

V—ON CONCEIVING THE INCONSISTENT

FRANCESCO BERTO

I present an approach to our conceiving absolute impossibilities—things which obtain at no possible world—in terms of *ceteris paribus* intentional operators: variably restricted quantifiers on possible and impossible worlds based on world similarity. The explicit content of a representation plays a role similar in some respects to the one of a *ceteris paribus* conditional antecedent. I discuss how such operators invalidate logical closure for conceivability, and how similarity works when impossible worlds are around. Unlike what happens with *ceteris paribus* counterfactual conditionals, the closest worlds are relevantly closest belief-worlds: closest to how things are believed to be, rather than to how they are. Also, closeness takes into account apriority and the opacity of intentional contexts.

I

Conceivability and Possibility. I believe that we can conceive and represent to ourselves the impossible. By ‘impossible’ I mean what obtains at no possible world, not just, for example, what is incompatible with the actual laws of nature. After Kripke, such absolute impossibility is often called ‘metaphysical’: the impossibility that water be XYZ, for instance. But I will allow logical and mathematical impossibilities to join the club, as most philosophers think that ‘ $2 + 2 = 4$ ’ or ‘If it snows then it snows’ fail at no possible world.

I take absolute modalities as objective, non-conceptual features of reality, irreducible to human conventions and mental undertakings. On the other hand ‘conceiving’ or ‘representing’ stand precisely for mental undertakings. I use the words as generics for a range of intentional activities involving the depiction of scenarios, situations or circumstances, and recorded in ordinary language by expressions like *imagining*, *mentally visualizing*, *picturing*, *envisaging*. I leave their serious investigation qua psychological events to cognitive scientists. We may just focus on what Yablo (1993) and, especially, Chalmers (2002) characterize as *positive* conceivability. When we

positively conceive that A , we don't just assume or suppose that A : we have something like a mental scenario verifying A . We represent to ourselves a setting, a configuration of objects and properties (not perforce *visually* imaginable, for instance, if the involved objects are abstract), correctly described by A . This kind of conceivability typically features in thought experiments, for example, zombie arguments in the philosophy of mind.

Some forms of conceivability may be a useful heuristic, providing *prima facie* evidence of possibility. But though I will not defend this claim here, I do not believe in the stronger thesis that conceivability, as roughly characterized above, automatically entails possibility—that we are prevented from peeking at absolute impossibilities.¹ The question for this paper is rather, assuming we can conceive the impossible, how are we to understand and model the fact? I reply: via modal sentential operators, understood as restricted quantifiers on worlds. The operators will be hyperintensional (unsurprisingly: they model states of the mind) because the set of worlds one quantifies over in their semantics is a larger totality than the set of possible worlds. It includes (absolutely) *impossible* worlds: ways things could not be, worlds which represent absolute impossibilities as obtaining (see Berto 2009 for an introduction, Berto 2012 for an application to Meinongianism and the philosophy of fiction).

§II introduces some simple semantic machinery and an initial 'naive conceivability' operator, some virtues of which are described in §III. §IV highlights a problem with the naive operator, exposes a solution proposed in the literature, and finds it insufficient. Then a hopefully more satisfactory view is drafted. §V introduces *ceteris paribus* conceivability operators interpreted as variably restricted world quantifiers: many worlds complying with the explicitly represented content are ruled out by considerations of (relevant, context-dependent) similarity or closeness. §§VI and VII explore how closeness works in the presence of impossible worlds. Unlike with ordinary counterfactuals, in *ceteris paribus* conceivability (what Lewis called) weak centring fails, for the selected worlds are closest to believed actuality rather than to actuality. Apriority and the opacity of intentional contexts also play a role in determining closeness.

¹ I think that attempts to rule out counterexamples via Kripkean redescrptions ('You think you can conceive of water not being H_2O , but you actually conceive of watery stuff not being H_2O' '), don't work, for example, with first-person thoughts and mathematical impossibilities (see Wright 2002).

II

Formalities and Naive \textcircled{R} . Let L be a sentential language with variables $p, q, r(p_1, p_2, \dots, p_n)$, negation \neg , conjunction \wedge , disjunction \vee , the conditional \rightarrow , the standard alethic modals \Box and \Diamond , a unary sentential operator \textcircled{R} , and parentheses. Sentential variables are atomic formulae. If A and B are formulae, so are $\neg A, A \wedge B, A \vee B, A \rightarrow B, \Box A, \Diamond A$, and $\textcircled{R}A$; outermost brackets are normally omitted. The only piece of novelty with respect to a normal modal sentential language is \textcircled{R} , which I will call the *naive representation* operator. ' $\textcircled{R}A$ ' is to be read as 'It is represented that A ', or 'It is conceived that A '.²

An interpretation for L is a sextuple $\langle P, I, E, @, R, \Vdash \rangle$. P is the set of possible worlds. I and E are two kinds of impossible worlds, the *intensionally* and *extensionally* impossible (what this distinction comes to, we will see). P, I and E are disjoint. $W = P \cup I \cup E$ is the totality of worlds *simpliciter*. $@ \in P$ is the actual world. $R \subseteq W \times W$. If $\langle w_1, w_2 \rangle \in R (w_1, w_2 \in W)$, we write ' $w_1 R w_2$ ' and claim that w_2 is representationally accessible (R -accessible) from w_1 . \Vdash is a pair $\langle \Vdash^+, \Vdash^- \rangle$ of relations between worlds and formulae: ' $w \Vdash^+ A$ ' means that A is true at world w , ' $w \Vdash^- A$ ' means that it is false there. The recursive truth conditions are as follows. The extensional vocabulary works thus for all $w \in P \cup I$ ('iff' stands for 'if and only if'):

