CrossMark

# Taming the runabout imagination ticket

**Francesco Berto**[1] ⓘ

**Abstract** The 'puzzle of imaginative use' (Kind and Kung in Knowledge through imagination, Oxford University Press, Oxford, 2016) asks: given that imagination is arbitrary escape from reality, how can it have any epistemic value? In particular, imagination seems to be logically anarchic, like a runabout inference ticket: one who imagines *A* may also imagine whatever *B* pops to one's mind by free mental association. This paper argues that at least a certain kind of imaginative exercise—*reality-oriented mental simulation*—is not logically anarchic. Showing this is part of the task of solving the puzzle. Six plausible features of imagination, so understood, are listed. Then a formal semantics is provided, whose patterns of logical validity and invalidity model the six features.

**Keywords** Epistemology of imagination · Mental simulation · Counterfactual thinking · Variably strict epistemic modals · Aboutness

## 1 Imagination as tonk?

In their beautiful, recently edited collection *Knowledge Through Imagination*, Amy Kind and Peter Kung state the *puzzle of imaginative use* (Kind and Kung 2016): given that imagination is arbitrary escape from reality, how can it enable us to learn about reality? This looks like a real dilemma, for both of the supposedly incompatible

---

✉ Francesco Berto
F.Berto@uva.nl

[1] Department of Philosophy, Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Oude Turfmarkt 141-147, 1012 GC Amsterdam, The Netherlands

features of imagination—arbitrariness, epistemic value—are well established. Here's Hume on arbitrariness:

> Nothing, at first view, is more unbounded than the thought of man, which not only escapes all human power and authority, but is not even restrained within the limits of nature and reality. To form monsters, and join incongruous shapes and appearances, costs the imagination no more trouble than to conceive the most natural and familiar objects. And while the body is confined to one planet, along which it creeps with pain and difficulty; the thought can in an instant transport us into the most distant regions of the universe; or even beyond the universe, into the unbounded chaos, where nature is supposed to lie in total confusion. (*Enquiry*, 2)

Yet Hume also thought, famously, that:

> 'Tis an establish'd maxim in metaphysics, that whatever the mind clearly conceives includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible. (*Treatise*, I, ii, 2)

Thus, imagination is, at least, a pathway to knowledge of absolute modalities.[1] Of course, Hume's overall view is consistent: imagination is completely unbounded when it deals with (recombinations of) matters of fact, but bound by absolute necessity as captured by relations of ideas.

However, this stance won't quite solve the puzzle. It is widely agreed in cognitive psychology (Markman et al. 2009) that imagination, taken as some kind of mental simulation (we'll get more precise soon), is of epistemic value also for more down-to-earth purposes than knowledge of absolute modalities. It allows us not only to improve future performance (that's on the know-how side), but also to make contingency plans by successfully anticipating future outcomes, or by learning from mistakes via the consideration of alternative courses of action (that's on the know-that side). We simulate alternatives to reality in our mind, to explore what would and would not happen if they were realized. It was forcefully argued by Ruth Byrne (2005) that such exercises of the mind have some logic: some things follow in the conceived situation, some do not. And we can often form reliable judgments around these issues. But how can this be, if imagination is governed by free associations of ideas of the kind highlighted by Hume? An imagination operator will look like Prior's *tonk*: given input $A$, it will output $B$, where $B$ is whatever one likes. By taking you anywhere, imagination will take you in no place of added epistemic value.

I think that imagination is not, in general, tonk-like. I think, that is, that once one has pinned down a certain meaning of 'imagination', it makes sense to speak of a logic of such activity. Showing that this is so is part of the project of solving the puzzle of imaginative use.

'Imagination' is highly ambiguous. What kind of mental activity is being targeted here? Starting from the more general: we will focus on propositional imagination. We

---

[1] Many (Byrne 2007; Fiocco 2007; Jago 2014; Kung 2014; Priest 2016) think that, *contra* Hume, we can imagine the absolutely impossible. I agree, see Berto and Schoonen (2017).

will deal with what happens in the mind of subjects who imagine that they jump a stream, that the furniture in the living room is arranged in such-and-such way, that *Wuthering Heights*' Heathcliff and Catherine meet for the last time, that the Nazis have nukes in 1944, that the streets of Amsterdam are full of bikers running around.

More specifically, we want to model imagination qua *reality-oriented mental simulation* (ROMS). This is the kind of mental activity most of the essays collected in (Kind and Kung 2016) focus on. The consensus seems to be that such activity can have epistemic value, insofar as the alteration of reality thereby represented is constrained by *topicality* and *minimal alteration* with respect to what one believes or knows actually to be the case. We shall soon be more precise on what this is supposed to mean: the following Section will list six features imagination, understood as ROMS, arguably has, among which topicality and minimal alteration figure prominently, and around which there seems to be some agreement among philosophers and cognitive psychologists. The Section after that will present a formal semantics. The Sections following it will show that the semantics complies with, and models, those features.

## 2 Six features of imagination qua ROMS

1. *Imagining is more substantive than assuming or supposing*, as when we make an assumption or supposition in a proof (Balcerak Jackson (2016) makes a forceful case for the difference; see also Kind 2016). What does 'more substantive' mean here? I refer to the celebrated works of Yablo (1993) and Chalmers (2002) on conceivability and imagination. They speak of a notion called by the latter 'positive conceivability':

   Positive notions of conceivability require that one can form some sort of positive conception of a situation in which *p* is the case. One can place the varieties of positive conceivability under the broad rubric of *imagination*: to positively conceive of a situation is to imagine (in some sense) a specific configuration of objects and properties. [...] Overall, we can say that *p* is positively conceivable when one can imagine that *p*: that is, when one can imagine a situation that verifies *p*. (Chalmers 2002, p. 150, notation modified.)

   Thus, imagining that *p* is more than merely tokening in the mind a *p*-sentence. It is to represent a *situation*—a configuration of objects, properties and relations—of which *p* is a truthful description: a situation that is a truthmaker for *p*.
   As we talk of representing in the mind, what sort of mental representation is involved here? This is our next issue.

