Non-Normal Worlds and Representation

Francesco Berto*

Abstract

World semantics for relevant logics include so-called *non-normal* or *impossible* worlds providing model-theoretic counterexamples to such irrelevant entailments as $(A \wedge \neg A) \rightarrow B$, $A \rightarrow (B \vee \neg B)$, or $A \rightarrow (B \rightarrow B)$. Some well-known views interpret non-normal worlds as information states. If so, they can plausibly model our ability of *conceiving* or *representing* logical impossibilities. The phenomenon is explored by combining a formal setting with philosophical discussion. I take Priest's basic relevant logic N_4 and extend it, on the syntactic side, with a representation operator, Ⓡ, and on the semantic side, with particularly anarchic non-normal worlds. This combination easily invalidates unwelcome "logical omniscience" inferences of standard epistemic logic, such as belief-consistency and closure under entailment. Some open questions are then raised on the best strategies to regiment Ⓡ in order to express more vertebrate kinds of conceivability.

1 Overview

Relevant logics are perhaps the most developed among paraconsistent logics, these being logical systems rejecting the principle *Ex contradictione quodlibet* (ECQ), according to which a contradiction entails everything (in 'object language' version, $(A \wedge \neg A) \rightarrow B$). Arguably, the most discussed kinds of formal semantics for relevant logics are world semantics. As specialists know, these include so-called *non-normal* or *impossible* worlds, often thought of as situations where the truth conditions of logical operators are different.

*Thanks to the audience at the Logica 2011 conference for comments and helpful feedback.

Non-normal worlds are crucial for providing model-theoretic counterexamples to ECQ as well as to other irrelevant entailments, such as $A \to (B \vee \neg B)$ and $A \to (B \to B)$.[1] They can thus help in modeling our capacity of reasoning non-trivially also in the face of inconsistent information. And such a capacity is widely attested, thus providing counterexamples to ECQ. For an often mentioned case: Bohr's atomic theory includes both the assumption that energy has the form of quanta, that is, discrete packs, and Maxwell's usual electromagnetic equations, which are inconsistent with that assumption.[2] Nevertheless, Bohr provided quite a successful theory. More importantly for our purposes: he did not infer arbitrary conclusions from his inconsistent assumptions – for instance, that electrons have the same electric charge as protons.

The main philosophical issue concerning world semantics for relevant logics has traditionally been the one of the intuitive reading of its worlds, and of the relations and operations defined on them. Some well-known views interpret these precisely as information states, or conduits thereof (see e.g. (Mares, 2004)). Given such an epistemically-driven reading, non-normal worlds may model our ability of *conceiving* or *representing* inconsistencies and broadly logical impossibilities. This is tightly connected to our aforementioned capacity of reasoning efficiently in inconsistent informational circumstances – if not a precondition of it. As Bohr knew he was making incompatible assumptions in his theory, for instance, he was arguably able to conceive those inconsistent suppositions as holding together. This did not lead him astray, though.

Supposing the non-normal worlds of relevant semantics are essen-

[1] Intuitively, a premise or conditional antecedent is irrelevant within an inference or a conditional, if it is of no utility in getting to the conclusion, or in grounding the consequent. The research program of relevant logic is based on the positive view that the intuition of relevance can be given formal substance, together with the negative view that classical logic legitimates irrelevant inferences – on the ground, for instance, of its admitting logically valid conditionals with no content connection between antecedent and consequent. At least part of the formal substance to the idea of relevance as content-connection is provided by the so-called Variable Sharing Property (VSP), also called weak or necessary condition of relevance. As far as the conditional goes, this states that if $A \to B$ is logically valid, then A and B must share some sentential variable. On this ground, ECQ and the two aforementioned formulas count as fallacies of relevance, not passing the VSP test. For a short and accessible introduction to relevant logic, see (Mares, 2004).

[2] For an account of this story, see (Brown, 1993).

tially realizations of *intentional* states, such as conceiving or representing,[3] this paper explores the phenomenon by combining a formal setting with philosophical discussion.

I proceed as follows: in section 2, I introduce the syntax of a first-order intensional language L and, in section 3, I present a model-theoretic semantics for it, which draws upon the relevant logic N_4 proposed in chapter 9 of (Priest, 2001). This combines techniques of many-valued and modal logics, including locally inconsistent and incomplete non-normal worlds, but has standard definitions of logical consequence and validity. Despite being simpler than the mainstream world semantics for relevant logics, the N_4 setting allows to model all the fetures of relevant systems that are significant for our purposes in a friendly formal setting. The language includes a representation operator, whose role is to capture our capacity of representing or conceiving inconsistencies and logical impossibilities.