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

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

$$w \Vdash^+ A \wedge B \text{ iff } w \Vdash^+ A \text{ and } w \Vdash^+ B$$

$$w \Vdash^- A \wedge B \text{ iff } w \Vdash^- A \text{ or } w \Vdash^- B$$

$$w \Vdash^+ A \vee B \text{ iff } w \Vdash^+ A \text{ or } w \Vdash^+ B$$

$$w \Vdash^- A \vee B \text{ iff } w \Vdash^- A \text{ and } w \Vdash^- B$$

As for the modals, for all $w \in P$:

² Intentional operators are often indexed to agents: in the usual notation of epistemic logics, ' $K_x A$ ' is to mean that cognitive agent x knows/believes that A . Since the agent subscript would not have done much work for our purposes, being just carried around as a parameter, I have omitted it.

$w \Vdash^+ \Box A$ iff for all $w_1 \in P$, $w_1 \Vdash^+ A$

$w \Vdash^- \Box A$ iff for some $w_1 \in P$, $w_1 \Vdash^- A$

$w \Vdash^+ \Diamond A$ iff for some $w_1 \in P$, $w_1 \Vdash^+ A$

$w \Vdash^- \Diamond A$ iff for all $w_1 \in P$, $w_1 \Vdash^- A$

The world quantifiers in these clauses are restricted to items in P : unrestricted necessity or possibility is truth at all or some possible world(s) (the standard modalities will not do much work: they are there mainly to be compared with $\textcircled{\ast}$). We have a strict conditional. For all $w \in P$:

$w \Vdash^+ A \rightarrow B$ iff for all $w_1 \in P$, if $w_1 \Vdash^+ A$ then $w_1 \Vdash^+ B$

$w \Vdash^- A \rightarrow B$ iff for some $w_1 \in P$, $w_1 \Vdash^+ A$ and $w_1 \Vdash^- B$

Truth and falsity conditions are given separately: as we want to model conceivable inconsistencies, we allow some formulae to be both true and false, ‘glutty’, at some worlds (and also, neither true nor false, ‘gappy’). We may not want this to happen at possible worlds, however. Then we may impose a Classicality Condition on them:

(CC) If $w \in P$, then for each variable p , either $w \Vdash^+ p$ or $w \Vdash^- p$, but not both.

An easy induction shows that the condition generalizes to all formulae: no gaps or gluts at possible worlds. Things get more interesting at impossible worlds. At points in I , formulae of the form $A \rightarrow B$, $\Box A$ and $\Diamond A$ are treated as atomic: truth-values are assigned to them directly, non-recursively, in an arbitrary way. At points in E , all complex formulae are treated as atomic and behave arbitrarily: a disjunction may be true there even though both disjuncts are false, etc. Hence the names: at intensionally impossible worlds, only the intensional operators (the conditional and modals) are anarchic; at the extensionally impossible ones, also the extensional vocabulary can be. Extensionally impossible worlds can fail to be closed under any non-trivial consequence relation, but they are accessible via R . For $w \in P$:

$w \Vdash^+ \textcircled{\ast} A$ iff for all $w_1 \in W$ such that wRw_1 , $w_1 \Vdash^+ A$

$w \Vdash^- \textcircled{\ast} A$ iff for some $w_1 \in W$ such that wRw_1 , $w_1 \Vdash^- A$ ³

Roughly, ' wRw_I ' is to intuitively mean that, at w_I , things are as they are conceived or represented to be at w . It is conceived or represented that A (at w), just in case A is true at all w_I that comply with the representation. Let $@A$ be my dreaming that I win the lottery; then an R -accessible w_I is a fine world at which my dream comes true. The operator works similarly to an ordinary restricted quantifier on worlds. The difference is only in the broader class of accessible worlds. Logical consequence is straightforward. If Σ is a set of formulae:

$\Sigma \models A$ iff, for every interpretation $\langle P, I, E, @, R, \Vdash \rangle$ if $@ \Vdash^+ B$ for all $B \in \Sigma$, then $@ \Vdash^+ A$

Logical consequence is truth preservation at the base world $@$ in all interpretations (logical validity is truth at $@$ in all interpretations: $\models A$ iff $\emptyset \models A$). We may as well define consequence as truth preservation at all possible worlds: $@$ is not special in this respect. But we must not trespass into impossible worlds. In settings of this kind, 'impossible worlds are only a figment of the agents' imagination: they serve only at epistemic alternatives. Thus, logical implication and validity are determined solely with respect to the standard worlds' (Fagin et al. 1995, p. 358).

Does the semantics misrepresent logical words? A world where $p \vee \neg p$ fails, one may say, is one where it does not express the proposition that either p or not- p . But this may be a problem, if at all, only if the failure happened at possible worlds. Intuitionists claim that instances of Excluded Middle can fail; some after Quine wonder whether they mean something different, say, by their negation. I doubt that the Quinean doubt is legitimate. Anyway, in the semantics above Excluded Middle is valid: its failure is a way things cannot be. A way things *can* be is someone conceiving failures of Excluded Middle, as the existence of intuitionists testifies. Impossible worlds, as 'figments of imagination', deal with such conceivings.