2. *Imagination has some mereological structure*. The Yablo and Chalmers account does not focus on the question: *how* do the relevant mental representations represent? Suppose that, as claimed by Kind (2001), imagination essentially involves pictorial, modal mental imagery. Based on experiments going back to the 70's and 80's (Kosslyn and Pomerantz 1977; Kosslyn 1980; Shephard and Cooper 1982), it has been argued that such imagery has quasi-spatial features: typically, we rep-

resent concrete objects and situations in three-dimensional egocentric space.[2] In his seminal book on mental representations Paivio (1986), p. 198, argues that representations of this kind are processed 'in parallel', not serially (as it happens with 'amodal', propositional or language-like representations), because they have mereological structure: one can, for instance, pictorially represent to oneself the arrangement of one's own living room and describe its contents from different viewpoints, mentally scanning the objects included there from top to bottom, or from left to right, mentally zooming into a corner, etc.

However, Williamson (2007, 2016) insists that imagination as mental simulation is not perforce tied to mental imagery. First, we can imagine situations involving abstract objects which are not visually imaginable. Second, we can imagine situations which are too complex for a representation of them to essentially involve imagery: 'Suppose a politician is trying to work out what his core supporters would do at the next election if he voted for gun control. He imagines their reactions, but doing so need not involve mental imagery' (Williamson 2016, p. 117).

But even Williamson thinks that imagination is *conjunctive*. His example concerns imagining a concrete scenario, but comes from a context in which he talks of the logical features of imagination in general: 'It would be hard to imagine that Mary is tall and thin without imagining that she is tall and imagining that she is thin' (Ibid, p. 121). For recall that we're not targeting mere supposition, or the tokening of a sentence in the head. Rather, we are dealing with Chalmersian positive conceivability: representing in the mind a situation that works as a truthmaker. Now it seems difficult for a situation to make true $A \wedge B$ without making true $A$ and making true $B$ and vice versa. Indeed, this is a generally agreed principle of the logic of (inexact) truthmaking (Fine 2014, 2015; Yablo 2014); and it mirrors the idea that situations themselves have can stand in parthood relations.

3. *Acts of imagination as mental simulation have a deliberate starting point*: we set out to target an explicit content, *that A*. So we should focus on specific acts of imagination voluntarily performed by a subject. It has been widely acknowledged (Wansing 2015; Langland-Hassan 2016; Van Leeuwen 2016; Williamson 2016) that imagination as mental simulation can be deliberate and voluntary in ways belief is not: you can imagine that all of London has been painted blue but, having overwhelming evidence of the contrary, you cannot make yourself believe it.

In their well-known cognitive model of mental simulation, Nichols and Stich (2003) have 'an initial premiss or set of premisses, which are the basic assumptions about what is to be pretended' (p. 24). This may be made up by the conceiver ('Now let us imagine what would happen if…'), or it may be given as an external instruction (think of going through a novel and take the sentences you read as your sequential input). In our formalism below, we represent the explicit input of an act of imagination as directly expressed by formulas. Such formulas index the imagination operators we will use.

---

[2] That the quasi-spatial features of mental imagery are essential to its representational role is, to be sure, disputed: this is the famous 'imagery debate' (Pylyshyn 1973, 2002), and I won't get into it.

Inputs are mere beginnings. They need to be developed. This is our next Feature.

4. *We integrate the explicit input A with background information we import, contextually, depending on A and what we know or believe*: once the initial input is in, Nichols and Stich (2003) claim, 'children and adults elaborate the pretend scenarios in ways that are not inferential at all', filling in the explicit instruction with 'an increasingly detailed description of what the world would be like if the initiating representation were true' (pp. 26–28). The additional details are borrowed from our knowledge or belief base (Van Leeuwen 2016, p. 95).

As we imagine the last meeting between Heathcliff and Catherine, we represent Heathcliff dressed as an an Eighteenth-Century country gentleman, not as a NASA astronaut. The text of *Wutering Heights* never says this explicitly, nor do we infer this from the text via sheer logic. Rather, we import such information into the represented situation, based of what we know: we know that the story is temporally located in the Eighteenth Century, and we assume, lacking information to the contrary from the text, that Heathcliff is dressed as we know country gentlemen were dressed at the time.

The modeling of imagination proposed below will stick to the tradition in epistemic and doxastic logic, of understanding intentional notions like knowledge, belief, and (cognitive) information as restricted quantifiers over possible worlds in a modal setting (Hintikka 1962). However, I propose to model this Feature 4 of imagination via modal operators interpreted as *variably* strict quantifiers over worlds. The variability of strictness will account for the contextual selection of the information we import in an act of imagination when we integrate its explicit input. The input will play a role similar to a variably strict conditional antecedent, in the style of the possible worlds semantics for counterfactuals due to Stalnaker (1968), Lewis (1973). This is plausible given that in acts of mental simulation, as Nichols and Stich told us, we represent 'what the world *would* be like *if* the initiating representation were true'.

The Lewis–Stalnaker semantics resorts, famously, to a similarity metric based on a total ordering of possible worlds (roughly: to establish whether $A$ counterfactually implies $B$, one checks whether the closest $A$-worlds are $B$-worlds). We will have one such metric in our framework below. One should, however, not think of it in terms of objective similarity—rather, in terms of our looking at the subjectively most *plausible* worlds where $A$ is true, as with the so-called 'grove spheres' (Grove 1988) used to model AGM-style belief revision (Alchourrón et al. 1985) in (static or dynamic) epistemic logic (van Ditmarsch et al. 2007; van Benthem and Smets 2015). This makes a lot of sense again, because imagination as ROMS is a sort of simulated belief revision. As Van Leeuwen (2016) has it, 'the things we […] imagine are typically things that could happen in the environment in which we live, given what we believe about that environment' (p. 93).