Section 4 provides a brief discussion of the distinction, embedded in the model, between two kinds of non-normal worlds, displaying different degrees of logical lawlessness and labeled, for reasons to be explained, as *extensionally* and *intensionally* impossible worlds.

In section 5, it is shown that the semantics makes of L's conditional a fully relevant (albeit weak) one, invalidating the fallacies of relevance and, in particular, ECQ.

Section 6 explains how the representation operator invalidates (the formulations in terms of it of) typical unwelcome inferences of epistemic logic gathered under the rubric of "logical omniscience", such as belief-consistency and closure under entailment. It is well known that logical omniscience phenomena make for highly idealized epistemic notions, not mirroring the actual condition of human beings as finite, fallible, and occasionally inconsistent cognitive agents. If we can conceive inconsistencies and impossibilities, non-normal worlds are natural candidates to model this condition: the content of a representational state is the set of worlds that make the representation true, that is, where things are as they are conceived or represented to be. This may include non-normal worlds where those inferences fail.

[3] I employ these two terms as generics for a range of broadly cognitive human activities, all involving the depiction of scenarios, situations, or circumstances, which count as their contents. I take a dim view on such intentional phenomena, and leave their serious investigation to philosophers of mind, cognitive scientists, or neuroscientists.

Finally, some open questions are raised in section 7, as to the best strategies to regiment the representation operator in order for it to express specific and more vertebrate kinds of conceivability; these should intuitively be closed under some (albeit weaker-than-classical) logical consequence relation and, more importantly, allow for *ceteris paribus* import of information from actuality.

2 Syntax of L

L consists of a fully standard first-order vocabulary with individual variables x, y, z (and, if more are needed, indexed ones, $x_1, ..., x_n$); individual constants: m, n, o (if more are needed, $m_1, ..., m_n$); n-place predicates: F, G, H ($F_1, ..., F_n$); the usual connectives, negation $\neg$, conjunction $\wedge$, disjunction $\vee$, the conditional $\rightarrow$; the two quantifiers, $\forall$ and $\exists$; the two standard alethic modal operators for necessity $\Box$ and possibility $\Diamond$; a unary sentential operator Ⓡ; round brackets as auxiliary symbols. Individual constants and variables are singular terms. If $t_1, ..., t_n$ are singular terms and P is any n-place predicate, $Pt_1, ..., t_n$ is an atomic formula. If A and B are formulas, $\neg A$, $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$, $\Box A$, $\Diamond A$ and ⓇA are; outermost brackets are normally omitted in formulas. If A is a formula and x is a variable, then $\forall xA$ and $\exists xA$ are formulas, closed and open formulas having their standard definitions.

The only piece of notational novelty is Ⓡ, which I shall call the *representation* operator. The intuitive reading of 'ⓇA' will be "It is represented that A", or "It is conceived that A".

3 Semantics for L

The semantics for L is largely down to Priest's work in non-standard intensional logic (see (Priest, 2001), (Priest, 2005)), with a few modifications. An interpretation is an ordered septuple $\langle \mathrm{P}, \mathrm{I}, \mathrm{E}, @, R, \mathrm{D}, v \rangle$, the intuitive reading of whose members is as follows. P is the familiar set of possible worlds; I and E are two sets of non-normal or impossible worlds of two kinds, the *intensionally* and *extensionally* impossible ones respectively (what this means, we will see soon); P, I and E are disjoint, $\mathrm{W} = \mathrm{P} \cup \mathrm{I} \cup \mathrm{E}$ is the totality of worlds *simpliciter*. @ is the obtaining world (or, better, its foster in the formalism). I assume, for prudence, that $@ \in \mathrm{P}$. $R \subseteq \mathrm{W} \times \mathrm{W}$; if $\langle w_1, w_2 \rangle \in \mathrm{R}$ ($w_1, w_2 \in \mathrm{W}$), I write this as '$w_1 R w_2$' and claim that w_2 is *representationally accessible* (R-accessible), from w_1 (what this means, we will also see soon). D is a non-empty set of objects. v is a function assigning denotations to the descriptive constant symbols of L, as follows:

If c is an individual constant, $v(c) \in \mathrm{D}$.
If P is an n-place predicate and $w \in \mathrm{W}$, $v(P, w)$ is a pair: $\langle v^+(P, w), v^-(P, w) \rangle$, with $v^+(P, w) \subseteq \mathrm{D}^n, v^-(P, w) \subseteq \mathrm{D}^n$.