³ Small technical point: the truth conditions are spelled thus for the sake of uniformity with the other operators. But to avoid gaps or gluts at possible worlds when $@$ is around, one would need to rephrase the falsity conditions; to rule out gluts, for example, one would need $w \Vdash^- @A$ iff $w \Vdash^+ A$.

III

Delivering Some Goods. Taken as quantifiers on possible worlds, knowledge and belief are closed under logical consequence, that is, with K the relevant operator:

Closure If KA and $A \models B$,⁴ then KB

As a special case, all valid formulae are known (believed):

Validity If $\models A$, then $\models KA$

Also, beliefs form a consistent set:

Consistency $\neg(KA \wedge K\neg A)$

Rantala (1982) proposes impossible worlds frames to deal with these ‘logical omniscience’ idealization problems. Rantala worlds, which correspond to our extensionally impossible worlds, can be seen as viable epistemic alternatives by imperfect or inconsistent cognitive agents. The same goes for naive \mathbb{R} . By allowing such worlds to be R -accessible, one dispenses with (the formulations with \mathbb{R} in place of K of) *Validity*, *Closure* and *Consistency*. As for *Closure*, for instance, in any interpretation, if $@ \Vdash^+ p$, $@ \Vdash^+ p \vee q$, so $p \models p \vee q$. But there can be interpretations with a $w \in E$ at which p holds, but $p \vee q$ fails. If $@Rw$, then $@ \Vdash^+ \mathbb{R}p$, but it is not the case that $@ \Vdash^+ \mathbb{R}(p \vee q)$. As for *Consistency*, just peek at contradictory worlds where both p and $\neg p$ are true and $@ \Vdash^+ \mathbb{R}p \wedge \mathbb{R}\neg p$.

IV

Syntactic Collapse or Weaker Closure? Why is \mathbb{R} naive, then? Because conceiving may demand more structure than it can afford. When one mentally represents a scenario (say, one engages in speculations on the next move of the financial markets), one accepts or rejects some things as holding in the conceived situation, beyond what is *explicitly* represented. The simple machinery above gives no hint on how this works. As noted by Mark Jago (2007, 2009), the lack of any restriction on accessible impossible worlds may make the worlds apparatus pointless. W hosts worlds not closed under

⁴ This is shorthand for ‘ $\{A\} \models B$ ’, and will be so in the following.

any non-trivial consequence relation. So worlds can correspond to arbitrary sets of formulae true at them. Then given w , take $S = \{w_I \mid wRw_I\}$, the set of worlds accessible from w . Let $C = \{A \mid w_I \Vdash^+ A \text{ for all } w_I \in S\}$, the set of formulae true at all of them. The conceiving agent's representational state looks like a merely syntactic structure: $w \Vdash^+ \textcircled{R}A$ if and only if $A \in C$, and C is just an arbitrary set of formulae. Now syntactic approaches to knowledge/belief representation have been developed (Eberle 1974; Moore and Hendrix 1979). But 'one gains very little intuition about knowledge by studying syntactic structures ... in these approaches knowledge is a primitive construct.... The semantics given to knowledge in Kripke structures *explains* knowledge as truth in all possible worlds' (Fagin et al. 1995, p. 345). The same holds if we replace 'knowledge' with 'representation'.

How about giving structure by closing \textcircled{R} under some weaker-than-classical consequence? The distinction between extensionally and intensionally impossible worlds comes into play. Rephrase the clauses for \textcircled{R} as follows. For $w \in P$:

$$w \Vdash^+ \textcircled{R}A \text{ iff for all } w_I \in P \cup I \text{ such that } wRw_I, w_I \Vdash^+ A$$

$$w \Vdash^- \textcircled{R}A \text{ iff for some } w_I \in P \cup I \text{ such that } wRw_I, w_I \Vdash^- A$$

Now conceivability only accesses impossible scenarios of a special kind, namely items in I , and one gets for \textcircled{R} Levesque's 'logic of implicit and explicit belief' (1984). This is basically the application of a simple paraconsistent logic, Belnap and Dunn's First Degree Entailment (FDE), to belief representation. Consistency still fails for \textcircled{R} : one can conceive contradictions. As expected in a paraconsistent context, $\textcircled{R}(A \wedge \neg A)$ does not entail $\textcircled{R}B$ for arbitrary B . Paraconsistent logics have been advertised as logics for cognitive inconsistency management, suitable to represent the epistemic states of non-ideal agents. But there are reasons of dissatisfaction with the approach. Worlds in $P \cup I$ correspond to situations of the world semantics for relevant logics: these can be inconsistent and incomplete, but are always adjunctive and prime. So we have:

$$(1) \textcircled{R}A, \textcircled{R}B \text{ F } \textcircled{R}(A \wedge B)$$

$$(2) \textcircled{R}A \text{ F } \textcircled{R}(A \vee B)$$

For some readings of \textcircled{R} , (1) may make sense: if one conceives that A and conceives that B , one is conceiving them together, $A \wedge B$ (we will come back to this issue below). I am not sure of (2), however. But the main problem, as stressed by Fagin and Halpern (1988), Jago (2007), is that logical omniscience backfires. Representation is closed under paraconsistent consequence: $\textcircled{R}A$, and B 's being an FDE-consequence of A , automatically give $\textcircled{R}B$. Now, paraconsistent arithmetic is well-developed non-classical mathematics, but we don't imagine the recondite paraconsistent consequences of its axioms.