The key epistemic role of imagination is to allow us to reliably form new conditional beliefs. Here's an example that draws on (Williamson 2016, p. 116): you may gain some advantage by jumping across a small river rather than undertaking a long walk around it, but you are not quite sure you will make it. Will you succeed if you jump the stream? You don't just blindly try, for that may be dangerous: you may fall in the water, or hurt yourself. So you mentally simulate the scenario:

you imagine jumping. You integrate the input importing beliefs concerning the distance, your physical abilities, past performance, etc. You let the story unfold. In the imagined scenario, you get to the other side. You add the conditional 'If I jump, I will get to the other side' to your belief stock.

Also according to the Nichols and Stich model (Nichols and Stich 2003, p. 62), an act of mental simulation is a form of subjunctive or counterfactual thought, whose main result is the import of newly formed conditional beliefs from what they call the 'possible worlds box' (where the pretence premise goes) to the 'belief box' (the long-term memory where our knowledge or belief base is stored). And the 'Updater' module in their boxology works in a way similar to a kind of simulated belief revision.

Why 'simulated'? Of course, we generally do not believe that the scenario we imagine is actual (you imagine jumping, but you don't believe you are actually jumping). However, we can come to actually believe that *if* the explicit imaginative input obtains, *then* what we have imagined in developing the simulation input will also obtain.

Looking at the most plausible worlds is connected to the idea that such exercises of imagination are 'reality-oriented': we stick to the minimal input-permitting alteration of how we take reality to be (Kind 2016). Reality-orientation comes up also in our list's next item.

5. *Imagination has topicality or relevance constraints*. We do not indiscriminately import unrelated contents into the conceived scenarios: '[We require] that the world be imagined as it is *in all relevant respects*' (Kind 2016, p. 153). What is imported is constrained by what is *on-topic* with respect to the input.

   Topicality is a distinguishing feature of ROMS as opposed to free-floating mental wandering: you know that Nuku'alofa is the capital of Tonga, but this is immaterial to your imagining Catherine and Heathcliff's adventures as per Emily Brontë's book, in so far as such adventures do not involve Tonga at all. The story is not *about* that. So you will not, in general, import such irrelevant content in your scenario.

   Such a topicality of ROMS will be represented in our framework by imposing topic-preservation constraints on our variably strict modal operators. Such constraints will take care of the mereological features of imagination listed as Feature 2 above, while ensuring that off-topic contents don't sneak into the imagined scenario.

   One point of (on-topic) belief importation from Features 4 and 5 is to fill in the explicit input with increasing (relevant) detail. We don't reach a maximally detailed representation, though. This brings to our last Feature.

6. *Imagination generally under-determines its contents*: an exercise of imagination involves adding content, suitably imported from our knowledge or belief base, to the starting input, in the business of building a coherent representation. However, we generally don't get into the full details, for the world is too detailed for our cognitive capacities.

   This means that we imagine things vaguely, but this does not entail that we imagine vague things. You imagine the moorland farmhouse of *Wuthering Heights* and you picture its windows and walls exposed to the stormy weather. You do not represent all the details, but you want the details to be there. The farmhouse is no vague

object in the scenario—one composed of an objectively indeterminate number of bricks. Either the number of bricks making up the house is even, or it is odd, but you do not represent it either way.

As we will see, Feature 6 will be modeled by the imaginative input having us look at a plurality of (most plausible) worlds, which will fill in the unspecified details in different ways.

On to the modeling!

## 3 Formal semantics

The formalism presented here draws on that of Berto (2017a), which is an enriched possible worlds semantics. The main novelty added in *this* paper, is that we take seriously the view of imagination as simulated belief revision by adding a plausibility metric on worlds, which is to represent belief entrenchment and dispositions to belief revision. The metric is in the style of Grove (1988).[3]

We use a sentential language $\mathcal{L}$ with an indefinitely large set $\mathcal{L}_{AT}$ of atomic formulas, $p, q, r$ ($p_1, p_2, \dots$), negation $\neg$, conjunction $\wedge$, disjunction $\vee$, a strict conditional $\prec$, round parentheses as auxiliary symbols ( , ), square parentheses [ , ], which we'll put to special use to form our imagination operators. We use $A, B, C, \dots$, as metavariables for formulas of $\mathcal{L}$. The well-formed formulas are items in $\mathcal{L}_{AT}$ and, if $A$ and $B$ are formulas:

$$\neg A \mid (A \wedge B) \mid (A \vee B) \mid (A \prec B) \mid [A]B$$

(We normally omit outermost brackets). We can just take $\mathcal{L}$ as the set of its well-formed formulas. We read '$[A]B$' as 'In an act of imagining with explicit input $A$, one imagines that $B$', or, briefly: 'Given input $A$, one imabines $B$'.

In the semi-formal metalanguage we have variables $w, w_1, w_2, \dots$, ranging over worlds, $x, y, z$ ($x_1, x_2, \dots$), ranging over *topics* (we'll come to what these are in a moment), and the symbols $\Rightarrow, \Leftrightarrow, \&, or, \sim, \forall, \exists$, read the usual way. A *frame* for $\mathcal{L}$ is a tuple $\mathfrak{F} = \langle W, \{R_A \mid A \in \mathcal{L}\}, \mathcal{T}, \oplus, t \rangle$, understood as follows:

- $W$ is a set of possible worlds.
- $\{R_A \mid A \in \mathcal{L}\}$ is a set of accessibilities between worlds: each $A \in \mathcal{L}$ has its own $R_A \subseteq W \times W$. These satisfy a number of conditions, to which we come soon.
- $\mathcal{T}$ is a set of *topics*.[4] We should understand topics as somewhat similar to cognitively understood Yablovian *subject matters* (Yablo 2014): they are the situations

---

[3] Besides Berto (2017a), a version of the semantics used in this paper, which resorts to a plausibility metric, can be found in Berto (2018). There, the relevant operators are interpreted precisely as a kind of conditional belief or (static) belief revision operators, in the context of a discussion of the AGM postulates for belief revision.