$\mathrm{D}^n = \{\langle \mathrm{d}_1, ..., \mathrm{d}_n \rangle | \ \mathrm{d}_1, ..., \mathrm{d}_n \in \mathrm{D}\}$, and $\langle \mathrm{d} \rangle$ is stipulated to be just d, so D^1 is D. To each pair of n-place predicate P and world w, v assigns a (positive) extension $v^+(P, w)$ and an anti-extension or negative extension, $v^-(P, w)$. The extension of P at w is to be thought of as the set of (n-tuples of) things of which P is true there, the anti-extension as the set of (n-tuples of) things of which P is false there. Such double extensions are to model inconsistencies – things being both true and false (truth value gluts; or also, neither true nor false – truth value gaps). On the other hand, one may sensibly want truth and falsity to be exclusive and exhaustive at possible worlds (this is part of what makes them possible, after all). We can recover the classical setting by imposing the following double clause – let us call it the Classicality Condition:

(CC) If $w \in \mathrm{P}$, for any n-ary predicate P: $v^+(P, w) \cap v^-(P, w) = \emptyset$, $v^+(P, w) \cup v^-(P, w) = \mathrm{D}^n$.

At possible worlds, extensions and anti-extensions are exclusive and exhaustive. We need the usual assignments of denotations to variables. If a is an assignment (a map from the variables to D), then v_a is the suitably parameterized denotation function, so that we have denotations for all singular terms:

If c is an individual constant, $v_a(c) = v(c)$.
If x is a variable, $v_a(x) = a(x)$.

Let us read '$w \Vdash_a^+ A$' as "A is true at world w (with respect to assignment a)", and '$w \Vdash_a^- A$' as "A is false at world w (with respect to assignment a)" (and an interpretation, but I will omit to mention it when no confusion arises). The truth and falsity conditions for atomic formulas are:

$$w \Vdash_a^+ Pt_1...t_n \text{ iff } \langle v_a(t_1), ..., v_a(t_n)\rangle \in v^+(P, w)$$
$$w \Vdash_a^- Pt_1...t_n \text{ iff } \langle v_a(t_1), ..., v_a(t_n)\rangle \in v^-(P, w)$$

The extensional vocabulary has straightforward clauses at all $w \in$ $\mathrm{P} \cup \mathrm{I}$:

$$w \Vdash_a^+ \neg A \text{ iff } w \Vdash_a^- A$$
$$w \Vdash_a^- \neg A \text{ iff } w \Vdash_a^+ A$$

$$w \Vdash_a^+ A \wedge B \text{ iff } w \Vdash_a^+ A \text{ and } w \Vdash_a^+ B$$
$$w \Vdash_a^- A \wedge B \text{ iff } w \Vdash_a^- A \text{ or } w \Vdash_a^- B$$

$$w \Vdash_a^+ A \vee B \text{ iff } w \Vdash_a^+ A \text{ or } w \Vdash_a^+ B$$
$$w \Vdash_a^- A \vee B \text{ iff } w \Vdash_a^- A \text{ and } w \Vdash_a^- B$$

$$w \Vdash_a^+ \forall x A \text{ iff for all } \mathrm{d} \in \mathrm{D},\ w \Vdash_{a(x/\mathrm{d})}^+ A$$
$$w \Vdash_a^- \forall x A \text{ iff for some } \mathrm{d} \in \mathrm{D},\ w \Vdash_{a(x/\mathrm{d})}^- A$$

$$w \Vdash_a^+ \exists x A \text{ iff for some } \mathrm{d} \in \mathrm{D},\ w \Vdash_{a(x/\mathrm{d})}^+ A$$
$$w \Vdash_a^- \exists x A \text{ iff for all } \mathrm{d} \in \mathrm{D},\ w \Vdash_{a(x/\mathrm{d})}^- A$$

'$a(x/\mathrm{d})$' stands for the assignment that agrees with a on all variables, except for its assigning d to x. As for the modals, we have the following for all $w \in \mathrm{P}$:

$$w \Vdash_a^+ \Box A \text{ iff for all } w_1 \in \mathrm{P},\ w_1 \Vdash_a^+ A$$
$$w \Vdash_a^- \Box A \text{ iff for some } w_1 \in \mathrm{P},\ w_1 \Vdash_a^- A$$

$$w \Vdash_a^+ \Diamond A \text{ iff for some } w_1 \in \mathrm{P},\ w_1 \Vdash_a^+ A$$
$$w \Vdash_a^- \Diamond A \text{ iff for all } w_1 \in \mathrm{P},\ w_1 \Vdash_a^- A$$

