VOLUNTARY IMAGINATION: A FINE-GRAINED ANALYSIS

ILARIA CANAVOTTO

Institute for Logic, Language and Computation, University of Amsterdam

FRANCESCO BERTO

Archè, University of St. Andrews
and

ALESSANDRO GIORDANI

Department of Philosophy, Catholic University of Milan

Abstract.   We study imagination as reality-oriented mental simulation (ROMS): the activity of simulating nonactual scenarios in one's mind, to investigate what would happen if they were realized. Three connected questions concerning ROMS are: What is the logic, if there is one, of such an activity? How can we gain new knowledge via it? What is voluntary in it and what is not? We address them by building a list of core features of imagination as ROMS, drawing on research in cognitive psychology and the philosophy of mind. We then provide a logic of imagination as ROMS which models such features, combining techniques from epistemic logic, action logic, and subject matter semantics. Our logic comprises a modal propositional language with non-monotonic imagination operators, a formal semantics, and an axiomatization.

§1. **Introduction.**   Imagination is "the faculty or action of forming new ideas, or images or concepts of external objects not present to the senses" (OED). This allows for the notion to cover a range of cognitive activities, from daydreaming, to freely following a train of thought, to hallucinating. We focus on one, labeled, for reasons that will become clear soon, *reality-oriented mental simulation* (ROMS): the episodic activity of simulating alternatives to reality in our mind, to investigate what would or is likely to happen, or what would have happened, if they were, or had been, realized. In cognitive psychology, ROMS is widely agreed to play important roles (Markman, Klein, & Surh, 2009): it allows us to improve performance, to make contingency plans by anticipating future outcomes, and to learn from mistakes by considering how we could have done better by acting differently. It is used to account for counterfactual thought (Byrne, 2005; Byrne & Girotto, 2009) and pretence (Nichols & Stich, 2003). In the philosophy of fiction, it can explain our engagement with fictional works (Walton, 1990; Currie, 1990). In epistemology, it is used to account for how we can come to know modal claims (Yablo, 1993; Chalmers, 2002; Williamson, 2007). In semantics

and the philosophy of language, it is invoked to explain how we evaluate conditionals (Stalnaker, 1968; Evans & Over, 2004).

We will focus on propositional imagination: one imagines that one jumps a stream, that Obama is tall and thin, that there are functional zombies devoid of phenomenal consciousness, that Scotland votes for independence, that *Wuthering Heights'* Heathcliff and Catherine meet for the last time, that the Nazis have nukes in 1944. We will provide a logical model of propositional imagination as ROMS and consider three connected questions concerning it:

(1) *What is the logic of such an activity*? The popular "Logic will get you from *A* to *B*, imagination will take you everywhere" sounds as if it says something good about imagination, evoking creativity and lateral thought. But it may actually say something bad about it: given some *A*, logic will take you only to the *proper B*s, namely, those that follow logically from *A*. But if imagination is governed by free association of ideas, then it may work as a runabout inference ticket, like Prior's *tonk*: given some *A*, it will take you to any *B* one likes. On the other hand, as persuasively argued by Byrne (2005), exercises of imagination as ROMS must have *some* logic: some things follow from the envisaging of a hypothesis, some not. Since Hintikka (1962), we have learned how to give logics for intentional states of the mind by taking these as modals: restricted quantifiers over possible worlds. One *X*s (knows, believes, is informed) that *A* just in case *A* holds throughout a selected set of worlds (those compatible with one's overall evidence, etc.). What about one's imagining that *A*? It seems that this *does* obey some logical rules, e.g., *Conjunction Commutation* (try to imagine that Obama is tall and thin without imagining that he's thin and tall) and *Elimination* (try to imagine that Obama is tall and thin without imagining that he's tall). The converse, *Conjunction Introduction*, may already be more controversial: does imagining that *A* and imagining that *B* entail imagining that *A* and *B* together? (We'll get back to this.) It seems that imagination can't be taken as a normal modal, at least for non-logically omniscient cognitive agents like us: imagining that *A* should not entail imagining all of *A*'s logical consequences, and we certainly don't imagine arbitrary logical truths whenever we engage in an act of mental simulation.

(2) *What is the relation between imagination and knowledge*? Connected to question (1) is the issue of how to reconcile imagination's apparent arbitrariness with its having some epistemic value: how can we get *knowledge through imagination*? In a recently edited collection bearing this title, Kind & Kung (2016) label this the "puzzle of imaginative use": if imagination is arbitrary escape from reality, how can it give us new knowledge—understanding of works of art, capacity to anticipate the future, etc.? Imagination seems to be voluntary in ways contrasting states, e.g., belief, are not: one can easily imagine that all of London is painted blue, while one can hardly make oneself believe it, given overwhelming contrary evidence. But if, given some input, one can imagine anything one wills (pending issues of imaginative resistance: see Gendler, 2000), then imagination cannot be conductive to knowledge.

(3) *What is voluntary in an exercise of imagination as mental simulation, and what is not*? A promising line of response to question (2) (Williamson, 2016; Langland-Hassan, 2016) relies on the idea that not *everything* in an exercise of imagination is voluntary, and that the involuntary component suffices, when things go well, to ensure new, reliably formed, and often true beliefs. But how to tell the two components from each other? The general distinction between voluntary and involuntary mental

processes is a conundrum in itself, but it seems to make intuitive sense and plays a key role at least in some mainstream general views in cognitive psychology. Thus, dual process theories of thought (Kahneman & Tversky, 1984; Stanovich & West, 2000; Evans & Over, 2004) distinguish between "System 1" and "System 2" processes on the basis of the former being largely automatic, the latter having to be activated and carried out by voluntarily overriding the former's workings, and by paying a cognitive cost.

To start addressing these problems, in the following section we build on literature in cognitive psychology and philosophy to lay out a general characterization of imagination as ROMS and single out its key features. In the sections after it, we introduce a logical language with imagination operators which, as we show, collectively model all of those features.

§2. Imagination as ROMS.   This list of features slightly modifies and complements the one initially considered in Berto (2018a). The items in the list have been proposed by researchers on imagination, mental simulation, and pretence, both in philosophy and in cognitive psychology. We will therefore refer to a number of works in both disciplines:

Feature 1. Imagination is agentive and episodic.   There is widespread agreement that acts of imagination as ROMS are started voluntarily by agents having a number of options given their situation, background knowledge and beliefs, cognitive abilities, etc. (Nichols & Stich, 2003). We decide to engage in one such act, carry it out for a while, often by controlling some aspects of what we imagine, and stop after some time. This has suggested to some authors, e.g. Wansing (2017), to model *agents voluntarily imagining something* (where the stress is on the imagined thing) by using techniques from STIT logics of agency (Belnap, Perloff, & Xu, 2001; Horty, 2001): an agent voluntarily imagines, say, that there are functional zombies if and only if she sees to it that she imagines that there are such things. Our concern will be to model *agents voluntarily engaging in imagination acts and deciding how to carry them out* (where the stress is on the acts themselves). We will borrow an idea from the logic of agency of Giordani (2018), which combines a dynamic logic based on action types with intuitions from STIT logics, and introduce an agentive modality corresponding to an equivalence relation between possible worlds: given a world w where we engage in an act of imagination, the worlds equivalent to w are like w except possibly for the fact that we carry out the imagination exercise in a different way (e.g., we decide to stop before its actual end). Equivalent worlds keep track of the options we have in conducting a ROMS.

Feature 2. Acts of imagination have deliberate starting points, given by an input. Such input is up to us. In their model of mental simulation, Nichols & Stich (2003) have "an initial premise or set of premises, which are the basic assumptions about what is to be pretended" (p. 24). This may be made up by the agent when engaging in predictions, e.g., when you guess what would happen if something were the case; or it may be taken on board via an external instruction, e.g., when you read a novel and take the explicit text as your input, or when you evaluate a conditional and start by taking the antecedent as input. Suppositional theories of conditionals in psychology as well as philosophy connect this to the so-called Ramsey test: see Evans & Over (2004). In our formalism below, we will represent the explicit input as directly expressed by formulas indexing imagination operators.

Feature 3. We integrate the explicit input with background information we import, contextually, depending on what we know or believe. The importance of background knowledge and beliefs in suppositional (e.g., conditional) thinking is increasingly acknowledged in the psychology of reasoning (Oaksford & Chater, 2010). Once the input of an act of mental simulation is in, Nichols & 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). When we imagine Watson talking with Holmes while walking through the streets of London, we represent Watson dressed as a nineteenth century gentleman, not as an astronaut. The text of the relevant novel need not say anything explicitly on how Watson is dressed, nor do we infer this from the explicit content via sheer logic. Rather, we import such information into the represented situation, based on what we know: we know that the story takes place in Victorian London and we assume, lacking information to the contrary from the text, that Watson is dressed as we know gentlemen were dressed at the time. We perform some *minimal alteration* to how we know or believe the world to be, or to have been, compatible with the initial explicit input, in a process somewhat similar to belief revision (Alchourrón, Gärdenfors & Makinson, 1985), whereby we perform a minimal change of our belief system needed to accommodate new information. Two issues here are: is this integration process voluntary or not? And how is it to be modeled logically?

