

# Impossible Worlds and the Logic of Imagination

Francesco Berto<sup>1</sup>

Received: 1 January 2017 / Accepted: 9 January 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

**Abstract** I want to model a finite, fallible cognitive agent who imagines that  $p$  in the sense of mentally representing a scenario—a configuration of objects and properties—correctly described by  $p$ . I propose to capture imagination, so understood, via variably strict world quantifiers, in a modal framework including both possible and so-called impossible worlds. The latter secure lack of classical logical closure for the relevant mental states, while the variability of strictness captures how the agent imports information from actuality in the imagined non-actual scenarios. Imagination turns out to be highly hyperintensional, but not logically anarchic. Section 1 sets the stage and impossible worlds are quickly introduced in Sect. 2. Section 3 proposes to model imagination via variably strict world quantifiers. Section 4 introduces the formal semantics. Section 5 argues that imagination has a minimal mereological structure validating some logical inferences. Section 6 deals with how imagination under-determines the represented contents. Section 7 proposes additional constraints on the semantics, validating further inferences. Section 8 describes some welcome invalidities. Section 9 examines the effects of importing false beliefs into the imagined scenarios. Finally, Sect. 10 hints at possible developments of the theory in the direction of two-dimensional semantics.

---

✉ Francesco Berto  
F.Berto@uva.nl

<sup>1</sup> Department of Philosophy and Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Amsterdam, The Netherlands

# 1 Introduction<sup>1</sup>

“Imagining” as well as “conceiving” refer in this work to intentional phenomena, intentionality being the feature of those mental states which are directed to objects and configurations thereof, situations, or circumstances. I rely on Chalmers’ (2002) characterization of a notion named *positive conceivability*: when we positively conceive that  $p$ , we do not just assume or suppose that  $p$ , as when we make an assumption in a mathematical proof. Rather, we represent in our mind a scenario: a state of affairs—a configuration of objects and properties—truthfully described by  $p$ .<sup>2</sup>

The human mind has the ability to conceive or imagine, in this sense, rich and detailed alternatives to actuality in order to extract information from them. This has a very pragmatic motivation. Because we cannot experience beforehand which scenarios are or will be actual for us to face in real life, we explore them in our mind, switching off the contribution of our current perceptual inputs. How will the far-East financial markets react if Greece defaults? What contingency plans will you adopt if you don’t get that research grant? Would Mr. Jones show the symptoms he shows, had he taken arsenic? A rich literature on “rational imagination” in cognitive science (Kahneman et al. 1982; Roeser and Olson 1993, 1995) shows how such mental activity improves our cognitive skills and practical performances: one can, for instance, learn from mistaken choices without actually making them, but by simulating them in one’s mind, exploring the consequences, and finding them unpalatable.

That we explore the consequences means that such exercises of imagination, as argued e.g. by Byrne (2005), have a logic: some things follow from the hypothesized scenario, some others do not. What kind of logical framework is suitable for investigating this phenomenon? One obvious place to look at is possible worlds semantics for epistemic and doxastic logics. But this mainstream approach faces a number of well-known issues, which have been grouped under the label of “logical omniscience”.

The logical study of intentionality flourished when authors like Hintikka (1962) realized that the techniques of possible worlds semantics could be applied to the analysis of intentional states like knowledge, belief, cognitive information. This was one of the success stories of philosophical logic, whose results were taken up by linguistics, computer science, and Artificial Intelligence [see Fagin et al. (1995), Meyer and van der Hoek (1995)]. The key insight is notorious: representational

<sup>1</sup> Various versions of this paper, or parts thereof, have been presented between 2014 and 2015 at the University of Lund, at the Archives Poincaré in Nancy, at the Northern Institute of Philosophy in Aberdeen, at the Munich Center for Mathematical Philosophy, at the University of Groningen and at the Institute for Logic, Language and Computation in Amsterdam. I am grateful to all those who provided comments and useful remarks, including three anonymous referees. The paper draws on ideas from Berto (2014); in particular, Sect. 4 relies on a formalism introduced in Sect. 2 of that work.

<sup>2</sup> Rationalists like Descartes made a lot of a distinction between conceiving and imagining (think of his famous example of the chiliagon); whereas empiricists like Hume blurred it. I will use “conceiving” and “imagining” broadly as synonyms for the aforesaid mental act of representing a scenario verifying a sentence or proposition. In particular, the imagined scenario need not be perforce visually imaginable: it may, for instance, involve abstract objects.

mental states are modelled as restricted (agent-indexed) quantifiers over worlds. Cognitive agent  $x$   $\mathbb{R}$ 's that  $p$ ,  $\mathbb{R}$  being the relevant representational mental state (knows, believes, is informed that), when  $p$  holds throughout a set of worlds compatible with  $x$ 's evidence, overall beliefs, etc. Accessibility relations single out the scenarios  $x$  entertains. Let  $R$  be one such accessibility: " $wRw_1$ " means "World  $w_1$  is an epistemic alternative for world  $w$ ". Read " $\mathbb{R}p$ " as "It is represented [believed, known, etc.] that  $p$ ". Then the (non-agent-indexed) truth conditions for  $\mathbb{R}$  are ("iff" = "if and only if"):

' $\mathbb{R}p$ ' is true at  $w$  iff  $p$  is true at all  $w_1$ , such that  $wRw_1$ .

Some authors have applied this framework specifically to the treatment of imagination as a modal operator (Niiniluoto 1985; Costa Leite 2010; Wansing 2015). However, if one characterizes representational mental states by using a standard possible worlds framework, these come out closed under logical consequence or entailment:

(Closure) If  $\mathbb{R}p$ , and  $p$  entails  $q$ , then  $\mathbb{R}q$ .

Agents represent (know, believe, imagine) all the logical consequences of what they represent. In particular, all logically valid formulae are represented:

(Validity) If  $p$  is valid, then  $\mathbb{R}p$ .

And mental states are perforce consistent:

(Consistency)  $\neg(\mathbb{R}p \wedge \mathbb{R}\neg p)$ .

Such principles hold in the weakest normal modal logic K (for Consistency, just add the seriality D-principle). They follow precisely from interpreting epistemic operators as quantifiers over possible (logically closed, maximally consistent) worlds. There is universal consensus (see e.g. Meyer and van der Hoek 1995, Sect. 2.5) that they deliver implausibly idealized mental states. We experience having (perhaps covert) inconsistent beliefs. Excluded Middle is (suppose) valid, but intuitionists do not believe it. We know basic arithmetic truths like Peano's postulates; and these entail (suppose) Goldbach's conjecture; but we don't know whether Goldbach's conjecture is true. The cognitive agency so modelled has little to do with human intelligence.