[4] These were called 'contents' in Berto (2017a, 2018). I'm not so happy with that terminology anymore: it has topic-preservation expressed as 'content containment', and this reminds one of some Kantian analyticity, which is misleading in this context. The terminology of contents was influenced by the fact that the semantics bears similarities with some 'logics of content containment' inspired by the work of Parry (1933) (Angell 1977; Fine 1986; Urquhart 1973), of which (Ferguson 2014) is an excellent survey and interpretation.

(the configurations of objects, properties and relations) the intentional states of the subject engaging in the imaginative act are *about*. We want a mereology of such things, capturing both the aforementioned Feature 2 (imagination has some kind of mereological structure), and Feature 5 (the importation of background information must respect topicality constraints).

- $\oplus$ is *topic fusion*, an algebraic operation on $\mathcal{T}$ satisfying, for all $xyz \in \mathcal{T}$:
    - (Idempotence) $x \oplus x = x$
    - (Commutativity) $x \oplus y = y \oplus x$
    - (Associativity) $(x \oplus y) \oplus z = x \oplus (y \oplus z)$

  Fusion shall be unrestricted, that is, $\oplus$ is always defined on $\mathcal{T}$: $\forall xy \in \mathcal{C}\ \exists z \in \mathcal{T}(z = x \oplus y)$. *Topic parthood*, $\leq$, can then be defined the usual way: $\forall xy \in \mathcal{T}(x \leq y \Leftrightarrow x \oplus y = y)$. Thus, it's a partial ordering—for all $xyz \in \mathcal{T}$:
    - (Reflexivity) $x \leq x$
    - (Antisymmetry) $x \leq y\ \&\ y \leq x \Rightarrow x = y$
    - (Transitivity) $x \leq y\ \&\ y \leq z \Rightarrow x \leq z$

  Then $\langle \mathcal{T}, \oplus \rangle$ is a join semilattice (we could have done things the other way around, having the partial ordering in the frames and defining fusion out of it, but this would have made little difference, and the algebraic setting might be more intuitive). We will also assume that $\mathcal{T}$ is complete: any set of topics $S \subseteq \mathcal{T}$ has a fusion $\oplus S$. Finally, we can think of all topics in $\mathcal{T}$ as built via fusions out of *atoms*, topics with no proper parts ($Atom(x) \Leftrightarrow \sim \exists y(y < x)$, with $<$ the strict order defined from $\leq$). $\langle \mathcal{T}, \oplus \rangle$ is needed to assign topics to formulas of $\mathcal{L}$, as follows.

- $t$ is a function, $t : \mathcal{L}_{AT} \to \mathcal{T}$, such that if $p \in \mathcal{L}_{AT}$, then $t(p) \in \{x \in \mathcal{T} | Atom(x)\}$: atomic topics are assigned to atomic formulas (this makes of our $\mathcal{L}$ an idealized language: grammatically simple sentences of ordinary language can be about intuitively complex contents). Next, $t$ is extended to the whole of $\mathcal{L}$: if $\mathfrak{At}A = \{p_1, \ldots, p_n\}$, the set of atoms in $A$, then $t(A) = \oplus \mathfrak{At}A = t(p_1) \oplus \cdots \oplus t(p_n)$: the topic of a formula is that of its atoms taken together.

Our little mereology of topics will allow us to model various *hyperintensional* distinctions. We want our acts of imagining to be, in general, capable of discriminating between logically or necessarily equivalent contents.[5] However, we don't (want to) get as fine-grained as the syntax of our language. In particular we have that $t(A) = t(\neg\neg A)$; but also that $t(A) = t(\neg A)$: the topic of a formula is that of its negation. And not only $t(A \wedge B) = t(B \wedge A)$, but also, e.g., $t(A \wedge B) = t(A) \oplus t(B) = t(A \vee B)$. These are often taken as key requirements for a good account of aboutness- or topic-inclusion (Yablo 2014; Fine 2015).[6]

---

[5] Though I will not argue for this here, I just take hyperintensionality to be a general feature of intentional states. Hence, it was not listed as a specific feature of imagination as ROMS above.

[6] I mention that very similar requirements are imposed on the notion of *awareness* in several awareness-based epistemic logics. Such logics aim at capturing awareness as the mental state of an intentional agent—roughly: what the agent hosts in its working memory, based on concepts in its possession. The seminal paper in this area is (Fagin and Halpern 1988), where the (essentially arbitrary) set of formulas an agent is aware of is treated syntactically. The more constrained notion of 'awareness determined by primitive propositions' (Schipper 2015) determines a semantics close to our mereology of topics.

A *model* $\mathfrak{M} = \langle W, \{R_A \mid A \in \mathcal{L}\}, \mathcal{T}, \oplus, t, \Vdash \rangle$ is a frame with an interpretation $\Vdash \subseteq W \times \mathcal{L}_{AT}$. Read '$w \Vdash p$', as usual, as meaning that $p$ is true at $w$, '$w \nVdash p$' as $\sim w \Vdash p$. The interpretation is extended into a recursive set of truth conditions for all of $\mathcal{L}$:

- (S¬) $w \Vdash \neg A \Leftrightarrow w \nVdash A$
- (S∧) $w \Vdash A \wedge B \Leftrightarrow w \Vdash A$ & $w \Vdash B$
- (S∨) $w \Vdash A \vee B \Leftrightarrow w \Vdash A$ *or* $w \Vdash B$
- (S≺) $w \Vdash A \prec B \Leftrightarrow \forall w_1(w_1 \Vdash A \Rightarrow w_1 \Vdash B)$
- (S[A]) $w \Vdash [A]B \Leftrightarrow \forall w_1(wR_Aw_1 \Rightarrow w_1 \Vdash B)$ & $t(B) \leq t(A)$

Read '$wR_Aw_1$' as meaning that $w_1$ is, from $w$'s viewpoint, one of the most plausible worlds where $A$ is true. The insight: when you imagine with input $A$, you look at the bunch of worlds representing the most plausible ways things would be, for all you know or believe, if $A$ were the case.