As for the first, some authors seem to agree that the involuntary component of imagination as ROMS comes into play exactly here. As Williamson (2016), p. 116, has it:

> Think of a hunter who finds his way obstructed by a mountain stream rushing between the rocks. He reaches the only place in the vicinity where jumping the stream might be feasible. [...] How should he try to determine whether he would succeed? [...] One *imagines* oneself trying. If one then imagines oneself succeeding, one judges that if one tried, one would succeed. If instead one imagines oneself failing, one judges that if one tried, one would fail. [...] When the hunter *makes himself* imagine trying to jump the stream, his imagination operates in *voluntary mode*. But he neither makes himself imagine succeeding nor makes himself imagine failing. Rather, *having forced the initial conditions*, he lets the rest of the imaginative exercise *unfold without further interference*. For that remainder, his imagination operates in *involuntary mode*. He imagines the antecedent of the conditional voluntarily, the consequent involuntarily. Left to itself, the imagination develops the scenario in a reality-oriented way, by default. [our emphasis]

Similarly, Langland-Hassan (2016) distinguishes between "guiding chosen" imaginings, "top down intentions [that] are key to initiating an imagining," and "lateral constraints [that] govern how it then unfolds" (p. 9), and which seem to operate in involuntary mode. If the additional details are borrowed from our knowledge or belief base, as Van Leeuwen (2016) and Nichols & Stich (2003) have it (the latter have a cognitive "belief box," from which contents are taken and imported into the mental simulation), this makes sense: for if beliefs are often formed and managed in largely

involuntary mode, it seems plausible for their importation to be essentially involuntary. Some research in cognitive psychology seems to support the view that imagination allows automatic, involuntary access to the knowledge deposited in implicit (long-term) memory, and that the results of imaginative exercises can themselves alter such memory (Kosslyn & Moulton, 2009).

On the logical side, it is natural to represent this integration of the initial input via background beliefs and knowledge by using modal operators that work as *variably strict* quantifiers over worlds: the input will play a role similar to a variably strict conditional antecedent, as per the mainstream possible worlds semantics for *ceteris paribus* conditionals due to Stalnaker (1968) and Lewis (1973) (Lewis marketed it only for counterfactuals or subjunctives, whereas Stalnaker proposed it for indicatives as well). This squares, too, with the aforementioned psychological insight that we evaluate conditionals by mentally representing the antecedent, developing this in our imagination, and seeing whether the consequent would, in some sense, follow (Oaksford & Chater, 2010).

Feature 4. Imagination has topicality and relevance constraints. We do not indiscriminately import unrelated contents into the imagined scenarios. As Amy Kind has it, "[We require] that the world be imagined as it is *in all relevant respects*" (Kind, 2016, p. 153). This is key to distinguishing imagination as ROMS from free-floating mental wandering. You know that Amsterdam hosts plenty of bikes, but this is immaterial to your imagining Watson and Holmes' adventures from Doyle's novels, insofar as Amsterdam and its bikes are not involved in them. So you will not import knowledge of this kind, even when it's perfectly consistent with the explicit input. This *topic-sensitivity* of ROMS will be modeled by imposing topic-preservation constraints on the outputs of imaginative exercises, which secure their complete relevance with respect to the starting point given by the explicit input. Our logical models will feature (formal representations of) *topics* to do the job—we will see what these are in due time.

Feature 5. The content of ROMS is goal-driven and question-based. This item was left unexplored in Berto (2018a), but it is of the greatest importance. Acts of ROMS have a *purpose* (Fraude-Koivisto, Wuerz & Gollwitzer, 2009), which can be understood via the question or issue the agent performing them aims to answer. When it's about imagining alternative endings for my jumping a stream, e.g., the issue may be, *Will I make it to the other side, or will I fall into the stream?* Now the same agent, with the same background knowledge and belief, and given the same input, can focus on different things depending on the goal of the exercise, and thereby end up imagining quite different scenarios. When the input is *I jump the stream*, and the question is, *Will I make it to the other side?*, one will focus, e.g, on one's weight and on the conditions of the ground, leaving behind issues less relevant for the question, such as, e.g., the kind of noises one makes while jumping. When the question is, instead, *Will I scare the cattle on the other side?*, the kind of noises one makes while jumping may become important.

On the psychological side, it seems that the choice of the goal to be pursued in an act of ROMS, hence of focusing on certain features of the imagined scenario rather than others, is, again, voluntarily chosen: the question *Will I make it to the other side?* is what we set out to answer in the act, and the input *I jump the stream* is how we make it start. At the same time, how the relevant features of the imagined scenario

are specifically represented seems to be partly involuntary and partly voluntary: while the representation of some of them is fixed by our background knowledge and beliefs, as when we know, e.g., that our weight is between 70 and 75 kg, the representation of others will be up to us. So, if we are planning a hike and we don't know yet whether it will be rainy or sunny, we may equally well opt for the worst case scenario, and represent the ground as wet, or for the best case scenario, and represent it as dry.

On the logical side: the issues one is pursuing in a ROMS exercise determine the salience of certain traits in the imagined scenario, and the fading in the background of others. So the phenomenon can be represented by having the issues determine a *partition* of the relevant possible worlds (those consistent with the input and the agent's background knowledge and beliefs), with cells in the partition corresponding to all possible specifications of the salient traits, such as, e.g., one's weighing 70 kg and the ground's being wet, one's weighing 70 kg and the ground's being dry, one's weighing 71 kg and the ground's being wet, and so on. This job can also be carried out by means of topics, which have been naturally linked to partitions of the set of possible worlds (Lewis, 1988; Yablo, 2014). We propose to model the process of specification of the salient traits by using modal operators that work as functions selecting one or more cells in the partition of the relevant possible worlds determined by the purpose or topic of the imaginative exercise. In a world where we start from the input *I jump the stream*, select the question *Will I make it to the other side?*, know that our weight is between 70 and 75 kg, and opt for being optimistic about the ground conditions, the function will select the union of all cells in the partition that include worlds in which our weight is between 70 and 75 kg and the ground is dry.

Let's wrap up with a general description of a ROMS episode: when the voluntary input comes in, two things happen: (1) in involuntary mode, we integrate the input with relevant background beliefs and knowledge and (2) in voluntary mode, we set the goal of the exercise, thus determining which traits in the imagined scenario are salient. We will call the (overall, partly voluntary) specification of these traits *specified initial scenario*, to distinguish it form the *basic initial scenario* that results from integration (1), independently from goal-setting (2) (e.g., when we receive the input *I jump a mountain stream*, we automatically picture the stream surrounded by nature, not by skyscrapers, regardless of whether the question we aim to answer is *Will I make it to the other side?* or *Will I scare the cattle?*). Next, the specified initial scenario unfolds in involuntary mode, generating new beliefs we may use in subsequent acts of mental simulation, like the belief that, if the ground is wet, we will not make it to the other side.

Figure 1 is a functional representation of the general structure of imagination as ROMS. Darker gray, lighter gray, and white boxes represent involuntary, partly voluntary, and voluntary processes respectively. The goal-driven process of specification of the basic initial scenario is decomposed into the processes of selection and specification of relevant traits. The arrow from the input to the development of the specified initial scenario is dashed to highlight that the overall process from the former to the latter is a mediated, rather than a direct, one. With respect to this general schema, the logic we present in the next sections aims at modeling, first, the processes of integration of the input and specification of the basic initial scenario and, second, the control the agent can exercise on them. It gives only a basic representation of involuntary unfolding in terms of what is necessary, or logically follows, given the imagined initial conditions and the topic of the input. It also makes sense to say that the specified initial scenario is
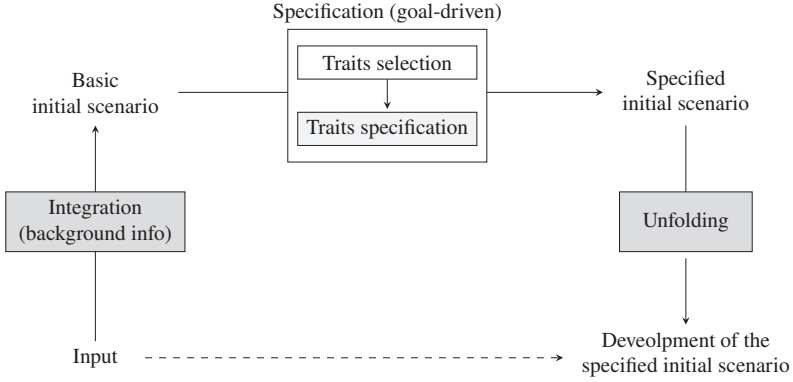
Fig. 1. General structure of ROMS.

unfolded in time, as episodes of ROMS can last for an amount of time and can involve representing actions and events, which themselves evolve in time. We conjecture that this may be best modeled by dynamic operators whose semantics is given in terms of model-transformations (van Benthem, 2007; Baltag & Smets, 2008; Baltag et al., 1998), leaving this to future work.

We now introduce a formal language featuring two kinds of imagination modalities and one agentive modality.

§3. Language. We work with a propositional language $\mathcal{L}$ based on a countable set Var_T of variables for *topics* and a countable set Var_P of variables for *propositions*. The set $Fm(\mathcal{L})$ of well-formed formulas of $\mathcal{L}$ is built according to the following rules in BNF:

$$\varphi ::= p_i \mid \neg\varphi \mid \varphi \wedge \alpha \mid \Box\varphi \mid [ag]\varphi \mid [im_\alpha]\varphi \mid [im_\alpha^\tau]\varphi \mid [\approx_\alpha^\tau]\varphi \,,$$

where $p_i \in Var_P$, $\tau \in Var_T$, and $\alpha \in Fm(\mathcal{L})$. The other propositional connectives and the dual modalities $\Diamond\varphi$, $\langle ag \rangle \varphi$, $\langle im_\alpha \rangle \varphi$, $\langle im_\alpha^\tau \rangle \varphi$, $\langle \approx_\alpha^\tau \rangle \varphi$ are defined as usual. We will interpret $\Box\varphi$ standardly as unrestricted necessity: truth at all possible worlds. The intended reading of the other four modals is as follows:

1. $[ag]$ is an agentive modality. It represents the imaginative options of agent *ag*. We read $[ag]\varphi$ as saying that φ is true at all worlds that are like the current one, up to the content imagined by *ag* given an input.
2. $[im_\alpha]$ picks out a set of worlds, selected given input α. We read $[im_\alpha]\varphi$ as saying that given α, the agent imagines that φ, or that α necessitates the truth of φ, independently of the selected topic. This corresponds to the original operator considered in Berto (2018a).
3. $[im_\alpha^\tau]$ also takes topics into account: it picks out a set of worlds, selected given input α and topic τ. We read $[im_\alpha^\tau]\varphi$ as saying that given α and τ, the agent imagines that φ, or that α and the chosen specification of τ together necessitate the truth of φ. As we will see, φ will turn out to be true at all worlds in the cells of the partition, determined by τ given input α, selected by the agent at the current world.

