There is more to negation than modality*

Michael De (Konstanz) and Hitoshi Omori (Kyoto)

Preprint: Forthcoming in the *Journal of Philosophical Logic*

Abstract

There is a relatively recent trend in treating negation as a modal operator. One such reason is that doing so provides a uniform semantics for the negations of a wide variety of logics and arguably speaks to a longstanding challenge of Quine put to non-classical logics. One might be tempted to draw the conclusion that negation is a modal operator, a claim Francesco Berto (Berto, 2015) defends at length in a recent paper. According to one such modal account, the negation of a sentence is true at a world x just in case all the worlds at which the sentence is true are *incompatible* with x. Incompatibility is taken to be the key notion in the account, and what minimal properties a negation has comes down to which minimal conditions incompatibility satisfies.

Our aims in this paper are twofold. First, we wish to point out problems for the modal account that make us question its tenability on a fundamental level. Second, in its place we propose an alternative, non-modal, account of negation as a contradictory-forming operator that we argue is superior to, and more natural than, the modal account.

1 Introduction

With the advent of Kripke semantics for modal logic, there has been a recent trend of treating what were traditionally taken as extensional, non-modal connectives, as modal connectives. This is especially true for the conditional, but also for the connective of present interest, negation. For instance, according to one such account, the negation of a sentence is true at a world w just in case all words where the sentence is true are incompatible with w. Incompatibility is taken to be the key notion in the account, and what minimal properties a negation has comes down to which minimal conditions compatibility satisfies. This semantics dates back to algebraic work done in the thirties by Garrett Birkhoff and John von Neumann (Birkhoff and von Neumann, 1936) on quantum logic, and subsequent work done in the seventies by Rob Goldblatt (Goldblatt, 1974) on ortho-logic using today's more familiar Kripke semantics. Not much was said then concerning the philosophical interpretation of the semantics.

There are of course other modal accounts of negation. Relevant logic, for instance, has the so-called Routley star due to Richard and Val Routley (Routley and Routley, 1972): the negation of a sentence is true at a world w just in case the sentence fails to be true

*This research was funded by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement nr 263227. Hitoshi Omori is a Postdoctoral Research Fellow of the Japan Society for the Promotion of Science (JSPS). We would like to thank Brendan Balcerak Jackson, Francesco Berto, Thomas Müller, Heinrich Wansing, and an anonymous referee for helpful discussion. We would also like to thank the participants of the Munich Centre for Mathematical Philosophy (MCMP) Colloquium.

at the so-called star-world of w. The philosophical interpretation of such a semantics is highly questionable, however, since it is unclear what sort of interpretation we ought to assign to the star function that takes a world to its star-counterpart. Richard Routley uses the following analogy as explanation, where he calls a star-world a "reverse world": "[a] reverse world is like the reverse side of something, e.g. a gramophone record" (Routley, 1980, p. 291), which B. J. Copeland rightly criticizes as telling us nothing concerning an appropriate interpretation for the semantics.[1] 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. It is on the basis of this alleged advantage that, in a recent paper, Francesco Berto (Berto, 2015) defends compatibility semantics as *the* correct account of negation. He argues further that it provides a compelling answer to Quine's infamous challenge summed up in his pithy claim that when the non-classicist "tries to deny the doctrine he only changes the subject" (Quine, 1970, p. 81).

The classicist and deviant all use the world 'not' and its cognates (or, of course, their translations into other languages), yet each seems to mean something different by it. How, then, can these rival logicians be said to genuinely disagree over the nature of negation (and contradiction, consistency, etc.) if they do not even mean the same thing by 'not'? And thus we arrive at Quine's famous charge of "change of logic, change of subject".

The problem with the Quinean challenge is that it can be put to any philosophical dispute over the nature of any notion or other, as least when those notions are sufficiently fundamental. One says that epistemic justification is infallible, the other says not. Maybe each means something different by 'justification' so that disagreement is only illusory. One says perception is veridical, the other says not. Maybe each means something different by 'perception', and again disagreement is only illusory. If the challenge is so easy to come by, why should we take it seriously? Shouldn't we instead take it as a starting point that there is genuine disagreement between classicist and rival, and then figure out from there how their disagreement could be genuine? While we would like to, we think the challenge needs addressing and that it is an important criterion in evaluating the adequacy of a characterization of a notion as fundamental as negation. Requiring that an account of negation meet the challenge ensures that the account does not beg the question against the deviant or classicist.

One way of meeting the Quinean challenge is to deny Quine's implicit claim that the meaning shared between classicist and rival is insufficient to allow for genuine disagreement. We have never seen a claim (let alone argument) to the effect that a necessary requirement on disagreement is that there be *precise* sharing of meaning amongst disputants. How could there be disagreement in such a case? Perhaps meaning is *never* precisely shared between disputants (a view once held by Nelson Goodman in (Goodman, 1949)): we would not then want to say that there is *never* genuine disagreement. The account of negation we offer in §5 provides an intuitive characterization according to which sufficient meaning is shared between classicist and deviant over the meaning of negation, though not necessarily over the meaning of related notions such as truth and falsity.

We will argue that the modal account of negation is implausible as an account of how we process or understand negated sentences, nor for providing an explanation as to when a negated sentence is true. 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.

[1] See (Copeland, 1983).

The alternative account we provide is more intuitive and so fundamental that there can be little dispute over what the Laws of Negation are.[2]

2 Preliminaries

For preliminary measures, we lay down the basic semantic machinery. The language L is propositional and contains the connectives $\sim$, $\wedge$, $\vee$, $\top$ and $\bot$.[3] We use metavariables A, B, etc. as ranging over object-level propositions.

A frame for L is a triple $\langle W, R, \sqsubseteq \rangle$, where W is a non-empty set, R is a binary relation on W, and $\sqsubseteq$ is a partial ordering on W. Intuitively, xRy is to be read as as 'y is compatible with x', and $x \sqsubseteq y$ as 'y includes all the information included in x'. Frames satisfy the following condition:

(Backwards) $\quad$ If xRy and $x_1 \sqsubseteq x$ and $y_1 \sqsubseteq y$, then $x_1 R y_1$.