(Unrestricted) necessity/possibility is truth at all/some *possible* world(s) (I am not making much use of the box and diamond in this work, but they can be usefully contrasted, within the model, with the behaviour of the representation operator). While we have the normal material conditional, say $A \supset B =_{df} \neg A \vee B$, our more vertebrate intensional conditional is the following. At all $w \in \mathrm{P}$:

$$w \Vdash_a^+ A \rightarrow B \text{ iff for all } w_1 \in \mathrm{P} \cup \mathrm{I} \text{ such that } w_1 \Vdash_a^+ A,\ w_1 \Vdash_a^+ B$$
$$w \Vdash_a^- A \rightarrow B \text{ iff for some } w_1 \in \mathrm{P} \cup \mathrm{I},\ w_1 \Vdash_a^+ A \text{ and } w_1 \Vdash_a^- B$$

So far everything works familiarly enough as far as worlds in P are concerned, the main change with respect to standard modal semantics being that truth and falsity conditions are spelt separately. But even this does not change much at possible worlds. The CC dictates that, at each $w \in \mathrm{P}$, any predicate is either true or false of the relevant object (or n-tuple thereof), but not both. That no atomic formula is both true and false or neither true nor false entails that no formula is, as can be checked recursively. Overall, there are no truth value gluts or gaps at possible worlds.[4] In particular, for instance, if $w \in \mathrm{P}$ then $w \Vdash_a^+ \neg A$ if and only if it is not the case that $w \Vdash_a^+ A$: at possible worlds negation works "homophonically", the classical way. And since $@ \in \mathrm{P}$, truth *simpliciter*, truth at the actual world, behaves in an orthodox way with respect to negation.

Things get more exciting at non-normal worlds. At points in I, v treats formulas of the form $A \rightarrow B$, $\Box A$, and $\Diamond A$ essentially as *atomic*: their truth values are not determined recursively, but directly assigned by v in an arbitrary way. At points in E, all formulas can be treated as atomic and behave arbitrarily: $A \vee B$ may turn out to be true even though both A and B are false, etc.[5] Hence the denominations for the

[4]A technical note which may be skipped without loss of continuity. Because of the world quantifiers in the clauses for $\rightarrow$ ranging on $\mathrm{P} \cup \mathrm{I}$, one needs, in fact, a couple of extra assumptions on the falsity conditions for $A \rightarrow B$ to rule out gaps and gluts from possible worlds, specifically: if $w \in \mathrm{P}$, then $w \Vdash_a^- A \rightarrow B$ iff: (1) For some $w_1 \in \mathrm{P} \cup \mathrm{I}$, $w_1 \Vdash_a^+ A$ and $w_1 \Vdash_a^- B$ and it is not the case that $w \Vdash_a^+ A \rightarrow B$; (2) (For some $w_1 \in \mathrm{P} \cup \mathrm{I}$, $w_1 \Vdash_a^+ A$ and $w_1 \Vdash_a^- B$) or it is not the case that $w \Vdash_a^+ A \rightarrow B$. A similar proviso is needed to rule out gaps and gluts for Ⓡ, given that its clauses, which we are about to meet, also allow access to non-normal worlds.

[5]Another technical note, skippable without loss of continuity. We want the syntax of various complex formulas to be semantically neglected at non-normal worlds: this is what "treating them as atomic" amounts to. But if, for instance, conditionals $A \rightarrow B$ are simply assigned arbitrary truth-values, we may have that $Fm \rightarrow Gm$ gets a different value from $Fn \rightarrow Gn$ even though m and n happen to denote the same thing, which would interact badly with the quantifiers. Priest fixes this as follows. Each formula, A, is associated with one of the form $M[x_1, ..., x_n]$, called the formula's *matrix*. One obtains the matrix of A by replacing each occurrence in it of a free term (either an individual constant, or a

two kinds of worlds: at intensionally impossible worlds, only the conditional and the modals are anarchic; at the extensionally impossible ones, also the extensional vocabulary behaves arbitrarily.[6]