4. $[\approx^\tau_\alpha]$ picks out the worlds indistinguishable from the current one, given input α and topic τ. We read $[\approx^\tau_\alpha]\varphi$ as saying that φ is true in all worlds which are like the current one given α and τ. As we will see, φ will turn out to be true at all worlds in the cell of the partition, determined by τ given input α, that contains the current world.

When the semantics for these is given, we'll see that the three modalities $[ag]$, $[im_\alpha]$, and $[im^\tau_\alpha]$ represent, respectively: that imagination as ROMS is agentive, as per **Feature 1**; that its acts are based on an explicit input, as per **Feature 2**; and that, besides the explicit input, they are driven by a purpose, as per **Feature 5**. $[im_\alpha]$ and $[im^\tau_\alpha]$ will also represent how we integrate the explicit input α by selecting background information depending on our knowledge and beliefs, as per **Feature 3**, and how we only select information that is on-topic relative to α, as per **Feature 4**.

The validities involving $[im_\alpha]$ and $[im^\tau_\alpha]$ will speak to our first issue—the question of the logic of imagination as ROMS. As for the other two, we need a preliminary definition:

DEFINITION 1 (**var**(φ)). *For any $\varphi \in Fm(\mathcal{L})$, let $Var(\varphi)$ be the set of propositional variables occurring in φ (where this set is defined in the standard way). Then, we define:*
$\mathbf{var}(\varphi) := \bigwedge_{p_i \in Var(\varphi)} (p_i \vee \neg p_i)$.

Hence, **var**(φ) is the conjunction of all the instances of the principle of excluded middle given by propositional variables occurring in φ. It is evident that $Var(\mathbf{var}(\varphi)) = Var(\varphi)$ and that, for any propositional variable p, $\mathbf{var}(p) = p \vee \neg p$. As it will become clear below (see Definition 5 and Lemma 1), using this kind of formulas—rather than the more familiar constant $\top$—is essential, in our setting, to express topic inclusion. This, in turn, is essential to read $\langle im_\alpha \rangle \mathbf{var}(\alpha)$ as saying that the content imagined by the agent given input α is not empty and, similarly, $\langle im^\tau_\alpha \rangle \mathbf{var}(\alpha)$ as saying that the content imagined by the agent given input α and topic τ is not empty. $\langle im_\alpha \rangle \mathbf{var}(\alpha)$ can, then, be used to say that the agent is *entertaining* an act of imagination based on input α, while $\langle im^\tau_\alpha \rangle \mathbf{var}(\alpha)$ to say that the agent is *processing* the input in light of topic τ. When combined with the agentive modal $[ag]$, these formulas allow us to address our second question, on the relation between imagination as ROMS and knowledge, and our third question, on its voluntary and involuntary components.

We now introduce our semantics and provide an axiomatization.

§4. Semantics. The semantics of Berto (2018a) included, besides the usual set of possible worlds, W, a set of possible topics, T. Everyone is familiar with the former, while we need to say something by way of introduction to the latter, which have already been invoked above.

One can understand topics as somewhat similar to Lewisian or Yablovian *subject matters* from aboutness theory (Lewis, 1988; Yablo, 2014). Aboutness is "the relation that meaningful items bear to whatever it is that they are *on* or *of* or that they *address* or *concern*" (Yablo, 2014, p. 1)—their subject matters, or topics. In works such as Lewis' or Yablo's, these are understood in relation to questions: the subject matter or

topic of sentence S in context c can be linked to the question(s) S can be an answer to in c. When the topic at issue is *the number of stars*, the corresponding question can be, *How many stars are there?* This determines a partition on W; two worlds end up in the same cell when they give the same answer to the question, that is, they agree on the number of stars: all zero-star worlds end up in one cell, all one-star worlds end up in another, and so on.[1]

Being about stuff—having a topic—is not only a feature of linguistic representations, but also, and perhaps more fundamentally, of mental ones. In particular, imagining is about stuff. In our setting, topics are what the mental states of imaginative agents are about, and the topic of an imaginative exercise is given via its specific purpose, which can also be understood via a question. As we saw in §2, when the goal is predicting whether one will make it to the other side if one jumps the stream, the question will be, *Will I make it to the other side?* The possible answers will depend on the specification of certain salient traits, which can also be spelled out as questions, like *How much do I weigh?* or *What are the conditions of the ground?* The set of all these questions determines a partition on W: two worlds end up in the same cell when they give the same answer to all the questions determined by the topic, that is, when they agree on our weight, the conditions of the ground, and so on. So, imaginative exercises have a topic (*Will I make it to the other side?*), which determines a set of questions (*How much do I weigh?*, *What are the conditions of the ground?*, etc.), which, in turn, determines a partition of the set of possible worlds. We'll see that the chosen answers, corresponding to a union of cells in the corresponding partition, are captured by the modality $[im_\alpha^\tau]\varphi$.

But *how* are topics like? What is their nature and structure? We don't need to say too much on this for our logical purposes, except that we need a recursion on them allowing us to come up with a compositional semantics for our language. Luckily, there's a natural mereology of topics at the sentential level: what a sentence is about can be (properly) included in what another one is about (Yablo, 2014, sec. 2.3), (Fine, 2015, sec. 3–5). Topics may be merged into wholes which inherit the proper features from the parts (Yablo, 2014, sec. 3.2). The Boolean connectives are topic-transparent— they have no subject matter of their own: φ has the same topic as $\neg\varphi$ ("Snow is not white" is about the color of snow, or how snow is like, or snow's whiteness, etc., just as "Snow is white"); $\varphi \wedge \psi$ and $\varphi \vee \psi$ have the same topic: a fusion of the subject matter of φ and that of ψ ("Obama is tall and thin" is about Obama's heights and figure, just as "Obama is tall or thin"). We extend transparency to our modals in a straightforward[2] way, and come up with the following.

---

[1] Partitions map to equivalence relations. Yablo (2014) proposes to generalize to divisions, which allow worlds to be in more than one cell and map to relations of similarity, for a question, e.g., *Where's a nice place to eat in London?*, can have more than one good answer. We won't complicate things accordingly in our framework.

[2] But not uncontroversial. "Obama is tall and thin" and "John imagines that Obama is tall and thin" seem to have different topics: only the latter is about John's mental states. Fortunately, for our logical purposes the difference only matters when one considers nested or higher-order imaginings (one imagines that one imagines that), which are beyond the scope of this work.

4.1. Topic models.

DEFINITION 2 (Topic model). *A topic model is a tuple $\langle T, \sqcup, t \rangle$, where*
– *$T \neq \varnothing$ is a set of topics;*
– *$\sqcup : T \times T \to T$ is a fusion operation that is*
  – *idempotent: $\tau_1 \sqcup \tau_1 = \tau_1$;*
  – *commutative: $\tau_1 \sqcup \tau_2 = \tau_2 \sqcup \tau_1$;*
  – *associative: $\tau_1 \sqcup (\tau_2 \sqcup \tau_3) = (\tau_1 \sqcup \tau_2) \sqcup \tau_3$;*
– *$t : Var_P \to T$ assigns topics to propositional variables.*

Topic inclusion is defined as $\tau_1 \sqsubseteq \tau_2 := \tau_1 \sqcup \tau_2 = \tau_2$.

COROLLARY 1. *$\sqsubseteq$ is a partial order of T, and $\tau_1 \sqcup \tau_2$ is the least topic including both τ_1 and τ_2.*

T is our set of possible topics; $\sqcup$ is the operation of fusion on T, merging topics into larger topics; t is a function assigning to each propositional variable p_i the topic associated with it.

DEFINITION 3 (Topic of a formula). *For any $\varphi \in Fm(\mathcal{L})$,*
$\mathbf{t}(\varphi) = \bigsqcup_{p_i \in Var(\varphi)} t(p_i)$.

A formula is about what its atomic variables, taken together, are about. Hence, the topic of a formula is completely determined by the topics of the variables occurring in it. The following corollary will be important later on.

COROLLARY 2. *$Var(\varphi) \subseteq Var(\psi)$ implies $\mathbf{t}(\varphi) \sqsubseteq \mathbf{t}(\psi)$.*

The previous definition justifies the following recursion on topics.

COROLLARY 3. *The topic of a complex formula is built as follows:*
  (1) $\mathbf{t}(\neg\varphi) = \mathbf{t}(\Box\varphi) = \mathbf{t}([ag]\varphi) = \mathbf{t}(\varphi)$.
  (2) $\mathbf{t}(\alpha \wedge \varphi) = \mathbf{t}([im_\alpha]\varphi) = \mathbf{t}([im_\alpha^\tau]\varphi) = \mathbf{t}([\approx_\alpha^\tau]\varphi) = \mathbf{t}(\alpha) \sqcup \mathbf{t}(\varphi)$.

Finally, we say that a topic is *consistent relative to* α if it is consistent with the topic associated with α, and not consistent relative to α otherwise:

– τ is a consistent topic w.r.t. α: $\tau \sqsubseteq \mathbf{t}(\alpha)$
– τ is an inconsistent topic w.r.t. α: $\tau \not\sqsubseteq \mathbf{t}(\alpha)$.