A model $\langle W, R, \sqsubseteq, \Vdash \rangle$ for L is a frame equipped with an interpretation function $\Vdash$ satisfying the Heredity Constraint:

(HC) $\quad$ If $x \Vdash A$ and $x \sqsubseteq y$, then $y \Vdash A$, for atomic A.

Here $x \Vdash A$ means that A holds at x and (below) $x \nVdash A$ means that A does *not* hold at x, where the metalanguage is understood as classical.[4] It is shown by induction that, given (Backwards) and the truth conditions given below, (HC) extends to arbitrary formulae.

The semantic clauses for the connectives are then given as follows:

(S$\wedge$) $\quad x \Vdash A \wedge B$ iff $x \Vdash A$ and $x \Vdash B$

(S$\vee$) $\quad x \Vdash A \vee B$ iff $x \Vdash A$ or $x \Vdash B$

(S$\top$) $\quad x \Vdash \top$

(S$\bot$) $\quad x \nVdash \bot$

(S$\sim$) $\quad x \Vdash {\sim}A$ iff $\forall y xRy, y \nVdash A$

Intuitively, (S$\sim$) says that the negation of a sentence is true at a world x just in case that sentence is not true at any world compatible with x.

A set Γ of formulae semantically entails A, written $\Gamma \models A$, iff for every model $\mathcal{M}$ of L and every world x of $\mathcal{M}$, we have $x \Vdash A$ whenever $x \Vdash B$ for all $B \in \Gamma$.

3 Is incompatibility a primitive notion?

Berto wishes to ground negation in terms of what he takes to be a more fundamental notion, viz. (in)compatibility. He says:

---

[2] As a referee presses, it might be a mistake to take the Quinean challenge seriously if it diverts attention from characterization negation to characterizing disagreement instead. However, we do not think meeting the Quinean challenge amounts to characterizing disagreement rather than negation. Meeting the challenge means giving a characterization of negation that makes room for disagreement and ensures that a given characterization of negation not beg the question against e.g. the deviant.

[3] Berto's language includes an additional connective for necessity for which we have no use. This gives us a slight simplification of the set-up. In particular, frames (hence models) for our purposes have only one accessibility relation for compatibility which we denote by R rather than R_N.

[4] We're suppressing the model subscripts on $\Vdash$; when being explicit we write $x \Vdash_{\mathcal{M}} A$ to mean that A holds at x in model $\mathcal{M}$.

> I take (in)compatibility as the primitive twofold notion grounding the origins of our concept of negation and of our usage of the natural language expression 'not'. Explanations stop when we reach concepts that cannot be defined in terms of other concepts, but only illustrated by way of example. A good choice of primitives resorts to notions we have a good intuitive grip of—and this is the case, I submit, with (in)compatibility (Berto, 2015, pp. 7–8).

These claims are to apply to both compatibility and incompatibility; i.e. that both are primitive and that both serve equally well in grounding our understanding of negation. He goes on:

> It is difficult to think of a more pervasive and basic feature of experience, than that some things in the world *rule out* some other things; or that the obtaining of this *precludes* the obtaining of that; or that something's being such-and-such *excludes* its being so-and-so. Not only rational epistemic agents and speakers of natural languages, but also animals, or sentient creatures generally, are acquainted with (in)compatibility (Berto, 2015, p. 8).

We have some doubts about these claims. First, it is not at all obvious that *both* compatibility and incompatibility are psychologically primitive notions and that, in particular, incompatibility is not a less primitive compound got from compatibility and negation.[5] A substantiation of such a *descriptive* claim requires empirical justification we have not been given. This is, in any case, a separate issue from the *normative* issue concerning whether both notions *ought to be* taken as primitive according to the best systematization of our total theory. In this sense it is quite plausible, for reasons of ideological economy, to take only one as primitive and to define its complement in terms of negation. We are not much concerned with the psychological claim and think that, for reasons given below, there is good reason to doubt the normative claim as well.

It is important to notice that if incompatibility is defined from compatibility and negation, (S$\sim$) becomes circular since the definiendum occurs in the definiens. For it says that a negation is true just in case the negand is *not* true at any compatible world. The version of (S$\sim$) given in terms of *in*compatibility (got by contraposing and removing double negations) would remain circular on the assumption that incompatibility is understood in terms of negation and compatibility. We therefore need a good reason for believing that incompatibility is at least as primitive a notion as compatibility if the former is to ground our understanding of negation.

In (Price, 1990), Huw Price gives us a "possible evolutionary story" about how our notion of incompatibility arose, and some reasons for thinking that incompatibility is a more primitive notion than sentential negation. The evolutionary story is quite speculative and we're not much convinced by it. Where we agree with Price (and Berto) is 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. Where we disagree with is that it is this very notion of incompatibility that grounds the truth conditions of negated statements. We

[5] In explaining incompatibility, using expressions like 'exclude', 'preclude', and 'rule out' instead of overtly negative expressions like 'not' (as Berto does in the quoted passage above) does not suggest that incompatibility is primitive. It is after all plausible, 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.

will come back to this point in a moment.[6]

One might wonder what the difference is between (S∼) and e.g. (S∧): both give truth conditions for an object-language connective in terms of the "same" corresponding meta-language connective. So if (S∼) is problematic, it seems the truth conditions for the other connectives are as well. 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 it is also why (S∼) is problematic as a grounding definition if negation cannot be eliminated from the right-side of the biconditional.

Let us suppose, for the sake of argument, that incompatibility is a primitive notion. We ask: does incompatibility help in explaining or elucidating when a negation is true? We can't see how, at least not in the usual cases. 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. An explanation of this sort is not, however, available to Berto since it explains incompatibility in terms of negation rather than the other way round.

