PY4669 Modal Logic & Metaphysics

Negation as (In)Compatibility

18th April 2024

I hereby declare that the attached piece of written work is my own work and that I have not reproduced, without acknowledgement, the work of another.

1 Introduction

Negation is an extensional connective in classical logic; in the standard possible world analysis, its truth value is captured in the familiar truth table and evaluated at the located world, without considering what happens in other worlds. However, there is certain negation that does care about the other worlds. For example, in Schrödinger's cat thought experiment, the cat is neither alive nor not-alive (dead) until the box is opened, and an observation is made. Here, the truth value of alive or not-alive at the current world (the box has not been opened) depends on the observation of the next accessible world (the box has been opened).

There have been attempts to treat negation as an intensional, modal connective. For instance, in relevant logic, the negation of A is defined as being true at a world w if and only if A is false at $w^{\times}$, the so-called Routley-Star world. The Routley-Star semantics for negation has been criticised for lacking an intuitively and philosophically appealing interpretation, since it does not provide an explanation of negation until itself is supplemented by an explanation. In contrast, (in)compatibility semantics grounds the meaning of negation in the concepts of compatibility and incompatibility (hereafter abbreviated as (in)compatibility), providing a more satisfactory philosophical explanation for negation from both metaphysical and phenomenological perspectives.

In addition to a more promising philosophical foundation, (in)compatibility semantics also offers theoretical advantages. Berto (2015) defends (in)compatibility semantics by arguing that it vindicates a pluralism of negation. This pluralism is facilitated by allowing us to understand the notion of world in different admissible ways, thereby yielding various definitions of negation. This form of pluralism is formally implemented by specifying the algebraic structure between worlds, such as accessibility relations and the ordering between worlds. For instance, by incorporating principles like Seriality and Star Postulate, we can obtain the Routley-Star negation mentioned previously.

In this essay, I will defend the (in)compatibility semantics for negation as a modal operator. This essay is organised as follows. Section two elucidates the motivations behind grounding negation in

(in)compatibility from linguistic, metaphysical, and phenomenological perspectives. Section three presents the formalisation of this idea and discusses the implications arising from this formal approach. Section four concludes the essay.

2 Motivation

Let us begin by examining the phenomena induced by negation in natural language, particularly focusing on the concept of opposites as identified in linguistics. There are three types of opposites: complementary, gradable, and relational.

Complementary opposites, such as alive and dead, follow a strict entailment pattern: A entails not-B and B entails not-A; moreover, not-A entails B and not-B entails A. This type is characterised by Boolean complementary where the negation of one necessarily implies the presence of the other.

Gradable opposites, like black and white, demonstrate a different pattern: A entails not-B and B entails not-A; but not-A does not entail B, nor does not-B entail A. These opposites exist on a spectrum, indicating that the absence of one does not automatically confirm the presence of its opposite. We can easily imagine a table that is neither black nor white (grey), but if a table is entirely painted white, then we cannot imagine a table that is entirely white also being black at the same time.

Relational opposites, such as parent and child, depend on additional contextual information to define the relationship. For instance, in a room with myself, my father, and my grandfather, my father can be both a parent (to me) and a child (to my grandfather). Here, being a parent does not entail not being a child. However, if we remove the information about the grandfather being in the room, then my father being my parent would entail him not being a child. When only my father and I are in the room, we cannot imagine a scenario where my father is both my parent and my child due to these incompatible roles. However, the introduction of my grandfather allows us to conceive of my father fulfilling both roles.

The examination of the last two opposites suggests that our understanding and use of negation in natural language is closely linked to our perception of compatibility and incompatibility. In other word, our brains initially detect an incompatibility, which is then articulated through negation in natural language. Based on this understanding, let us revisit the first type of opposites. At first glance, "alive or dead" appears to be a pair of complementary opposites, as the state of a living being must reside within one of these two. However, our imagination introduces the concept of a zombie, prompting the question: is a zombie alive or dead? We cannot immediately assert that a zombie is not alive or a zombie is not dead based on intuition alone, because the concept of a zombie does not evoke a sense of incompatibility between being alive and being dead. Similarly, with Schrödinger's

cat, we can conceive of the cat being both alive and dead. Until we open the box to observe, these two states are not incompatible, so we would not immediately claim that the cat is not alive or the cat is not dead.

