two accessibility relations on worlds. When $\langle x, y \rangle \in P$ we write this as xPy and claim that y is *possible* relative to x. When $\langle x, y \rangle \in C$ we write this as xCy and claim that

x is *compatible* with *y*. $\sqsubseteq$ is to be thought of as an information ordering, as in the Kripke semantics for intuitionistic logic [23]. Here, worlds are understood as representing the epistemic states of the idealized mathematician (the 'creative subject' of Brouwer's). More generally, the official reading of '$x \sqsubseteq y$' is that world *y* retains at least all the information in world *x*.

Now unless $\sqsubseteq$ boils down to identity, the points in our frames cannot be taken as the ordinary, maximally consistent possible worlds of standard modal semantics. It doesn't make sense to claim that one *such* world can properly include the information carried by another *such* world. These worlds are maximally informative: the one way for *y* to retain at least all the information in *x* is for *y* to *be x*.

On the other hand, Barwise and Perry's situation semantics [1, 2, 32] already showed the usefulness of *situations*, taken as partial states of reality; and for such partial items, non-trivial informational inclusion is not only natural: it is an essential aspect of what it is for situations to be partial states. To use a familiar example: the situation consisting of GR's living room in Melbourne does not carry information about the weather in Sydney, whereas the situation in Australia as a whole does carry that information; and the latter situation properly includes the former.

In standard possible worlds semantics, the proposition expressed by formula *A* is taken as $|A| \in \mathcal{P}(W)$, the set of worlds where *A* is true. But when points in frames can stand in non-trivial information-inclusion relations, one should take the set of propositions in a frame $\mathfrak{F}$ — call that set $Prop(\mathfrak{F})$, a subset of the power set of worlds $\mathcal{P}(W)$ — as including only sets closed upwards with respect to $\sqsubseteq$: $X \in Prop(\mathfrak{F})$ only if $x \in X \ \& \ x \sqsubseteq y \Rightarrow y \in X$ [33, 35].

A frame becomes a *model* $\mathfrak{M} = \langle W, P, C, \sqsubseteq, \Vdash \rangle$ when one adds an interpretation, $\Vdash$, relating worlds to formulas: we write '$x \Vdash A$' to mean that *A* holds at *x*, '$x \nVdash A$' to mean that *A* fails at *x*. We will only deal with *admissible* interpretations where, for each $p \in \mathcal{L}_{\mathcal{AT}}$, $|p| = \{x \in W \mid x \Vdash p\} \in Prop(\mathfrak{F})$, satisfying the so-called Heredity Constraint on atoms [14, 30]. For each $x, y \in W$:

- (HC) $x \Vdash p \ \& \ x \sqsubseteq y \Rightarrow y \Vdash p$

The HC makes obvious sense: if all the information in *x* is retained in *y*, and *p* holds at *x*, then *p* must also hold at *y*. HC generalizes to all formulas of $\mathcal{L}$ once the semantic clauses for the connectives are given. These go as follows. For all $x \in W$:

- (S$\wedge$) $x \Vdash A \wedge B \Leftrightarrow x \Vdash A \ \& \ x \Vdash B$
- (S$\vee$) $x \Vdash A \vee B \Leftrightarrow x \Vdash A \ or \ x \Vdash B$
- (S$\top$) $x \Vdash \top$
- (S$\perp$) $x \nVdash \perp$
- (S$\square$) $x \Vdash \square A \Leftrightarrow \forall y(xPy \Rightarrow y \Vdash A)$
- (S$\neg$) $x \Vdash \neg A \Leftrightarrow \forall y(xCy \Rightarrow y \nVdash A)$

D&O claim that they 'have no use' [12, p. 3, fn] for our penultimate item, the positive modal of necessity with its usual accessibility relation, *P*. We do. Highlighting a number of obvious connections and dualities between it and our last item, the negative modal with its compatibility relation, *C*, will help understanding.

Here's a first connection. The following conditions on all x, y, x_1, $y_1 \in W$ make $\sqsubseteq$ interact properly with relative possibility and compatibility:

- (*Forwards*) xPy & $x_1 \sqsubseteq x$ & $y \sqsubseteq y_1 \Rightarrow x_1 P y_1$
- (*Backwards*) xCy & $x_1 \sqsubseteq x$ & $y_1 \sqsubseteq y \Rightarrow x_1 C y_1$

Forwards is just a familiar condition on positive modalities from normal modal logics. *Backwards* is found in a number of works on negation as a modal [13; 15, see, for example, 16]. Technically, they allow the Heredity Constraint to generalize by straightforward induction to all formulas of $\mathcal{L}$: for each A and for all $x, y \in W$: $x \Vdash A$ & $x \sqsubseteq y \Rightarrow y \Vdash A$. Also, $|A| = \{x \in W \mid x \Vdash A\} \in Prop(\mathfrak{F})$. Intuitively, both make a lot of sense:

- *Forwards*: if xPy, that is, y is possible relative to x, then everything necessary at x holds at y: this is just what the clause for necessity (S□) says. Then if $x_1 \sqsubseteq x$, whatever is necessary at x must already be such at x_1, because the former preserves the information supported by the latter. And if $y \sqsubseteq y_1$, then anything holding at y must hold at y_1 for the same reason. Then whatever is necessary at x_1 holds at y_1, therefore y_1 is possible relative to x_1, $x_1 P y_1$.
- *Backwards*: if xCy, that is, x is compatible with y, then nothing ruled out at x holds at y: this is just what the clause for negation (S¬) says. Then if $x_1 \sqsubseteq x$, anything ruled out at x must already be such at x_1, because the former preserves the information supported by the latter. And if $y_1 \sqsubseteq y$, then anything ruled out at y must be ruled out at y_1 for the same reason. Then nothing ruled out at x_1 holds at y_1, therefore x_1 is compatible with y_1, $x_1 C y_1$.

Finally, we define logical consequence in a frame $\mathfrak{F}$ as truth preservation at all points x in $\mathfrak{F}$ in all admissible interpretations (that is, in all the relevant models based on the frame). Given a set Σ of formulas:

- $\Sigma \vDash B \Leftrightarrow$ For all models $\mathfrak{M}$ on $\mathfrak{F}$: $x \Vdash A$ for all $A \in \Sigma \Rightarrow x \Vdash B$

(For single-premiss entailment we write $A \vDash B$ for $\{A\} \vDash B$.)