The idea of having complex formulas behave as atomic at some worlds comes from the classic (Rantala, 1982), where non-normal worlds were introduced to make logical omniscience fail for epistemic operators. I use non-normal worlds for similar, but more general, purposes. Such worlds are to be accessible via the binary R when the truth conditions for Ⓡ are at issue. At $w \in$ P:

$w \Vdash_a^+$ ⓇA iff for all w_1 such that wRw_1, $w_1 \Vdash_a^+ A$
$w \Vdash_a^-$ ⓇA iff for some w_1 such that wRw_1, $w_1 \Vdash_a^- A$

The semantics for Ⓡ is similar to the ordinary binary accessibility semantics for the standard modal operators. 'wRw_1' ("w_1 is R-accessible from w"), should be read as the claim that, at w_1, things are as they are conceived or represented to be at w. So it is represented that A (at w) just in case A is true at all w_1 where things are as they are represented to be. For instance, if ⓇA is your dreaming that you win the lottery, (an R-accessible) w_1 is a fine world at which your dream comes true. The difference with the usual binary accessibility

variable free in A), from left to right, with a distinct variable $x_1, ..., x_n$, in this order, these being indexed as the least variables greater than all the variables bound in A in some canonical ordering. One gets back the original formula from its matrix via a number of reverse substitutions (which may be zero: a formula may already be its own matrix, if it has the proper structure).What happens at non-normal worlds is, in fact, the following: v assigns there to each matrix M of the relevant kind (a conditional or modal matrix at points in I, any matrix at points in E) pairs of subsets of D^n, that is, extensions and anti-extensions: if w is a non-normal world and M the relevant matrix, $v(M, w) = \langle v^+(M, w), v^-(M, w)\rangle$, with $v^+(M, w), v^-(M, w) \subseteq D^n$. Next, if $M[x_1, ..., x_n]$ is a matrix and $t_1, ..., t_n$ the susbstitutable terms, we have the following truth conditions for its substitution instances:

$w \Vdash_a^+ M[t_1, ..., t_n]$ iff $\langle v_a(t_1), ..., v_a(t_n)\rangle \in v^+(M, w)$
$w \Vdash_a^- M[t_1, ..., t_n]$ iff $\langle v_a(t_1), ..., v_a(t_n)\rangle \in v^-(M, w)$

See (Priest, 2005) pp. 26-9 for a proof that the matrix semantics works as expected. For the sake of brevity, I will keep talking of "treating complex formulas as atomic" at non-normal worlds; but it is this matrix procedure that is implied.

[6](Priest, 2005) calls our extensionally impossible worlds *open worlds*, meaning that they are not closed under any non-trivial consequence relation; but they deserve to be called impossible if any world does.

for modalities is in the broader set of accessible worlds: representation allows us to intend impossibilities.

The definitions of logical consequence and validity are standard. If S is a set of formulas:

S $\vDash A$ iff for every interpretation $\langle \mathrm{P}, \mathrm{I}, \mathrm{E}, @, R, \mathrm{D}, v\rangle$, and assignment a, if $@ \Vdash_a^+ B$ for all $B \in \mathrm{S}$, then $@ \Vdash_a^+ A$.

As for logical validity:

$\vDash A$ iff $\emptyset \vDash A$, i.e., for every interpretation $\langle \mathrm{P}, \mathrm{I}, \mathrm{E}, @, R, \mathrm{D}, v\rangle$, and assignment a, $@ \Vdash_a^+ A$.

4 Two Kinds of Worlds

There are collateral, but philosophically interesting, reasons for flagging items in I among the non-normal worlds, that is, worlds less anarchic than those in E, where only the intensional logical vocabulary behaves in a deviant fashion. The distinction between intensionally and extensionally impossible worlds mirrors the presence of two positions in the current debate on the subject. The first may be labeled as the "Australasian stance". In the Australasian approach, worlds are constituents of interpretations of some relevant logic or other, which imposes to them some logical structure: they are closed under a relevant consequence relation, weaker than classical consequence relation (see e.g. (Mares, 1997), (Restall, 1997)). Since this position draws especially on the conception of non-normal worlds as worlds where "logical laws may fail or be different", it is naturally allied to the idea that, at the (admissible) non-normal worlds, only intensional operators, such as a relevant conditional, behave in non-standard fashion. After all, it is the conduct of such operators that concerns the laws of logic. The truth conditions for conjunction, disjunction, or the quantifiers, should thus remain the same as in ordinary, possible worlds.[7]

The more radical view may be labeled the "American stance", since it reflects the opinion of some north-American impossible worlds theorists. The American stance focuses on the definition of non-normal

[7]For similar considerations, see e.g. (Priest, 2001), ch. 9.

worlds as "ways things could (absolutely) not be", and adopts what we may call an unrestricted comprehension principle for them. Roughly: for any way the world could not be, there is some impossible world which is like that. This can deliver particularly anarchic worlds, not closed under any non-trivial notion of logical consequence (see e.g. (Vander Laan, 1997), (Zalta, 1997)).