So far, we have learnt that, in natural language, negation implies an inherently incompatible binary relationship that is connected to psychological conditions. We have ample reason to ground negation in (in)compatibility, viewing it as a psychologically primitive notion. Berto (2015) argues:

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.

He further notes that on the cognitive side, even newborns and animals understand (in)compatibility before they grasp what negation is, hence (in)compatibility is more primitive than negation. De & Omori (2018) agree with this argument but further note that the truth conditions of negation should not be grounded in (in)compatibility. Berto & Restall (2019) respond that the semantics discussed in Berto (2015) do not ground truth conditions in (in)compatibility. The truth conditions are based on our understanding of the notion of the world, which further defines the algebraic structure of modal logic with partial order. Thus, grounding negation in (in)compatibility and its truth conditions in our understanding of the world notion which represented by an algebraic structure, i.e., grounding negation in (in)compatibilities between algebraically structured worlds, is promising in metaphysics perspective.

In a metaphysically oriented view, negation arises from inherent incompatibilities between states of affairs. This concept aligns with classical phenomenologists, who noted experiences of incompatibility as central to understanding negation. Phenomenologically, negation can be viewed as a means of expressing exclusion. Husserl discussed the conflict experienced when two incompatible perceptions occur simultaneously; we experience a sense of conflict when I try to mentally superimpose two colours on the same physical space. Furthermore, as illustrated by Sartre in his depiction of Pierre's absence from a café, negation is intimately tied to experiences rather than merely logical or linguistic constructs. Sartre contrasts different judgments of experience of absence such as 'Pierre is not here' and 'Paul Valéry is no longer here', suggesting that authentic judgments of negation are grounded in genuine experiences of incompatibility.

After finding sufficient philosophical motivation to ground negation in (in)compatibility, Berto & Restall (2018) further assert that because (in)compatibility is modal, negation is a modal operator. It

is not immediately obvious why (in)compatibility should be considered a modality. In brief, when we consider a negated sentence, we are attempting to conceive incompatibility. This process of conceiving involves imagining a set of possible worlds that we take to verify the negated sentence. In doing so, we're essentially exploring the realm of possibility. Since this conceivability entails possibility, we say that (in)compatibility is modal. Since negation is grounded in this (in)compatibility modality, it therefore functions as a modal operator.

3 Formalisation and its implications

We first lay down the mathematical formalisation of the (in)compatibility semantics. The sentential language $\mathcal{L}$ contains 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 $\Box$ and $\neg$, the nullary connectives $\top$ and $\bot$, round brackets as auxiliary symbols used to indicate grouping, and $A, B, C, A_1, A_2, \ldots$ as metavariables for formulae. We also use the set-theoretic notation and symbols as the usual way.

A frame for $\mathcal{L}$ is a quadruple $\mathfrak{F} = \langle W, P, C, \sqsubseteq \rangle$, where W is a nonempty set of worlds, P and C are two accessibility relations on worlds, namely, $P, C \subseteq W \times W$, and $\sqsubseteq$ is a partial ordering on W. We use $x, y, z, x_1, x_2, \ldots$ as metavariables ranging over worlds in W. When $\langle x, y \rangle \in P$ we abbreviate it as xPy and read it as world y is possible relative to world x. When $\langle x, y \rangle \in C$ we abbreviate it as xCy and read it as world x is compatible with world y. $x \sqsubseteq y$ is read as world y retains all the information included in world x. In order for $\sqsubseteq$ to interact with the two accessibility relations as we would expect, the following two conditions are needed. For all $x, y, x_1, x_2 \in W$:

(Forwards) $\quad xPy \text{ and } x_1 \sqsubseteq x \text{ and } y_1 \sqsubseteq y \Rightarrow x_1 P y_1$

(Backwards) $\quad xCy \text{ and } x_1 \sqsubseteq x \text{ and } y_1 \sqsubseteq y \Rightarrow x_1 C y_1$

(Forwards) and (Backwards) both embody the idea that even after a reduction in the amount of information, the two types of accessibility relations should still be maintained. For instance, in (Backwards), if x is compatible with y, then anything ruled out at x is also ruled out at y. Anything ruled out at x is also ruled out at x_1 since x_1 maintains the information from x. Likewise, things being ruled out at y are carried over to y_1 for the same reason. Therefore, nothing that is ruled out at x_1 occurs at y_1, making x_1 compatible with y_1.