Now that we have a frame semantics, we focus on objections by D&O that refer specifically to *this* set-up.

4 Looking at Worlds

One first objection involves a comparison between (S∧) and (S¬). D&O consider the following possible criticism of their own position: if the compatibility clause for negation is problematic, then the standard one for conjunction is as well, because 'both give the truth conditions for an object-language connective in terms of the "same" corresponding meta-language connective' [12, p. 6]. But (S∧), of course, is not problematic; so (S¬) isn't either. D&O's reply:

> The difference between (S¬) and (S∧), however, is that the latter is intended to provide mere truth conditions for object-language sentences and a *definition* of conjunction that would "ground the origins of our concept" and its usage in natural

language. This is why homophonic truth conditions will do for (S∧) but not for (S¬), and is also why (S¬) is problematic as a grounding definition if negation cannot be eliminated from the right-hand side of the biconditional. (p. 6)

There are a number of things to say in response to this passage. To begin with, it gives more evidence that D&O believe the Australian Plan to be after a reduction by definition of negation. It isn't.

Next, even the sense in which the homophonic clause for conjunction gives a definition of '∧' is controversial. *If* one accepts a truth-conditional account of meaning to begin with, or claims that truth conditions are at least part of what makes for the meaning of an expression (which is controversial anyway, as testified, e.g., by competing inferentialist accounts), then of course the semantic clauses for the connectives must tell something about their meaning.

However, as Tarski taught us, the definition one is after when one gives recursive semantic clauses for a formal language like our $\mathcal{L}$ above is, rather, the one of *truth in $\mathcal{L}$*. As the official Tarskian wisdom has it, the definition is materially adequate when we can infer from it the various instances of the T-schema for formulas of $\mathcal{L}$. Homophonic clauses, on the other hand, presuppose some understanding of the meaning of the connective used in the metalanguage: see e.g. [46]. One who lacks the concept of conjunction will not come to understand it by looking at the truth conditions for conjunctive formulas given in (S∧). The account assumes that we have some grasp of conjunction (as we, of course, do). The same goes for negation and (S¬).

For a final remark this ballpark: as mentioned by D&O in another passage quoted two sections above, one can twist the clause for (S¬) removing the metalinguistic, sentential negation in its right hand side and phrasing the truth conditions using incompatibility (whose name of course includes 'in-', etc. etc.). That's just another way this kind of semantics is presented in the literature:

- $(S^1\neg)\ x \Vdash \neg A \Leftrightarrow \forall y(y \Vdash A \Rightarrow xIy)$

The negation of A is true at point x iff any point making A true is incompatible with x. This is the form used, for instance, in [16].

$(S^1\neg)$ is useful in the debate around the Australian Plan: it helps understanding by packing the controversial bit in the incompatibility relation, I, and allowing to exploit, again, the obvious duality with the box. Compare $(S^1\neg)$ with (S□) and ask yourself: is the latter a definition of the box, reducing it to notions that do not involve necessity and possibility? Of course not: the accessibility relation on its right hand side stands for (relative) *possibility*. If one does not have some grasp of what possibility is to begin with, one will not come to understand boxes and diamonds by being shown their truth conditions in biconditionals that involve *y's being possible relative to x*. Does that disqualify clauses like (S□) from having a valuable role in an explanatory account of the concepts of necessity and possibility? Of course not: all of contemporary modal logic with Kripke semantics testifies to the usefulness of Kripke semantics in the analysis of necessity. Though it is not a reduction of the modal to the non-modal, it is an explication of the truth conditions of modal claims in terms of

a particular kind of modal claim—relative possibility between worlds. The same sort of conceptual work is done in the Australian Plan, for $(S^1\neg)$.

D&O also claim, however, that incompatibility does not explain what it is for a negation to be true at a point, *even if* one grants for the sake of the argument that it be a primitive notion:

> It's true that Sam is not a gram heavier than she actually is, even though she easily could have been. And since she could easily have been a gram heavier, there are worlds where she is that are very similar to our own. Indeed, these worlds seem compatible with ours, if we are going by our intuitive notion of compatibility. And yet, on the modal account of negation, all the worlds compatible with ours are ones where Sam is not a gram heavier than she actually is, no matter how similar they are to ours. Why are all the compatible worlds like this? To emphasize, our intuitive understanding of (in)compatibility does not tell us that these worlds are incompatible with ours. If there is any kind of explanation as to why these worlds should be incompatible with ours, we can only see that it must ultimately appeal to negation. Worlds where it is true that we are a gram heavier than we actually are are incompatible with our own because they make true *the negation* of a sentence that is here true. [12, p. 6]

Now of course the 'explanation as to why these worlds should be incompatible to ours', in the Australian Plan, runs the other way around with respect to the one proposed by D&O. Given that Sam weighs n grams at point x, a point y can be incompatible with x by having Sam weigh $n + 1$ grams: one cannot simultaneously have two different weights. Point x, then, makes 'It is not the case that Sam weighs $n+1$ grams' true by ruling out that Sam is $n+1$ grams, that is, by being incompatible with any point making 'Sam is $n+1$ grams' true, as mandated by the Plan's semantics.

Now why would this go against D&O's 'intuitive notion' or 'understanding' of (in)compatibility? The one way D&O try to provide some content for this alleged intuition, is by invoking the intuitive similarity between incompatible scenarios. Compatibility and similarity are, however, independent from each other. Let l be the state of affairs consisting of this chair's being light blue all over; d, the state of affairs consisting of this chair's being a darker blue all over; r, the state of affairs consisting of that table's being red all over. Given some relevant similarity metric, l is more similar to d than r is: l and d involve the same object, r, a different one; l and d involve two shades of color closer on the color scale than l's is to r's. Still, l and d are incompatible: the chair cannot be simultaneously lighter and darker all over. Instead, l and r are compatible: a chair's being light blue does not rule out a table's being red.

Unsurprisingly, then: in a Kripke-style frame semantics different relations are used in the semantic clauses for different modal operators. Comparative similarity between points is likely to show up in the clause for a variably strict conditional, as in the standard semantics for counterfactuals of Stalnaker [43] and Lewis [25]; relative possibility, in the clause for the box of positive modality; and, if the Australian Plan is right, compatibility in the clause for negation.