Wansing (2015) nicely discusses Niiniluoto (1985) and Costa Leite (2010)'s works on imagination in this respect. Wansing himself uses neighbourhood semantics from minimal models (see Chellas 1989, Part III) for his own logic and semantics of imagination. This allows several logical closure properties to fail for it: for instance, that one imagines that if  $p$  then  $q$ , and one imagines that  $p$ , does not entail that one imagines that  $q$ . However, it is still the case that if  $p$  is equivalent to  $q$  (the two come out true in the same worlds in all interpretations, so ' $p$  iff  $q$ ' is logically valid), and one imagines that  $p$ , one imagines that  $q$  and vice versa. This may be questioned. Even in logics much weaker than classical logic, e.g., weak relevant logics,  $p$  is equivalent to  $p \vee (p \wedge q)$ . However, one may imagine that  $p$

without imagining that  $p \vee (p \wedge q)$ ; for instance,  $q$  may include expressions for concepts the conceiving agent simply does not possess.<sup>3</sup>

Logical omniscience is clearly connected to the topic of *hyperintensionality*. One can take H as a hyperintensional operator when  $Hp$  and  $Hq$  may have different truth values, even if  $p$  and  $q$  are necessarily or logically equivalent. Now intentional states seem to draw distinctions between intensionally (necessarily) equivalent contents:  $\mathbb{R}p$  may differ in truth value from  $\mathbb{R}q$  even when  $p$  and  $q$  are logically equivalent. The possible worlds apparatus can only draw intensional, not hyperintensional, distinctions. Thus it cannot easily model conceivability and connected doxastic and informational notions. Among the approaches to hyperintensionality in the logical literature, possibly the most interesting are Tichy's Transparent Intensional Logic (Duží et al. 2010) and structuralist accounts of content (King 1996). Each faces troubles (see e.g. Ripley 2012; Jago 2014, for a set of thorough objections to structuralism).

*This* work aims at modelling imagination as a hyperintensional mental state, while retaining the thought that similar states are restricted quantifiers on worlds, thereby preserving the key insight of world semantics. The two core ideas behind the approach are: (1) to expand the worlds apparatus by adding so-called non-normal or impossible worlds; and (2) to model acts of imagining or conceiving as variably strict world quantifiers. The first idea has been already explored in epistemic and doxastic logic, but never applied to imagination<sup>4</sup>; I introduce it in Sect. 2 below. The second idea is new; I introduce it in Sect. 3.

## 2 Impossible Worlds

If possible worlds are ways things could be, then non-normal or impossible worlds are ways things could not be: they represent some absolute impossibility as being the case.<sup>5</sup> What we take to be absolutely impossible depends on what we take to be absolutely necessary, that is, to hold across the total modal space of possible worlds. Logical and mathematical necessity are at times taken as candidates for absoluteness. To them, some add metaphysical necessity (e.g., the necessity that Hesperus be Phosphorus). Not much hinges on this in what follows. The impossible worlds we are going to employ only represent failures of logical necessity. And, as

<sup>3</sup> I should mention another feature of Wansing's approach: it combines neighbourhood semantics with a "stit" logic of agency (see Belnap et al. 2001; Horty 2001). I find this strategy very promising: it allows to model the *agentive* role of imagination, namely the idea that acts of imagination are what agents voluntarily set out to do. It might be that a stit framework is embeddable in the one I propose below.

<sup>4</sup> Though I found an anticipation in cognitive science research by Nichols and Stich (2003). They propose a model of imagination and mental simulation based on a "possible world box", where we store the contents of our acts of imagining, and which gets integrated via the importation of relevant beliefs into the imagined scenarios. In a footnote, they claim: "We are using the term 'possible world' more broadly than it is often used in philosophy [...], because we want to be able to include descriptions of worlds that many would consider *impossible*. For instance, we want to allow that the Possible World Box can contain a representation with the content *There is a greatest prime number.*" (Ibid: 28).

<sup>5</sup> For a quick introduction, see Berto (2013). For an application to the ontology of fiction, see Berto (2008).

we will see, the formal framework I will propose can be fine-tuned in such a way that the logic at issue is fully classical. Then the relevant impossibilities turn out to be violations of classical logic (more precisely, of the normal propositional modal logic S5).<sup>6</sup>

The idea of using non-normal worlds to model intentional states has been pursued in epistemic logic since Rantala (1982), and later on by authors such as Priest (2005) and Jago (2014). These worlds are understood as viable epistemic alternatives for limited and fallible cognitive agents. The intentional operators characterised via them are still taken, as in the standard approach, as modals: (restricted) quantifiers over worlds. But by accessing non-normal worlds in the truth conditions of the relevant  $\mathbb{R}$ , one easily refutes Validity, Closure and Consistency. For instance, Closure: take a non-normal world  $w$  where  $p$  holds, but  $p \vee q$  fails. If  $w$  is accessible (to the relevant agent), we have  $\mathbb{R}p$  without  $\mathbb{R}(p \vee q)$ , although  $p$  logically entails  $p \vee q$ . For Consistency: access a non-normal world where both  $p$  and  $\neg p$  hold to get  $\mathbb{R}p$  and  $\mathbb{R}\neg p$ . Wansing (1990) proved that non-normal worlds semantics provide a very comprehensive framework for epistemic logics, within which other approaches, for instance, syntactic ones in which the cognitive states of agents are taken as sets of sentences (Eberle 1974; Fagin and Halpern 1988), can be recaptured.

All of this has already been done. In this paper I want to apply the techniques of impossible worlds semantics specifically to the logic of imagination for finite and fallible conceiving agents. I approach this via an issue which is symmetric to the problem of logical omniscience. On the one hand, our imagination should sometimes be inconsistent, and/or not closed under entailment: we do not conceive everything that follows from what we explicitly imagine, and we can occasionally have inconsistent conceptions. But on the other hand, it is another manifest fact of our inner mental life that we do imagine things not logically entailed by what is explicitly included in the mental act of imagining a scenario. I think that impossible worlds can help with this as well. The first question we need to ask is: what does “explicit” mean here?

### 3 Ceteris Paribus Imagination

When we engage in a conscious act of imagination whereby we conceive a scenario, such an act has some deliberate basis: we set out to target a given content. Call such content *explicit*. Byrne (2005) reports a number of experiments carried out by cognitive scientists, showing how imagination has such a deliberate component, whereby we focus on a limited number of non-actual possibilities, directly represented in our mind (a similar view is advocated in Nichols and Stich 2003). For instance:  $x$  reads one of Arthur Conan Doyle’s novels, portraying Sherlock Holmes as a man who is variously active in London, so-and-so dressed, doing this and that. On the basis of the input overtly given in the text,  $x$  starts forming a mental representation of the situation described there.

<sup>6</sup> Thanks to an anonymous referee for prompting me to clarify this.

When we engage in such exercises we do not limit ourselves to the information we explicitly represent in our minds, or to what we can unpack from it via pure logic. London actually is in the UK and normally endowed men have kidneys, even if Doyle's stories (assume) do not claim this explicitly. Now  $x$  can take such facts as holding throughout the represented situation, absent information to the contrary:  $x$  does imagine Holmes as a normally endowed man with kidneys and as living in the UK. This integration is typical: we do not conceive such additional details by inferring them logically from the explicitly given content, rather by importing background information we already have, and which we retain in the non-actual scenario we build a mental representation of.

Then in such exercises of imagination we do both less and more than applying end-to-end a fixed set of logical rules of inference. We do less, in the sense that we don't draw all the logical consequences of what we explicitly conceive. That is the non-omniscience side of the story. But we also do more, because we enlarge our imagined scenarios by importing what does not follow logically from their explicit content. As Timothy Williamson claimed in *The Philosophy of Philosophy*, we should then avoid looking for smooth logical rules governing such exercises in their entirety:

Calling [the relevant conceiving] “inferential” is no longer very informative. [...] To call the new judgment “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 judgment cannot be derived from the prior beliefs or suppositions purely by the application of general rules of inference. (Williamson 2007: 147 and 151)

What Williamson is in the business of explaining here is his approach to the epistemology of metaphysical modality, which has been labelled as “counterfactual” (Vaidya 2015). According to Williamson, we come to know absolute (metaphysical) necessities via counterfactual thoughts, so that such knowledge can be explained as a special case of our coming to know things via counterfactual imagination. Williamson does not adopt a specific formal semantics for counterfactuals, but in the mainstream one—the possible worlds framework of conditional logics by Stalnaker (1968) and Lewis (1973)—these are understood as variably strict modal conditionals. Similarly, we propose to model imagination via modal operators interpreted as variably strict quantifiers over worlds, possible and impossible. Adding impossible worlds accounts for our imagining absolute impossibilities and inconsistencies and for the lack of (classical) logical closure of our mental states. The variability of strictness is to account for the (highly contextual) selection of the information we import in a representational mental act when we integrate its explicit content. The explicit content itself will play a role similar to that of a variably strict, or *ceteris paribus*, conditional antecedent. What is actually imagined, then, is what holds in worlds where the antecedent holds *and* further information imported from actuality holds, too.