A frame becomes a model $\mathfrak{M} = \langle W, P, C, \sqsubseteq, \Vdash \rangle$ if we add an interpretation function to it, satisfying the Heredity Constraint. For each atomic formula $p \in \mathcal{L}_{\mathcal{AT}}$:

(HC) $\quad x \Vdash p \text{ and } x \sqsubseteq y \Rightarrow y \Vdash p$

Here $x \Vdash p$ means that an atomic formula p holds at a world x. The (HC) tell us that if an atomic formula p holds in a world x, then it also holds in any world y that contains x in terms of information. This aligns with our intuition: if I am observing from my dormitory window in St Andrews and see that it is raining, someone observing from Liverpool would not have access to information supporting the fact that it is raining outside my window in St Andrews. However, if observed from the perspective of the entire UK, because the information about the UK includes information about St Andrews, it should also encompass information about the situation outside my dormitory window in St Andrews. It is shown by induction that given (Backwords), (HC) generalises to arbitrary formulae of $\mathcal{L}$ constructed by the semantic clauses for the connectives given as follows. For all $x \in W$:

(S$\wedge$) $$x \Vdash A \wedge B \iff x \Vdash A \text{ and } x \Vdash B$$

(S$\vee$) $$x \Vdash A \vee B \iff x \Vdash A \text{ or } x \Vdash B$$

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

(S$\bot$) $$x \Vdash \bot$$

(S$\Box$) $$x \Vdash \Box A \iff \forall y \in W(xPy \Rightarrow y \Vdash A)$$

(S$\neg$) $$x \Vdash \neg A \iff \forall y \in W(xCy \Rightarrow y \nVdash A)$$

and (HC) generalises; for each formula A of $\mathcal{L}$ and all $x, y \in W$:

(HC') $$x \Vdash A \text{ and } x \sqsubseteq y \Rightarrow y \Vdash A$$

In (S$\neg$), $y \nVdash A$ means that the formula A does not hold at the world y. Intuitively, (S$\neg$) states that a sentence's negation $\neg A$ is true in a world x if and only if that sentence A is **not** true in all worlds that are compatible with world x. De & Omori (2018) note that the definition of the truth condition of negation, as presented in (S$\neg$), employs negation itself, specifically on the right-hand side of the iff in the form of $\nVdash$. They think this apparent circularity in the definition is problematic. Upon contrasting (S$\neg$) with (S$\wedge$), they observe that although the definition of the truth condition of conjunction similarly incorporates the term "and", it corresponds to the conjunction device in our natural language, not the logical operator being defined. In contrast, in (S$\neg$), the negation used within the definition appears to refer to the negation operator that is being defined. Berto & Restall (2019) reference Tarski's truth schema to argue that inductive definitions such as (S$\neg$) and (S$\wedge$) are valid. They further compare (S$\neg$) with (S$\Box$) that also presents in the standard possible world analysis, pointing out that the definition of modality in (S$\Box$) involves the notion of possibility through the possible relations among worlds, and this is the same pattern used in defining (S$\neg$). However, there appears to be a formal difference between (S$\neg$) and (S$\Box$); the former contains $\nVdash$, while the latter does not. Berto & Restall further clarify that (S$\neg$) can be reformulated from a compatibility-based form

to an incompatibility-based, and the latter aligns closely with the form of (S$\Box$):

$$(S^1\neg) \qquad x \Vdash \neg A \iff \forall y \in W(y \Vdash A \Rightarrow xIy)$$