It is no good, thus, to plainly invoke similarity in an attempt to establish or undermine claims of incompatibility. One needs to back up such an invocation with

arguments, on a case by case basis. But D&O give no argument, aside from invoking intuitions which, we claim, have counterexamples.

Besides contesting the truth conditions for negation proposed in the Plan's semantics, D&O deem such semantics 'implausible as an account of how we process and understand negation' (p. 3). Here's why:

It does not seem that when we go about determining whether a negation $\neg A$ is true, we think about all the A-worlds there are (non-recursively enumerably many!) and then we see somehow and all at once that each is incompatible with our own. So it does not seem that we ever understood negation in terms of a primitive notion of compatibility between worlds that would have "grounded the origins of our concept". (pp. 6-7, notation adjusted for consistency with ours.)

But if, as per the Australian Plan's **Idea 2**, negation is a modal (which cannot be ruled out beforehand, on pain of a *petitio*), that's exactly *not* the way we go about determining whether a negation is true. If Kripke-style frame semantics were committed to this being the way we evaluated modal claims, *no* broadly Kripkean or worlds semantics for *any* modality could ever be right. This is too much to swallow. The obvious duality with relative possibility makes this plain. Think about someone objecting to the standard Kripke semantics for the box along the line D&O pursue:

It does not seem that when we go about determining whether a modality $\Box A$ is true, we think about all the A-worlds there are (non-recursively enumerably many!) and then we see somehow and all at once that each is possible relative to our own. So it does not seem that we ever understood necessity in terms of a primitive notion of relative possibility between worlds that would have "grounded the origins of our concept".

This gets the entire Kripkean story wrong. Whether the points in frames endowed with accessibility relations represent classical (maximally consistent) possible worlds, or intuitionistic constructions as in the Kripke semantics for intuitionistic logic, or situations from situation semantics, or something else, these are the semanticist's tools. Lay people have for the most part never heard of Kripkean frames or worlds semantics. Of course, they use 'necessarily', 'if it were the case that . . . , then', and 'not', generally competently. But this is no objection to a frame-theoretic treatment of such items, à la Kripke. Kripke himself, and the other logicians working in the same frame-theoretic tradition, introduced points and accessibilities to provide a semantic analysis (not perforce, as we have seen, a reductive definition) of various modal notions. No semantics of this kind is committed to the additional psychological claim that, when we go about determining whether a sentence involving a modal like 'necessarily', or 'if it were the case that . . . , then it would be the case that', interpreted as a (variably or constantly) strict universal quantifier over points in frames, we need to run through a plurality, or perhaps an infinity, of worlds in our head (all the relatively possible ones, or the closest accessible ones) before deciding whether

the sentence is true. To expect that the model theoretic semantics describes some process of cognition is to ask it to perform tasks beyond its remit.[8]

5 Logical Pluralism?

One who claims, following **Idea 1**, that negation is grounded, in the sense clarified above, in compatibility, and who provides a frame semantics like the one given above to account for **Idea 2** (negation is a modal), is liable to be asked further questions concerning the logical features of negation. By answering such questions, one is liable to incur in certain commitments, which could be criticized. But one may also refuse to take a stance on specific questions; which is likely to prompt the further issue whether the professed neutrality is coherent with the proposed approach to negation.

Here, too, one must be careful to distinguish several issues. A first one is connected to *logical pluralism*. FB's paper, [8], targeted by D&O, entertains a certain pluralism for negation, connected to the logical pluralism explored by GR in various works. The connection is the following. Beall and Restall [4, 5] have proposed a logical pluralism centered on a model-theoretic characterization of the notion of logical consequence, called Generalized Tarski's Thesis:

- (GTT) An argument is valid$_x$ if and only if for every case$_x$ in which the premises hold, so does the conclusion.

The key thought is that 'case' is ambiguous here – hence the subscript x – and can be made precise in different ways, resulting in different notions of validity – hence the subscript on 'valid'. Not all ways of making 'case' precise are admissible, but more than one is. Different admissible precisifications of 'case' originate different, equally legitimate notions of logical consequence. It is one thing for cases to be *worlds* in the traditional sense of maximally consistent ways things could be or have been; it is another for cases to be *situations*, allowing for incompletess, and possibly inconsistency. (FB spoke of 'worlds' rather than 'cases' but, as explained above, we take this as a merely terminological matter: in *that* sense, worlds, or cases, are points of evaluation in our frame semantics.) Analogously, that FB paper claimed, the notion of world in (S¬) can be made precise in different ways, which mandate different kinds of logical behaviour for the information-inclusion relation $\sqsubseteq$ of our semantics. For instance, if the points are understood as classical, maximally consistent worlds (and a pair of additional assumptions on compatibility are made), ¬ behaves classically (in particular, it satisfies Double Negation Elimination and Excluded Middle).

But even if logical pluralism (of this kind) is wrong, this is no objection to the Australian Plan as such. It might be that, contra Beall and Restall, there is only one sensible way of characterizing the notion of *case* showing up in GTT, and that this

---

[8]The analogy holds not only with necessity and other modal operators, but also the quantifiers. When determining whether $\forall x\, Fx$ is true, we need not think about *all* of the possible assignments of the values of the variable x in Fx. Regardless, the standard Tarskian semantics for the quantifiers has a role in explaining the logic of Peano Arithmetic, for example.

fixes logical consequence up to uniqueness. This is as such no problem for the two key ideas of the Plan on negation: first, that the meaning of negation is grounded in such concepts as compatibility and incompatibility; and second, that, since these are modal notions, negation is a modal as well. Nor would a hypothetical victory of logical monism be an objection to the way of making the semantics precise that resorts to frame semantics, taking negation as a restricted quantifier on points of evaluation.

It is a *still* different issue what the properties of the restriction should be. It is a claim made in FB's *Mind* paper, that whatever logical pluralism there is for negation should come from different ways of fine-tuning the notion of point (case, world), which entail different kinds of behaviour for $\sqsubseteq$; whereas any acceptable negation must comply with the features of the compatibility relation that grounds negation. Thus, if the restricting relation C in our semantics must have feature f, and C's having f makes a certain negation-involving inference valid, any negation-like operator proposed in the literature not validating that inference is no real negation. This brings no specific commitment yet, on what can make for the relevant f's. FB does take (few) commitments around this topic (we'll come to this below). But even if these are wrong, this is no objection to the Plan either. The features of compatibility may be different from what FB thinks they are, but this by itself brings no trouble to the Plan's key **Idea 1** and **Idea 2**.