Indeed, it is that way round that the order of explanation seems to go. It is natural to say that all the gram-heavier-worlds are incompatible with ours *because* it is true at our world that Sam is *not* a gram heavier than she is. Without negation, we have faintly an idea of when two *worlds* are incompatible. We do have an idea of when two events or properties are incompatible, e.g. when they are determinates of the same determinable, but that is not the notion of compatibility that is relevant to the modal account.[7]

It does not seem that when we go about determining whether a negation $\sim A$ is true, we think about all the A-worlds there are (non-recursively enumerably many!) and then

[6]For Price, negation's primary function is to express when two—let us call them *states of affairs*—are incompatible. You say you're going to talk to Fred in the kitchen, we say he's in the garden, to which you say thanks and set off for the kitchen. If you don't see the implicit incompatibility between Fred's being in the garden and his being in the kitchen, we can make it explicit by uttering 'But he's *not* in the kitchen'. Notice that we are talking about incompatibility between states of affairs here, not worlds, assuming worlds are typically not just single states of affairs (if sometimes they may be). Also notice that this story works only when there is a salient contrast between two (or more) states of affairs. Our uttering 'There's no beer left' does not seem to signal or express any implicit incompatibility between the state of there being beer and some one *other* state. Which other state? For this and other reasons, we remain unconvinced by Price's claims concerning the primary function of negation.

[7]We can think of worlds being compatible with each other in certain respects and incompatible in certain other respects; that is, (in)compatibility can be viewed as a *ternary* relation between states and respects. Or, we could think of negated statements being context-sensitive and (in)compatibility remaining two-place, with the quantifier in (S∼) being restricted by which respects are salient. Both of these options seem to give us a response to some of the tension just noted in saying that all the worlds where Sam is a gram heavier than she actually is are incompatible with ours, period. But no matter how we restrict the quantifier, many of the issues raised above still stand. For instance, why at all do worlds need to enter into the truth conditions for negated statements?

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".[8] We think the burden lies with the modalist to more plausibly establish that at least a good number of negated sentences are so understood. Perhaps the claim ought to be understood normatively, so that it says that we *ought* to understand negation in terms of compatibility rather than that we in fact do. Perhaps we ought to because the compatibility semantics provides a correct account of negation and our usage of it in natural language. That is a different claim that needs argument, and one that is in any case undermined by what we say in §4.

To emphasize, the claim is that incompatibility is the primitive notion grounding our concept and usage of negation. As far as we understand this claim and that it is intended as descriptive, we simply find it hard to believe. We agree that in certain—but not all—cases we see quite quickly that two things (events, etc.) are incompatible without any apparent processing of negation. What we disagree with is that we are also doing this with respect to *worlds* when we evaluate the truth or falsity of a negated sentence. How we might extend the intuitive idea that certain ordinary events, actions, etc. are incompatible to the idea that worlds (or situations, information states, etc.) are incompatible—*without recourse to negation*—is not at all clear to us. Do we really have a strong intuitive grip on what it means for two worlds to be (in)compatible? We have our doubts.

Despite these challenges for the compatibility account, we shall for the moment forget any descriptive claims and focus on whether the modal account furnishes negation with extensionally adequate truth conditions.

4 Properties of (in)compatibility

4.1 Contraposition

It is easy to prove, using only (S∼), that

(Contra) $$\text{If } A \models B, \text{then } {\sim}B \models {\sim}A.$$

It is thus a minimal feature on the incompatibility account that negation satisfy contraposition, at least in this rather weak form. This requirement rules out a significant number of well-known (alleged) negations: for instance the paraconsistent negations of Newton da Costa's C-systems (da Costa, 1974), negation in Nelson's constructive logic (Nelson, 1949; Gurevich, 1977; Odintsov, 2008), the paraconsistent negation of Priest's Logic of Paradox (**LP**) (Priest, 1979), and the paraconsistent and paracomplete de Morgan negation of Belnap-Dunn logic, **BD** (also known as **FDE**) (Belnap, 1976).[9] The problem is that contraposition and weakening (both in arrow form) yield that the negation of every formula follows from any contradiction or, worse, in the presence of double negation elimination, the dreaded ECQ—that every formula follows from any contradiction. This means that any paraconsistent logic with weakening, such as the da Costa systems, must give up contraposition. On the other hand, one could also give up weakening as Berto recommends on the grounds that doing so is independently motivated (think of relevance logic). But weakening too is independently motivated, so whether we accept or reject it cannot follow simply from whether it or its rejection has independent motivation.

[8] See also (Wansing, 2016).

[9] It is worth noting that (Contra) is admissible in **BD** but not **LP**, even though the former can be seen as a natural generalization of the latter to a paracomplete or partial logic.

The main worry with (Contra), however, is that it commits paraconsistency to a rejection of weakening. It is not, however, part and parcel of paraconsistency that weakening fail. Even if it were true that the only viable version of paraconsistency is weakening-free, that claim should follow from a correct account of negation. The two issues are simply independent of each other.

There is one further issue concerning (Contra) that involves making a distinction between kinds of worlds. A certain style of semantics marks a distinction between what are sometimes called 'normal' worlds and 'non-normal' worlds, where the normal worlds are taken to be possible or well-behaved, and the non-normal worlds impossible. For example, some of the Lewis systems and relevant logics are standardly given a semantics marking such a distinction. Defining negation in terms of incompatibility on such a semantics will not guarantee that (Contra) hold.[10] So if some understanding of 'world' warrants making a normal/non-normal distinction of just this sort, (Contra) will not come out as minimal law of negation. While we do not find this troubling, it does speak against the compatibility account by stripping negation of one of its few core principles (according to that account). One might wonder whether, in the end, there is enough left of negation to provide a satisfactory characterization of the notion.

4.2 Symmetry and double negation introduction

Berto claims that

> (in)compatibility must be symmetric: whatever ontological kinds a and b belong to, it appears that if a rules out b, then b has to rule out a; that if a's obtaining is incompatible with b's obtaining, then b's obtaining must also be incompatible with a's obtaining; etc (Berto, 2015, p. 17).