The topic associated with α at a world w is a maximal topic with respect to all the topics consistent with α at w. The idea is that $\mathbf{t}(\alpha)$ corresponds to the most specific question α can be an answer to, so that a topic τ is consistent with α only if it corresponds to a question whose answers are disjunctions of answers to the question associated with $\mathbf{t}(\alpha)$.[3]

---

[3] See Lewis (1988), pp. 161–163 for an extensional implementation of this idea. There, the topic of a proposition on the seventeenth century corresponds to the most specific question concerning that century, i.e., to the partition that gives all the possible complete answers to the question itself. The question asks which cell of the partition includes the actual world and the cells represent the most specific exclusive descriptions of the seventeenth century. Therefore, any other proposition on the seventeenth century turns out to be a union of disjunctive cells in that partition.

4.2. Models. Full models for $\mathcal{L}$ involve topic models together with a set of functions interpreting the modalities introduced above.

DEFINITION 4 (Model for $\mathcal{L}$). *A model for $\mathcal{L}$ is a tuple*

$$\mathcal{M} = \left\langle W, \{\mathcal{T}_w\}_{w \in W}, R, \{C_\alpha\}_{\alpha \in Fm(\mathcal{L})}, \{C_\alpha^\tau\}_{\substack{\alpha \in Fm(\mathcal{L}) \\ \tau \in Var_T}}, \{I_\alpha^\tau\}_{\substack{\alpha \in Fm(\mathcal{L}) \\ \tau \in Var_T}}, V \right\rangle$$

consisting of the following elements:

- *a set $W \neq \varnothing$ of possible worlds;*
- *for each $w \in W$, a topic model $\mathcal{T}_w = \langle T_w, \sqcup_w, t_w \rangle$; and*
- *a function $R : W \to \wp(W)$;*
  (If the agent at w produces a certain imaginative content based on an input, $R(w)$ is the set of worlds that are like w up to the content imagined by the agent given the input, i.e., where the agent could produce a different imaginative content based on the same input. $R(w)$ gives the agent's set of options when she receives an input, selects a topic, and picks a specification of the selected topic.)
- *for each $\alpha \in Fm(\mathcal{L})$, a function $C_\alpha : W \to \wp(W)$;*
  (If the agent is entertaining an imagination act starting from input α at w, $C_\alpha(w)$ is the set of worlds consistent with the input and the background knowledge and beliefs of the agent at w, i.e., the set of relevant worlds given input α at w. This represents the content imagined by the agent at w given input α. It corresponds to the set-selection functions adopted in Berto (2018a)'s semantics for imagination operators, and makes of them variably strict, nonmonotonic modals.[4])
- *for each $\alpha \in Fm(\mathcal{L}), \tau \in Var_T$, a function $C_\alpha^\tau : W \to \wp(W)$;*
  (If the agent is entertaining an imagination act starting from input α and with topic τ at w, $C_\alpha^\tau(w)$ is the union of the cells selected by the agent at w in the partition of $C_\alpha(w)$ determined by τ. This represents the content imagined by the agent at w given input α and topic τ.)
- *for each $\alpha \in Fm(\mathcal{L}), \tau \in Var_T$, a function $I_\alpha^\tau : W \to \wp(W)$;*
  (If w belongs to a set of relevant worlds given input α, $I_\alpha^\tau(w)$ is the equivalence class of w in the partition induced by τ on that set of relevant worlds.)
- *$V : Var_P \to \wp(W)$ is a standard valuation.*

In case you're feeling a bit lost concerning the behavior of C_α, C_α^τ, and I_α^τ: suppose w_0 is a world where an act of ROMS, based on input α, is performed. Given the input, a set $C_\alpha(w_0)$ of worlds is selected: these are all the worlds consistent with α and with the agent's belief system at w_0. Also, a topic τ is chosen: this corresponds to the set of questions the agent performing the ROMS based on input α at w_0 aims to answer. This topic determines a partition, represented by I_α^τ, of $C_\alpha(w_0)$. Cells $C_\alpha^\tau(w_0)$ in this

---

[4] Which worlds are selected, exactly? Our C_α will have few constraints in this respect, but one can intuitively understand them as the *most plausible* worlds where the input α is true, given the agent's background knowledge and beliefs. The idea is the one of minimal alteration of how we believe reality to be (given the input), flagged in §2, ROMS feature 3. In Berto (2018b), this is modeled via a system of nested spheres of worlds in the style of Lewis (1973)'s semantics for counterfactuals, but interpreted as representing not degrees of objective similarity, but degrees of subjective plausibility, as in the (Grove, 1988) reformulation of the spheres popular in doxastic logic and belief revision literature. We keep things simpler here by not adding the spheres to our models.
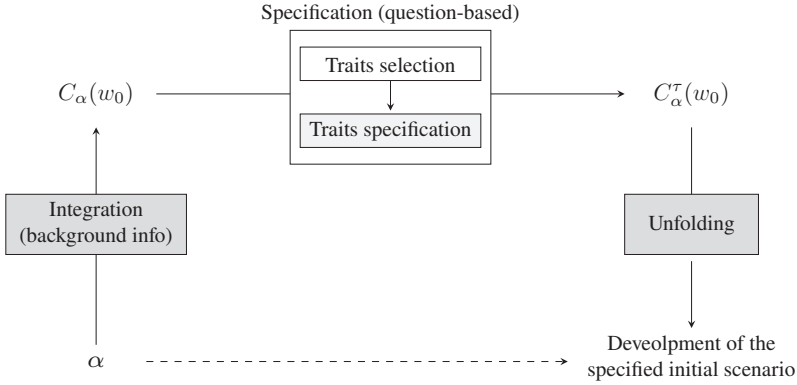
Fig. 2. Representation of a ROMS at a world w_0 of a model for $\mathcal{L}$.

partition are then picked out: these represent the answers the agent gives to the set of questions determined by topic τ, given input α, at w_0.

Thus, $C_\alpha(w_0)$ and $C_\alpha^\tau(w_0)$ represent, respectively, just our *basic initial scenario* and *specified initial scenario* in an act of ROMS, based on input α and driven by topic τ, in which the agent engages at w_0 (look at Figure 2). Function I_α^τ is needed to represent the partition of $C_\alpha(w_0)$ induced by τ: without it, we wouldn't be able to connect topics with partitions, and so to explicitly model the selection of cells in a partition.

As for function R, consider again world w_0. As shown in Illustration 1, since the agent is free to select different cells in the partition determined by τ, there is a world w_1 that is like w_0 up to the imaged content in which different cells are picked out. If α is *I jump the stream* and τ is *Will I make it to the other side?*, then w_0 can be a world in which the agent represents herself as weighing either 74 or 75 kg and the ground as wet and w_1 a world in which the agent represents herself as weighing either 74 or 75 kg and the ground as dry.

Illustration 1: Same topic, different cells (in grey).
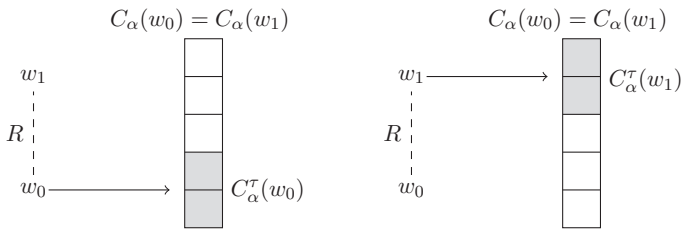
Also, as shown in Illustration 2, since the agent is free to select the topic itself, there is a world w_2 that is like w_0 up to the imaged content in which the selected topic—hence the corresponding partition—is different. If α is *I jump the stream*, τ is *Will I make it to the other side?*, and τ' is *Will I scare the cattle on the other side?*, then w_0 can be a world in which the agent has the purpose to answer τ and, to this end, represents herself as weighing either 74 or 75 kg and the ground as wet and w_2 a world in which

the agent has the purpose to answer τ' and, to this end, represents herself as holding back a scream.

Illustration 2: Different topic, different cells (in grey).
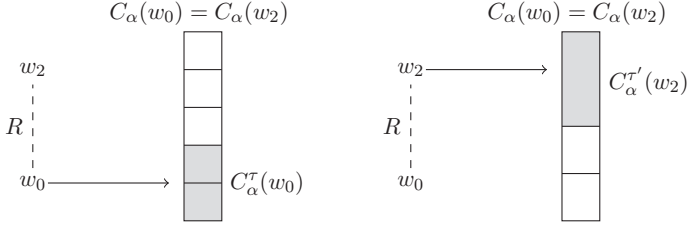
To actually ensure that the items in our models work as depicted in the diagrams, we need them to satisfy a number of conditions, to which we now turn.

4.3. Conditions on models. First, worlds indiscernible with respect to the imaginative options of the agent must share the same topic model. Next, R is an equivalence relation: it returns worlds that are indistinguishable up to the imaginative contents the agent can produce. Overall:

Conditions on $\mathcal{T}$ and R:
>   Condition $\mathcal{T}$: $v \in R(w) \Rightarrow \mathcal{T}_w = \mathcal{T}_v$
>   Condition $R(1)$: $w \in R(w)$
>   Condition $R(2)$: $v \in R(w) \Rightarrow R(v) = R(w)$.