However, D&O, take as a trouble for a semantics in this ballpark, that it delivers, all on its own, no clear verdict on the features of (in)incompatibility:

> We also do not find the arguments in favor of or against various constraints on incompatibility compelling enough to allow us to comfortably say that such and such are the Laws of Negation. If something is to qualify as an adequate account of negation, it should be fairly clear according to that account what the laws of negation are. (p. 3)

If 'clear' means 'determined once and for all' (we take it D&O don't mean that compatibility semantics is just obscure), we plainly deny the consequent of the last quoted claim. What the laws of negation are has been the subject of debate for millennia. Any putative inferential feature of the connective has been called into question by someone: from the various Contrapositions and De Morgan laws, to Non-Contradiction, Excluded Middle, Double Negation Elimination and Introduction, and the very truth-functionality of the operator [17, see]. The Plan promises a unified framework in which such debates can be *carried out*, while overcoming the Quinean 'change of subject' threat that any deviation from classicality entails that the parties are talking past each other. It grounds negation on compatibility and explains a good deal of disagreement on negation as disagreement on the features of compatibility.

Go back to the duality between positive and negative modalities: in the standard Kripkean frame semantics for modal logics, we take modal m as a quantifier over worlds, restricted from the standpoint of a given world by an accessibility relation endowed with intuitive meaning. We then turn questions about features of m into questions about features of accessibility. Is m factive? Well, does each point look at itself? When one m's that A, does one m that one m's that A? Well, is accessibility

transitive? (And so on: this procedure is so well known that it hardly needs rehearsing.) It is widely agreed that the relation of relative possibility used in the standard semantics of positive modals like the box and diamond has intuitive meaning and helps to provide an enlightening account of such modals. This does not entail that the intuitive meaning will ensure a unique reply to any question one may ask about the relation itself. Nor would this fact be taken by anyone as an issue for the standard, Kripkean possible worlds semantics. Nobody nowadays would raise a claim like 'If something is to qualify as an adequate account of m, it should be fairly clear according to that account what the laws of m are' as an objection to the standard framework. To the contrary: before the frame semantics, some would ask: 'Ok, which one between **S3**, **S4**, etc., is the right system for necessity?'. Now we don't ask that anymore. As a popular logic textbook has it:

> These notions [normal, positive modals] are highly ambiguous. [. . .] Which [normal] system is correct? There is, in fact, no single answer to this question, since there are many different notions of necessity [. . .] the first thing that one needs to do is distinguish among them. [30, pp. 20, 46]

Is this indiscriminate pluralism on m? ('Any system of conditions on accessibility gives an equally acceptable m?'). Not quite: we may have different modals characterized by different (sets of) conditions. Still, given condition k, we ask: is k good as a characterization of that positive modal? (Is the k giving the characteristic validities for **S5** good for metaphysical necessity? Is the one giving the characteristic validities of **GL** good for provability? Etc.) We have some grasp of a certain m (metaphysical necessity, alethic necessity, epistemic commitment, etc.) and we argue on which system of conditions on the relevant accessibility best captures it. It is generally agreed that, by transforming our original question this way, the Kripkean frame semantics with accessibility has brought one of the most celebrated advancements in 20th Century philosophical logic.

This said, both of us did make, in print, claims on the features of compatibility and incompatibility, some of which have been challenged by D&O. We have something to say about these, too.

6 Features of (In)Compatibility

6.1 Symmetry

The only feature of (in)compatibility on which FB takes a resolute positive stance in the *Mind* paper is Symmetry. GR is on the same page in [33]. Whatever kinds of entities a and b are, it seems that if a is incompatible with b, then b has to be incompatible with a (if a's obtaining rules out that of b, b's obtaining rules out that of a, etc).

If compatibility is symmetric, it is easy to show that Double Negation Introduction turns out to be valid in our semantics:

- (DNI) $A \vDash \neg\neg A$

Indeed, a correspondence result from [34, p. 264] shows that DNI holds just in case compatibility is symmetric.

Is it? D&O repeat a point due to Hartonas and Dunn: 'The state of my son's practising his saxophone prevents my reading, but the state of my reading does not one wit prevent his practising the saxophone' [16, p. 32], and argue that prevention provides an example of a non-symmetric compatibility relation.

One should careful here: of course, prevention is not, in fact, a compatibility relation at all. If anything, prevention is an *in*compatibility relation, and if incompatibility is not symmetric, then its complement relation of compatibility also fails to be symmetric, so this is how Dunn and Hartonas' example could be developed. Given a semantics involving situations, it is plausible that there is a non-symmetric prevention relation between those situations, such that situation *a* (involving my son playing the saxophone) prevents situation *b* (involving me reading), while *b* does not in turn prevent *a*. On these points we may agree.

However, this does not itself address the question of whether situations *a* and *b* are compatible. On one reading of compatibility, it is clear that they are not: situations *a* and *b* fail to be compatible because *a* prevents *b*. On a straightforward reading, no two situations are compatible when one prevents the other. There is a non-symmetric positive relation between situations in the vicinity of compatibility: we can say that *c* *permits d* if and only if *c* does not *prevent d*, and permitting is not, here, a symmetric relation. Is permitting a relation of compatibility? It seems to us that it is not. If *b* permits *a* but nonetheless, *a* does not permit *b*, then *b* and *a* are not compatible: they do not fit together. In other words, the presence of non-symmetric prevention (and permission) relations does not mean that compatibility relations need to also be seen as non-symmetric.

6.2 Reflexivity

It is easily shown that, if one assumes $\forall x(xCx)$ (compatibility is reflexive), our semantics validates the Explosion principle or *Ex Contradictione Quodlibet*, in the form:

- (ECQ) $A \wedge \neg A \vDash \bot$

This gives us that a contradiction entails everything, via the fact that, by (S⊥), $\bot \vDash B$ for any B, plus the transitivity of entailment. ECQ is, notoriously, rejected by paraconsistent logicians, and a first group of claims made by D&O around the issue of Reflexivity has to do with the interpretation of paraconsistency.