The closest to this proposal I found in the philosophical literature is Lewis' (1978) famous paper on truth in fiction. Here the key idea was 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 what is *implicitly* so” (Sainsbury 2010: 76). The explicit fictional content corresponds to the explicit content of our imagined scenarios, and works, in Lewis’ approach, too, like the antecedent of a *ceteris paribus* conditional. Williamson similarly claims:

We seem to have a prereflective tendency to minimum alteration in imagining counterfactual alternatives to actuality, reminiscent of the role that similarity between possible worlds plays in the Lewis–Stalnaker semantics. (Williamson 2007: 151)

Later on I will come to the topic of world similarity. We first need to make the idea of *ceteris paribus* imagination formally more precise.

### 4 Formal Semantics

Take a sentential language  $L$  with atoms  $p, q, r$  ( $p_1, p_2, \dots$ ), negation  $\neg$ , conjunction  $\wedge$ , disjunction  $\vee$ , conditional  $\rightarrow$ , modals  $\square$  and  $\diamond$ , square and round brackets [ and ], ( and ). The round brackets are auxiliary symbols for scope disambiguation. The square brackets allow the formation of sententially indexed modals. I use meta-variables  $A, B, C$  ( $A_1, A_2, \dots$ ) for formulas of  $L$ . The well-formed formulas are the atoms and, if  $A$  and  $B$  are formulas:

$$\neg A | (A \wedge B) | (A \vee B) | (A \rightarrow B) | \square A | \diamond A | [A]B$$

(Outermost brackets are usually omitted). Things of the form “[ $A$ ]” are to be thought of as modal operators indexed by formulas (the idea, in the context of conditional logics, goes back to Chellas 1975; see also Segerberg 1989). Take a bunch of acts of imagining, performed by a given conceiving agent on specific occasions. Suppose each is characterized by an explicit content, to be directly expressed by a formula of  $L$ : this is what we intentionally set out to imagine when we engage in the act. Let the set of formulas expressing possible acts be  $K$ . Each  $A \in K$  has its own operator,  $[A]$ . One can read “[ $A$ ]  $B$ ” as: “It is imagined in act  $A$  that  $B$ ”; or, less briefly and more accurately: “It is imagined in the act whose explicit content is  $A$ , that  $B$ ”. We can call our  $[A]$ ’s “imagination operators”.<sup>7</sup>

An interpretation for  $L$  is a tuple  $\langle P, I, @, \{R_A | A \in K\}, \Vdash \rangle$ .  $P$  is the set of possible worlds;  $I$  is the set of impossible worlds;  $P \cap I = \emptyset$ ;  $W = P \cup I$  is the totality of worlds;  $@ \in P$  is the actual world;  $\{R_A | A \in K\}$  is a set of binary accessibilities on  $W$ ,  $R_A \subseteq W \times W$ : each  $A \in K$  fixes its own accessibility,  $R_A$ . Finally,  $\Vdash$  is a pair  $\langle \Vdash^+, \Vdash^- \rangle$  of relations between items in  $W$  and formulas: “ $w \Vdash^+ A$ ” says that  $A$  is true at world  $w$ , “ $w \Vdash^- A$ ” says that  $A$  is false there. Each interpretation relates

<sup>7</sup> Intentional operators are usually agent-indexed: in the notation of epistemic logics, “ $K_x A$ ” means that cognitive agent  $x$  knows/believes that  $A$ . But the subscript would not have done much work in our setting, which is essentially single-agent: imagination takes place in the private of one’s mind. So we are omitting it. This does not rule out possible developments of the theory into a multi-agent logic of imagination, as in Nichols and Stich (2003)’s “mindreading” perspective, where agents try to imagine what other agents are imagining.

atoms to truth and falsity at worlds and is extended to the whole L as follows. For all  $w \in P$ :

- (S $\neg$ 1)  $w \Vdash^+ \neg A$  iff  $w \Vdash^- A$
- (S $\neg$ 2)  $w \Vdash^- \neg A$  iff  $w \Vdash^+ A$

- (S $\vee$ 1)  $w \Vdash^+ A \vee B$  iff  $w \Vdash^+ A$  or  $w \Vdash^+ B$
- (S $\vee$ 2)  $w \Vdash^- A \vee B$  iff  $w \Vdash^- A$  and  $w \Vdash^- B$

- (S $\wedge$ 1)  $w \Vdash^+ A \wedge B$  iff  $w \Vdash^+ A$  and  $w \Vdash^+ B$
- (S $\wedge$ 2)  $w \Vdash^- A \wedge B$  iff  $w \Vdash^- A$  or  $w \Vdash^- B$

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

- (S $\Box$ 1)  $w \Vdash^+ \Box A$  iff for all  $w_1 \in P$ ,  $w_1 \Vdash^+ A$
- (S $\Box$ 2)  $w \Vdash^- \Box A$  iff for some  $w_1 \in P$ ,  $w_1 \Vdash^- A$
- (S $\Diamond$ 1)  $w \Vdash^+ \Diamond A$  iff for some  $w_1 \in P$ ,  $w_1 \Vdash^+ A$
- (S $\Diamond$ 2)  $w \Vdash^- \Diamond A$  iff for all  $w_1 \in P$ ,  $w_1 \Vdash^- A$

Unrestricted necessity/possibility at possible worlds is truth at all/some possible world(s). We have a strict conditional. For all  $w \in P$ :

- (S $\rightarrow$ 1)  $w \Vdash^+ A \rightarrow B$  iff for all  $w_1 \in P$ , if  $w_1 \Vdash^+ A$ , then  $w_1 \Vdash^+ B$ .
- (S $\rightarrow$ 2)  $w \Vdash^- A \rightarrow B$  iff for some  $w_1 \in P$ ,  $w_1 \Vdash^+ A$  and  $w_1 \Vdash^- B$ .

Finally we come to [A]. For  $w \in P$ :

- (S[A]1)  $w \Vdash^+ [A]B$  iff for all  $w_1 \in W$  such that  $wR_A w_1$ ,  $w_1 \Vdash^+ B$
- (S[A]2)  $w \Vdash^- [A]B$  iff for some  $w_1 \in W$  such that  $wR_A w_1$ ,  $w_1 \Vdash^- B$

Read “ $wR_A w_1$ ” as saying that  $w_1$  realizes the content of an intentional state obtaining at  $w$ : things are at  $w_1$  as they are represented at  $w$ , in the act of imagination whose explicit content is expressed by  $A$ . Because the  $R_A$ ’s access impossible worlds, as we will see, we can have various hyperintensional distinctions we want acts of imagination to be able to make.