5 Relevant Conditional

Having world quantifiers range on P $\cup$ I in the semantic clauses for $\rightarrow$ makes of it a relevant conditional, in the sense of fulfilling the aforementioned Variable Sharing Property. In particular, the arrangement above makes irrelevant entailments like $A \rightarrow (B \rightarrow B)$ fail – take a $w \in$ I where A is true but $B \rightarrow B$ is not. The failure is in the spirit of the "illogical" features of non-normal worlds: these are situations where laws of logic, like the law of sentential identity, may fail. EFQ as $(A \wedge \neg A) \rightarrow B$, and $A \rightarrow (B \vee \neg B)$, also fail (take a non-normal $w \in$ I where A is both true and false but B is untrue for the former, one where A is true but B is neither true nor false for the latter).

The conditional counts as a weak one by relevantist standards (it does not satisfy minimal contraposition, for instance). This may or may not be a problem, depending on what one expects from a conditional. A stronger setting can be obtained by adding to the interpretations for L a ternary relation on worlds and providing the semantics for a conditional in terms of it, as per the classical approach of (Routley & Meyer, 1973). This would complicate matters here, though. Our main concern is the representation operator Ⓡ, to which I now turn.

6 The (Non-)Logic of Representation

The traditional debate in epistemic logic concerns the logical principles that should characterize the epistemic operators at issue, so as to mirror at best the corresponding intuitive notions. Some views are straightforward, for instance, knowledge being factive: if K_c is *cognitive agent c knows that*, it should sustain the entailment from K_cA to A for any A.

Other inferences are more controversial. Must K_c allow the entailment from K_cA to K_cK_cA? While this turns on issues concerning our intuitions about knowledge, it is not difficult to vindicate the inference, if we like it, by tampering with accessibility between worlds (in this case, just have it be transitive). But the failure of some basic logical inferences in epistemic and intentional contexts is more difficult to handle. This is the cluster of problems gathered under the well-known label of "logical omniscience". When modeled in standard possible world semantics, knowledge (or belief) turns out to be closed under entailment:

(Cl) $A \rightarrow B, K_c A \vDash K_c B$

Also, all valid formulas turn out to be known (believed):

(Val) If $\vDash A$, then $\vDash K_c A$

And beliefs form a consistent set:

(Cons) $\vDash \neg(K_c A \wedge K_c \neg A)$

Taken together, these principles deliver an idealized notion of knowledge (belief), not mirroring the status of fallible and occasionally inconsistent cognitive agents.[8] Now Rantala's non-normal worlds were proposed to deal with these phenomena: despite being logically impossible, and not closed under any non-trivial consequence relation, they can be seen as viable epistemic alternatives by imperfect or inconsistent cognitive agents.

A similar story is to be told for Ⓡ. If we can conceive and represent impossibilities, the content of our representational state is the set of worlds that make our representation true, that is, where things are as they are conceived or represented to be; and this has to include non-normal worlds. Given the way things were set up above, non-normal worlds have no effect at the actual world @ on formulas not including Ⓡ. By allowing such worlds to be R-accessible in the evaluation of formulas including it, though, one can eliminate any unwelcome closure feature, thereby dispensing with (the formulations with Ⓡ in

[8] E.g. I know Peano's axioms as basic truths of arithmetic, and Peano's axioms entail (let us suppose) Goldbach's conjecture; but I do not know whether Goldbach's conjecture is true. With other intentional states such as belief or desire, also broad consistency is at stake.

place of K_c of) (Cl), (Val), and (Cons).

As for (Cl), for instance: assume $\vDash A \rightarrow B$. Then at all worlds in P ∪ I where A holds, B holds. But there can be a non-normal world, w, at which A holds and B fails. If $@Rw$, then we can have that $@ \Vdash_a^+ ®A$, but it is not the case that $@ \Vdash_a^+ ®B$. Similarly for consistency: when the relevant R-accessible worlds are inconsistent worlds where both A and $\neg A$ are true, we can have $@ \Vdash_a^+ ®A \wedge ®\neg A$.