FB claimed in the *Mind* paper that the intuition that all points (worlds, or whatnot) must be self-compatible has been countered by paraconsistent logicians. Such a bare plural was not meant to be interpreted as: 'by *all* paraconsistent logicians', of course – only some of them. (Bare plurals, as is well known, are ambiguous between generic and existential readings.) D&O retort that:

> Paraconsistent logicians have not said much if anything about compatibility, let alone whether it is intuitively reflexive. What they have countered is ECQ, the inference that from a contradiction, anything follows. Most paraconsistentists do

not endorse a compatibility semantics and those that endorse a semantics which is formally analogous, such as the Routley star, do not view the relation in question that governs negation as one of compatibility. (p. 10)

We come back to D&O's final remark on the lack of (perceived) connection between the Routley star and compatibility in the coming section. Let's now quibble a bit over how one should count paraconsistent logicians. We notice that, for instance, Dunn [16], Mares [27], Dunn and Zhou [13], and both of us, have worked on relevant logic, which we all endorse, if not as The One True Logic (some of us are pluralists), as one valuable logical theory. Relevant logic is paraconsistent, so we all count as paraconsistentists to the extent that we count as relevantists. Now the aforementioned works *do* say a lot about compatibility, indeed some endorse the Australian Plan's compatibility semantics as the favoured semantics for negation in a relevant logic setting.

As for Reflexivity in particular: in the *Mind* paper, FB does not take a stance on the issue whether compatibility is reflexive. But GR does explicitly reject it in [33], precisely on the basis of the need for inconsistent or self-incompatible points in a compatibility-based semantics for nonexplosive logics. We are thus happy for one of us, GR, to contribute a truthmaker for 'Some paraconsistent logicians have countered the intuition that all points (worlds, or whatnot) must be self-compatible'.

As a pluralist, GR does not take these self-incompatible points to be *possible worlds* (more on this below), but as for inconsistent situations, GR has variously argued that these are all self-incompatible, and in fact that it is the self-incompatiblity of these situations that makes them impossible, in the sense of not being included in any possible world.

Now for the substantive issue: *is* compatibility reflexive? Besides arguing *ad hominem* that paraconsistentists who endorse the Australian Plan will be in a predicament if it is, D&O also give one substantive argument in favour. It goes thus. D&O (Ibid) claim, correctly, that what logical properties negation has in the Australian Plan's semantics depends not only on the features of the compatibility relation C showing up in (S¬), but also on those of the information-inclusion relation, $\sqsubseteq$, which, in spite of not showing up in (S¬), does show up elsewhere (in particular, in *Backwards*, which, as we know, is needed for $\sqsubseteq$ to properly interact with C). They then introduce the following condition linking information-inclusion to compatibility:

- (LINK) $x \sqsubseteq y \Rightarrow xCy$

If all the information in x is preserved in y, then x is compatible with y. Assuming Symmetry (which both of us like), we also have 'a version of (LINK) of the form $x \sqsubseteq y \Rightarrow yCx$' (Ibid, notation adjusted for consistency with ours). Now $\sqsubseteq$ is naturally thought of as a partial order, thus reflexive, so 'it follows immediately that compatibility is reflexive' (Ibid).

The argument, however, is question-begging. Acceptance of LINK presupposes acceptance of Reflexivity, that is, of the idea that all points are self-compatible – which FB doesn't endorse, and GR rejects. For suppose y is not self-compatible, which cannot be ruled out beforehand on pain of a *petitio*. Then it may well be the case that for some x, $x \sqsubseteq y$ but it is not the case that xCy. Indeed, y may be

self-incompatible precisely because it encompasses information from x incompatible with further information y itself supports.

So far, thus, we have not been presented with good arguments for Reflexivity. Now on to the previously postponed issue, namely the connection between compatibility and the Routley star.

6.3 Maximal Compatibility and the Star

Assume again that compatibility is symmetric. Add Seriality, $\forall x \exists y(xCy)$ (every point is compatible with some point), and Convergence, that is, the idea that if x is compatible with anything, then there will be a maximally informative point x is compatible with: if $\exists y(xCy)$, then $\exists y(xCy \ \& \ \forall z(xCz \Rightarrow z \sqsubseteq y))$. Call this maximally informative point x^*. Symmetry gives us $x \sqsubseteq x^{**}$. By imposing the converse condition, $x^{**} \sqsubseteq x$, we validate Double Negation Elimination. Via the antisymmetry of $\sqsubseteq$, $x^{**} = x$. Our clause for negation (S¬) now simplifies into:

- (S*) $x \Vdash \neg A \Leftrightarrow x^* \nVdash A$

For $x^* \nVdash A$ precisely if $y \nVdash A$ for all y compatible with x, because xCy just in case $y \sqsubseteq x^*$: x^* is a 'cover all' for each point y compatible with x.

(S*) is the Routley star semantic clause for de Morgan negation, which gets the name because (besides Double Negation Introduction and Elimination) it satisfies all the de Morgan Laws, but differs from Boolean negation by not being explosive. The star was introduced in [42] as a period two operation mapping each point to its maximally compatible peer. Notice that negation keeps being a modal in this setup, for in general x need not be the same as x^*. Negation stops being a modal if we impose that this indeed be the case: for all x, $x = x^*$. Then (S*) boils down to Boolean, classical negation, which is explosive.

D&O present the Routley star semantics for negation as in a certain way alternative to compatibility semantics:

> The philosophical interpretation of the [star] semantics is highly questionable, however, since it is unclear what sort of interpretation we ought to assign to the star function that that takes a world to its star-counterpart [...] It would seem that it is here where (in)compatibility semantics has the advantage since presumably we can attach both clear intuitive and philosophical meaning to the notion of (in)compatibility." (p. 2)

But the star semantics is not just another modal account with respect to the compatibility semantics. We have in fact just shown (following GR's [33], which D&O do not cite) that the star semantics *is* but the compatibility semantics for negation – once the appropriate conditions have been added to the latter. So if, as granted by D&O, 'presumably we can attach clear intuitive and philosophical meaning to the notion of (in)compatibility'; and we can attach clear and intuitive meaning to the additional conditions; then we can attach clear and intuitive meaning to the star semantics as well. The philosophical interpretation of the latter is no more 'highly questionable' than that of the compatibility semantics plus that of the appropriate conditions. The star semantics is just a special case of a compatibility semantics.