Symmetry of compatibility implies Double Negation Introduction:

(DNI) $$A \models {\sim}{\sim}A.$$

Now consider the following counterexample to symmetry 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" (Dunn, 1999, p. 32). So it appears we have one situation being incompatible with another but not conversely.

In response to this counterexample Berto argues "the two situations still count as symmetrically incompatible with each other: for if my son's playing the sax entails that I get too distracted to read the paper, then by the uncontroversial Minimal Contraposition my not being too distracted to read the paper entails that my son can't be playing the sax" (Berto, 2015, p. 17). Now we cannot see the relevance of contraposition to showing that the original two situations are indeed symmetric as regards incompatibility. An implication like 'If my son plays the sax, I'm not reading a technical paper' does not hold as a matter of *logical* entailment. So the fact that such an implication holds does not allow us to infer the contrapositive via (Contra), which applies only to logical entailments. To draw the contrapositive, one would need to argue for a form of contraposition involving some conditional that adequately captures the implication in question. But as the conditional in question looks *counterfactual*, we will not expect the required form of contraposition to be available. So we cannot appeal to contraposition to get symmetry in this case.

[10] Such an example is given by the Routley-star—hence compatibility—semantics for **LP** on which (Contra) fails. See (Beall, 2009, p. 9) for details.

The real question is whether prevention in the sense employed in the counterexample is a relation of incompatibility. If it is, it looks like we have a genuine counterexample regardless of whether we can infer the contrapositive from it. The only thing we could draw from the counterexample (which is a conditional sentence, recall) and its contrapositive is that the conditional in question does not capture the relevant notion of compatibility. To rebut the counterexample, one needs to argue that prevention in the given sense is not a relation of incompatibility and that that relation *is* captured by an implication of the form $A \to \neg B$ which states that (the situation expressed by) A is incompatible with (the situation expressed by) B, and hence by contraposition and DNI for $\to$, that B is incompatible with A. The problem is that such an explanation of incompatibility better not hold in general, for otherwise we appear to have an explanatorily useful account—*in terms of negation!*—of when two events are incompatible with each other (i.e. when some implication holds), undermining any claims of the primitiveness of incompatibility. On the other hand, if such an explanation does not hold in general, what is to convince us that it holds in the particular Dunn-Hartonas case at hand?[11]

There are, in addition, stronger reasons for rejecting symmetry we shall discuss shortly.

4.3 Reflexivity and paraconsistency

On compatibility semantics, paraconsistency and reflexivity of compatibility stand and fall together. Suppose R is reflexive and that $x \Vdash A \wedge {\sim}A$. Then by (S$\wedge$), $x \Vdash A$ and $x \Vdash {\sim}A$. By (S$\sim$) and the reflexivity of R, $x \nVdash A$—contradiction. Hence, if R is reflexive, paraconsistency is ruled out. (The other direction also follows.)

Berto asks

> [i]s compatibility reflexive?. . . the view that worlds must be self-compatible is sufficient to give full-fledged intuitionistic negation. . . [a]nd this seems another strong intuition. . . this strong intuition has been strongly countered by paraconsistent logicians (Berto, 2015, pp. 18–19).

Here we have a fallacy of presupposing one's own view. 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.

If the intuition behind reflexivity is so strong, what are the reasons for rejecting it? Berto provides the following. There are various legitimate senses of 'world' (hence his pluralism about negation). If worlds are taken to be e.g. information states, then worlds may be self-incompatible.[12] Thus reflexivity fails under some senses of 'world'. So compatibility shouldn't be, in general, reflexive.

[11] One could rebut the counterexample in other ways. For example, one might think that if one's son's playing the sax really prevents one's reading a paper, then surely one's reading a paper must prevent one's son's playing the sax. For how could one be reading a paper while one's son is playing the sax if the latter prevents the former? One can't, so it looks like prevention is symmetric after all, despite initial appearances. As Berto notes, the asymmetry may be explained by causal asymmetry since one's son's playing the sax causes one not to read a paper, but not conversely. Thanks to [anonymous] for pressing us on this response to the Dunn-Hartonas counterexample. We think there are better counterexamples in any case, as discussed below.

[12] Here Berto also cites the actual world as being self-incompatible for a dialetheist like Graham Priest. But why would the actual world be incompatible with itself? What does that mean? We guess the idea is that the actual world makes certain contradictions true, pace Priest, so it must be incompatible with itself, assuming compatibility semantics. Aside from question-beggingness, here we are explaining self-

Even if we grant that information states may contain inconsistent information, it is unclear to us in what sense information states are self-incompatible. We would not say of an information state that supports A and not-A that it is self-incompatible, at least if we are going by our intuitive understanding of the notion. The fact that a state supports a contradiction gives us no reason to think that it is self-incompatible in the intuitive sense. There are clearly theoretical senses according to which states can be self-incompatible, but these senses are not available to Berto who is working with an intuitive understanding of incompatibility. And if worlds, taken under any sense of the word, are not self-incompatible, then the paraconsistentist has strong reasons for rejecting compatibility semantics.

Furthermore, if the argument against reflexivity were good, a parallel argument shows that symmetry isn't a minimal constraint on compatibility either. Under certain determinations of 'world', one world may be compatible with another but not conversely. E.g. it seems plausible that if worlds can be self-incompatible (hence inconsistent) then such worlds may be compatible with self-compatible (hence consistent) worlds. But it also seems that under some determinations of 'world', some consistent worlds cannot be compatible with inconsistent ones, even when some inconsistent ones are compatible with it. For consider the following scenario.

Let w be a maximal, consistent world that doesn't make A true, and let w' be an inconsistent world that makes $A \wedge {\sim}A$, hence (by (S$\wedge$)) A, true. It seems perfectly plausible that w' be compatible with w. The converse, however, could not be true, for by maximality w makes ${\sim}A$ true, so any world compatible with it can't make A true. (Notice if that one wants to reject this argument, one would have to show that the situation just described can never be the case. We can't see how anyone could show that!)