(S[A]) can be equivalently expressed using set-selection functions (as in Lewis 1973, pp. 57–60). Each $A \in \mathcal{L}$ has a function $f_A : W \to \mathcal{P}(W)$ selecting the bunch of most plausible $A$-worlds, $f_A(w) = \{w_1 \in W | wR_Aw_1\}$. If $|A| = \{w \in W | w \Vdash A\}$, we can tidy up the clause:

- (S[A]) $w \Vdash [A]B \Leftrightarrow f_A(w) \subseteq |B|$ & $t(B) \leq t(A)$

The two formulations are equivalent ($wR_Aw_1 \Leftrightarrow w_1 \in f_A(w)$). However, either formulation is at times handier than the other to make specific points.

In order for it being true that, given explicit input $A$, one imagines $B$ in a certain act of mental simulation, thus, we ask two things at once. First, we have a truth-conditional requirement: that $B$ be true throughout the most plausible worlds where $A$ is true. This corresponds to our aforementioned Feature 4: given input $A$, one develops the imagined scenario by looking at what is most plausibly going to happen in a situation realizing the input. Second, we have a topicality requirement: $B$ must be fully on topic with respect to what $A$ is about. This corresponds to Feature 5. We then characterize our $f_A$'s (thus, $R_A$'s) via a system of spheres capturing plausibility or belief entrenchment. This is done by a function, $, assigning to each $w$ a finite set of subsets of $W$ (the spheres), $\$(w) = \{S_0^w, S_1^w, \ldots, S_n^w\}$, with $n \in \mathbb{N}$, which shall be nested: $S_0^w \subseteq S_1^w \subseteq \cdots \subseteq S_n^w = W$. For each $A \in \mathcal{L}$ and $w \in W$, $f_A(w)$ works thus: if $|A| = \emptyset$, then $f_A(w) = \emptyset$. Otherwise, $f_A(w) = S_i^w \cap |A|$, where $S_i^w \in \$(w)$ is the smallest sphere such that $S_i^w \cap |A| \neq \emptyset$. Because there's a finite amount of spheres around each $w$, Lewis' 'Limit Assumption' is satisfied: the existence of a smallest such $S_i^w$ for each $w \in W$ and $A \in \mathcal{L}$ is automatically guaranteed.[7] On the other hand, Lewis' 'Weak Centering' fails: we do not demand that $w \in S_0^w$, that is, the reference world be in the innermost sphere. That's because the spheres express subjective plausibility for the agent located at $w$, not objective similarity to $w$.

The system satisfies various conditions, such as:[8]

- (C1) $f_A(w) \subseteq |A|$

---

[7] Though I will not argue it here, I think the Limit Assumption makes perfect sense when one wants to model plausibility and the conditional arrangement of beliefs.

[8] *Proof*: check Priest (2001, pp. 91–92).

- (C2) $f_A(w) \subseteq |B|$ & $f_B(w) \subseteq |A| \Rightarrow f_A(w) = f_B(w)$
- (C3) $|A| \neq \emptyset \Rightarrow f_A(w) \neq \emptyset$
- (C4) $f_A(w) \cap |B| \neq \emptyset \Rightarrow f_{A \wedge B}(w) \subseteq f_A(w)$

C1 and especially C2 will play a role in the following.

Logical consequence gets the usual characterization as truth preservation at all worlds of all models (based on our system of spheres given by \$—I will omit mentioning this in the following). If $\Sigma$ is a set of formulas:

$\Sigma \vDash B \Leftrightarrow$ in all models $\mathfrak{M} = \langle W, \{R_A \mid A \in \mathcal{L}\}, \mathcal{T}, \oplus, t, \Vdash \rangle$ and for all $w \in W$: $w \Vdash A$ for all $A \in \Sigma \Rightarrow w \Vdash B$

For single-premise entailment, we write $A \vDash B$ for $\{A\} \vDash B$. Logical validity, $\vDash A$, truth at all worlds of all models, is then $\emptyset \vDash A$, entailment by the empty set.

The logic induced by the semantics for the extensional operators is just classical propositional, and $\prec$ is a strict S5-like conditional. All the novelty is in the $[A]$'s, to which we now turn: the following Sections will illustrate how these model the features of imagination as ROMS listed above.

## 4 Success, parthood, non-monotonicity

Condition C1 gives us:

(Success) $\vDash [A]A$[9]

The content of the 'pretence premise' or initial imaginative input $A$ is imagined, as per Feature 3 (it's called 'Success' for it reminds us the Success Postulate from AGM-style belief revision.)

The next validities and invalidities involve conjunction. Two validities have imagination operators behave in a 'fully conjunctive' way. The first is:

(Simplification) $[A](B \wedge C) \vDash [A]B$  $[A](B \wedge C) \vDash [A]C$[10]

The second is:

(Adjunction) $\{[A]B, [A]C\} \vDash [A](B \wedge C)$[11]

Simplification and Adjunction take care of the mereological nature of imagination as per Feature 2 above: when one imagines—in Chalmers' 'positive conceivability' sense—the whole of a scenario, one imagines its parts (using Williamson's example: when one imagines that Mary is tall and thin, one imagines that Mary is tall). Vice versa, when one imagines the parts of a certain situation all together in a single act (notice that in Adjunction the explicit input $A$ has to stay constant from the premises

---

[9] *Proof*: by C1, for any $w$ and $w_1$, $wR_Aw_1 \Rightarrow w_1 \Vdash A$, and of course $t(A) \leq t(A)$.

[10] *Proof*: we do the first one (for the second, replace $B$ with $C$ appropriately). Let $w \Vdash [A](B \wedge C)$. By (S[A]), for all $w_1$ such that $wR_Aw_1$, $w_1 \Vdash B \wedge C$, thus by (S$\wedge$), $w_1 \Vdash B$. Also, $c(B \wedge C) = c(B) \oplus c(C) \leq c(A)$, thus $c(B) \leq c(A)$. Then, by (S[A]), $w \Vdash [A]B$.