The truth and falsity conditions for [A] above may also be expressed using set-selection functions, similarly to what Lewis does in *Counterfactuals* (see Lewis 1973: 57–59). These work as follows: each  $A \in K$  comes with a function,  $f_A$ . This takes as input the world  $w$  where the act of imagination obtains and outputs the set of worlds accessed via that act:  $f_A(w) = \{w_1 \in W \mid wR_A w_1\}$ . Let  $|A|$  be the set of worlds where  $A$  is true. Then for  $w \in P$ :

- (S[A]1)  $w \Vdash^+ [A]B$  iff  $f_A(w) \subseteq |B|$
- (S[A]2)  $w \Vdash^- [A]B$  iff  $f_A(w) \cap |B| \neq \emptyset$

So  $[A]B$  is true (false) at  $w$  iff  $B$  is true at all worlds (false at some world) in a set selected by  $f_A$ . Because  $wR_A w_1$  iff  $w_1 \in f_A(w)$ , the formulation in terms of accessibility and the one in terms of set-selection functions are perfectly



interchangeable. However, it will sometimes be easier to make a point using either formulation rather than the other.

The recursive truth conditions have been defined for points in  $P$ . At points in  $I$  all complex formulas are treated as if they were atoms, in that they are related to truth values directly, not recursively. So for  $w \in I$  we can have, for instance,  $w \Vdash^+ A \vee B$  even if it is not the case that  $w \Vdash^+ A$  or that  $w \Vdash^+ B$  (impossible worlds can be nonprime);  $w \Vdash^+ A$  together with  $w \Vdash^+ \neg A$  (impossible worlds can be inconsistent); etc.

Logical consequence is truth preservation at the base-actual world in all interpretations. Where  $S$  is a set of formulas:

$S \models A$  iff, in every interpretation  $\langle P, I, @, \{R_A \mid A \in K\}, \Vdash \rangle$ , if  $@ \Vdash^+ B$  for all  $B \in S$ , then  $@ \Vdash^+ A$ .

As a special case, logical validity is truth at the actual world in all interpretations:

$\models A$  iff  $\emptyset \models A$ , that is, in every interpretation  $\langle P, I, @, \{R_A \mid A \in K\}, \Vdash \rangle$ ,  $@ \Vdash^+ A$ .

There is nothing telling  $@$  apart from any other possible world with respect to the definitions of logical truth and consequence. However,  $@$ 's being flagged in our interpretations will come handy for later considerations. What matters is that  $@ \in P$ , for “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: 358). If we understand impossible worlds as “worlds where logic may be different”, this sounds natural: we want to define logical consequence with respect to worlds where logic is not different.<sup>8</sup>

The attentive reader will already have guessed why truth and falsity conditions have been given separately in the semantics: we want to model imaginable inconsistencies, so we allow some formulas to be both true and false, or “glutty”, at some worlds (also, neither true nor false, or “gappy”). However, one may not want this to happen at possible worlds: their being maximally consistent is what makes them possible, one may say. To accommodate this, one can restrict attention to interpretations of  $L$  which comply with a Classicality Condition on possible worlds:

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

The CC generalizes, by easy induction on their complexity, to formulas not including imagination operators. To extend the CC to the whole language, though,

<sup>8</sup> It is sometimes claimed that impossible worlds semantics alters the meaning of the logical vocabulary (a nice discussion is in Jago 2007). This is connected to the understanding of non-normal worlds as points where “logic may be different”: a world where a contradiction,  $p \wedge \neg p$ , is true, one may say, is one where that formula does not express the proposition *that p and not-p*. However, the semantics above does validate the Law of Non-Contradiction and furthermore (when fixed in a way we are about to explain) provides no counterexamples to it. A world where a contradiction is true is a way things could not be according to the (amended) semantics; but a way things *could* be is someone’s conceiving a contradiction, that is, imagining a scenario in which it obtains, and impossible worlds have the role of modelling such acts of imagination. Dialetheists like Graham Priest (1987) believe that the actual world is inconsistent, and it is controversial whether they are thereby automatically misunderstanding the meaning of negation (or that of conjunction). Williamson (2007) defends the view that they are not.

we need to rephrase (S[A]2), the falsity conditions for the latter, for their clauses allow one to look at impossible worlds. This is easily done: for instance, to avoid glutty acts of imagination bouncing back into real world inconsistencies, one just needs:  $w \Vdash^- [A]B$  iff not  $w \Vdash^+ [A]B$ .

Once this is done, the logic induced by the semantics for the connectives other than the imagination operators is simply the normal propositional modal logic S5. However, the framework is flexible enough to model also weaker non-classical logics as the base logic, e.g., (modal extensions of) paracomplete and paraconsistent logics (this is one motivation for phrasing the semantics in terms of truth and falsity conditions separately: see Priest (2005), Ch. 1, where the same strategy is adopted).

What we will discuss from now on, is how the logic of our imagination operators is to work. The framework for the discussion will be model-theoretic: we will investigate the opportunity of adding conditions on interpretations, involving the  $R_A$ 's or  $f_A$ 's—the indexed accessibilities—for those operators.

## 5 The Mereology of Imagination

The first conditions I want to propose for our imagination operators involve conjunction. It seems to me that, when one imagines that a conjunction is the case, one also imagines each conjunct: you cannot imagine that Sherlock Holmes is a bachelor and lives in London without imagining that Sherlock Holmes is a bachelor. The following condition on accessibilities:

(*Simplification*) If  $w \in P$ , then if  $wR_A w_1$  then (if  $w_1 \Vdash^+ B \wedge C$ , then  $w_1 \Vdash^+ B$  and  $w_1 \Vdash^+ C$ )

... Gives the following validities:

$$\begin{aligned} [A](B \wedge C) &\models [A]B \\ [A](B \wedge C) &\models [A]C \end{aligned}$$

(Proof: suppose  $@ \Vdash^+ [A](B \wedge C)$ . By (S[A]1), for all  $w \in W$  such that  $@R_A w$ ,  $w \Vdash^+ B \wedge C$ . Because  $@ \in P$ , Simplification applies: for any such  $w$ ,  $w \Vdash^+ B$  and  $w \Vdash^+ C$ . By (S[A]1) again,  $@ \Vdash^+ [A]B$  and  $@ \Vdash^+ [A]C$ ).

The companion constraint to Simplification is:

(*Adjunction*) If  $w \in P$ , then if  $wR_A w_1$  then (if  $w_1 \Vdash^+ B$  and  $w_1 \Vdash^+ C$ , then  $w_1 \Vdash^+ B \wedge C$ )

This gives us:

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

(Proof: suppose  $@ \Vdash^+ [A]B$  and  $@ \Vdash^+ [A]C$ . By (S[A]1), for all  $w \in W$  such that  $@R_A w$ ,  $w \Vdash^+ B$  and  $w \Vdash^+ C$ . Because  $@ \in P$ , Adjunction applies: for any such  $w$ ,  $w \Vdash^+ B \wedge C$ . By (S[A]1) again,  $@ \Vdash^+ [A](B \wedge C)$ ).

If Simplification and Adjunction are accepted, imaginative accessibility is *de facto* limited to worlds which are, so to speak, fully well-mannered with respect to

conjunction. However, Adjunction might look especially controversial. Is it so that, when one imagines in one act  $[A]$  that  $B$  and that  $C$ , one automatically imagines that  $B \wedge C$ ? This is the counterpart of a question that has been asked for counterfactuals, namely whether different counterfactuals with the same antecedent allow for conjunction of their consequents, given the role consequents play in fixing the context of evaluation. We can in fact adapt to our imagination operators the very story that Quine (1960): 222 used for making the point concerning counterfactuals.