V

Conceiving as Ceteris Paribus Activity. I still think there is something to the distinction between implicit and explicit representation, though it should not be developed as per Levesque's logic. Start with the insight that what holds in a representation may exceed both what is explicitly represented and what is entailed by mere logic. Conceiving often rules out worlds, not by way of logic, but via further content imported from actuality. I imagine Sherlock Holmes as a man living in London. At \textcircled{R} , London is in Europe, and normally a man has lungs. Conan Doyle's stories say nothing against this, so I take it as holding throughout the worlds that comply with the representation: I do imagine, if implicitly, Holmes as living in Europe and as having lungs, although this is not entailed by the explicit content. We want 'It is conceived that A ' to hold at w just in case A holds at the worlds where the explicit content obtains *and* some information from w is preserved.

This brings conceiving in the vicinity of *ceteris paribus* conditionals. The explicit content of a representation may play a role similar to a conditional antecedent. Cognitive science hosts a vast literature on 'counterfactual imagination' (see, for example, Byrne 2005). People conceive alternatives to reality by mentally erasing some of its aspects. Some scenarios complying with the explicit content are ruled out because they bring gratuitous changes with respect to actuality (or *believed* actuality: we will see that the difference matters).⁵ Stal-

⁵ Something similar happens in Lewis's treatment of truth in fiction (1978), where a non-factive fictional operator 'according to such-and-such a story' works as a restricted world quantifier. The thought is that 'we can help ourselves to the notion of what is *explicitly* so according to the fiction and use the notion of possible worlds to extend outwards and define

naker (1968) and Lewis (1973) famously proposed understanding *ceteris paribus* conditionals in terms of closeness or similarity between worlds. The conceiving we are interested in will often be ‘counterpossible’: the explicit representational contents similar to *ceteris paribus* conditional antecedents may hold at no possible world. In the Lewis-Stalnaker approach any counterpossible conditional is trivially true, but there are good accounts of counterpossible conditionals (Read 1995; Nolan 1997; Brogaard and Salerno 2013) which don’t make them all trivially true, precisely by resorting to impossible worlds.

World similarity is notoriously vague and context-dependent. But Lewis argued that this is no defect of his account, which just mirrors the vagueness of *ceteris paribus* conditionals themselves:

Comparative similarity is not ill-understood. It is vague ... in a well-understood way. Therefore it is just the sort of primitive that we must use to give a correct analysis of something that is itself undeniably vague. ... I am not one of those philosophers who seek to rest fixed distinctions upon a foundation quite incapable of supporting them. I rather seek to rest an unfixed distinction upon a swaying foundation, claiming that the two sway together rather than independently. (Lewis 1973, pp. 91–2)

Ceteris paribus conditionals are variably strict: as different accessibility relations are determined by different formulae (or propositions), they are irreducible to combinations of standard modal and truth-functional operators (Lewis 1973, p. 23). Correspondingly, rather than talking of representation in the abstract, we resort to a set K of *representational acts* performed on specific occasions. These are characterized by their explicit content, to be directly expressed by formulae, so we can just take K as a set of formulae.⁶ Each $A \in K$ determines its own accessibility, R_A . Each $A \in K$ comes with its own operator, say $[A]$. So we add the square brackets to our language L , and if A and B are formulae, $[A]B$ is now a formula of L .

Next, an interpretation is now a tuple $\langle P, I, E, @, \{R \mid A \in K\}, \Vdash \rangle$. $\{R \mid A \in K\}$ is a set of accessibilities on W , the other items are as be-

what is *implicitly* so’ (Sainsbury 2010, p. 76). The explicit fiction works in various respects like a *ceteris paribus* conditional antecedent.

⁶ For certain purposes, we may want representational acts to be identified in the formalism by the explicitly represented content *and* by the context where the act takes place. We will return to this.

fore. The truth conditions are, for $w \in P$:

$w \Vdash^+ [A]B$ iff for all $w_1 \in W$ such that $wR_A w_1$, $w_1 \Vdash^+ B$

$w \Vdash^- [A]B$ iff for some $w_1 \in W$ such that $wR_A w_1$, $w_1 \Vdash^- B$

One may reformulate using set-selection functions, similarly to Lewis (1973, pp. 57–9).⁷ Each $A \in K$ comes with its own function, f_A , mapping each world w where the representing takes place to the set of worlds the act allows access to: $f_A(w) = \{w_1 \in W \mid wR_A w_1\}$. The two are equivalent: $wR_A w_1$ if and only if $w_1 \in f_A(w)$. Let $|A|$ be the set of worlds where A is true. Then we get, for $w \in P$:

$w \Vdash^+ [A]B$ iff $f_A(w) \subseteq |B|$

$w \Vdash^- [A]B$ iff $f_A(w) \cap |\neg B| \neq \emptyset$

$[A]B$ is true (false) at w just in case B is true at all worlds (false at some world) in a set determined by f_A .⁸ All the rest, including the definitions of logical consequence and validity, is as before. Next, we must address two main questions: Which constraints may one want or *not* want for the various R_A or f_A ? And how is closeness to be understood in the presence of impossible worlds?