7 Constraints

By accessing non-normal worlds of any kind on the one hand, and by not having constraints on its R-accessibility relation on the other, ® has quite a poor logic – one may indeed wonder whether it is worth being called a *logic* at all. What is doing the interesting work here, though, is not the logic but the semantics. I am interested in the general form of the latter, and representability or conceivability had better be, generally speaking, quite anarchic.

In order to have ® express specific intentional operators under the generic umbrella of conceivability, say, *mentally representing a scenario* as opposed to *hallucinating*, we may nevertheless demand more structure. When one mentally represents a scenario, say, engaging in speculations on the next move of the financial markets, one's representation must have some more or less minimal coherence, that is, be closed under some, however weaker-than-classical, notion of logical consequence. This is proved by the fact that people meaningfully argue on how things are, and on what follows from what, in the relevant scenarios, that is, they accept or reject some things as holding in the situations at hand. Even when we represent to ourselves the impossible, we generally believe that we can draw inferences from what we explicitly represented.

One way to achieve this would be to place appropriate constraints on R-accassibility. We could then have ® model different species of representation depending on the constraints at issue. If there is something like *truthful representation* which is factive, we stipulate its R to be reflexive. Conversely, we may have *make-believe* representations such that the world w where the representing takes place is ruled out as a candidate for realizing them (as per the proviso to much fiction: "Any resemblance with real people or actual facts is merely accidental"). To have Ⓡ express something like "It is represented as holding *purely fictionally* that A", we stipulate R to be irreflexive.

Another way would be to make subdistinctions between non-normal worlds of various kinds. One may then allow only worlds that are closed under some form of entailment to be R-accessible, for instance, worlds in I. This gives us interesting results: representation then only accesses "typical" worlds of relevant logics, which are occasionally inconsistent or incomplete, and can also violate some logical laws, but are nevertheless adjunctive and prime (conjunction and disjunction behave standardly there). Then Ⓡ becomes closed under relevant entailment: if $\vDash A \rightarrow B$, and $@ \Vdash_a^+ ⓇA$, then $@ \Vdash_a^+ ⓇB$; thus, this kind of "relevant conceivability" brings a form of logical omniscience for *relevant* consequences of what is represented. However, inconsistent representation is still allowed, i.e., (Cons) fails, as well as (Val), i.e., not all logically valid formulae are represented.[9]

The need for further constraints is apparent when the representational act at issue is *fictional* representation, that is, the conceiving of situations described in fictional works, tales, stories, myths, etc. Sherlock Holmes is represented (at @), by Doyle and his readers, as a detective living in Baker Street, gifted with acute observational and logical skills, etc. Things are as they are represented at the worlds that make the relevant representational characterization true. But *which* are the relevant R-accessible worlds? That is: under which conditions does a world count as such that things are at it as they are represented? We want the relevant representations to be closed under some notion of logical consequence, so that if ⓇA, and B is a consequence of A, then ⓇB. In general, then, things represented in a certain way may well have *further* properties besides those they are explicitly represented as having. Some such properties will just follow on the basis of the entailments mandated by the logic for Ⓡ. For instance, from the fact that Tolkien represents Gandalf as a friend of Bilbo and Bilbo as a pipe-smoker, we can infer that Gandalf is represented as being friends with a pipe-smoker even though (let us suppose) Tolkien never says that explicitily.

On the other hand, what holds in a representation in many cases goes beyond both what is explicitly represented and what is entailed

[9] The closest antecedent to this in the literature, as far as I know, is Levesque's logic of explicit and implicit belief - see (Levesque, 1984).

by logical implication. For while making inferences on what does or does not hold in a representation, we often import information from actuality, which we want to retain when assessing what goes on in a certain represented situation. What the relevant information is depends on our background knowledge of reality; but may also depend on our beliefs (even mistaken beliefs!). The import can rely on *ceteris paribus* and default clauses. Again, the case of fictional representation makes the point evident, and has been extensively studied, e.g., in (Lewis, 1978), (Proudfoot, 2006). Doyle never explicitly represents (let us suppose) Holmes as living in Europe, or as having lungs. We are inclined to take these things as holding at all worlds that realize Doyle's characterization of Holmes, though, for we integrate the explicit representation with information imported from actuality. Now Doyle certainly characterizes Holmes as a man living in London. At the actual world, London is in Europe and, if something is a normally endowed man, then it has lungs. Doyle says nothing against this, so, absent contrary indications from the author, the import is legitimate.