The situation one sets out to imagine,  $[A]$ , involves Caesar the Roman emperor being in command of the US troops in the Korean war. Given the same explicit content as input (say, a time travel science fiction story), one can imagine Caesar using the atom bomb,  $B$ , or one can imagine that he resorts to catapults,  $C$ . If one imports into the representation information concerning the weapons available in the Twentieth Century, one can imagine that Caesar drops the bomb in Korea,  $[A]B$ . If one rather allows the Roman military apparatus to step in, one can imagine Caesar dropping stones to the Reds via catapults,  $[A]C$ . However, one would not thereby have  $[A](B \wedge C)$ , Caesar employing both the bomb and catapults. Of course, one *can* imagine that as well (making the scenario even weirder). But the point is that it should not come automatically, as a logical entailment.

However, I believe that something has gone wrong in this reconstruction of the situation. We are considering individual acts of imagining, but we individuate them only via their explicit content. But different acts of imagining the same explicit content can trigger the import of different background information depending on contexts (the time and place at which the cognitive agent performs the act, the status of its background information, etc.). Now it seems clear that there is a shift in context in the Quinean example. So I think that Adjunction can be maintained by fixing some contextual parameter. The formalism may represent this, if wanted, by adding a set of contexts to the interpretations, variables ranging on them in the language, and by directly indexing representational acts with contexts:  $[A]_x$ ,  $[A]_y$ , for instance, will stand for two distinct acts with the same explicit content,  $A$ , performed in contexts  $x$  and  $y$ . Once the adjunctive inference is parameterized to same-indexed contents, it should work fine.

But this is just the technical fix. The philosophical point behind Simplification and Adjunction is that imagined contents (of the kind we are addressing), however logically anarchic, have some minimal mereological structure. Let me explain what is meant by this.

Recall that we are not modelling the mere assumption or supposition of some content  $p$ , but more substantive (in Chalmers' jargon) positive conceivability—someone's bringing to one's mind a mental scenario: a state of affairs, or a configuration of objects and properties, which verifies  $p$ . One can assume or suppose that  $p$  in a proof, without perforce representing in the mind a state of affairs verifying  $p$ . One may then suppose (in this sense) a conjunction without supposing the conjuncts separately, or vice versa.

But configurations of objects and properties, or states of affairs, seem to allow for constituent parts (again, they need not be *spatial* parts: the scenarios may involve abstract objects having no spatial extension, or properties which do not involve spatiality). A state of affairs such that object  $o$  is  $P$  and object  $o$  is  $Q$  includes as

constituents the state of affairs such that  $o$  is  $P$ , and the state of affairs such that  $o$  is  $Q$ . Next, it seems to me, when one positively imagines a whole scenario or state of affairs, it appears that one automatically imagines its constitutive parts. I cannot imagine that Holmes is a bachelor living in London without imagining Holmes as a bachelor, and without imagining him as living in London, for the last two contents are just constituents of the scenario I am imagining. Vice versa, if I imagine Holmes as a bachelor and I imagine Holmes as living in London in the very same act of imagination (that is, modulo the fixing of context as per above), then I imagine both things together, for they are constitutive parts of the same content.

If we can conceive blatantly impossible scenarios like a round square (which might be controversial, but let us assume it for the sake of the argument), the same may hold for them. If I can somehow positively conceive something's being round and square, the relevant scenario will be a configuration of objects and properties, such that something has the property of being round, and that thing has the property of being square. But then I will have already conceived that the thing is round, and I will have already conceived that the thing is square (and vice versa, modulo the contextual indexing to a single act of imagination). The proposal that positive imagination be well-behaved with respect to conjunction is a way to mirror this fact in our formal semantics: positive imagination involves representing a scenario or a state of affairs, and scenarios or states of affairs can have parts. When we conceive the whole of a scenario, we conceive its parts. When we conceive the parts in the same act of imagination, we conceive the whole.

I should anyway add that, if one still does not agree, the formal framework of Sect. 4 can easily accommodate non-simplifying and/or non-adjunctive exercises of imagination: one simply does not accept Simplification and/or Adjunction as conditions on interpretations. Much (though not all) of the subsequent discussion on other inferential features of our imagination operators goes on unaltered.<sup>9</sup>

## 6 The Under-Determinacy of Imagination

If some dual story with respect to what has been said for  $\wedge$  could be told for  $\vee$ , we may limit accessibility for our imagination operators to impossible worlds which are well-behaved with respect to disjunction too. Our impossible worlds, then, would be like the worlds used in relevant logics such as Belnap and Dunn's First Degree Entailment (Dunn 1976; Belnap 1977): worlds which can be locally glutty or gappy, but are always adjunctive and prime.

Now it is intuitive that the contents of our acts of imagination ought to be under-determined. When one imagines a situation, one does not normally represent all its

<sup>9</sup> Thanks to an anonymous referee for pressing me on this. It is worth remarking that some literature from cognitive psychology corroborates the view that mental imagery has a mereological structure implemented in the mind (though the point is controversial: see e.g. Pylyshyn 2002). This can also be a quasi-spatial structure, when one imagines physical situations. There are experiments which show that, on average, it takes more time to explore in the mind the parts of an imagined situation with constituents relatively apart from each other, or which are themselves conceived as mereologically complex, than those of more compact or simpler situations (see e.g. Shepard and Cooper 1982; Kosslyn 1980, 1994).

details. If one sets out to imagine how the European leaders are debating on the future of Greece in Brussels, one thinks of a complex situation involving bilateral meetings, conferences, etc. One does not imagine the details, but one wants the details to be there, in a sense. One does not imagine Brussels brick by brick, but that does not mean that in one’s mind Brussels is a *de re* vague object with an indeterminate number of bricks. That is, one would not allow moving from the under-determinacy of one’s representation to the under-determinacy of the represented objects themselves. Either the overall number of bricks composing the Brussels buildings is even, or it is odd; but one does not imagine Brussels either way. So we need to be able to have representational acts such that  $[A](B \vee C)$  does not entail  $[A]B \vee [A]C$ . Again, our operators mirror what happens with *ceteris paribus* conditionals. As Stalnaker claimed, the “situations determined by the antecedents of counterfactual conditionals are like the imaginary worlds created by writers of fiction”, in that “in both cases, one purports to represent and describe a unique determinate [...] world, even though one never really succeeds in doing so” (Stalnaker 1981: 95).

However, to achieve this we do not need nonprime worlds such that  $w \Vdash^+ A \vee B$  but neither  $w \Vdash^+ A$  nor  $w \Vdash^+ B$ . What does justice to under-determinacy in imagination is the plurality of worlds accessed via  $R_A$  or  $f_A$ : different worlds fill up the unspecified details in different ways. You imagine Angela Merkel signing documents in Brussels,  $A$ . Merkel is either left-handed,  $L$ , or right-handed,  $R$  (or ambidextrous), and you import this by default in the imagined scenario: for all worlds,  $w$ , such that  $@ R_A w$ ,  $w \Vdash^+ R \vee L$ ; thus,  $@ \Vdash^+ [A](R \vee L)$ . But you have no idea which one actually is the case. So there will be accessible worlds that do not have Merkel left-handed, and accessible worlds that do not have her right-handed. Then we have neither  $@ \Vdash^+ [A]R$  nor  $@ \Vdash^+ [A]L$ .