Besides, the appropriate conditions themselves are easily interpreted: we just did it at the beginning of this subsection. The issue with them is not what their intuitive meaning is, but whether they *hold* in a given model. The Australian Plan's taking no stance on questions of this kind, as we already argued above, is no problem for it. However, in [33] GR has considered some reasons for not liking Convergence. In particular, if x is a consistent and very incomplete point (say, the situation of some small part of our world), Convergence guarantees that there is a y collecting up everything compatible with x. This may be wildly inconsistent on all the very many A's that x has nothing to say about, for if $x \nvDash A \vee \neg A$, then $y \vDash A \wedge \neg A$. One may, on the other hand, say that the advantages of validating all of de Morgan's Laws compensates the admission of certain odd points in our frames. This is another way in which the semantics of the Australian Plan helps to turn old questions in the foundations of logic (in particular, concerning constructive versus nonconstructive accounts of negation) into clearer and more manageable ones, phrased in terms of (in)compatibility and the extendability of information states.

7 The American Plan, Redux

So much for D&O's arguments against the Australian Plan. Now let us have a look at the American Plan, with its independent accounts of *truth* and *falsity*. We don't deny that American Plan models for different logical systems are interesting and formally useful. However, using the American Plan comes with its own costs, which should be noted. We will consider just four:

Coordination between truth and falsity conditions. Why is it that the falsity conditions for $\wedge$, $\vee$, $\exists$, $\forall$ are the *obvious* de Morgan duals of the truth conditions for these concepts? The semantics allows for the characterisation of a connective (let's call it *interjunction*) with the truth conditions for $\wedge$ and the falsity conditions of $\vee$:

- $A \sharp B$ is *true* iff A is *true* and B is *true*.
- $A \sharp B$ is *false* iff A is *false* and B is *false*.

Is interjunction a sensible logical connective? The American Plan makes it just as available in terms of its basic semantic machinery as any old two-place connective. If $\sharp$ is meaningful, what does it mean? In particular, what does $p \sharp \neg p$ mean? Is it expressible in natural language? If interjunction is meaningless, why is it meaningless? Isn't it strange that something can be 'defined' in models, which does not make sense? The same questions can be asked for '$\flat$', the dual of interjunction, which has the truth conditions of disjunction and the falsity conditions of conjunction.[9]

---

[9]As remarked by a Referee, the connectives $\sharp$ and $\flat$ are of interest in theories of bilattices [19]. The question here, however, is whether they make sense for everyone seeking to follow the American Plan. If you wish to allow for truth value gaps or truth value gluts to accommodate the paradoxes [3, e.g.], but you take it that statements in some restricted language take purely classical (non-glut, non-gap) values, then $\sharp$ and $\flat$ are inappropriate, for if p is true only or false only, then $p \sharp \neg p$ in neither true nor false, and $p \flat \neg p$ is both true and false—with no paradoxes in sight.

The formal machinery of the American Plan gives us a great deal of freedom to define concepts. *Any* American Plan model, in the absence of extra restrictions on semantic evaluations, allows for such concepts, which may go far beyond what is called for in a semantics for a given collection of concepts. If we think of the frame of points as giving a 'semantic field' of possible evaluations of sentences, then the American Plan, as far as D&O have told us, allows for *any* pair $\langle E, A \rangle$ (*extension*, *anti-extension*) of sets of points to be the semantic value of a sentence. This freedom means that the old boundaries are to be revisited. Is it merely a convention or a coincidence that the truth and falsity conditions of conjunction, disjunction, the quantifiers, etc, are coordinated in the usual way? If not, what explains this?

There is no such phenomenon in arbitrary Australian Plan models. The semantic value of a sentence is merely its extension: the set of points at which it is true – closed under the information-inclusion relation, $\sqsubseteq$, if present. The interaction between truth and falsity is determined by the compatibility relation on the underlying frame. If the frame allows for a connective such as interjunction, this is down to the behaviour of the compatibility relation. Given that there are Australian Plan models where compatibility is Boolean (xCy iff $x = y$), there are Australian Plan models where such odd connectives cannot be defined. The coordination between truth and falsity conditions for connectives such as $\wedge$, $\vee$ and the quantifiers is given a uniform explanation on Australian Plan models, in terms of the behaviour of compatibility. No such explanation is given in the American Plan as it stands.

Complexity of truth and falsity conditions. A related cost of using the American Plan as one's model for giving a truth-conditional semantics is the added complexity of giving independent truth and falsity conditions for concepts.

Consider the insight in the Lewis–Stalnaker analysis of counterfactuals as variably strict modals: $A > B$ is true at a point x when B is true at the A-points nearest to x. Question: when is $A > B$ *false* at a point? On an American Plan semantics, nearly any answer to this question is formally compatible with its truth conditions.[10] This does not make any answer equally good. How is such a question to be addressed? Given the complexity of giving American Plan models for relevant logics [41], giving American Plan models for counterfactuals will also be complex.

In the Australian Plan, giving the truth conditions for $A > B$ at points in a frame determines the interaction between the counterfactual conditional and negation, and hence the falsity conditions for $A > B$. No such answer is given in an American

---

[10]Of course, not every account of the falsity conditions for a connective will be compatible with other constraints on the language as a whole. For example, logics like first degree entailment FDE, Priest's logic of paradox, LP, and Kleene's three valued logic K_3, all have the property that if the atomic formulas have classical truth values, then so do any complex expressions constructed out of those values. (Notice, this is not satisfied in logics with connectives such as ♯ and ♭.) Conditions like this place constraints on the falsity conditions for a connective. D&O do not tell us whether conditions like this are to be satisfied or not, and nothing in the American Plan by itself tells us whether constraints like these ought to be respected.