Let us consider one last argument in favor of reflexivity, and hence against compatibility semantics (assuming the plausibility of paraconsistency). One of Berto's central claims is that, in determining which core laws a negation satisfies, we need only look at which minimal constraints compatibility satisfies. But the core properties of negation also depend on which properties the *inclusion relation* has because of (Backwards), which is assumed to be a minimal condition on the semantics. So it is two relations, not one, that figure in determining which are the core laws of negation.

If this is right then the following condition seems to us to be in just as good standing as symmetry and (Backwards):

(LINK) $$x \sqsubseteq y \Rightarrow xRy.$$

(LINK) tells us that if one world is informationally included in another, then the other is compatible with its subworld. Assuming symmetry, we also have a version of (LINK) of the form $x \sqsubseteq y \Rightarrow yRx$ which says that subworlds are compatible with worlds they are subworlds of. How could a world not be compatible with any of its subworlds or superworlds (unless they were self-incompatible, which is the issue currently in question)? We can't think of any sense of 'world' under which (LINK) could fail.[13]

Now since $\sqsubseteq$ is reflexive,[14] it follows immediately that compatibility is reflexive. So we have an argument for reflexivity relying only on one plausible connection between inclusion and compatibility that has at least as much intuitive force as the already accepted (Backwards).

incompatibility in terms of negation—specifically, in terms of whether or not a sentence and its *negation* are true. Yet the modalist needs the explanation to go the other way round.

[13] In passing, Berto refers to (LINK) in fn. 27, but he makes no further mention of it, including the fact that it entails reflexivity of compatibility.

[14] Berto never says so but we assume so given his notation, and we can anyway define such a relation in terms of the proper inclusion relation and identity and substitute that in for $\sqsubseteq$ in (LINK).

Given the surmounting grounds for its acceptance, reflexivity looks like a minimal constraint on compatibility, and paraconsistency is therefore ruled out on compatibility semantics. While some might take this as a virtue of the account, we mention as an *ad hominem* that Berto himself is a paraconsistentist, so he should find this conclusion as repugnant as we do.

5 An alternative, non-modal account of negation

We now wish to present a rival account of negation.[15] This is by no means the only alternative, and we will not be able to argue for our preferred account over others due to space constraints, but we find it most convincing when it comes to (i) explaining our understanding of negation, (ii) clarifying the difference between nonclassical logics, and (iii) meeting the Quinean challenge we discussed in §1. It just so happens that we here find ourselves in good company with one of the most prominent researchers on compatibility semantics, J. Michael Dunn.

> Tim Smiley once good-naturedly accused me of being a kind of lawyer for various non-classical logics. He flattered me with his suggestion that I could make a case for anyone of them, and in particular provide it with a semantics, no matter what the merits of the case [...] But I must say that my own favourite is the 4-valued semantics. I am persuaded that '$\neg\phi$ is true iff ϕ is false', and that '$\neg\phi$ is false iff ϕ is true'. And now to paraphrase Pontius Pilate, we need to know more about 'What are truth and falsity?'. It is of course the common view that they divide up the states into two exclusive kingdoms. But there are lots of reasons, motivated by applications, for thinking that this is too simple-minded. (Dunn, 1999)

It could not be clearer that Dunn is here endorsing an account of negation as a contradictory-forming operator. It is this account of negation on which we wish to further expand.[16]

[Contradictories] Let A and B be sentences. Then,

- A and B are *contraries* if one of them is false whenever the other is true;
- A and B are *subcontraries* if one of them is true whenever the other is false;
- A and B are *contradictories* if they are contraries and subcontraries.

Perhaps a more familiar recasting of these notions is to say that two sentences are contraries if both can*not* be true together and that they are subcontraries if both can*not* be false together. It is important to emphasize that these two separate formulations of contrariety and subcontrariety are not equivalent unless we make certain controversial assumptions concerning falsity and untruth. In particular, the formulation we prefer is more tolerant to non-classical negations and, more importantly, does not employ negation in the metalanguage, and so cannot be charged with circularity.

[Negation] Let $\sim$ be a sentential operator. Then $\sim$ is a negation iff it is contradictory-forming, i.e. iff $\sim A$ and A are contradictories.

The above definition is equivalent to the conjunction of the following two biconditionals:

[15] For an up-to-date survey on various accounts of negation, see (Horn and Wansing, 2015).

[16] See (Wansing, 2006) for other *syntactic* accounts of contraries, subcontraries and contradictories. For some reasons for preferring a semantic over syntactic characterization of negation, see (De, 2011, Chap. 2).

- $\sim A$ is true iff A is false;
- $\sim A$ is false iff A is true.

Indeed, we can decompose these conditions into the following two pairs:

- If $\sim A$ is true then A is false, and if A is true then $\sim A$ is false;
- If $\sim A$ is false then A is true, and if A is false then $\sim A$ is true.

The first characterizes contraries and the second, subcontraries. Defining negation as a contradictory-forming operator therefore provides a very natural account of negation as an operator that flip-flops truth and falsity.[17]

With this account of negation as a contradictory-forming operator, do we secretly smuggle in the notion of incompatibility? Is saying that two sentences are contraries nothing more than saying that truth and falsity are incompatible? Not on our account. We have defined contrariety in terms of truth and falsity *without* appealing to incompatibility and with it being consistent that truth and falsity are compatible. In other words, on our account, two sentences can be contraries even if truth and falsity are compatible. Our definitions of contrariety, subcontrariety and contradictoriness do not, moreover, require a prerequisite understanding of incompatibility.

Berto argues that the modal account he endorses gives rise to a pluralism about negation. One already gets this from logical pluralism since there are different logics with negations satisfying different properties, so there must be a plurality of negations. Given that we've said nothing about the relation between truth and falsity, do we get pluralism about negation on our account, supposing there are a variety of legitimate relations holding between truth and falsity?