[11] *Proof*: let $w \Vdash [A]B$ and $w \Vdash [A]C$, that is, by (S[A]): for all $w_1$ such that $wR_Aw_1$, $w_1 \Vdash B$ and $w_1 \Vdash C$, so by (S$\wedge$) $w_1 \Vdash B \wedge C$. Also, $c(B) \leq c(A)$ and $c(C) \leq c(A)$, thus $c(B) \oplus c(C) = c(B \wedge C) \leq c(A)$. Thus, $w \Vdash [A](B \wedge C)$.

to the conclusion), one imagines the whole (when one imagines in a single act that Mary is tall and that Mary is thin, one imagines that Mary is tall and thin).

A conjunction-involving invalidity displays the contextual nature of the importation of background beliefs in acts of imagination, as per Feature 4. Our operators are 'nonmonotonic':

$$[A]B \nvDash [A \wedge C]B^{12}$$

The inference is not off-topic, for if $t(B) \leq t(A)$, then also $t(B) \leq t(A) \oplus t(C) = t(A \wedge C)$. What makes it invalid is the variability in strictness: $f_A(w)$ need not be the same as $f_{A \wedge C}(w)$. Enlarging the explicit input can change the bunch of worlds we look at as the most plausible: as you imagine that Mary walks across Poznan, you imagine her in Poland by importing your (relevant) background belief that Poznan is in Poland. But if you imagine Mary walking across Poznan and that the city has been annexed by Germany, you will not imagine Mary in Poland.

## 5 Topicality, indeterminacy

When one imagines in an act whose explicit input is $A$, that $B$, one does not thereby imagine a disjunction between the latter and an unrelated $C$, although any formula logically entails the disjunction between itself and something else. This would violate our topicality requirements, as per Feature 5. The agent need not be aware of that disconnected $C$ at all, or that $C$ might be irrelevant to the mental simulation exercise. Thus we need, and we get:

$$[A]B \nvDash [A](B \vee C)^{13}$$

Given the input that Mary is giving a party in Paris, you imagine that she is happy. You do not thereby come to imagine that either she is happy or Poznan has been annexed by Germany, for the added disjunct is off-topic.

Disjunction comes in also for the modeling of Feature 6: imagination can be 'non-prime' because it generally under-determines its contents. You imagine that Mary landed in New Zealand. You integrate the explicit input on the basis of your beliefs by representing her either in the North Island or in the South Island, but you are not sure which one it is. Thus, you don't imagine her to be in either rather than the other. So we need, and we get:

$$[A](B \vee C) \nvDash [A]B \vee [A]C^{14}$$

The different worlds one looks at in the imagination act will fill in the unspecified details in different ways. Among the most plausible $A$-worlds there will be a world where $B$ but not $C$, and a world where $C$ but not $B$.

---

[12] *Countermodel*: let $W = \{w, w_1\}$, $w R_p$-accesses nothing, $w R_{p \wedge r} w_1$, $w_1 \nVdash q$, $c(p) = c(q) = c(r)$. Then $w \Vdash [p]q$, but $w \nVdash [p \wedge r]q$.

[13] *Countermodel*: let $W = \{w, w_1\}$, $w R_p w_1$, $w_1 \Vdash q$, $t(p) = t(q) \neq t(r)$. Then $t(q) \leq t(p)$, so by (S[A]), $w \Vdash [p]q$. But $t(q \vee r) = t(q) \oplus t(r) \nleq t(p)$, thus $w \nVdash [p](q \vee r)$.

[14] *Countermodel*: let $W = \{w, w_1, w_2\}$, $w R_p w_1$, $w R_p w_2$, $w_1 \Vdash q$ but $w_1 \nVdash r$, $w_2 \Vdash r$ but $w_2 \nVdash q$, $t(p) = t(q) = t(r)$. Then by (S$\vee$), $w_1 \Vdash q \vee r$ and $w_2 \Vdash q \vee r$, so for all $w_x$ such that $w R_p w_x$, $w_x \Vdash q \vee r$. Also, $t(q \vee r) = t(q) \oplus t(r) \leq t(p)$, thus by (S[A]), $w \Vdash [p](q \vee r)$. However, $w \nVdash [p]q$ and $w \nVdash [p]r$ for both $p$ and $r$ fail at some $R_p$-accessible world. Thus by (S$\vee$), $w \nVdash [p]q \vee [p]r$.

## 6 Hyperintensionality

Some invalidities highlight the hyperintensional features of imagination as an intentional state. One's imagining that $B$ given $A$ is not entailed by the corresponding strict conditional:

$$A \rightarrow B \nvDash [A]B^{15}$$

The strict conditional is 'irrelevant', in the sense criticized in the relevant logic programme (Anderson and Belnap 1975; Anderson et al. 1992): even when all the $A$-worlds are $B$-worlds (thus, all the most plausible $A$-worlds are $B$-worlds), $B$ may have little to do with $A$. This failure is in line with Feature 5 again.

Similarly, we do not imagine off-topic logical validities when we imagine something:

$$\nvDash [A](B \rightarrow B)$$
$$\nvDash [A](B \lor \neg B)^{16}$$

However, we do imagine the topic-preserving logical validities that comply with the explicit input. As is easy to show:

$$\vDash [A](A \rightarrow A)$$
$$\vDash [A](A \lor \neg A)$$

Some find these validities good (they are on-topic, after all!), some find them bad. It might be that the only way to get rid of them, then, is by adding so-called non-normal or impossible worlds where truths of logic may fail (Kiourti 2010; Nolan 2013; Jago 2014).

I close with two validities (one good, one possibly bad) warranted by the important condition C2 above. That condition was added by brute force in Berto (2017a), whose semantics had no plausibility metric on worlds; but it is automatically validated by our system of spheres. The possibly bad validity points at an open problem of the current approach.