Intuitively, we should exclude from the R-accessible worlds that matter in evaluating what holds in the representation those worlds that, despite making true what is explicitly represented, add gratuitous changes with respect to actuality: we must exclude worlds that differ from @ more than required. Holmes is represented by Doyle as walking through London; we infer that Holmes is represented as walking through a European city. All worlds where Holmes walks through London but London is in Africa must be ruled out, for that would be a departure from actuality not mandated by what Doyle explicitly represents. London's being in Europe has to be held fixed across the worlds where things are as they are represented. This means that, to some extent, representations (of this kind) are about the real world as well. For what holds in a representation depends on what holds at the R-accessible worlds, where things are as represented. And which worlds these are depends also, to some extent, on how our reality is.

Even if this is worked out in a satisfactory way, it does not mean that we can expect precise answers to all the questions we may ask concerning a represented situation. Is Holmes, as characterized in Doyle's stories, right-handed or left-handed? Doyle does not say. And, intuitively, it is not the case that worlds where Holmes is left-handed in general differ gratuitously from @ more than worlds where he is right-handed, or vice versa. Representation typically under-represents.

Providing a detailed account of the workings of the representation operator, especially of how one is to select the worlds that are relevant to address what holds in a certain representation, is overall a difficult issue. Part of the difficulty is similar to the one of the standard treatment of counterfactuals *à la* Stalnaker-Lewis, where a counterfactual "If it were the case that A, then it would be the case that B" is true just in case the world(s) most similar to the actual world that make(s) the antecedent true, make(s) the consequent true as well. We need to invoke some notion of similarity between worlds, having to take into account worlds with minimal differences from actuality in certain respects. And this notion is notoriously slippery. The task becomes exceptionally tricky when we have to consider the intentions and beliefs of those who do the representing. Sometimes, for instance, an author of a work of fiction can make claims that, later on, turn out to be false in the story, or can make claims that are subtly ironic, etc.

What the appropriate constraints on R-accessibility are to be for the various species of representational activities is a difficult issue, and I am happy to leave it open here. Besides the similarities there is, in fact, a philosophical disanalogy between Ⓡ and more traditional epistemic and intentional notions. That we are fallible and at times inconsistent as cognitive agents may be seen as a defect due to our finite and imperfect nature – when it's about knowing and, perhaps, believing. This is not so when it's about imagining and conceiving: in this case, logical fantasy is, generally speaking, a gift (or so I view it).

Francesco Berto
Department of Philosophy and Northern Institute of Philosophy
University of Aberdeen
King's College, High Street
Aberdeen AB24 3UB
Scotland UK
e-mail: f.berto@abdn.ac.uk

References

Brown, B. (1993). Old Quantum Theory: a Paraconsistent Approach. *Proceedings of the Philosophy of Science Association*, *2*, 397–

441.

Levesque, H. (1984). A Logic of Implicit and Explicit Belief. *Proceedings of the National Conference on Artificial Intelligence (AAAI-84)*, 198–202.

Lewis, D. (1978). Truth in Fiction. *American Philosophical Quarterly*, *15*, 37–46.

Mares, E. (1997). Who's Afraid of Impossible Worlds? *Notre Dame Journal of Formal Logic*, *38*, 516–26.

Mares, E. (2004). *Relevant logic. a philosophical interpretation.* Cambridge: Cambridge University Press.

Priest, G. (2001). *An introduction to non-classical logic.* Cambridge: Cambridge University Press.

Priest, G. (2005). *Towards non-being. the logic and metaphysics of intentionality.* Oxford: Oxford University Press.

Proudfoot, D. (2006). Possible Worlds Semantics and Fiction. *Journal of Philosophical Logic*, *35*, 9–40.

Rantala, V. (1982). Impossible Worlds Semantics and Logical Omniscience. *Acta Philosophica Fennica*, *35*, 106–15.

Restall, G. (1997). Ways Things Can't Be. *Notre Dame Journal of Formal Logic*, *38*, 583–96.

Routley, R., & Meyer, R. (1973). The Semantics of Entailment. In H. Leblanc (Ed.), *Truth, syntax and modality* (pp. 194–243). Amsterdam: North-Holland.

Vander Laan, D. (1997). The Ontology of Impossible Worlds. *Notre Dame Journal of Formal Logic*, *38*, 597–620.

Zalta, E. (1997). A Classically-Based Theory of Impossible Worlds. *Notre Dame Journal of Formal Logic*, *38*, 640–60.