I_α^τ must be an equivalence relation as well: given a w, I_α^τ returns a set of worlds indistinguishable from w with respect to input α and topic τ, that is, which give the same answers as w to the questions determined by τ, given α. In addition, we need a condition ensuring that worlds indistinguishable with respect to input α and topic $\tau_1 \sqcup \tau_2$ be indistinguishable with respect to input α and both topics τ_1 and τ_2. Overall:

Conditions on I_α^τ:
>   Condition $I_\alpha^\tau(1)$: $w \in I_\alpha^\tau(w)$
>   Condition $I_\alpha^\tau(2)$: $v \in I_\alpha^\tau(w) \Rightarrow I_\alpha^\tau(v) = I_\alpha^\tau(w)$
>   Condition $I_\alpha^\tau(3)$: $I_\alpha^{\tau_1 \sqcup \tau_2}(w) \subseteq I_\alpha^{\tau_1}(w) \cap I_\alpha^{\tau_2}(w)$.

The following group of conditions goes thus: the one on C_α accounts for the idea that, if v is a world that is like w up to what the agent images given an input, then the agent's background belief system must be the same at v and w, and so v and w must agree on which worlds are consistent with both α and the agent's background belief system. The four conditions on C_α^τ ensure that this interacts the right way with the other items of our models:

Conditions on C_α and C_α^τ:
>   Condition C_α: $v \in R(w) \Rightarrow C_\alpha(v) = C_\alpha(w)$
>   Condition $C_\alpha^\tau(1)$: $C_\alpha^\tau(w) \subseteq C_\alpha(w)$
>   Condition $C_\alpha^\tau(2)$: $x \in C_\alpha(w) \Rightarrow \exists v(v \in R(w) \text{ and } x \in C_\alpha^\tau(v))$
>   Condition $C_\alpha^\tau(3)$: $x \in C_\alpha^\tau(w) \Rightarrow I_\alpha^\tau(x) \subseteq C_\alpha^\tau(w)$
>   Condition $C_\alpha^\tau(4)$: $C_\alpha^{\tau_1 \sqcup \tau_2}(w) \subseteq C_\alpha^{\tau_1}(w) \cap C_\alpha^{\tau_2}(w)$.

Recall that $C_\alpha^\tau(w)$ gives the content imagined at w viewed as the cells picked by the agent in the partition corresponding to the selected topic τ, given input α. Thus, $C_\alpha^\tau(1)$ states that what is imagined at w given an input and a selected topic is a specification of what is imagined given the input. $C_\alpha^\tau(2)$ states that any world consistent with input α and the agent's background beliefs at w represents an answer the agent can give to the questions corresponding to any topic she can select at w. $C_\alpha^\tau(3)$ states that no answer the agent can give to the questions determined by a topic, given a certain input, can take apart worlds that are indistinguishable given that topic and input. Finally, $C_\alpha^\tau(4)$ has it that the cells selected by the agent in the partition determined by topic τ_1 (τ_2) given input α at w include the cells selected by the agent in the partition determined by the topic $\tau_1 \sqcup \tau_2$ given the same input at w.

The following propositions show that the conditions we imposed on models for $\mathcal{L}$ are sufficient to capture the intuitive interpretation codified by the diagrams in Illustrations 1 and 2.

PROPOSITION 1. I_α^τ *induces a partition on* $C_\alpha(w)$.

Proof.   Suppose $x \in C_\alpha(w)$. Since $x \in I_\alpha^\tau(x)$, by $I_\alpha^\tau(1)$, every scenario in $C_\alpha(w)$ is contained in some I_α^τ-cell. Suppose now that $I_\alpha^\tau(x_1) \cap I_\alpha^\tau(x_2) \neq \varnothing$, for some $x_1, x_2 \in C_\alpha(w)$. Then there is y such that $y \in I_\alpha^\tau(x_1)$ and $y \in I_\alpha^\tau(x_2)$, and so $I_\alpha^\tau(x_1) = I_\alpha^\tau(y) = I_\alpha^\tau(x_2)$, by Condition $I_\alpha^\tau(2)$.   □

PROPOSITION 2. C_α^τ *is refined by the partition induced by* I_α^τ *on* $C_\alpha(w)$.

Proof.   Suppose $x \in C_\alpha(w)$. Then $x \in C_\alpha^\tau(v)$, for some $v \in R(w)$ by $C_\alpha^\tau(2)$, and so every scenario in $C_\alpha(w)$ is contained in some C_α^τ-cell. Suppose now that $C_\alpha^\tau(v_1) \cap C_\alpha^\tau(v_2) \neq \varnothing$, for some $v_1, v_2 \in R(w)$. Then there is y such that $y \in C_\alpha^\tau(v_1)$ and $y \in C_\alpha^\tau(v_2)$, and so $I_\alpha^\tau(y) \subseteq C_\alpha^\tau(v_1) \cap C_\alpha^\tau(v_2)$, by Condition $C_\alpha^\tau(3)$.   □

We now give the truth conditions for formulas of $\mathcal{L}$ in our models.

4.4. *Truth.*

DEFINITION 5 (Truth at a world in a model).   *Where* φ *is a formula of* $\mathcal{L}$, $\mathcal{M}$ *a model for* $\mathcal{L}$, *and* w *a world in* $\mathcal{M}$, *the relation* $\mathcal{M}, w \models \varphi$ *of truth of* φ *at* w *in* $\mathcal{M}$ *is recursively defined as:*

$$\mathcal{M}, w \models p \Leftrightarrow w \in V(p)$$
$$\mathcal{M}, w \models \neg\varphi \Leftrightarrow \mathcal{M}, w \not\models \varphi$$
$$\mathcal{M}, w \models \varphi \wedge \psi \Leftrightarrow \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \psi$$
$$\mathcal{M}, w \models \Box\varphi \Leftrightarrow \forall v(v \in W \Rightarrow \mathcal{M}, v \models \varphi)$$
$$\mathcal{M}, w \models [ag]\varphi \Leftrightarrow \forall v(v \in R(w) \Rightarrow \mathcal{M}, v \models \varphi)$$
$$\mathcal{M}, w \models [im_\alpha]\varphi \Leftrightarrow \forall v(v \in C_\alpha(w) \Rightarrow \mathcal{M}, v \models \varphi) \text{ and } \mathbf{t}_w(\varphi) \sqsubseteq_w \mathbf{t}_w(\alpha)$$
$$\mathcal{M}, w \models [im_\alpha^\tau]\varphi \Leftrightarrow \forall v(v \in C_\alpha^\tau(w) \Rightarrow \mathcal{M}, v \models \varphi) \text{ and } \mathbf{t}_w(\varphi) \sqsubseteq_w \mathbf{t}_w(\alpha)$$
$$\mathcal{M}, w \models [\approx_\alpha^\tau]\varphi \Leftrightarrow \forall v(v \in I_\alpha^\tau(w) \Rightarrow \mathcal{M}, v \models \varphi).$$

The indexing of $[im_\alpha]$ and $[im_\alpha^\tau]$ by α deals with **Feature 2**: the starting point of imagination acts comes with the input, deliberately chosen. For $[im_\alpha]\varphi$ to be true we require two constraints:
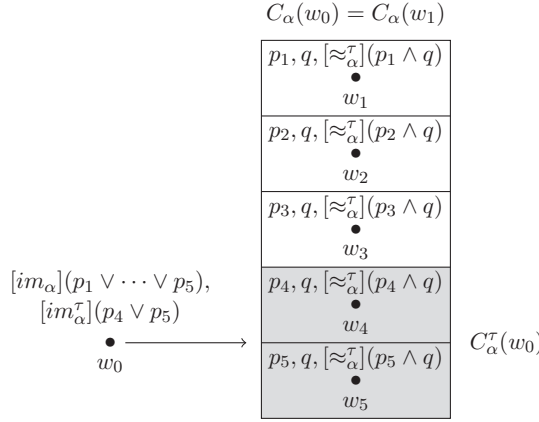
$$C_\alpha(w_0) = C_\alpha(w_1)$$

Fig. 3. Difference between imagination modalities.

(i) φ must be true in the worlds selected given input α, where the variability in the sets selected given different inputs captures the variability of information or beliefs imported, relative to such inputs, in accordance with **Feature 3**[5];

(ii) φ must be fully on-topic with respect to the input, thereby capturing the relevance or topicality constraints of a proper act of imagination, in accordance with **Feature 4**.

For $[im_\alpha^\tau]\varphi$ to be true, besides the same constraints, we require φ to be true throughout the worlds selected given input α *and* topic τ, that is, the issue or question addressed in the imaginative exercise, which contributes to fix what it's about, in accordance with **Feature 5**. The definition also makes evident the difference between $[\approx_\alpha^\tau]$ and $[im_\alpha^\tau]$: indiscernibility with respect to inputs and topics is an objective feature characterizing worlds, so the truth of $[\approx_\alpha^\tau]\varphi$ is independent of conditions concerning topics. Imaginability with respect to inputs and topics depends on the imaginative act and choices of the agent, so the truth of $[im_\alpha^\tau]\varphi$ also depends on conditions on topics.