Although our accessible impossible worlds need not misbehave with respect to disjunction in order to model under-determinacy, they need to misbehave for another reason. If the truth and falsity conditions for disjunction above held for all worlds, the following would be valid (while it shouldn’t):

$$[A]B \models_? [A](B \vee C)$$

(Proof: suppose  $@ \Vdash^+ [A]B$ . By (S[A]1), for all  $w \in W$  such that  $@ R_A w$ ,  $w \Vdash^+ B$ . If (S $\vee$ 1) held at all worlds, we would have  $w \Vdash^+ B \vee C$ . And by (S[A]1) again,  $@ \Vdash^+ [A](B \vee C)$ ). Now this cannot be: when one imagines in the act whose explicit content is  $A$ , that  $B$ , one does not thereby imagine a disjunction between  $B$  and an arbitrary  $C$  (you imagine Holmes in London and you imagine him in England, but it does not follow that you imagine that either Holmes is in England or Watson is a jelly fish).

## 7 Further Conditions, and the Issue of Similarity

The Lewis–Stalnaker semantics for *ceteris paribus* conditionals is based on a notion of closeness between worlds, which is universally understood as similarity. Roughly: a conditional of this kind, let us say, a subjunctive “If it were the case that

$A$ , then it would be the case that  $B$ ", is true at  $w$  iff the world(s) most similar to  $w$  where  $A$  holds also make true  $B$ . However, world similarity has been variously criticized as a desperately vague and context-dependent notion. Some have tried to make the notion viable in Artificial Intelligence (Delgrande 1988). But things seem to get worse when worlds representing absolute impossibilities are around. How does similarity work for *them*? Supposing mathematical necessity is unrestricted, is a world where the Axiom of Choice fails closer than one where Fermat's Last Theorem is false? One can find some work on the topic (Nolan 1997; Brogaard and Salerno 2013; Bjerring 2014), but there is as yet no mainstream approach to the issue.

Jago (2014)'s theory of impossible worlds allows a distinction between *obvious* and *subtle* logical impossibilities via a total ordering of impossible worlds with respect to the degree of complexity of the logical truths violated at them. I think that Jago's techniques may be used to provide some kind of logical similarity metric for impossible worlds. The approach has found critics: see Bjerring (2013) (Bjerring 2014 also has a promising positive proposal: an alternative framework for an extended Lewis–Stalnaker semantics including impossible worlds).

However, I would like to leave this idea as work to be developed in a further paper. We can discuss some additional conditions for the semantics of our imagination operators, without taking a stance on world similarity. We can impose direct constraints on the  $R_A$ 's or  $f_A$ 's without presupposing a metric for world closeness, or nested similarity spheres, similarly to what happens with the *ceteris paribus* conditionals of weak conditional logics such as  $C^+$  [see Priest (2008): 87–90].

Here is one basic constraint:

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

Possible worlds only access worlds where the explicit content obtains. Obtaining gives this logical validity:

$\models [A]A$

It is obvious that one imagines what one explicitly imagines.<sup>10</sup> As a special case:

$\models [A \wedge B](A \wedge B)$

And via Simplification:

$\models [A \wedge B]A$

Next, one condition I am inclined to buy is, as we will call it, the Principle of Imaginative Equivalents:

(PIE) If  $f_A(w) \subseteq |B|$  and  $f_B(w) \subseteq |A|$ , then  $f_A(w) = f_B(w)$ .

<sup>10</sup> The restriction to items in  $P$  in Obtaining is needed because, otherwise, with nested conceiving (one imagines that one imagines that), we get  $\models [A]([B]B)$ : whatever one explicitly imagines, one imagines that what one explicitly imagines, one imagines—which sounds bad.

If all the worlds selected by  $f_A$  make  $B$  true and vice versa, then  $A$  and  $B$  are “imaginative equivalents”: when we set out to imagine either, we look at the same set of circumstances. (PIE) validates the following inference:

$$(Substitutivity) [A]B, [B]A, [B]C \models [A]C$$

(Proof: suppose  $[A]B, [B]A, [B]C$  are all true at  $@$ . By (S[A]1), the first entails that  $f_A(@) \subseteq |B|$ , the second that  $f_B(@) \subseteq |A|$ , the third that  $f_B(@) \subseteq |C|$ . Applying (PIE) to the first two of these, we get  $f_A(@) = f_B(@)$ . Thus,  $f_A(@) \subseteq |C|$ , from which, by (S[A]1) again, we get  $@ \Vdash^+ [A]C$ ).

Substitutivity says that two imaginative equivalents  $A$  and  $B$  can be replaced *salva veritate* with each other as the explicit content of a representational act. Given the number of hyperintensional distinctions we can make in our imagination, there might be very few imaginative equivalents for a given agent. But suppose that *bachelor* and *unmarried man* are for you imaginative equivalents: you are so aware of their meaning the same thing, that you can’t imagine someone being one thing without imagining him being the other. Thus,  $[A]B$ : when you explicitly imagine Holmes as unmarried, you imagine him as a bachelor; and  $[B]A$ : when you explicitly imagine Holmes as a bachelor, you imagine him unmarried. Suppose  $[B]C$ : as you explicitly imagine Holmes as a bachelor, you imagine him happy; it follows that the same happens when you imagine him unmarried.

(PIE) licenses another inference I am tentatively inclined to accept. I will call it Special Transitivity<sup>11</sup>:

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

(Proof: suppose (1)  $@ \Vdash^+ [A]B$  and (2)  $@ \Vdash^+ [A \wedge B]C$ . From (1) and  $[A]A$  (secured by Obtaining as valid), via Adjunction, we get  $@ \Vdash^+ [A](A \wedge B)$ .  $[A \wedge B]A$  is valid (from  $\models [A \wedge B](A \wedge B)$ , via Simplification), so  $@ \Vdash^+ [A \wedge B]A$ . Applying (S[A]1), to the last two we get  $f_A(@) \subseteq |A \wedge B|$  and  $f_{A \wedge B}(@) \subseteq |A|$ . By (PIE),  $f_A(@) = f_{A \wedge B}(@)$ . From (2), by (S[A]1) again,  $f_{A \wedge B}(@) \subseteq |C|$ . From this and the previous identity,  $f_A(@) \subseteq |C|$ . From this via (S[A]1) again,  $@ \Vdash^+ [A]C$ ).

Now Special Transitivity has some very good instances.  $[A]B$ : when you imagine yourself winning the lottery, you imagine yourself having a lot of money.  $[A \wedge B]C$ : when you imagine yourself winning the lottery and having a lot of money, you imagine yourself happy. Thus,  $[A]C$ : when you imagine yourself winning the lottery, you imagine yourself happy. And I can think of no intuitive counterexamples. But that does not mean none could be found, of course (hence the “tentatively” above).

Having looked at what is (tentatively) valid in our semantics, let us see what is (interestingly) invalid. Quite a lot, it turns out.

<sup>11</sup> General Transitivity fails for our operators, just as it does for *ceteris paribus* conditionals. Its failure is a consequence of the failure of Antecedent Strengthening, to which we are about to come.