Plan model. If truth and falsity are as independent as is allowed in American Plan semantics, then truth conditions do not (by themselves) give falsity conditions.[11]

Not fully utilising the strengths of point semantics. At an abstract level, a strength of a frame semantics, such as Kripke models for modal logics or Routley–Meyer models for relevant logics, is the set-theoretic representation of logical concepts. Sentences are recursively assigned sets of points. Entailment is subsethood. Conjunction is intersection. Modalities, positive and negative, are closure operators mediated by accessibility relations (or other set-theoretical operations). The picture is formally powerful and philosophically salient. It differs from algebraic semantics were we evaluate sentences as taking various values in a many-valued algebra (three, four, or many more). These kinds of models are also formally useful and philosophically salient.

But an American Plan worlds semantics is a hybrid of both of these. As far as we can see, there is no principle given as to why we have stopped at four distinct semantic values at each world. If formulas can be *true*, *false*, *both* and *neither* at worlds,[12] why not also allow for modal values of necessity and possibility? Why do we analyse them in terms of truth and falsity at other worlds and not think of them as different semantic statuses at this world? The American Plan is a halfway house between a worlds semantics and a many-valued logic, which seems to miss out on some benefits of either approach on its own.

Drawing the wrong distinctions among concepts. The American Plan, as D&O would have it, counts intuitionist negation and the ortho-negation of quantum logics as failing to be genuine negations. To be a negation, on their view, is to satisfy the constraints of first degree entailment or stronger logics. As proponents of the Australian Plan, we can see the distinction that is being drawn here — all of these logics are equally well modelled using the Routley Star, which as we have seen, is a kind of Australian Plan semantics — but it would be a genuine cost to divide the line between genuine negations and other kinds of negative operators *here*, and to leave out ortho-negation and intuitionist negation, venerable claimants to the title of an analysis of *negation*. Friends of the Australian Plan can agree that intuitionistic negation is a genuine constructive analysis *of negation*, and not some foreign modal notion. It is a genuine cost to say that the core notion of negation is such that by definition our constructivist and intuitionist colleagues are making systematic conceptual errors. To take the core of the concept of negation to be given by the American Plan, one has to pay this cost, by giving some account of how it is that these notions miss the mark.

This is particularly pressing given the understanding that D&O have of negation as a contradictory-forming operator (see their *Definition 1*). In ortholattices,

---

[11] A Referee reminds us that this flexibility allows the proponent of the American Plan to give a simple semantics for connexive logics (see, for example Priest [30, Section 9.7]). We have the flexibility—for example—to allow $A > B$ to be *false* when and only when $A > \neg B$ is *true*. In effect, we identify $\neg(A > B)$ with $A > \neg B$. If $\bot > A$ is true for every A, then $\bot > A$ and $\neg(\bot > A)$ are both true.

[12] Each sentence has one of these four different statuses whether we think of these as four distinct values in an explicitly four-valued semantics, or a two-valued *relational* semantics in which formulas may be related to the two values 0 and 1 [30]. In this case, there are still four different semantic statuses, even if we prefer to think of 0 and 1 as the only two truth values.

ortho-negation is exactly a contradictory forming operator in D&O's sense. In intuitionistic logics, whether negation forms contradictories depends on what is meant by *falsity*. We can see, however, that in intiuitionistic logic, negation is a contradictory forming operator in a stronger sense than allowed by D&O: for any sentence A, we have $A, \neg A \vdash$ (A and its negation are *inconsistent*), and $\neg A$ is the weakest sentence with that property.

8 Conclusions

We have defended the Australian Plan, with its **Idea 1** and **Idea 2**, as a coherent and natural account. It is the picture of negation that you get when you use the tools of a point semantics, analysing semantic values of sentences as sets of points, and entailment between sentences in terms of the subset relation. Once we move beyond thinking of points as consistent and complete worlds, the Australian Plan gives an intuitive view of the semantic behaviour of negation. Not only is this a defensible formal model; the core notion, the binary relation of compatibility between points, is natural when it comes to understanding how our concept of negation is grounded. This grounding can be specified both metaphysically (as a robust notion of compatibility between situations) and pragmatically (in terms of norms governing the clash between assertion and denial). So the Australian Plan is not only a productive tool when it comes to the formal semantics of negation: it is also eminently defensible on philosophical grounds.[13]

References

1. Barwise, J., & Etchemendy, J. (1990). Information, infons, and inference. In Cooper, R., Mukai, K., Perry, J. (Eds.) *Situation theory and its applications*, (Vol. 1, number 22 pp. 33–78). Stanford.
2. Barwise, J., & Perry, J. (1983). *Situations and attitudes. Bradford books*. MIT Press.
3. Beall, J.C. (2009). *Spandrels of truth*. Oxford: Oxford University Press.
4. Beall, J.C., & Restall, G. (2000). Logical pluralism. *Australasian Journal of Philosophy*, *78*, 475–493. http://consequently.org/writing/pluralism.
5. Beall, J.C., & Restall, G. (2006). *Logical pluralism*. Oxford: Oxford University Press.
6. Berto, F. (2008). Adynaton and material exclusion. *Australasian Journal of Philosophy*, *86*, 165–90.
7. Berto, F. (2014). Absolute contradiction, dialetheism, and revenge. *Review of Symbolic Logic*, *7*, 193–207.
8. Berto, F. (2015). A modality called "negation". *Mind*, *124*(495), 761–793.
9. Birkhoff, G., & von Neumann, J. (1936). The logic of quantum mechanics. *The Annals of Mathematics*, *37*(4), 823–843. http://www.jstor.org/stable/1968621.