VI

Ceteris Paribus Invalidities. One constraint we should have is:

Obtaining If $w \in P$, then $f_A(w) \subseteq |A|$

When w is possible, it sees via R_A only worlds where the explicit content *obtains* (we will come to the restriction to possible worlds below). This gives:

$\models [A]A$

What I explicitly conceive, I conceive. A constraint one may find plausible limits accessibility to adjunctive worlds (a point hinted at

⁷ Set-selection functions require what Lewis called the Limit Assumption: there can never be endlessly closer and closer worlds to a given world. Lewis rejects it, but I believe it to be justified. What matters is similarity in *relevant* respects. For each world w and antecedent A , there may well be A -worlds so close to w that any further closeness is irrelevant: see Bennett (2003, pp. 177–80) on this.

⁸ To avoid gaps and gluts at $w \in P$ we need clauses as for @ above. I skip the details.

above for \textcircled{B}): for $w \in P$, $wR_A w_I$ only if w_I is such that, if $w_I \Vdash^+ A$ and $w_I \Vdash^+ B$, then $w_I \Vdash^+ A \wedge B$. This gives:

Adjunction $[A]B, [A]C \models [A](B \wedge C)$

If I imagine in one representation $[A]$ that B and that C , I also imagine that $B \wedge C$. This might look controversial given a popular example by Quine, stressing the role that context and the consequent play in counterfactuals. The imagined situation, $[A]$, is ancient Roman Caesar commanding troops in the Korean war. Do we imagine him using the atomic bomb, B , or rather catapults, C ? You can have $[A]B$, Caesar dropping the bomb, if you hold fixed the weapons available during the Korean war (and Caesar's stubbornness). Or, you have $[A]C$, Caesar dropping stones via catapults, if you stick to the Roman military apparatus. You'd hardly infer that $[A](B \wedge C)$, you imagine Caesar employing both the bomb and catapults (you can imagine *that* too; but it shouldn't be an automatic entailment). But maybe *Adjunction* can be maintained on the ground that $[A]$ is what is conceived in a single representational act, if particular acts settle their own context.⁹

Invalid inferences for *ceteris paribus* conditionals correspond to invalidities for our operators. These ought to be 'non-monotonic', in that the following fails:

$[A]B \models [A \wedge C]B$

I can imagine myself in my hometown, Venice, and I will imagine myself in Italy. But If I imagine myself in Venice and that Venice has been displaced in the middle of Aberdeenshire, I will not imagine myself in Italy. Importing additional information into the explicit content of a representational act turns it to a different one. Other invalidities are special, and due to impossible worlds. Here's an important one:

$A \rightarrow B \models [A]B$

That all the possible A -worlds are B -worlds does not imply that, in a representational act whose explicit content is given by A , we automatically conceive that B : we may look at impossible A -worlds

⁹ This may be better represented in the formalism by explicitly indexing to contexts: $[A]_x$ and $[A]_y$ will stand for two distinct acts with the same explicit content, A , associated with contexts x and y . Then we'll restrict *Adjunction* to same-indexed operations (I leave you to fill in the details).

where B fails. So irrelevant strict conditionals holding because the antecedent is impossible, or the consequent necessary, do not imply the corresponding irrelevant conceivings. Although $\vDash (A \wedge \neg A) \rightarrow B$, $[A \wedge \neg A]B$ fails. We don't conceive any B whatsoever just by explicitly conceiving an inconsistency (look at an inconsistent $w \Vdash^+ p \wedge \neg p$ where it is not the case that $w \Vdash^+ q$). Similarly, although $\vDash A \rightarrow (B \vee \neg B)$, we don't have $[A](B \wedge \neg B)$. In general, thanks to impossible worlds, we have neither $\Box B \vDash [A]B$ nor $\neg \Diamond A \vDash [A]B$.¹⁰

Background information in *ceteris paribus* representation may comply in *some* respects with what Lewis (1973), after Goodman, called 'cotenableity'. Roughly: take cotenableity as the relation that, by holding between some information and a formula, A , makes the information eligible to be imported into the representation whose explicit content is given by A . $[A]B$ intuitively holds at a world when the explicit content A , plus a *ceteris paribus* clause, say C_A , determined by A and cotenable with A at that world, delivers B . If I am right about our conceiving the impossible, A can be absolutely impossible. But once we have conceived it, cotenableity marks the familiar aspects of reality we still don't want to 'undo'. We hold them constant relative to A : they make us feel at home, even if A has us look at the strange.

However, firstly, C_A is not an ordinary set of inference premisses. As Williamson stresses in *The Philosophy of Philosophy*, our 'naive physics'—our intuitive view of causation, our expectations on ordinary objects' movements and persistence, etc.—is highly cotenable. We cannot easily dispense with it in much of our imagination. But naive physics is not stored in our mind as a set of premisses: clearly identifiable, linguistically expressible principles (for details see Williamson 2007, p. 145).