For instance, if truth and falsity are exclusive and exhaustive, negation is *classical*. If they're not exhaustive, negation is *paracomplete*, and if they're not exclusive, negation is *paraconsistent*. But truth and falsity, we assume, relate in one and only one way. As such, we get no pluralism about negation on our account. Precisely how truth and falsity relate is a matter of further philosophical discussion that would take us too far afield, so we defer the reader to the extensive literature on the topic.[18]

We now turn to some of the more formal details of our account. A *model* is a pair $\langle \Vdash, \dashV \rangle$, consisting of a truth assignment $\Vdash$ and a falsity assignment $\dashV$ to atoms. Since $\Vdash$ and $\dashV$ are uniquely extended to arbitrary formulae, we use this notation also for the relations holding between non-atomic sentences.

On the present account of negation as a contradictory-forming operator, we need only take truth and falsity as primitives. There is no commitment to metaphysically controversial items like worlds, a relation of containment between them, nor a contentiously primitive notion of incompatibility. While we do have *two* primitive assignment functions, a single truth assignment $\Vdash$ can be obtained (that is, one of the primitives may be taken as redundant) by constraining the relation between $\Vdash$ and $\dashV$.[19] We do not mean to imply that no semantics should be committed to worlds, and certainly the present account of

[17]Note here that we may take truth and negation as primitive, and define falsity in terms of truth of negation. The problem with doing so, however, is that it settles too many questions concerning the relation between the two truth values, and whether they are *both sui generis* notions or whether one is definable from the other. There will be many cases where we will want to distinguish falsity from untruth (e.g. when there are truth-value gaps), and taking each as primitive allows us to do so.

[18]For a rejection of gaps, see e.g. (Priest, 2006b, §4.7), and for a rejection of gluts, see e.g. (Lewis, 1982).

[19]E.g. if $\{A : \Vdash A\} \cap \{A : \dashV A\} = \varnothing$ (there are no gluts) and $\{A : \Vdash A\} \cup \{A : \dashV A\} = S$ for S is the set of all sentences of the language (there are no gaps), then we need only one of either $\Vdash$ or $\dashV$.

negation can be relativized to worlds for languages with modal operators. We are rather claiming that an account of *negation* should not be committed to worlds.

The following facts hold for $\sim$.

- For any A, $\sim\sim A \models A$ and $A \models \sim\sim A$;
- All the de Morgan laws are valid;
- For some A and B, $A \not\models \sim A$ and $\sim B \not\models B$.[20]

The last condition has been claimed by Wolfgang Lenzen (Lenzen, 1996) and João Marcos (Marcos, 2005) to be a necessary condition on negation.

In the modal account of negation, intuitionistically valid de Morgan laws are obtained once symmetry of R together with usual truth conditions for conjunction and disjunction are assumed, and the remaining law becomes valid after adding a further condition on R. In contrast, the de Morgan laws on our account corresponds to the usual falsity conditions for conjunction and disjunction. That is, we obtain the de Morgan laws if and only if the following conditions hold:

- $A \wedge B$ is false iff A is false or B is false;
- $A \vee B$ is false iff A is false and B is false.

Similarly, once we enrich the language with additional connectives, the behavior of the connectives with respect to our negation is completely determined by the falsity conditions for the connectives in question. For example, how a conditional $\rightarrow$ interacts with $\sim$ depends only on the falsity conditions assigned to $\rightarrow$. We find this a pleasing result.

6 Some examples and observations

We now turn to the original Quinean problem of "change of logic, change of subject". Which logics on our account can be said to have genuine negations? In other words, amongst which rival logicians can there be said to be genuine disagreement over the nature of negation?

6.1 Classical negation

Dialetheists such as Graham Priest famously argue that classical negation is unintelligible. He says e.g. that "[i]f one takes it that a dialetheic solution to the semantic paradoxes is correct, one must deny the coherence of Boolean negation" (Priest, 2006a, p.88).[21] We do not think matters are straightforward here, but it is, in any case, worth noting that classical negation can be *conservatively* added to paraconsistent logics such as **BD**, the relevant logic **R**, and even Priest's preferred propositional logic **LP**. The same is true concerning richer theories such as a naive theory of truth or of sets, provided care is taken in formulating delicate principles like naive comprehension or the T-schema.[22] We have not yet found any compelling reason to question the coherence of classical negation, and indeed we find it a legitimate concept with interesting implications for dialetheic solutions

[20] As with compatibility semantics, semantic consequence is defined in terms of the preservation of truth over $\Vdash$.

[21] See (Priest, 2006a, Chapter 5) for a sustained attack on classical negation.

[22] See (Omori, 2015) concerning these matters.

to the paradoxes. It should be no surprise, then, that it turns out to be a contradictory-forming operator.

Suppose we define a notion of contradictoriness just like the one given in Definition 5 except that we replace there 'false' by 'untrue', where it is assumed that truth and untruth are exclusive and exhaustive. Call the notion *C-contradictoriness*. Then we can go on to define classical negation as follows.

[Classical negation] Let A and B be sentences. Then, B is the *classical negation* of A iff A and B are C-contradictories. In the following, we write B as $\neg A$. The above is equivalent to the beloved and familiar definition of classical negation: $\neg A$ is true iff A is untrue. It follows that A negation is classical iff untruth and falsity are identified.

We take it to be quite controversial to assume that falsity is untruth. There are a number of reasons for rejecting such an identification that we won't delve into here (though in passing we mention untrue sentences that are intuitively not false because they are grammatically defective, contain an irreferential singular term, or express a gappy proposition or no proposition at all.) We think truth and falsity are equally *sui generis* notions and should therefore be treated as such.

6.2 Intuitionistic "negation"

As is well-known, intuitionistic "negation" does not satisfy DNE, and so it is not a contradictory-forming operation. It is therefore *not* a negation on our account. Given that constructivism is a plausible view, and more so on the assumption that it can make genuinely negative claims,[23] we think it would be an unfortunate state of affairs if intuitionistic logic couldn't be expanded to include legitimate negation. Fortunately it can.