The difference between the three modalities $[im_\alpha]\varphi$, $[im_\alpha^\tau]\varphi$, and $[\approx_\alpha^\tau]\varphi$ is illustrated by Figure 3. As in Illustration 1, suppose that α is the input *I jump the stream* and τ the topic *Will I make it to the other side?* In addition, let p_1 be *I weigh 71 kg*, p_2 be *I weigh 72 kg*, ..., p_5 be *I weigh 75 kg*, and q be *There are plenty of bikes in Amsterdam*. Given her background belief system, upon receiving input α, the agent selects worlds $w_1, \ldots, w_5$ in which she weighs between 71 and 75 kg and there are plenty of bikes in Amsterdam: letting $[\![\varphi]\!]^{\mathcal{M}} = \{w \in W \mid \mathcal{M}, w \models \varphi\}$ be the set of worlds where φ is true in model $\mathcal{M}$, we have both $C_\alpha(w_0) \subseteq [\![p_1 \vee p_2 \vee p_3 \vee p_4 \vee p_5]\!]^{\mathcal{M}}$ and $C_\alpha(w_0) \subseteq [\![q]\!]^{\mathcal{M}}$. Yet, while the topic of each of $p_1, \ldots, p_5$ (hence that of the disjunction $p_1 \vee p_2 \vee p_3 \vee p_4 \vee p_5$) is included in the topic of α, the topic of q is not. In fact, an answer to the most specific question corresponding to *I jump the stream* includes information about the agent's weight, but not about bikes in Amsterdam. This is why $[im_\alpha](p_1 \vee p_2 \vee p_3 \vee p_4 \vee p_5)$ is

---

5 In particular, the operators are non-monotonic: $[im_\alpha]\varphi$ can be true while $[im_{\alpha \wedge \beta}]\varphi$ is false. Given *I jump the stream*, you imagine that you fall into the water; but given *I jump the stream and I'm endowed with large operating wings*, you imagine that you comfortably make it to the other side.

true at w_0 but $[im_\alpha]q$ is not. Now, the model represents a situation in which, upon receiving input α, the agent chooses to focus on topic τ and to specify this topic by selecting, among the $C_\alpha(w_0)$-scenarios, only those in which she weighs either 74 or 75 kg, i.e., $C_\alpha^\tau(w_0) \subseteq [\![p_4 \vee p_5]\!]^{\mathcal{M}}$ (as before, $C_\alpha^\tau(w_0)$-scenarios are in gray). Given the facts about the topics of p_4 and p_5 mentioned above, we thus have that $[im_\alpha^\tau](p_4 \vee p_5)$ is true at w_0, even if $[im_\alpha](p_4 \vee p_5)$ is not: the selection and specification of a topic constrains the act of imagination at w_0 in a stronger way that the mere reception of input α. Finally, if we shift focus from world w_0, at which the agent is engaged in the imagination act, to world w_4, which is part of the specified imagined scenario, we can see that all worlds that are like w_4 with respect to input α and topic τ are such that the agent weighs 74 kg and there are bikes in Amsterdam. Since $I_\alpha^\tau(w_4) \subseteq [\![p_4]\!]^{\mathcal{M}}$ and $I_\alpha^\tau(w_4) \subseteq [\![q]\!]^{\mathcal{M}}$, we thus have that $[\approx_\alpha^\tau]p_4$ and $[\approx_\alpha^\tau]q$ are both true at w_4.

At this point, the special role of the formulas met in Section 3 can also be precisely characterized:

– $\langle im_\alpha \rangle \mathbf{var}(\alpha)$ states the occurrence of an imagination act based on input α:
$\mathcal{M}, w \models \langle im_\alpha \rangle \mathbf{var}(\alpha) \Leftrightarrow C_\alpha(w) \neq \varnothing$
– $\langle im_\alpha^\tau \rangle \mathbf{var}(\alpha)$ states the occurrence of an imagination act based on input α and processed in light of topic τ:
$\mathcal{M}, w \models \langle im_\alpha^\tau \rangle \mathbf{var}(\alpha) \Leftrightarrow C_\alpha^\tau(w) \neq \varnothing.$

Finally, let us select the class of models which are appropriate to represent ROMS, i.e., where imagination acts are successful:

DEFINITION 6 (Appropriate model). *A model $\mathcal{M}$ for $\mathcal{L}$ is said to be appropriate for imagination acts only if the following Success condition is satisfied:*

Success: $v \in C_\alpha(w) \Rightarrow \mathcal{M}, v \models \alpha.$

Success has it that all the α-selected worlds are α-worlds. It ensures that α is true of any imagined scenario based on input α, no matter under which topic α is specified. Thus, given input α, one always imagines that α. (We mention that this may be controversial if imaginative resistance is taken seriously (Gendler, 2000))

§5. VI: the logic of voluntary imagination. The logic of voluntary imagination **VI**, as defined by the class of all appropriate models for $\mathcal{L}$, is an extension of the logic **I** of imagination introduced in Berto (2018a). In this and the next sections we show that this logic is axiomatisable, that it is a conservative extension of **I**, and that its expressive power allows us to better understand some important traits of ROMS.[6]

5.1. Axiomatization. The logic **VI** can be axiomatised by extending the system of axioms proposed by Giordani (2019) for **I**. The set of axioms and rules characterizing **VI** is the following:

[**Group 1**] Axioms and rules on $\Box$: standard $S5$ axioms and rules.
[**Group 2**] Axioms and rules on $[ag]$: standard $S5$ axioms.
[**Group 3**] Axioms and rules on $[\approx_\alpha^\tau]$: standard $S5$ axioms.

---

[6] On the logical side, it would also be interesting to consider whether the system **VI** has the finite model property. Yet, the presence of topic models associated to possible worlds makes this is a nontrivial problem, which we decided to leave to future work.

[**Group 4**] Axioms on $[im_\alpha^\tau]$:

$[im_\alpha^\tau]$**1**: $[im_\alpha^\tau]\varphi \wedge [im_\alpha^\tau]\psi \rightarrow [im_\alpha^\tau](\varphi \wedge \psi)$

$[im_\alpha^\tau]$**2**: $[im_\alpha^\tau]\varphi \wedge [im_\alpha^\tau](\varphi \rightarrow \psi) \rightarrow [im_\alpha^\tau]\psi$

$[im_\alpha^\tau]$**3**: $[im_\alpha^\tau]\varphi \rightarrow [im_\alpha^\tau]\mathbf{var}(\varphi)$

$[im_\alpha^\tau]$**4**: $[im_\alpha^\tau]\mathbf{var}(\varphi)$, if $Var(\varphi) \subseteq Var(\alpha)$

$[im_\alpha^\tau]$**5**: $[im_\alpha^\tau]\mathbf{var}(\varphi) \wedge [im_\varphi^\tau]\mathbf{var}(\psi) \rightarrow [im_\alpha^\tau]\mathbf{var}(\psi)$.

[**Group 5**] Axioms on topics composition:

Topic 1: $[\approx_\alpha^{\tau_1}]\varphi \vee [\approx_\alpha^{\tau_2}]\varphi \rightarrow [\approx_\alpha^{\tau_1 \sqcup \tau_2}]\varphi$;

Topic 2: $[im_\alpha^{\tau_1}]\varphi \vee [im_\alpha^{\tau_2}]\varphi \rightarrow [im_\alpha^{\tau_1 \sqcup \tau_2}]\varphi$.

[**Group 6**] Bridge axioms:

Bridge 1: $\Box\varphi \rightarrow [ag]\varphi \wedge [\approx_\alpha^\tau]\varphi$

Bridge 2: $[im_\alpha]\varphi \leftrightarrow [ag][im_\alpha^\tau]\varphi$

Bridge 3: $[im_\alpha^\tau]\mathbf{var}(\varphi) \rightarrow [ag][im_\alpha^\tau]\mathbf{var}(\varphi)$

Bridge 4: $[im_\alpha^\tau]\varphi \rightarrow [im_\alpha^\tau][\approx_\alpha^\tau]\varphi$

Bridge 5: $\Box(\alpha \rightarrow \varphi) \wedge [im_\alpha^\tau]\mathbf{var}(\varphi) \rightarrow [im_\alpha^\tau]\varphi$.

Note that, in light of **Bridge 2**, $[im_\alpha]$ turns out to be definable in terms of $[ag][im_\alpha^\tau]$. This corresponds to the fact that, according to the conditions on C_α and C_α^τ, $C_\alpha(w)$ is precisely the set of scenarios that are contained in $C_\alpha^\tau(v)$ for some scenario accessible from w, i.e., $C_\alpha(w)$ is the set that is partitioned by τ. We decided to introduce C_α and C_α^τ separately, and then to include Bridge 2 as an axiom, to highlight the role of C_α in the act of imagination and focus on its significance in selecting the imagined scenarios. The following theorems are direct consequences of **Bridge 2** and the axioms on operators $[ag]$ and $[im_\alpha^\tau]$. In particular, they are crucial to prove completeness.

$[im_\alpha]$**T1**: $[im_\alpha]\varphi \wedge [im_\alpha]\psi \rightarrow [im_\alpha](\varphi \wedge \psi)$

$[im_\alpha]$**T2**: $[im_\alpha]\varphi \wedge [im_\alpha](\varphi \rightarrow \psi) \rightarrow [im_\alpha]\psi$

$[im_\alpha]$**T3**: $[im_\alpha]\varphi \rightarrow [im_\alpha]\mathbf{var}(\varphi)$

$[im_\alpha]$**T4**: $[im_\alpha]\mathbf{var}(\varphi)$, if $Var(\varphi) \subseteq Var(\alpha)$

$[im_\alpha]$**T5**: $[im_\alpha]\mathbf{var}(\varphi) \wedge [im_\varphi]\mathbf{var}(\psi) \rightarrow [im_\alpha]\mathbf{var}(\psi)$.

Furthermore

$[im_\alpha]$**T6**: $[im_\alpha]\varphi \rightarrow [im_\alpha^\tau]\varphi$

$[im_\alpha]$**T7**: $[im_\alpha]\varphi \rightarrow [ag][im_\alpha]\varphi$

$[im_\alpha]$**T8**: $\langle im_\alpha \rangle \varphi \rightarrow [ag]\langle im_\alpha \rangle \varphi$.

Theorems $[im_\alpha]$**T1**- $[im_\alpha]$**T5** show that the logic of $[im_\alpha]$ is analogous to the logic of $[im_\alpha^\tau]$, while the final three theorems provide insight on the connection between $[im_\alpha]$ and $[im_\alpha^\tau]$ and the connection between $[im_\alpha]$ and $[ag]$.