and the reading of this dual is that a sentence's negation $\neg A$ is true in a world x if and only if any worlds that makes sentence A true are incompatible with world x. The advantage of this form is that it integrates the negation from $\nVdash$ into the accessibility relation, thus encapsulating the contentious use of negation, as highlighted by De & Omori, within the controversial concept of compatibility, transforming it into the singularly contentious notion of incompatibility. On the other hand, the duality of compatibility and incompatibility demonstrated by Berto & Restall also addresses the concerns raised by De & Omori regarding which of the two notions, compatibility or incompatibility, is primitive, and one can be defined by the other. Berto & Restall suggest that these are merely two perspectives, analogous to the wave-particle duality in physics, representing different manifestations of the same underlying reality. This explanation clarifies why we use the term (in)compatibility to refer to both compatibility and incompatibility concurrently, as under the framework of (in)compatibility semantics, they converge into a single notion.

Finally, we define logical consequence in a frame $\mathfrak{F}$ as truth preservation across all worlds x in all the relevant models based on frame $\mathfrak{F}$. Given a set Σ of formulae:

$$\Sigma \vDash A \iff \forall \mathfrak{M} \text{ on } \mathfrak{F} \ (\forall x \in W \in \mathfrak{M} \ (\forall B \in \Sigma \ (x \Vdash B) \implies x \Vdash A))$$

In other words, a set Σ of formulae entails A if and only if for every model $\mathfrak{M}$ based on frame $\mathfrak{F}$ (namely, for every admissible interpretations) and for every world x of W of $\mathfrak{M}$, A holds at x whenever every formula of Σ holds at x.

3.1 Inauthentic negations

Since negation is grounded in (in)compatibility under this semantics, the minimal logical properties that negation possess hinge on the minimal constraint that (in)compatibility satisfies and also on those of the information-inclusion relation, $\sqsubseteq$. In Berto (2015), it is argued that (in)compatibility must be symmetric. If world x is incompatible with world y, then world y must also be incompatible with world x. This symmetry is justified by the intuition that if the information contained in world x renders it incompatible with some information in world y, then that particular information in world y must also be incompatible with the conflicting information in world x. Similarly, if the information in world x is compatible with the information in world y, then the information in world y will not conflict with that in world x, making world y also compatible with world x. Hence, (in)compatibility relation is inherently symmetric. Berto proved that if the (in)compatibility relation is symmetric,

then the semantics must validate Double Negation Introduction:

(DNI) $$A \models \neg\neg A$$

On the other hand, Berto points out that the compatibility relation should not be reflexive,

(Reflexivity) $$\forall x(xCx)$$

because having a reflexive compatibility relation would lead our semantics to validate the infamous Explosion Principle or Ex Contradictione Quodlibet:

(ECQ) $$A \wedge \neg A \models \bot$$

which is rejected by paraconsistent logicians like Berto & Restall. The idea that the compatibility relation is not reflexive, meaning that a world is not compatible with itself, might initially seem counterintuitive. However, there is a compelling intuition to support this viewpoint. Consider a scenario where an intelligence agency maintains a central database into which information gathered by its spies is entered. Subsequent automated reasoning processes on this database yield inferences that assist with intelligence operations. It is conceivable that two spies could submit entirely incompatible pieces of information into this database. Over time, this database might become saturated with incompatible intelligence. Nonetheless, reasoning within this database of contradictory information should not lead to any conclusions. Our rejection of the notion that a world is compatible with itself stems from the paraconsistent logicians' understanding of the notion of world, which they consider to be filled with various incompatible pieces of information. This understanding of the world under (in)compatibility semantics effectively invalidates (ECQ), aligning with paraconsistent expectations for reasoning; they seek a logical system that facilitates sound reasoning in a world perennially brimming with incompatible information.