If truth and falsity are exclusive but not exhaustive, then the constructive negation of **N3** (sometimes called 'strong negation' or 'constructible falsity', cf. (Nelson, 1949)) turns out a legitimate negation on our account.[24] And if truth and falsity are neither exclusive nor exhaustive, then the negation of $\mathbf{N4}^{\perp}$ (cf. (Odintsov, 2008)) turns out legitimate on our account. So intuitionstic logic, though lacking a negation of its own, can be conservatively extended by a legitimate negation in our sense, viz. the well-known strong negations of Nelson logic.[25]

What, then, should we call the more familiar intuitionistic "negation" often defined as the implication to absurdity, if it not a negation? We think it more appropriate to call such an operator a *negative modality* since it is an intensional or modal operator sharing certain characteristics with negation. Since it is not a contradictory-forming operator, however, it is not, strictly speaking, a negation.

There is another negation in the context of intuitionistic logic that makes for an interesting test case. In (De, 2013) and (De and Omori, 2014), intuitionistic logic is extended by a weak empirical negation for expressing sentences of the form 'A has not yet been decided', a sentence which is not expressible using the usual, stronger intuitionistic negation. Let $M = \langle W, \sqsubseteq, a, V\rangle$ be a pointed Kripke structure for intuitionistic logic with $\sqsubseteq$ a preorder on W and $a \in W$ the point. Then the empirical negation $\sim A$ of A is true at a world iff A is true at the point:

$$\text{(em}\sim) \qquad x \Vdash \sim A \text{ iff } a \nVdash A.$$

[23] We are not here discounting the school of negationless constructivism.

[24] The truth and falsity conditions for negation will obviously have to be relativized to worlds, assuming a world semantics for **N3**.

[25] It is here interesting to note that we cannot assume truth and falsity to be exhaustive without the conditional collapsing into the classical material conditional, thereby forgoing constructiveness. Failure of excluded middle remains a key feature of constructive logic even when negation is strong.

Truth in a model is there defined as truth at the point, and consequence as preservation of truth in a model.

On this semantics, if a sentence is true, its negation is false (i.e. untrue), and if it is false, its negation is true, and conversely. On our account, empirical negation is then a legitimate negation. Indeed, it is indistinguishable from classical negation construed as a contradictory-forming operator: both are contradictory-forming and falsity is identified with untruth. But empirical negation *is* distinct from the familiar classical negation $\neg$ defined by

$$x \Vdash \neg A \text{ iff } x \nVdash A,$$

since classical negation collapses the logic to classical logic whereas $\sim$ does not; indeed, it conservatively extends intuitionistic logic. This is true despite the fact that $\sim A$ and $\neg A$ are *logically equivalent*: one is true in a model iff the other is. The catch, clearly, is that the logic so defined is not closed under substitution of provable equivalents as the two negations come apart when embedded under other connectives such as the conditional.

6.3 Other logical systems

One of the most well-known paraconsistent logics is the three-valued logic **LP**. As we saw earlier, its alleged negation is illegitimate according to the compatibility account, due to the failure of (Contra). It is, however, a legitimate negation on our account, vindicating genuine disagreement between dialetheist and classicist, contra a host of naysayers.[26]

Sometimes, **LP** is compared to Strong Kleene Logic ($\mathbf{K}_3$) because the former is obtainable from the latter by taking the intermediate value of $\mathbf{K}_3$ to be designated. Note that the negation of $\mathbf{K}_3$ is deemed a negation on Berto's account since it can be given a compatibility semantics via the Routley star.[27] Berto's account of negation seems therefore sensitive to which values are taken as designated. We find it curious that one and the same operator should and should not count as a negation depending on which values are taken to be truth-like, a curiosity that does not arise on our account.[28]

Note also that if we consider our account of negation in the context of **BD** and its extensions, it follows that de Morgan negation is *the* negation satisfying Definition 5. We find this a nice result since in many cases it legitimizes talk of *the* negation of a sentence. However, when we add other negations to the language such as classical negation, they

[26] Such naysayers include e.g. David Lewis (Lewis, 1982) and Hartley Slater (Slater, 1995). Note that Slater rejects the paraconsistent negation of **LP** as being genuine on the grounds that is not contrary-forming, but in so doing he presupposes a classicist understanding of contrariety. In response to Slater, Jean-Yves Béziau (Béziau, 2006) agrees with Slater in that no paraconsistent negation can be a contradictory-forming operator, but disagrees with the claim that negation must be a contradictory-forming operator. The reason for this is that according to Béziau's definition of a contradictory-forming operator, only classical negation is contradictory-forming. This illustrates a critical point of departure from our view, according to which some paraconsistent negations are contradictory-forming.

[27] The Routley star semantics for $\mathbf{K_3}$ is obtainable by adding one further condition to the Routley star semantics for **BD**. See (Priest, 2008, Chapter 8) for details.

[28] One could deny that the many-valued semantics is relevant here when negation is given a Routley star reading, and that the star semantics takes precedence over the many-valued one. The compatibility account is then not sensitive to which values are taken to be designated, because the correct semantics does not employ three or more values and the notion of a designated value apart from truth. As we see it, however, compatibility semantics should tell us when an operator, *inferentially construed*, is a negation or not: it is iff it satisfies the core principles given by the compatibility account. The problem is that this conflicts with the intuitive idea that if such an operator has an intuitive semantics relative to which it corresponds to some operation o, then o too should count as a negation *no matter which values are taken as designated* (equivalently, no matter whether we think of truth and falsity as being exclusive or exhaustive). This conflict simply does not arise on our account.

may not be uniquely determined according to Definition 5 when falsity is understood in the appropriate sense—for classical negation, that means understanding falsity as untruth. In such a case one logic may be said to have a multiplicity of legitimate negations.[29]

According to our account of negation as a contradictory-forming operator, some paraconsistent logics lack negation. This follows from the fact that at least one of the laws of double negation fails for these logics. Such systems include the C-systems of da Costa and **CLuN** of Batens (Batens, 1980). Note here that negation is *required* in stating paraconsistency, so any logic lacking a genuine negation cannot be paraconsistent by definition. We can obtain a genuine negation for these negationless logics by adding either Double Negation Introduction (DNI) or Double Negation Elimination (DNE). For example, **CLuN** has a genuine negation when we add both (DNE) and (DNI). Note also that some of the Brazilian and Belgian systems already have legitimate negations on our account, such as **LFI1** (Carnielli et al., 2000) and **CLuNs** (Batens and De Clercq, 2004), being equivalent expansions of **LP** whose negation is genuine by our lights.