## 7 Equivalents in imagination, limited transitivity

Condition C2 was called the Principle of Imaginative Equivalents (PIE) in Berto (2017a). It limits the effects of the hyperintensional nature of imagination and greatly strengthens the logic. To rehearse, for all $A, B \in \mathcal{L}$ and $w \in W$:

- (C2-PIE) $f_A(w) \subseteq |B|$ & $f_B(w) \subseteq |A| \Rightarrow f_A(w) = f_B(w)$

When all the most plausible $A$-worlds make $B$ true and vice versa, $A$ and $B$ are *equivalent in imagination*: as we imagine either, we look at the same worlds. This gives the following validity:

---

[15] *Countermodel*: let $W = \{w, w_1\}$, $wR_p w_1$, $w \nVdash p$, $w_1 \Vdash q$, $t(p) \neq t(q)$. By C1, $w_1 \Vdash p$. Now $|p| \subseteq |q|$, thus $w \Vdash p \rightarrow q$. But although $f_p(w) \subseteq |q|$, $t(q) \nleq t(p)$, thus $w \nVdash [p]q$.

[16] *Countermodel*: we do the former: let $W = \{w\}$, $t(p) \neq t(q)$. Then although (trivially) $f_p(w) \subseteq |q \rightarrow q|$, $t(q \rightarrow q) = t(q) \nleq t(p)$. Thus by (S[A]), $w \nVdash [p](q \rightarrow q)$.

(Substitutivity) $\{[A]B, [B]A, [A]C\} \vDash [B]C$[17]

Substitutivity has it that 'equivalents in imagination' $A$ and $B$ can be replaced *salva veritate* as modal indexes in $[\,\cdot\,]$. This sounds plausible. $[A]B$ and $[B]A$: as you imagine that Timmy is a whoodchuck, you imagine that Timmy is a groundhog and vice versa. Suppose $[A]C$: as you imagine that Timmy is a woodchuck, you imagine him with a dense coat of fur. Then the same happens when you imagine that he is a groundhog, $[B]C$.

But while general transitivity fails for our imagination operators (that's just a consequence of their variable strictness), C2-PIE also validates a kind of Restricted Transitivity:

(RT) $\{[A]B, [A \wedge B]C\} \vDash [A]C$[18]

Now some instances of RT sound plausible. $[A]B$: as you imagine that Mary is giving a party in Paris, you imagine that she's happy. $[A \wedge B]C$: as you imagine Mary happily giving a party in Paris, you imagine she is surrounded by friends. Thus, $[A]C$: as you imagine that Mary is giving a party in Paris, you imagine that she's surrounded by friends.

However, there may be counterexamples around. The issue has to do with cases where $C$ easily pops to mind given $B$ alone, but is only dimly related to $A$—for then RT (acting a bit like a Cut rule in a logical calculus) washes the bridging $B$ away in the conclusion. Here's a situation suggested by Claudio Calosi, that may do. $[A]B$: given the input that I am wearing a red shirt in Pamplona, I imagine that I am being chased by bulls. $[A \wedge B]C$: given the input that I am being chased by bulls on the streets of Pamplona while wearing a red shirt, I imagine that I die on the street. But it's not the case that $[A]C$: Given that I am wearing a red shirt in Pamplona, I don't imagine that I die on its streets.

If RT has to go, then it might be that there is some problem with the plausibility metric on worlds validating C2-PIE, thus with the idea that imagining that $A$ involves looking at the most plausible worlds where $A$ is the case. Or, it might be that there is some problem with the conjunctive nature of imagination as per Feature 2—for notice that Simplification and Adjunction are crucially involved in the proof of RT, though not in that of Substitutivity. In Berto (2017a), I suggested that if one resorts to an extended semantics that uses impossible worlds (of a non-adjunctive kind) besides possible ones, one can have C2-PIE without having RT because Simplification and/or Adjunction can fail in such a framework. On the other hand, the proof of this [carried out in Berto (2017b)] is in an impossible worlds setting that does not have a metric on worlds, and the issue of world orderings involving impossible worlds is a open one.

---

[17] *Proof*: suppose $w \Vdash [A]B, w \Vdash [B]A, w \Vdash [A]C$. By (S[A]), these entail, respectively, (1) $f_A(w) \subseteq |B|$ and $t(B) \leq t(A)$, (2) $f_B(w) \subseteq |A|$ and $t(A) \leq t(B)$, (3) $f_A(w) \subseteq |C|$ and $t(C) \leq t(A)$. From (1) and (2) we get $f_A(w) = f_B(w)$ (by C2-PIE) and $t(A) = t(B)$ (by antisymmetry of topic parthood). From these and (3) we get $f_B(w) \subseteq |C|$ and $t(C) \leq t(B)$. Thus by (S[A]), $w \Vdash [B]C$.

[18] *Proof*: suppose (1) $w \Vdash [A]B$ and (2) $w \Vdash [A \wedge B]C$. From (1), $\vDash [A]A$ and Adjunction we get $w \Vdash [A](A \wedge B)$, thus, by (S[A]), $f_A(w) \subseteq |A \wedge B|$ and $t(A \wedge B) \leq t(A)$. Also, $w \Vdash [A \wedge B]A$ (from $\vDash [A \wedge B](A \wedge B)$ and Simplification) thus, by (S[A]) again, $f_{A \wedge B}(w) \subseteq |A|$ and (of course) $t(A) \leq t(A \wedge B)$. Thus, $f_A(w) = f_{A \wedge B}(w)$ (by C2-PIE) and $t(A \wedge B) = t(A)$ (by antisymmetry of topic parthood). Next, by (2) and (S[A]) again, $f_{A \wedge B}(w) \subseteq |C|$ and $t(C) \leq t(A \wedge B)$. Therefore, $f_{A \wedge B}(w) = f_A(w) \subseteq |C|$ and $t(C) \leq t(A) = t(A \wedge B)$. Thus by (S[A]) again, $w \Vdash [A]C$.

Here's one final thought on the issue of RT.[19] One may make it fail by reinterpreting those operators as only requiring some increase in plausibility for the relevant *A*-worlds, probabilistically understood in a quantitative or degree-theoretic way. This can be motivated by reading the operators as something along the lines of "That *A* would make it more plausible or more likely that *B*", with plausibility behaving in a tholdy way. Then RT will fail: even when *A* makes it likely enough that *B*, and $A \wedge B$ makes it likely enough that *C*, *A* may fail to make *C* likely enough.

# References

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic*, *50*, 510–30.