Secondly, what background is imported is constrained by what is *relevant* with respect to the explicit content. I know that Phnom Penh is in Cambodia, but this is immaterial to my imagining Holmes's performances as per Conan Doyle's novels. So I need not import the information in my imagining Holmes in action: my representation is not *about* that. This may have a formal counterpart in our setting, beyond the aforementioned irrelevancies. I suspect a general

¹⁰ The restriction to possible worlds in *Obtaining* is there because, otherwise, if we have nested conceivings (I conceive that I conceive that ...), for any A and B we get $[A]([B]B)$: whatever I explicitly conceive, I conceive that what I explicitly conceive, I conceive—which sounds bad.

relevance or ‘aboutness’ property to hold for *ceteris paribus* conceivability: $[A]B$ holds only when A and B share a sentential variable, so B is at least partly about what A is about.

Thirdly, and most importantly, what is cotenable in our conceiving is not made only of truths. There may also be believed falsehoods: ‘what people do not change when they create a counterfactual alternative depends on their beliefs’ (Byrne 2005, p. 10). Importation of false beliefs may be pervasive. As also remarked by Williamson (2007), naive physics, despite being extremely cotenable, is just false of our world. So the selected worlds are not precisely the closest to how things stand (at the world where one conceives). They must be closest to how things are believed to be (there).¹¹ The job of the actual world @ is to fix, not how things actually are, but the beliefs that are actually held (the more true beliefs we hold, the smaller the difference between the two jobs).¹² Then one constraint we should *not* want is (the counterpart in our framework of) what Lewis called *Weak Centring*:

Weak Centring If $w \in |A|$, then $w \in f_A(w)$

If w realizes the explicit content, A , of an act of representing, then it is one of the closest A -worlds. Even restricted to possible worlds, *Weak Centring* validates this:

$A, [A]B \vDash B$

If the explicit content, A , of an act obtains, and it is represented in this act that B , then B also obtains. This will not do, for by importing cotenable false beliefs we can fail to access actuality also when the explicitly represented content is actually true. I imagine Crispin Wright working in Aberdeen, $[A]$, while believing that Aberdeen is in England (may the Scots forgive me). I import the cotenable belief and imagine that Crispin works in an English city, $[A]B$. A obtains:

¹¹ This may relate our operators to (*ceteris paribus*) *indicative* conditionals, whose evaluation has a strong doxastic component. Without addressing this complex issue, I notice that various authors, such as Mares (2004, pp. 125–6), Bennett (2003, pp. 172–5), Brogaard and Salerno (2013), believe the difference between *ceteris paribus* indicatives and subjunctives to be smaller than supposed in much literature.

¹² Similarly, one account of truth in fiction in Lewis (1978) proposes that the selected worlds be the closest, not to how things actually are, but to how they are believed to be by the storyteller and her intended audience. Lewis is concerned with the distinction between shared and idiosyncratic beliefs. This is no problem for us: we work on conceiving in the privacy of a single thinker’s mind.

Crispin does work in Aberdeen. But it doesn't follow that B obtains, Crispin works in an English city. Our false beliefs will affect the faithfulness to actuality of what we conceive, even when the explicitly conceived content is faithful.

Finally, another constraint we don't want is our set-selection functions singling out a single world, that is, $f_A(w)$ being generally a singleton. When I conceive of a situation, I normally do not represent all its details: it would be too cognitively difficult a task. I try to foresee the developments of the financial markets: I imagine a realistic situation involving brokers, displacements of real and virtual money, and so on. I cannot imagine this in full detail, but I want the details to be there. I would not allow the inference from the incompleteness of my representation to the 'de re' incompleteness of the represented objects themselves. I don't imagine London brick by brick; this doesn't mean that, in my imagination, the City is a vague object with a *de re* indeterminate number of bricks.

Now underspecificity is about accessing a plurality of worlds: '[in] imaginative experiments ... we imagine only some salient features, and thereby cover an infinite class of worlds all in one act of imagining' (Lewis 1986, p. 114). Different worlds fill up the unspecified details in different ways. I imagine the Prime Minister signing autographs for some City bankers. Let A express the explicit content. I have no idea whether the PM is left-handed or right-handed (or ambidextrous). But I don't imagine the PM as a *de re* underdetermined object as far as his dominant hand goes. The PM is an ordinary man, and normally endowed men are either left-handed or right-handed (or ambidextrous): the worlds that realize my representation must comply, or they won't be close enough. So the PM is imagined as either left-handed, L , or right-handed, R (or ambidextrous): for all worlds, w , such that $@R_A w, w \Vdash L \vee R$. Thus, $@\Vdash^+ [A] (L \vee R)$. But some closest worlds don't have the PM left-handed, some don't have him right-handed: we have neither $@\Vdash^+ [A] L$ nor $@\Vdash^+ [A] R$.

VII

Closeness, Apriority and Opacity. That our background beliefs, as Williamson claims, 'seem to function more like patterns of inference than like general principles' (2007, p. 147), suggests that we should

not look for seamless logical rules governing our *ceteris paribus* conceiving:¹³

Calling [the relevant conceiving] ‘inferential’ is no longer very informative. ... To call the new judgement ‘inferential’ simply because it is not made independently of all the thinker’s prior beliefs or suppositions is to stretch the term ‘inferential’ beyond its useful span. At any rate, the judgement cannot be derived from the prior beliefs or suppositions purely by the application of general rules of inference. (Williamson 2007, pp. 147, 151)

Our conceiving (of this kind) is based, rather, on imaginative simulation, where much work is done by fast similarity judgements. As we represent that *A*, we import background beliefs via pre-reflective adjudications on which situations are relevantly closest to how we believe the world to be. Now, world closeness may be even more difficult to handle when impossible worlds are around.