10. Brandom, R.B. (1994). *Making it explicit*. Harvard University Press.
11. Brandom, R.B. (2000). *Articulating reasons: an introduction to inferentialism*. Harvard University Press. ISBN 0674001583.
12. De, M., & Omori, H. (2017). There is more to negation than modality. *Journal of Philosophical Logic*, 1–19. ISSN 1573-0433, https://doi.org/10.1007/s10992-017-9427-0.
13. Dunn, J.M., & Zhou, C. (2005). Negation in the context of gaggle theory. *Studia Logica*, *80*, 235–64.
14. Dunn, J.M. (1993). Partial-gaggles applied to logics with restricted structural rules. In Schroeder-Heister, P., & Došen, K. (Eds.) *Substructural logics*: Oxford University Press.
15. Dunn, J.M. (1994). Star and perp: two treatments of negation. In Tomberlin, J.E. (Ed.) *Philosophical perspectives*, (Vol. 7 pp. 331–357). Atascadero: Ridgeview Publishing Company. http://www.jstor.org/stable/2214128.
16. Dunn, J.M. (1996). Generalised ortho negation. In Wansing, H. (Ed.) *Negation: a notion in focus* (pp. 3–26). Berlin: Walter de Gruyter.
17. Novaes, C.D. (2007). Contradiction: the real philosophical challenge for paraconsistent logic. In Béziau, J.Y., Carnielli, W., Gabbay, D. (Eds.) *Handbook of paraconsistency* (pp. 477–92). London: College Publications.
18. Fine, K. (2009). The question of ontology. In Chalmers, D., Manley, D., Wasserman, R. (Eds.) *Metametaphysics* (pp. 157–77). Clarendon.
19. Fitting, M. (1991). Bilattices and the semantics of logic programming, (Vol. 11 pp. 91–116). ISSN 0743-1066, https://doi.org/10.1016/0743-1066(91)90014-G, http://www.sciencedirect.com/science/article/pii/074310669190014G.
20. Goldblatt, R. (1974). Semantic analysis of orthologic. *The Journal of Philosophical Logic*, *3*(1–2), 19–35. https://doi.org/10.1007/BF00652069. Reprinted as Chapter 3 of Mathematics of Modality, Robert Goldblatt. Mathematics of Modality. Number 43 in CSLI Lecture Notes. CSLI Publications, 1993. http://standish.stanford.edu/bin/detail?fileID=458253745.
21. Grim, P. (2004). What is a contradiction? In Priest, G., Beall, J.C., Armour-Garb, B. (Eds.) *The law of non-contradiction* (pp. 49–72). Clarendon.
22. Incurvati, L., & Schlöder, J.J. (2017). Weak rejection. *Australasian Journal of Philosophy*, *95*(4), 741–760. https://doi.org/10.1080/00048402.2016.1277771.
23. Kripke, S. (1965). Semantical analysis of intuitionistic logic. In Crossley, J., & Dummett, M.A.E. (Eds.) *Formal Systems and Recursive Functions* (pp. 92–130). Amsterdam: North-Holland Publishing.
24. Kripke, S. (1980). *Naming and necessity*. Oxford: Blackwell.
25. Lewis, D.K. (1973). *Counterfactuals*. Oxford: Blackwell.
26. Mares, E.D. (1995). A star-free semantics for R. *Journal of Symbolic Logic*, *60*, 579–590.
27. Mares, E.D. (2004). *Relevant logic. A philosophical interpretation*. Cambridge: Cambridge University press.
28. Meyer, R.K., & Martin, E.P. (1986). Logic on the australian plan. *The Journal of Philosophical Logic*, *15*(3), 305–332. https://doi.org/10.1007/BF00248574.
29. Price, H. (1990). Why 'not'? *Mind*, *99*, 221–38.
30. Priest, G. (2008). *An introduction to non-classical logic*, 2nd edn. Vol. 2008. Cambridge: Cambridge University Press.
31. Restall, G. (1993). Four-valued semantics for relevant logics (and some of their rivals). *Journal of Philosophical Logic*, *24*, 139–69.
32. Restall, G. (1995). Information flow and relevant logics. In Seligman, J., & Westerståhl, D. (Eds.) *Logic, language and computation: the 1994 Moraga proceedings* (pp. 463–477).
33. Restall, G. (1999). Negation in relevant logics (how i stopped worrying and learned to love the Routley star). In Gabbay, D., & Wansing, H. (Eds.) *What is negation?* (pp. 53–76). Dordrecht: Kluwer.
34. Restall, G. (2000). *An introduction to substructural logics*. Routledge.
35. Restall, G. (2000). Defining double negation elimination. *The Logic Journal of the IGPL*, *8*(6), 853–860. http://jigpal.oxfordjournals.org/cgi/content/abstract/8/6/853.
36. Restall, G. (2005). Multiple conclusions. In Hájek, P., Valdés-Villanueva, L., Westerståhl, D. (Eds.) *Logic, methodology and philosophy of science: proceedings of the twelfth international congress* (pp. 189–205): KCL Publications. http://consequently.org/writing/multipleconclusions.
37. Restall, G. (2009). Truth values and proof theory. *Studia Logica*, *92*(2), 241–264. http://consequently.org/writing/tvpt/.

38. Ripley, D. (2013). Paradoxes and failures of cut. *Australasian Journal of Philosophy*, *91*(1), 139–164. https://doi.org/10.1080/00048402.2011.630010.
39. Routley, R., & Meyer, R.K. (1972). The semantics of entailment II. *Journal of Philosophical Logic*, *1*, 53–73.
40. Routley, R., & Meyer, R.K. (1973). The semantics of entailment I. In Leblanc, H. (Ed.) *Truth, syntax, and semantics* (pp. 194–243). North-Holland.
41. Routley, R. (1984). The American plan completed: alternative classical-style semantics, without stars, for relevant and paraconsistent logics. *Studia Logica*, *43*(1–2), 131–158. https://doi.org/10.1007/BF00935746.
42. Routley, R., & Routley, V. (1972). Semantics of first degree entailment. *Noûs*, *6*(4), 335–359. http://www.jstor.org/stable/2214309.
43. Stalnaker, R. (1968). A theory of conditionals. In Rescher, N. (Ed.) *Studies in logical theory* (pp. 98–112). Oxford: Blackwell.
44. Tahko, T. (2009). The law of non-contradiction as a metaphysical principle. *Australasian Journal of Logic*, *7*, 32–47.
45. Tennant, N. (1999). Negation, absurdity and contrariety. In Gabbay, D., & Wansing, H. (Eds.) *What Is negation?* (pp. 199–222). Dordrecht: Kluwer.
46. Tye, M. (1990). Vague objects. *Mind*, *XCIX*(396), 535. https://doi.org/10.1093/mind/XCIX.396.535.
47. Wansing, H. (2001). Negation. In Goble, L. (Ed.) *The Blackwell guide to philosophical logic* (pp. 415–36). Oxford: Blackwell.
48. Wansing, H. (2008). Constructive negation, implication, and co-implication. *Journal of Applied Non-Classical Logics*, *18*(2-3), 341–364.
49. Wiggins, D. (2001). *Sameness and substance renewed*. Cambridge: Cambridge University Press.