THEOREM 1. *The axiom system **VI** is sound and strongly complete with respect to the class of all appropriate models for $\mathcal{L}$.*

Soundness is not difficult and is left as exercise. Completeness is based on a canonical model construction, which in turn is based on the construction of a canonical topic model. Thus we need a way to define topics in terms of linguistic entities. The idea is to identify a topic with an equivalence class of formulas, i.e., those which are about the same topic. Hence, we start by defining a suitable relation between formulas of *being about the topic of*. Consider, first, the following lemma.

LEMMA 1. $M, w \models [im_\alpha] \operatorname{var}(\varphi) \Leftrightarrow t_w(\varphi) \sqsubseteq_w t_w(\alpha)$.

Proof. By the definition of var (φ), for all v, we have that $M, v \models \operatorname{var}(\varphi)$. Hence, for all w, $M, w \models [im_\alpha] \operatorname{var}(\varphi) \Leftrightarrow t_w(\varphi) \sqsubseteq_w t_w(\alpha)$, by the truth conditions relative to $[im_\alpha]$. □

Given Lemma 1, we introduce:

DEFINITION 7 (Aboutness). *For any $\varphi, \alpha \in Fm(\mathcal{L})$,*
– φ is about the topic of α: $\varphi \leq_T \alpha := [im_\alpha] \operatorname{var}(\varphi)$
– φ and α are about the same topic: $\varphi =_T \alpha := [im_\alpha] \operatorname{var}(\varphi) \wedge [im_\varphi] \operatorname{var}(\alpha)$.

The following theorems are then derivable from theorems $[im_\alpha]$**T1** – $[im_\alpha]$**T5**.

1. $\vdash_{\mathbf{VI}} \alpha =_T \alpha$
2. $\vdash_{\mathbf{VI}} \alpha =_T \varphi \to \varphi =_T \alpha$
3. $\vdash_{\mathbf{VI}} \alpha =_T \varphi \wedge \varphi =_T \psi \to \alpha =_T \psi$
4. $\vdash_{\mathbf{VI}} \alpha_1 =_T \varphi_1 \wedge \alpha_2 =_T \varphi_2 \to (\alpha_1 \wedge \alpha_2) =_T (\varphi_1 \wedge \varphi_2)$
5. $\vdash_{\mathbf{VI}} \alpha =_T (p_1 \wedge \cdots \wedge p_n)$, if $Var(\alpha) = \{p_1, ..., p_n\}$
6. $\vdash_{\mathbf{VI}} \alpha \leq_T \varphi \leftrightarrow \alpha =_T (\alpha \wedge \varphi)$.

The topic of a formula α at a world, viewed as a maximally **VI**-consistent set, is just the set of formulas that are about the same topic as α at that world: $T(\alpha, w) := \{\psi : \alpha =_T \psi \in w\}$. Thus, by virtue of the previous theorems

(i) $T(\alpha, w) = T(\varphi, w) \Leftrightarrow \alpha =_T \varphi \in w$,
(ii) $T(\alpha, w) = T(\alpha \wedge \varphi, w) \Leftrightarrow \alpha \leq_T \varphi \in w$.

This allows us to define the canonical topic model.

DEFINITION 8 (Canonical topic model). *Let w be a maximally **VI**-consistent set. The w-canonical topic model is the triple $\mathcal{T}_w = \langle T_w, \sqcup_w, t_w \rangle$ such that*

– $T_w = \{T(\alpha, w) : \alpha \in Fm(\mathcal{L})\}$
– $\sqcup_w$ is such that $T(\alpha, w) \sqcup_w T(\varphi, w) = T(\alpha \wedge \varphi, w)$
– $t_w(p_i) = T(p_i, w)$.

Theorems 1–4 ensure that $\sqcup_w$ is well-defined and that it is idempotent, commutative and associative, so that $\langle T_w, \sqcup_w, t_w \rangle$ is indeed a topic model. Finally, the definition of $\mathbf{t}_w$ is consistent with the definition of t_w. To see that, let $Var(\alpha) = \{p_1, \ldots, p_n\}$. Then

$\vdash_{\mathbf{VI}} \alpha =_T (p_1 \wedge \cdots \wedge p_n)$, by point 6
$T(\alpha, w) = T(p_1 \wedge \cdots \wedge p_n, w)$, by the definition of T
$T(\alpha, w) = T(p_1, w) \sqcup_w \cdots \sqcup_w T(p_n, w)$, by the definition of $\sqcup_w$
$T(\alpha, w) = t_w(p_1) \sqcup_w \cdots \sqcup_w t_w(p_n)$, by the definition of t_w
$T(\alpha, w) = \mathbf{t}_w(\alpha)$, by the definition of $\mathbf{t}_w$.

With canonical topic models in place, let us now take a **VI**-consistent set $X \subseteq Fm(\mathcal{L})$ and let us extend it to a maximally **VI**-consistent set w_0. We now build the canonical model for **VI** around w_0.

DEFINITION 9 (Canonical model for **VI** around w_0). *The canonical model for **VI** around w_0 is the tuple*

$$\mathcal{M}_C = \left\langle W, \{\mathcal{T}_w\}_{w \in W}, R, \{C_\alpha\}_{\alpha \in Fm(\mathcal{L})}, \{C_\alpha^\tau\}_{\substack{\alpha \in Fm(\mathcal{L}) \\ \tau \in Var_T}}, \{I_\alpha^\tau\}_{\substack{\alpha \in Fm(\mathcal{L}) \\ \tau \in Var_T}}, V \right\rangle$$

such that

- $W = \{w : w \text{ is maximally } \textbf{VI}\text{-consistent and } \{\varphi : \Box\varphi \in w_0\} \subseteq w\}$
- $\mathcal{T}_w$ *is the* w-*canonical topic model*
- R *is such that* $v \in R(w) \Leftrightarrow \{\varphi : [ag]\varphi \in w\} \subseteq v$
- C_α *is such that* $v \in C_\alpha(w) \Leftrightarrow (\exists x, \tau)(x \in R(w) \text{ and } v \in C_\alpha^\tau(x))$
- C_α^τ *is such that* $v \in C_\alpha^\tau(w) \Leftrightarrow \{\varphi : \exists\psi(\Box(\psi \to \varphi) \wedge [im_\alpha^\tau]\psi \in w)\} \subseteq v$
- I_α^τ *is such that* $v \in I_\alpha^\tau(w) \Leftrightarrow \{\varphi : [\approx_\alpha^\tau]\varphi \in w)\} \subseteq v$
- V *is such that* $w \in V(p_i) \Leftrightarrow p_i \in w$.

The axioms of **VI** ensure that, for any $w \in W$, $\alpha \in Fm(\mathcal{L})$, and $\tau \in Var_T$,

- $R(w) \subseteq W$, immediate by **Bridge 1**;
- $I_\alpha^\tau(w) \subseteq W$, immediate by **Bridge 1**;
- $C_\alpha^\tau(w) \subseteq W$, immediate by the logic of $\Box$, axiom $[im_\alpha^\tau]\textbf{4}$, and **Bridge 5**;
- $C_\alpha(w) \subseteq W$, by definition and the previous facts.

LEMMA 2 (Canonicity Lemma). *The canonical model for* **VI** *around* w_0 *is a model for* $\mathcal{L}$.

Proof. It is not difficult to see that $\mathcal{M}_C$ satisfies the conditions on $\mathcal{T}$, R, I_α^τ, given the axioms of Group 2, Group 3, **Topic 1**, **Bridge 3**, and Theorem **T6**. Hence, let us focus on the conditions on the functions C_α and C_α^τ modeling imaginative acts.

Condition C_α: $v \in R(w) \Rightarrow C_\alpha(v) = C_\alpha(w)$.

Suppose $v \in R(w)$. Since $v \in R(w) \Rightarrow R(v) = R(w)$, $\exists x, \tau(x \in R(v)$ and $v \in C_\alpha^\tau(x))$ if and only if $\exists x, \tau(x \in R(w)$ and $v \in C_\alpha^\tau(x))$, and so $x \in C_\alpha(v)$ if and only if $x \in C_\alpha(w)$, by the definition of C_α.

Condition C_α^τ (1): $C_\alpha^\tau(w) \subseteq C_\alpha(w)$.

Suppose $x \in C_\alpha^\tau(w)$. Then $w \in R(w)$ and $x \in C_\alpha^\tau(w)$, since R is reflexive, and so $x \in C_\alpha(w)$, by the definition of C_α.

Condition C_α^τ (2): $x \in C_\alpha(w) \Rightarrow \exists v(v \in R(w)$ and $x \in C_\alpha^\tau(v))$.

Straightforward, by the definition of C_α.

Condition C_α^τ (3): $x \in C_\alpha^\tau(w) \Rightarrow I_\alpha^\tau(x) \subseteq C_\alpha^\tau(w)$.

Suppose $x \in C_\alpha^\tau(w)$ and $v \in I_\alpha^\tau(x)$, so that $\{\varphi : [\approx_\alpha^\tau]\varphi \in x\} \subseteq v$. We show that $v \in C_\alpha^\tau(w)$. So, suppose there is ψ s.t. $\Box(\psi \to \varphi) \wedge [im_\alpha^\tau]\psi \in w$. Then, by the logic of $\Box$, $[\approx_\alpha^\tau]$, and **Bridge 4**, $\Box([\approx_\alpha^\tau]\psi \to [\approx_\alpha^\tau]\varphi) \wedge [im_\alpha^\tau][\approx_\alpha^\tau]\psi \in w$. Since $x \in C_\alpha^\tau(w)$, it follows that $[\approx_\alpha^\tau]\varphi \in x$, and so, since $v \in I_\alpha^\tau(x)$, $\varphi \in v$, whence the conclusion.

Condition C_α^τ (4): $C_\alpha^{\tau_1 \sqcup \tau_2}(w) \subseteq C_\alpha^{\tau_1}(w) \cap C_\alpha^{\tau_2}(w)$.