Anderson, A. R., & Belnap, N. (1975). *Entailment: The logic of relevance and necessity I*. Princeton: Princeton University Press.

Anderson, A. R., Belnap, N., & Dunn, J. M. (1992). *Entailment: The logic of relevance and necessity II*. Princeton: Princeton University Press.

Angell, R. B. (1977). Three systems of first degree entailment. *Journal of Symbolic Logic*, *47*, 147.

Balcerak Jackson, M. (2016). On imagining, supposing and conceiving. In A. Kind & P. Kung (Eds.), *Knowledge through imagination* (pp. 42–60). Oxford: Oxford University Press.

Berto, F. (2017a). Aboutness in imagination. *Philosophical Studies*,. https://doi.org/10.1007/s11098-017-0937-y.

Berto, F. (2017b). Impossible worlds and the logic of imagination. *Erkenntnis*, *82*, 1277–97.

Berto, F. (2018). Simple hyperintensional belief revision. *Erkenntnis*,. https://doi.org/10.1007/s10670-018-9971-1.

Berto, F., & Schoonen, T. (2017). Conceivability and possibility: Some dilemmas for humeans. *Synthese*,. https://doi.org/10.1007/s11229-017-1346-7.

Byrne, R. (2005). *The rational imagination. How people create alternatives to reality*. Cambridge, MA: MIT Press.

Byrne, A. (2007). Possibility and imagination. *Philosophical Perspectives*, *21*, 125–44.

Chalmers, D. (2002). Does conceivability entail possibility? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility* (pp. 145–99). Oxford: Oxford University Press.

Fagin, R., & Halpern, J. (1988). Belief, awareness and limited reasoning. *Artificial Intelligence*, *34*, 39–76.

Ferguson, T. M. (2014). A computational interpretation of conceptivism. *Journal of Applied Non-Classical Logic*, *24*, 333–67.

Fine, K. (1986). Analytic implication. *Notre Dame Journal of Formal Logic*, *27*, 169–79.

Fine, K. (2014). Truthmaker semantics for intuitionistic logic. *Journal of Philosophical Logic*, *43*, 549–77.

Fine, K. (2015). Angellic content. *Journal of Philosophical Logic*,. https://doi.org/10.1007/s10992-015-9371-9.

Fiocco, M. O. (2007). Conceivability, imagination and modal knowledge. *Philosophy and Phenomenological Research*, *74*, 364–80.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic*, *17*, 157–170.

Hintikka, J. (1962). *Knowledge and belief. An introduction to the logic of the two notions*. Ithaca, NY: Cornell University Press.

Jago, M. (2014). *The impossible. An essay on hyperintensionality*. Oxford: Oxford University Press.

Kind, A., & Kung, P. (Eds.). (2016). *Knowledge through imagination*. Oxford: Oxford University Press.

[19] Kindly suggested by an anonymous referee, who should also be thanked for suggesting a nice title for this paper.

Kind, A. (2001). Putting the image back in the imagination. *Philosophy and Phenomenological Research*, *62*, 85–109.

Kind, A. (2016). Imagining under constraints. In A. Kind & P. Kung (Eds.), *Knowledge through imagination* (pp. 145–59). Oxford: Oxford University Press.

Kiourti, I. (2010). *Real impossible worlds: The bounds of possibility*. PhD thesis, University of St Andrews.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S., & Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, *9*, 52–76.

Kung, P. (2014). You really do imagine it: Against error theories of imagination. *Nous*, *50*, 90–120.

Langland-Hassan, P. (2016). On choosing what to imagine. In A. Kind & P. Kung (Eds.), *Knowledge through imagination* (pp. 61–84). Oxford: Oxford University Press.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

Markman, K. D., Klein, W., & Surh, J. A. (Eds.). (2009). *Handbook of imagination and mental simulation*. New York: Taylor and Francis.

Nichols, S., & Stich, S. P. (2003). *Mindreading. An integrated account of pretence: Self-awareness, and understanding other minds*. Oxford: Oxford University Press.

Nolan, D. (2013). Impossible worlds. *Philosophy Compass*, *8*, 360–372.

Paivio, A. (1986). *Mental representation*. Oxford: Oxford University Press.

Parry, W. T. (1933). Ein axiomensystem fr eine neue art von implikation (analitische implikation). *Ergebnisse eines Mathematischen Kolloquiums*, *4*, 5–6.

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

Priest, G. (2016). Thinking the impossible. *Philosophical Studies*,. https://doi.org/10.1007/s11098-016-0668-5.

Pylyshyn, Z. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*, 1–25.

Pylyshyn, Z. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*, 157–82.

Schipper, B. (2015). Awareness. In H. van Ditmarsch, J. Halpern, W. van der Hoek, & B. Kooi (Eds.), *Handbook of epistemic logic* (pp. 79–146). London: College Publications.

Shephard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford: Blackwell.

Urquhart, A. (1973). A semantical theory of analytical implication. *Journal of Philosophical Logic*, *2*, 212–19.

van Benthem, J., & Smets, S. (2015). Dynamic logics of belief change. In H. van Ditmarsch, J. Halpern, W. van der Hoek, & B. Kooi (Eds.), *Handbook of epistemic logic* (pp. 313–93). London: College Publications.

van Ditmarsch, H. P., van der Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic*. Dordrecht: Springer.

Van Leeuwen, N. (2016). The imaginative agent. In A. Kind & P. Kung (Eds.), *Knowledge through imagination* (pp. 85–111). Oxford: Oxford University Press.

Wansing, H. (2015). Remarks on the logic of imagination. A step towards understanding doxastic control through imagination. *Synthese*. https://doi.org/10.1007/s11229-015-0945-4.

Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.

Williamson, T. (2016). Knowing by imagining. In A. Kind & P. Kung (Eds.), *Knowledge through imagination* (pp. 113–23). Oxford: Oxford University Press.

Yablo, S. (1993). Is conceivability a guide to possibility? *Philosophy and Phenomenological Research*, *53*, 1–42.

Yablo, S. (2014). *Aboutness*. Princeton: Princeton University Press.