## 8 Invalidities

*Ceteris paribus* conditionals are, notoriously, non-monotonic in that Antecedent Strengthening fails from them: a counterfactual “If it were the case that  $A$ , then it would be the case that  $B$ ” does not entail “If it were the case that  $A \ \& \ C$ , then it would be the case that  $B$ ”. Our imagination operators inherit such a nice feature, for the following is invalid in the semantics:

$$[A]B \not\models [A \ \& \ C]B$$

(Counterexample: suppose  $@$   $R_p$ -accesses nothing. Then  $@ \Vdash^+ [p]q$ . Let  $@ R_{p \wedge r} w$  ( $w$  may be possible, or not),  $w \Vdash^+ p \ \& \ r$  but not  $w \Vdash^+ q$ . Then it is not the case that  $@ \Vdash^+ [p \ \& \ r]q$ ).<sup>12</sup> An act of imagination (in a given context) is individuated by its explicit content. And one cannot indiscriminately import further information into the explicit content itself without turning it into a different act. You imagine that you fail your logic class, and you will imagine yourself in a sad mood. But if you imagine failing your logic class and that everyone else has failed, so that the exam needs to be re-taken with an easier array of exercises, your mood will not be that sad in such a scenario. What does the trick is the fact that  $f_A(w)$  need not be the same as  $f_{A \wedge C}(w)$ . The variability in the strictness of our operators is the essential tool securing such non-monotonicity of our exercises of imagination.

Next, here is one invalidity essentially involving the hyperintensional features of our operators:

$$A \rightarrow B \not\models [A]B$$

(Counterexample: let all  $w \in P$  be such that not  $w \Vdash^+ p$ . Then  $@ \Vdash^+ p \rightarrow q$ . For some  $w_1 \in I$ , let  $@ R_p w_1$ ,  $w_1 \Vdash^+ p$  but not  $w_1 \Vdash^+ q$ . Then not  $@ \Vdash^+ [p]q$ ). Recall that the premise is an intensional (strict) conditional: all the possible  $A$ -worlds are  $B$ -worlds. However, in an act of imagination whose explicit content is given by  $A$ , we do not automatically imagine that  $B$ : as our act is hyperintensional, that is, it discriminates between various absolute impossibilities, we may look at impossible  $A$ -worlds where  $B$  fails. In particular, strict conditionals which logicians in the tradition of relevant logics (see Mares 2004 for a nice introduction) call “irrelevant”, such as conditionals which hold just because the antecedent is impossible, or the consequent necessary, do not imply the corresponding irrelevant conceivings. In our semantics, this is fine:

$$\models (A \ \& \ \neg A) \rightarrow B$$

However, this fails:

$$\not\models [A \ \& \ \neg A]B$$

<sup>12</sup> I use sentential letters,  $p, q, \dots$ , for invalidity arguments, for to show invalidity we present a counter-model assigning truth values to particular formulas. I use the schematic meta-variables for formulas,  $A, B, \dots$ , for validity arguments, for there one argues for the validity of any instance of the relevant schema (as in Priest (2008): 10–11).



(Just let  $w \in I$ , @  $R_{p \wedge \neg p} w$ ,  $w \Vdash^+ p \wedge \neg p$  but not  $w \Vdash^+ q$ ). That we explicitly imagine an inconsistent scenario does not mean that we trivialize our act of imagination. Similarly, although we have:

$$\models A \rightarrow (B \rightarrow B)$$

the counterpart for imagination fails:

$$\not\models [A](B \rightarrow B)$$

(Let  $w \in I$ , @  $R_p w$ ,  $w \Vdash^+ p$  but not  $w \Vdash^+ q \rightarrow q$ ). In general, we can discriminate between logical or absolute necessities and we do not conceive them automatically, independently of what we (explicitly) conceive. Thanks to impossible worlds, we have:

$$\Box B \not\models [A]B$$

$$\neg \Diamond A \not\models [A]B$$

## 9 Modus Ponens in Imagination

To discuss the plausibility of one further constraint, we need to introduce the notion of *cotenability*. This is the connection that, by holding between some information and a formula,  $A$ , makes the information eligible to be imported into the act of imagination whose explicit content is given by  $A$  (the term was used by Lewis (1973) in the context of counterfactuals; he took it from Goodman).  $[A]B$  will hold (at a world) when the explicitly imagined content,  $A$ , plus a *ceteris paribus* clause, say,  $C_A$ , dependent on  $A$  and cotenable with  $A$  (at that world), entails  $B$ .  $C_A$  is not an ordinary premise or set of premises, but works rather like a catch-all *ceteris paribus* clause: it captures the background information we hold fixed relative to  $A$ , and which we can import into our imagined scenario.<sup>13</sup>

Now so far we generically referred to what is imported in our exercises of *ceteris paribus* imagination as “information”. But we should add now that our cotenable information is not made only of truths,<sup>14</sup> because “what people do not change when they create a counterfactual alternative depends on their beliefs” (Byrne 2005: 10), and believed falsities may get involved. Therefore, (the counterpart in our framework of) what Lewis (1973) called Weak Centering should not hold in our semantics:

$$(Weak\ Centering)\ \text{If } w \in |A|, \text{ then } w \in f_A(w).$$

<sup>13</sup> What background is imported is constrained by what is relevant with respect to the explicit content. Such relevance is difficult to pin down formally, but the intuitive insight is clear. Although you know that the city of Brussels is composed of more than 1000 bricks, this is irrelevant with respect to your imagining that you fail your logic class. Your mental representation is not about that. So you need not import information about those bricks.

<sup>14</sup> Information may be factive, as claimed by Floridi (2005). If so, it would not even be appropriate to label as “information” what is imported in our exercises of imagination. The issue is controversial, however. We can safely adopt a weak conception of semantic information as meaningful, well-formed data which need not be truthful.

This entails that if a world  $w$  realizes the explicit content of an act of imagination  $[A]$ , then  $w$  is one of the worlds in the set outputted by the selection function for  $A$ . Even restricted to possible worlds, Weak Centering validates what we may call *imaginative modus ponens*:

$$A, [A]B \models ? B$$

If the explicit content of an act of imagination actually obtains, and it is represented in that act that  $B$ , then  $B$  also obtains. This is wrong. We can import false but cotenable relevant beliefs into our representation, as part of the  $C_A$ . And this can make us imagine falsities although  $A$  gets things right. For instance, you imagine that Merkel is signing treatises in Brussels,  $[A]$ , but you mistakenly believe Brussels to be in France. you import the (relevant, cotenable) belief and you imagine Merkel signing treatises in France,  $[A]B$ .  $A$  is true, but it does not follow that  $B$  is.

Then even if we manage to impose a plausible closeness metric on worlds in the form of a system of nested similarity spheres, we already know that this will not be an even weakly centred system. The role of @ (for things imagined at @) is just to fix the beliefs that are actually held (by the relevant conceiving agent), rather than what actually is the case. Once this is done, @ steps out of the picture.

## 10 Primary Versus Secondary Conceivability?

I close by mentioning a question for further research, which I leave open for the time being: are such imagination operators actually closer to subjunctive *ceteris paribus* conditionals, or rather to indicatives? This is a major question which deserves separate treatment in a further paper. I will limit myself to some tentative considerations here.

It has often been remarked that the logic of the two kinds of conditionals is very similar (both kinds fail Antecedent Strengthening, Transitivity, etc.). A key difference is that what is cotenable with respect to indicatives is not made of facts, but of beliefs (see Bennett 2003: 175–176). This would bring our *ceteris paribus* imagination operators closer to indicatives, for the failure of imaginative *modus ponens* above is due precisely to cotenability for them being tied to beliefs which might be false.