If we accept impossible worlds at all, some of them will be more similar to actuality than others. The trivial world where everything is true has to be remote from @, much more than a slightly inconsistent world just like @, but where some meson is both green and anti-green. This is intuitive enough, but are there more substantial constraints? One may accept Nolan’s Strangeness of Impossibility Condition (SIC): any possible world is closer to @ than any impossible world (Nolan 1997, p. 550). One may also take intensionally impossible worlds as closer than extensionally impossible ones. This picture (inspired by Priest 2005, p. 22) has some intuitive force: Kripke’s impossible worlds for non-normal modal logics (1965), or the occasionally inconsistent but still adjunctive and prime worlds of paraconsistent logic, may be less deviant than Rantala worlds.

The key aspect of closeness for a *ceteris paribus* account of conceivability, though, has to do with apriority and opacity. This does not show up in the Lewis-Stalnaker semantics; it was first clearly spotted, I think, in Brogaard and Salerno’s account of counterpossible conditionals (2013) (but see also Jago 2009). One may imagine Hesperus as not being Phosphorus, or water as failing to be H₂O. Take [water ≠ H₂O]. The *Obtaining* constraint gives:

¹³ This does not rule out different kinds of conceiving where logical inferences play a key role. One clever approach to this is in Jago (2007, 2009), where some impossible worlds are accessed, others are ruled out, via the step-by-step application of rules of inference to the represented content.

(1) [water \neq H₂O] water \neq H₂O

However, we may want this to fail:

(2) [water \neq H₂O] water \neq water

I make no commitment on our capacity of conceiving something as plainly different from itself; but conceiving that water or Phosphorus are not self-identical should not follow by sheer logic from conceiving that water \neq H₂O, or that Hesperus \neq Phosphorus. The ancients took Hesperus as distinct from Phosphorus, but had no doubts on Phosphorus's self-identity. When we 'imaginatively develop' our conceiving that Hesperus is not Phosphorus, or that water is not H₂O, 'we are committed to the explicit denial of no logical truth' (Williamson 2007, p. 174). Williamson takes the failure of the entailment between (the counterfactual counterparts of) (1) and (2) as undermining the non-trivial treatment of counterpossibles: the view requires counterfactuals to create opaque contexts (and this is too bad). Brogaard and Salerno accept that this is indeed the case (and claim that it's not bad).¹⁴ But even if Williamson is right that substitutivity should hold in counterfactual contexts because these are not about intentional features, this is certainly not so for our conceivability operators. They model intentional phenomena: their hyperintensionality is no surprise.

Now this says something about world closeness. In order for (2) to fail, we need closest impossible worlds where water is not H₂O—say, it is XYZ—but water is self-identical. Even while conceiving water as not being H₂O, the self-identity of water is an a priori truth to be retained. Appeal to apriority explains the failure of substitutivity, for 'It is a priori that' creates opaque contexts: we cannot get from 'It is a priori that water = water' to 'It is a priori that water = H₂O'. Impossible worlds' closeness should be 'partially epistemic' (Brogaard and Salerno 2013, p. 654) and take apriority on board. I agree, at least for *ceteris paribus* conceivability.

Brogaard and Salerno follow two-dimensionalism à la Chalmers in taking apriority as context-dependent, but unlike what happens in, for example, Chalmers (2001), they don't stipulate that all math-

¹⁴ That substitutivity works in *most* counterfactual contexts may be explained by closeness (and our usually focusing on possible antecedents), given the SIC: when a counterfactual antecedent is possible, impossible worlds will be too far away. Substitutivity may fail only if impossible worlds are accessed.

emational and logical truths are a priori accessible to conceiving subjects. They call their notion ‘a priori*’ to mark the difference (a similar picture of a ‘non-ideal epistemic space’, in fact, is already in Jago 2009). Then they propose the following heuristics. Given impossible worlds w_1 and w_2 , w_1 is closer than w_2 when:

(a) w_1 does not contain a greater number of sentences formally inconsistent with the relevant background facts (held fixed in the context) than w_2 does. And if w_1 and w_2 contain the same number of sentences formally inconsistent with the relevant background facts (held fixed in the context):

(b) w_1 preserves a greater number of a priori* implications between sentences than w_2 does.

(Brogaard and Salerno 2013, p. 655)

I suggest a similar view for our purposes, with two provisos. First, we should replace ‘background facts’ with ‘background beliefs’: the cotenable background of our conceiving can host false beliefs. Second, while spelling out points (a) and (b) separately can be useful, many a priori* implications and statements are but more cotenable claims—the more so the more obvious they are. Closeness is not only about preserving a greater *number* of items, but also about retaining what is more obvious: the more evident the a priori* statements violated in an impossible world are, the more far-fetched it is, all things considered. What is obvious, of course, varies with context and speaker, as does the relevant similarity. But in particular, ‘ H_2O is H_2O [or “water is water”] is a priori* for all minimally rational speakers’ (Brogaard and Salerno 2013, p. 656): it’s self-evident a priori background we just tend not to give up, while ‘water is H_2O ’ is not a priori. This is how, bracketing the issue whether we can conceive something as plainly non-self-identical, impossible worlds where water is not H_2O but is self-identical can be closer than worlds where water is neither H_2O nor self-identical.¹⁵

¹⁵ This work has been developed within the 2013–15 AHRC project The Metaphysical Basis of Logic: The Law of Non-contradiction as Basic Knowledge (grant ref. AH/K001698/1). A version of the paper was presented in September 2013 at the Modal Metaphysics Workshop in Bratislava. I am grateful to the audiences there and at the Aristotelian Society meeting for many helpful comments and remarks.