In Berto (2015), it is also demonstrated that within our semantics, negation must satisfy Contraposition:

(Contraposition) $$A \models B \Rightarrow \neg B \models \neg A$$

Therefore, in our semantics, where negation is grounded on the binary relation of (in)compatibility that is symmetric but not reflexive, nothing can qualify as a negation unless it satisfies both (DNI) and (Contraposition). Berto (2015) illustrates that da Costa negation cannot be considered a negation within our semantics because it fails to satisfy (Contraposition). Surprisingly, Routley-star negation qualifies as a negation under this framework. When we incorporate seriality in the (in)compatibility

relation

(Seriality) $\forall x \exists y (xCy)$

and add Star Postulate

(Star Postulate) $\forall x \exists y (xCy \wedge \forall z (xCz \Rightarrow x \sqsubseteq y))$

into our semantics, (S¬) can be simplified to:

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

While we struggle to find an intuitive explanation for the Routley-star world x^*, the inclusion of non-intuitive inferences like this Routley-star negation within our semantics should not be viewed as a flaw. Instead, this characteristic of the semantics represents an embodiment of *negation pluralism*.

3.2 Negation pluralism

Negation pluralism is comprised of the following claims:

(1) (S¬).

(2) (S¬) is schematic because the notion of world is ambiguous there, and an instance of (S¬) comes from disambiguate what is world by restricting accessibility relations $\langle P, C \rangle$ and information-inclusion relation $\sqsubseteq$.

(3) Admissible instances of (S¬) satisfy (DNI) and (Contraposition) but dissatisfy (ECQ).

(4) A theory of negation is given by an admissible instance of (S¬).

(5) There is more than one admissible instance of (S¬).

This was proposed in Berto (2015), adapting from Beall-Restall's logic pluralism. It is noted that in the formalisation of (in)compatibility semantics, the notion of a "world" is frequently mentioned but not formally defined. Up to this point, we have been using the term "world" to refer to elements in the set WW, but these elements are actually just points defined within the algebraic structure of (in)compatibility semantics. These points are connected by two accessibility relations and an information-inclusion relation, providing the basic requirements for expressing negation as (in)compatibility. We are still unclear about the ultimate form of negation based on (in)compatibility, as our understanding of the relationships between points in the algebraic structure is still evolving. (In)compatibility

semantics offers us a canvas upon which we can sketch a more detailed form of negation grounded in (in)compatibility, by defining more concretely the algebraic relations between these points.

While this canvas of (in)compatibility semantics showcases numerous possible forms of negation such as the aforementioned Routley-star negation, it does not assert that there must be multiple negations. Agreeing with the idea of negation as (in)compatibility, even negation monism can use this canvas to describe their One True negation. Such flexible semantic machinery provides a stable starting point for the development of negation theory. As mentioned in Berto (2015), the primary purpose of proposing this semantic machinery is not to suggest what he believes the formal aspects of the correct negation should be, but to provide a fundamental point of dialogue for philosophers and logicians who often argue incessantly due to a lack of a unified theoretical foundation and misunderstandings of each other's concept of negation.

4 Conclusion

The (in)compatibility semantics offers a compelling framework for understanding negation as a modal operator, grounded in (in)compatibility. This approach aligns with both metaphysical and phenomenological perspectives, providing a philosophically satisfying explanation for the nature of negation. Through mathematical formalisation, the semantics reveal minimal properties of negation while accommodating negation pluralism, allowing for multiple admissible instances of negation within the framework. By providing a stable foundation for exploring the diverse facets of negation theory, (in)compatibility semantics offers a promising avenue for further philosophical and logical inquiry into the nature and function of negation in our cognition and natural language.

Bibliography

Berto, F. (2015). A modality called ?negation? *Mind*, *124*(495), 761–793. https://doi.org/10.1093/mind/fzv026

Berto, F., & Restall, G. (2019). Negation on the australian plan. *Journal of Philosophical Logic*, *48*(6), 1119–1144. https://doi.org/10.1007/s10992-019-09510-2

De, M., & Omori, H. (2018). There is more to negation than modality. *Journal of Philosophical Logic*, *47*(2), 281–299. https://doi.org/10.1007/s10992-017-9427-0

Horn, L. R., & Wansing, H. (2022). Negation. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2022). Metaphysics Research Lab, Stanford University.

Kinkaid, J. (2020). What would a phenomenology of logic look like? *Mind*, *129*(516), 1009–1031. https://doi.org/10.1093/mind/fzaa031