7 Concluding remarks

We have argued against compatibility semantics as an account of negation that does sufficient justice to meeting the Quinean challenge and to providing an intuitively plausible account of our understanding and usage of negation in natural language. We think negation as a contradictory-forming operator better meets Quine's challenge and is more intuitive in terms of how we understand and evaluate negated sentences. The reason we think it better meets Quine's challenge lies in the fact that, what accounts for shared meaning amounts rival logicians is just how much they agree on concerning the *definition* and nature of negation. Where rivals diverge comes down to what might be considered the lowest level at which disagreement could arise, viz. the matter of whether or not truth and falsity are exclusive or exhaustive. On the compatibility account, disagreement can arise over less fundamental matters such as those concerning the nature of worlds and what compatibility *between worlds* amounts to. On these matters there is simply much more to disagree about, leaving open the question as to whether rivals can be said to share sufficient meaning to allow for genuine disagreement. If two parties to an alleged debate mean something substantially different by their respective uses of 'world' (are they maximal, prime, concrete, self-compatible, gappy, glutty, constructions out of linguistic or informational entities. . . ?), is it really all that clear that they are not speaking past each other despite their agreement on S∼ as a definition of negation? We have our doubts.

References

Batens, D. (1980). Paraconsistent extensional propositional logics. *Logique et Analyse*, 90–91:195–234.

Batens, D. and De Clercq, K. (2004). A rich paraconsistent extension of full positive logic. *Logique et Analyse*, 185-188:227–257.

Beall, J. (2009). *Spandrels of Truth*. Oxford University Press.

Belnap, N. (1976). How a computer should think. In Ryle, G., editor, *Contemporary aspects of philosophy*, pages 30–55. Oriel Press.

[29] For more details on this, see (De and Omori, 2015).

Berto, F. (2015). A modality called ‘negation’. *Mind*, 124(495):761–793.

Béziau, J.-Y. (2006). Paraconsistent logic! (a reply to Slater). *Sorites*, 17:17–25.

Birkhoff, G. and von Neumann, J. (1936). The logic of quantum mechanics. *Annals of mathematics*, 37(4):823–842.

Carnielli, W., Marcos, J., and de Amo, S. (2000). Formal inconsistency and evolutionary databases. *Logic and Logical Philosophy*, 8:115–152.

Copeland, B. J. (1983). Pure semantics and applied semantics. *Topoi*, 2(2):197–204.

da Costa, N. C. (1974). On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, 40(4):497–510.

De, M. (2011). *Negation in context*. PhD thesis, University of St Andrews, Scotland.

De, M. (2013). Empirical negation. *Acta Analytica*, 28(1):49–69.

De, M. and Omori, H. (2014). More on empirical negation. In *Advances in modal logic*, volume 10, pages 114–133. College publications.

De, M. and Omori, H. (2015). Classical negation and expansions of Belnap-Dunn logic. *Studia Logica*, 103(4):825–851.

Dunn, J. M. (1999). A comparative study of various model-theoretic treatments of negation: a history of formal negation. In Wansing, H., editor, *What is Negation?*, pages 23–51. Kluwer Academic Publishers.

Goldblatt, R. I. (1974). Semantic analysis of orthologic. *Journal of Philosophical Logic*, 3(1):19–35.

Goodman, N. (1949). On likeness of meaning. *Analysis*, 10(1):1–7.

Gurevich, Y. (1977). Intuitionistic logic with strong negation. *Studia Logica*, 36(1/2):49–59.

Horn, L. R. and Wansing, H. (2015). Negation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/negation/, Spring 2015 edition.

Lenzen, W. (1996). Necessary conditions for negation operators. In Wansing, H., editor, *Negation: A Notion in Focus*, Perspective in Analytical Philosophy, pages 37–58. Walter de Gruyter.

Lewis, D. (1982). Logic for equivocators. *Noûs*, 16:431–441.

Marcos, J. (2005). On negation: Pure local rules. *Journal of Applied Logic*, 3:185–219.

Nelson, D. (1949). Constructible falsity. *Journal of Symbolic Logic*, 14:16–26.

Odintsov, S. P. (2008). *Constructive Negations and Paraconsistency*. Dordrecht: Springer-Verlag.

Omori, H. (2015). Remarks on naive set theory based on **LP**. *The Review of Symbolic Logic*, 8(2):279–295.

Price, H. (1990). Why 'not'? *Mind*, 99:221–238.

Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic*, 8(1):219–241.

Priest, G. (2006a). *Doubt truth to be a liar.* Oxford University Press.

Priest, G. (2006b). *In Contradiction: A Study of the Transconsistent.* Oxford University Press, 2nd edition.

Priest, G. (2008). *Introduction to non-classical logics: from ifs to is.* Cambridge University Press, second edition.

Quine, W. V. (1970). *Philosophy of Logic.* Prentice Hall Inc.

Routley, R. (1980). *Exploring Meinong's jungle and beyond.* McMaster University.

Routley, R. and Routley, V. (1972). The semantics of first-degree entailment. *Noûs*, 6(4):335–390.

Slater, B. H. (1995). Paraconsistent logics? *Journal of Philosophical Logic*, 24(4):451–454.

Wansing, H. (2006). Contradiction and contrariety: Priest on negation. In Malinowski, J. and Pietruszczak, A., editors, *Essays in Logic and Ontology*, pages 81–93. Amsterdam/New York, NY: Rodopi.

Wansing, H. (2016). On split negation, strong negation, information, falsification, and verification. In Bimbó, K., editor, *J. Michael Dunn on Information Based Logics*, pages 161–189. Springer.