Department of Philosophy
University of Amsterdam
Oude Turfmarkt 141-7
 1012 GC Amsterdam
 Netherlands
f.berto@uva.nl

Northern Institute of Philosophy
University of Aberdeen
King's College
High Street
 Aberdeen AB24 3UB
 UK
f.berto@abdn.ac.uk

REFERENCES

- Bennett, Jonathan 2003: *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.
- Berto, Francesco 2009: 'Impossible Worlds'. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/impossible-worlds/>.
- 2012: *Existence as a Real Property: The Ontology of Meinongianism*. Dordrecht: Springer.
- Brogaard, Berit, and Joe Salerno 2013: 'Remarks on Counterpossibles'. *Synthese*, 190, pp. 639–60.
- Byrne, Ruth M. J. 2005: *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA: MIT Press.
- Chalmers, David J. 2001: 'The Nature of Epistemic Space'. <http://consc.net/papers/espace.pdf>. Revised version in Andy Egan and Brian Weatherson (eds.), *Epistemic Modality*. Oxford: Oxford University Press, 2011.
- 2002: 'Does Conceivability Entail Possibility?'. In Gendler and Hawthorne 2002, pp. 145–99.
- Eberle, Rolf A. 1974: 'A Logic of Believing, Knowing and Inferring'. *Synthese*, 26, pp. 356–82.
- Fagin, Ronald, and Joseph Y. Halpern 1988: 'Belief, Awareness and Limited Reasoning'. *Artificial Intelligence*, 34, pp. 39–76.
- Yoram Moses and Moshe Y. Vardi 1995: *Reasoning About Knowledge*. Cambridge, MA: MIT Press.
- Gendler, Tamar Szabó, and John Hawthorne (eds.) 2002: *Conceivability and Possibility*. Oxford: Oxford University Press.

- Jago, Mark 2007: 'Hintikka and Cresswell on Logical Omniscience'. *Logic and Logical Philosophy*, 15, pp. 325–54.
- 2009: 'Logical Information and Epistemic Space'. *Synthese*, 167, pp. 327–41.
- Kripke, Saul A. 1965: 'Semantical Analysis of Modal Logic II: Non-normal Modal Propositional Calculus'. In J. W. Addison, Leon Henkin and Alfred Tarski (eds.), *The Theory of Models: Proceedings of the 1963 Symposium*, pp. 202–20. Amsterdam: North-Holland.
- Levesque, Hector J. 1984: 'A Logic of Implicit and Explicit Belief'. *Proceedings of the National Conference on Artificial Intelligence*, AAAI-84, pp. 198–202.
- Lewis, David 1973: *Counterfactuals*. Oxford: Blackwell.
- 1986: *On the Plurality of Worlds*. Oxford: Blackwell.
- 1978: 'Truth in Fiction'. *American Philosophical Quarterly*, 15, pp. 37–46.
- Mares, Edwin D. 2004: *Relevant Logic: A Philosophical Interpretation*. Cambridge: Cambridge University Press.
- Moore, Robert C., and Gary G. Hendrix 1979: 'Computational Models of Belief and the Semantics of Belief Sentences'. *SRI Artificial Intelligence Center Technical Note* 187. Menlo Park, CA: SRI International.
- Nolan, Daniel 1997: 'Impossible Worlds: A Modest Approach'. *Notre Dame Journal of Formal Logic*, 38, pp. 535–72.
- Priest, Graham 2005: *Towards Non-Being: The Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press.
- Rantala, Veikko 1982: 'Impossible World Semantics and Logical Omniscience'. In Ilkka Niiniluoto and Esa Saarinen (eds.), *Acta Philosophica Fennica*, 35: *Intensional Logic: Theory and Application*, pp. 106–15.
- Read, Stephen 1995: *Thinking About Logic: An Introduction to the Philosophy of Logic*. Oxford: Oxford University Press.
- Sainsbury, R. M. 2010: *Fiction and Fictionalism*. London and New York: Routledge.
- Stalnaker, Robert C. 1968: 'A Theory of Conditionals'. Reprinted in William L. Harper, Robert Stalnaker and Glenn Pearce (eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time*, pp. 41–55. Dordrecht: Reidel, 1981.
- Williamson, Timothy 2007: *The Philosophy of Philosophy*. Oxford: Blackwell.
- Wright, Crispin 2002: 'The Conceivability of Naturalism'. In Gendler and Hawthorne 2002, pp. 401–39.
- Yablo, Stephen 1993: 'Is Conceivability a Guide to Possibility?'. *Philosophy and Phenomenological Research*, 53, pp. 1–42.