Indicatives are connected to subjective probabilities, or degrees of belief, so much so that according to some (including Bennett himself) one cannot even give a truth-conditional semantics for them. But even if indicative *ceteris paribus* conditionals lack truth values (which is controversial anyway), one should not suspect that our *ceteris paribus* operators themselves lack genuine truth conditions, and thus that the prospect of a truth-conditional semantics for them as sketched here is hopeless. For authors like Bennett, indicatives lack truth values for they report or describe nothing, although they express something about the (conditional) belief arrangements of those who utter them. But a formula of the form  $[A]B$  is exactly a report of the mental state of the relevant conceiving agent: it reports that the agent

imagines that  $B$  (in a certain context) in the act of imagination whose explicit content is given by  $A$ ; and such a report may be true or false.

I suspect that our variably strict operators may behave in a way more similar to indicatives, or to subjunctives, depending on how the *ceteris paribus* worlds are selected on the basis of some similarity metric for them. It might be, that is, that two different kinds of similarity or closeness are in play here. None of this surfaced in this paper, because I have introduced no similarity metric.

I conjecture that one may impose two different similarity structures, which would account for two different kinds of conceivability or imagination in the sense of Chalmers' (2002): a primary conceivability where we imagine a certain scenario as a candidate for actuality, and which works in a way more similar to indicative *ceteris paribus* conditionals; and a secondary conceivability where we imagine a certain scenario as counterfactual, and which works closely to subjunctive conditionals in the sense of the relevant worlds similarity structure, although it differs (at least) in that Weak Centering is lacking. If such a development of the semantics presented above is feasible, it may nicely connect the framework to mainstream debates about conceivability and two-dimensional semantics (Garcia-Carpintero and Macia 2006). Whether the development is feasible hinges on the crucial issue flagged above: how to account for world similarity in models of *ceteris paribus* imagination that use possible and impossible worlds. I will thus not venture here into a more detailed discussion of indicative vs. subjunctive and primary vs. secondary, because the world similarity problem has been intentionally left out of this paper, in view of subsequent work.

**Acknowledgements** This paper is published within the project “The Logic of Conceivability”, funded by the European Research Council (ERC CoG), grant number 681404.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Belnap, N. D. (1977). A useful four-valued logic. In J. M. Dunn & G. Epstein (Eds.), *Modern uses of multiple-valued logics* (pp. 8–37). Dordrecht: Reidel.
- Belnap, N. D., Perloff, M., & Xu, M. (2001). *Facing the future*. Oxford: Oxford UP.
- Bennett, J. (2003). *A philosophical guide to conditionals*. Oxford: Oxford UP.
- Berto, F. (2008). Modal meinongianism for fictional objects. *Metaphysica*, 9, 205–218.
- Berto F. (2013). Impossible Worlds. *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/impossible-worlds/>.
- Berto, F. (2014). On conceiving the inconsistent. *Proceedings of the Aristotelian Society*, 114, 21–103.
- Bjerring, J. C. (2013). Impossible worlds and logical omniscience: An impossibility result. *Synthese*, 190, 24–2505.
- Bjerring, J. C. (2014). On counterpossibles. *Philosophical Studies*, 168, 53–327.
- Brogaard, B., & Salerno, J. (2013). Remarks on counterpossibles. *Synthese*, 190, 60–639.
- Byrne, R. (2005). *The rational imagination*. Cambridge, MA: MIT Press.

- Chalmers, D. (2002). Does conceivability entail possibility? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility* (pp. 145–199). Oxford: Oxford UP.
- Chellas, B. (1975). Basic conditional logic. *Journal of Philosophical Logic*, 4, 53–133.
- Chellas, B. (1989). *Modal logic: An introduction*. Cambridge: Cambridge UP.
- Costa-Leite, A. (2010). Logical properties of Imagination. *Abstracta*, 6, 16–103.
- Delgrande, J. P. (1988). An approach to default reasoning based on a first-order conditional logic. *Artificial Intelligence*, 36, 63–90.
- Dunn, J. M. (1976). Intuitive semantics for first-degree entailment and ‘Coupled Trees’. *Philosophical Studies*, 29, 68–149.
- Duží, M., Jespersen, B., & Materna, P. (2010). *Procedural semantics for hyperintensional logic*. Dordrecht: Springer.
- Eberle, R. A. (1974). A logic of believing, knowing and inferring. *Synthese*, 26, 82–356.
- Fagin, R., & Halpern, J. Y. (1988). Belief, awareness and limited reasoning. *Artificial Intelligence*, 34, 39–76.
- Fagin, R., yHalpern, J., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge, MA: MIT press.
- Floridi, L. (2005). Is information meaningful data? *Philosophy and Phenomenological Research*, 70, 351–370.
- Garcia-Carpintero, M., & Macia, J. (Eds.). (2006). *Two-Dimensional Semantics*. Oxford: Oxford UP.
- Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. Ithaca, NY: Cornell UP.
- Horty, J. F. (2001). *Agency and deontic logic*. Oxford: Oxford UP.
- Jago, M. (2007). Hintikka and Cresswell on logical omniscience. *Logic and Logical Philosophy*, 15, 54–325.
- Jago, M. (2014). *The impossible: An essay on hyperintensionality*. Oxford: Oxford UP.
- Kahneman, D., Slovic, P., & Twersky, A. (1982). *Judgment under uncertainty*. Cambridge: Cambridge UP.
- King, J. (1996). Structured propositions and sentence structure. *Journal of Philosophical Logic*, 25, 495–521.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard UP.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly*, 15, 37–46.
- Mares, E. (2004). *Relevant logic*. Cambridge: Cambridge UP.
- Meyer, R., & van der Hoek, W. (1995). *Epistemic logic for AI and computer science*. Cambridge: Cambridge UP.
- Nichols, S., & Stich, S. (2003). *Mindreading. An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Oxford UP.
- Niiniluoto, I. (1985). Imagination and fiction. *Journal of Semantics*, 4, 22–209.
- Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic*, 38, 72–535.
- Priest G. (1987), *In contradiction. A study of the transconsistent*, Dordrecht: Martinus Nijhoff, 2nd expanded ed, Oxford: Oxford University Press, 2006.
- Priest, G. (2005). *Towards non-being*. Oxford: Oxford UP.
- Priest, G. (2008). *An introduction to non-classical logic*, 2nd edn, Cambridge: Cambridge UP.
- Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25, 157–182.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Rantala, V. (1982). Impossible world semantics and logical omniscience. *Acta Philosophica Fennica*, 35, 106–115.
- Rescher, N. (Ed.). (1968). A theory of conditionals. *Studies in logical theory* (pp. 98–112). Blackwell: Oxford.
- Ripley, D. (2012). Structures and circumstances: Two ways to fine-grain propositions. *Synthese*, 189, 97–118.
- Roese, N. J., & Olson, J. (1993). The structure of counterfactual thought. *Personality & Social Psychology*, 19, 312–319.

- Roese, N. J., & Olson, J. (Eds.). (1995). *The social psychology of counterfactual thinking*. New York: Taylor & Francis.
- Sainsbury, M. (2010). *Fiction and fictionalism*. London & New York: Routledge.
- Seegerberg, K. (1989). Notes on conditional logic. *Studia Logica*, 48, 157–168.
- Shepard, R. N., Cooper, L. A., et al. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Stalnaker, R. (1981). A defence of conditional excluded middle. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 87–105). Dordrecht: Reidel.
- Vaidya A. (2015). The epistemology of modality. *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/modality-epistemology>.
- Wansing, H. (1990). A general possible worlds framework for reasoning about knowledge and belief. *Studia Logica*, 49, 523–539.
- Wansing H. (2015). Remarks on the logic of imagination. A step towards understanding doxastic control through imagination. *Synthese*, On Line First 23 October 2015. <http://link.springer.com/article/10.1007>.
- Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.