Straightforward, by **Topic 2**. □

LEMMA 3. *For all* $w \in W$, $\varphi, \alpha \in Fm(\mathcal{L})$, *and* $\tau \in Var_T$,

$\forall v(v \in C_\alpha^\tau(w) \Rightarrow \varphi \in v)$ *and* $\varphi \leq_T \alpha \in w \Leftrightarrow [im_\alpha^\tau]\varphi \in w$.

Proof. See Giordani (2019), Lemma 3.12 and Lemma 3.14. □

LEMMA 4 (Truth Lemma). *For all* $w \in W$ *and* $\varphi \in Fm(\mathcal{L})$,

$\mathcal{M}_C, w \models \varphi \Leftrightarrow \varphi \in w$.

Proof. We only consider the cases concerning the imaginative modalities, since the proof of the other ones is standard.

Case 1: $\mathcal{M}_C, w \models [im_\alpha^\tau]\varphi \Leftrightarrow [im_\alpha^\tau]\varphi \in w$.

$\mathcal{M}_C, w \models [im_\alpha^\tau]\varphi$
$\Leftrightarrow \forall v(v \in C_\alpha^\tau(w) \Rightarrow \mathcal{M}_C, v \models \varphi)$ and $\mathbf{t}_w(\varphi) \sqsubseteq_w \mathbf{t}_w(\alpha)$
$\Leftrightarrow \forall v(v \in C_\alpha^\tau(w) \Rightarrow \mathcal{M}_C, v \models \varphi)$ and $\varphi \leq_T \alpha \in w$, by the definition of $\mathbf{t}_w$
$\Leftrightarrow \forall v(v \in C_\alpha^\tau(w) \Rightarrow \varphi \in v)$ and $\varphi \leq_T \alpha \in w$, by inductive hypothesis
$\Leftrightarrow [im_\alpha^\tau]\varphi \in w$, by Lemma 3.

Case 2: $\mathcal{M}_C, w \models [im_\alpha]\varphi \Leftrightarrow [im_\alpha]\varphi \in w$.

$\mathcal{M}_C, w \models [im_\alpha]\varphi$
$\Leftrightarrow \forall v(v \in C_\alpha(w) \Rightarrow \mathcal{M}_C, v \models \varphi)$ and $\mathbf{t}_w(\varphi) \sqsubseteq_w \mathbf{t}_w(\alpha)$
$\Leftrightarrow \forall v(v \in C_\alpha(w) \Rightarrow \mathcal{M}_C, v \models \varphi)$ and $\varphi \leq_T \alpha \in w$, by the definition of $\mathbf{t}_w$
    and by the definition of C_α
$\Leftrightarrow \forall x, v(x \in R(w)$ and $v \in C_\alpha^\tau(w) \Rightarrow \mathcal{M}_C, v \models \varphi)$ and $\varphi \leq_T \alpha \in w$
$\Leftrightarrow \forall x(x \in R(w) \Rightarrow \mathcal{M}_C, x \models [im_\alpha^\tau]\varphi)$ and $\varphi \leq_T \alpha \in w$
$\Leftrightarrow \mathcal{M}_C, w \models [ag][im_\alpha^\tau]\varphi$ and $\varphi \leq_T \alpha \in w$
$\Leftrightarrow [ag][im_\alpha^\tau]\varphi \in w$ and $\varphi \leq_T \alpha \in w$, by inductive hypothesis
$\Leftrightarrow [im_\alpha]\varphi \in w$, by **Bridge 2**, the definition of $\leq_T$ and $[im_\alpha]$**T3**. $\qquad \square$

Finally, we show that the canonical model is an appropriate model for $\mathcal{L}$.

LEMMA 5. $\mathcal{M}_C$ *satisfies Success, i.e., for any* $w, v \in W$ *and* $\alpha \in Fm(\mathcal{L})$,

$$v \in C_\alpha(w) \Rightarrow \mathcal{M}_C, v \models \alpha.$$

Proof. First, note that

$\vdash_{\mathbf{VI}} \Box(\alpha \to \alpha) \wedge [im_\alpha^\tau]\mathbf{var}(\alpha) \to [im_\alpha^\tau]\alpha$, by **Bridge 5**
$\vdash_{\mathbf{VI}} [im_\alpha^\tau]\mathbf{var}(\alpha) \to [im_\alpha^\tau]\alpha$, by the logic of $\Box$
$\vdash_{\mathbf{VI}} [im_\alpha^\tau]\alpha$, by $[im_\alpha^\tau]$**4**:
$\vdash_{\mathbf{VI}} \Box(\alpha \to \alpha) \wedge [im_\alpha^\tau]\alpha$, again by the logic of $\Box$.

Hence, for every $w \in W$, $\alpha \in \{\varphi : \exists\psi(\Box(\psi \to \varphi) \wedge [im_\alpha^\tau]\psi \in w)\}$. Thus, if we assume $v \in C_\alpha^\tau(w)$, we conclude $\alpha \in v$. $\qquad \square$

Since w_0 extends an arbitrary **VI**-consistent set, Lemmas 2, 4, and 5 suffice to conclude that every **VI**-consistent set is satisfiable in an appropriate model for $\mathcal{L}$, and so that **VI** is strongly complete with respect to the class of all such models.

5.2. Comparison with the basic system.

THEOREM 2. *VI is a conservative extension of* **I**[7] :

For all $X \subseteq Fm(\mathcal{L}_I)$ *and* $\varphi \in Fm(\mathcal{L}_I)$, $X \Vdash_I \varphi \Leftrightarrow X \Vdash_{VI} \varphi$,

where $Fm(\mathcal{L}_I)$ *is the set of formulas of the language* $\mathcal{L}_I$ *of* **I**.

Proof. Since **I** is complete and the axioms and rules of **I** are sound with respect to the class of appropriate models for **VI**, all the valid propositions of the basic system

---

[7] For the definition of system **I** see Giordani (2019). Essentially, the logic **I** of imagination consists of the axioms for $\Box$, axiom **Bridge 4** and axioms corresponding to Theorems $[im_\alpha]$**T1**– $[im_\alpha]$**T5**.

are valid in the present system. In addition, all the invalid propositions of the basic system are invalid in the present system, since to every model of **I** corresponds a model of the present system validating the same propositions in the restricted language of **I**. To be sure, let $\mathcal{M} = \langle W, \mathcal{T}, \{C_\alpha\}_{\alpha \in Fm(\mathcal{L})}, V \rangle$ be a model of **I**. Then,

$$\mathcal{M}' = \left\langle W, \{\mathcal{T}_w\}_{w \in W}, R, \{C_\alpha\}_{\alpha \in Fm(\mathcal{L})}, \{C_\alpha^\tau\}_{\substack{\alpha \in Fm(\mathcal{L}) \\ \tau \in Var_T}}, \{I_\alpha^\tau\}_{\substack{\alpha \in Fm(\mathcal{L}) \\ \tau \in Var_T}}, V \right\rangle$$

is a model satisfying the same propositions in $Fm(\mathcal{L}_\mathbf{I})$ as $\mathcal{M}$, where

$\mathcal{T}_w = \mathcal{T}$, for all $w \in W$
R is such that $R(w) = \{w\}$, for all $w \in W$
I_α^τ is such that $I_\alpha^\tau(w) = \{w\}$, for each α and τ, and for all $w \in W$
C_α^τ is such that $C_\alpha^\tau(w) = C_\alpha(w)$, for each α and τ, and for all $w \in W$. $\qquad\square$

Hence, as expected, the basic system of the logic of imagination can be obtained from the logic of voluntary imagination by avoiding modeling the voluntary activity of topic selection, as represented by C_α^τ.

§6. Imagination and knowledge. Let us now consider issues concerning paradigmatic cases of imagination acts. The aim here is to illustrate the expressive power of logic **VI** in reference to the general questions proposed in the introduction. Hence, we are going to use our fine-grained framework to capture important distinctions involved in those cases. A more extensive and systematic study of the implications of the framework is left for future work.

Case 1. *The possibility of exporting $[im_\alpha]$ with respect to conjunction.*

$[im_\alpha]$ **T1**: $\quad [im_\alpha]\varphi \wedge [im_\alpha]\psi \to [im_\alpha](\varphi \wedge \psi)$, derivable in **VI**.

This concerns the logic of imagination. In Berto (2018a) the following problem is proposed, building on a famous Quinean example:

> The explicit input indexing $[im_\alpha]$ involves Caesar being in command of the U.S. troops in the war of Korea. We can imagine him using bombs, φ, importing in the representation the weapons available in the Korean war, or we can imagine him using catapults, ψ, importing the military apparatus available to Caesar. However, one would not thereby infer $[im_\alpha](\varphi \wedge \psi)$, Caesar's employing both bombs and catapults. [pp. 1879–1880, *notation adapted*]

The pursued solution is based on the observation that acts of imagination are contextually determined, so that the same explicit input can trigger different acts in different circumstances. Hence, relative to the quoted case, we might assume that the conclusion seems to be paradoxical due to a contextual shift and that, once the context is fixed, no paradox arises. Thus, we can obtain a straightforward way out by indexing acts of imagination with respect to contexts. This is a mere suggestion in that paper. Here's how we can make it work in our framework: Caesar employing bombs and Caesar employing catapults are different cells in a partition connected with the topic *weapons used by Caesar*. Let this be τ. We can represent the case as $[im_\alpha^\tau]\varphi \wedge \langle ag \rangle [im_\alpha^\tau]\psi$ and note that the possibility of exporting $[im_\alpha]$ with respect to conjunction is not touched by this example, since

$$[im_\alpha^\tau]\varphi \wedge \langle ag \rangle [im_\alpha^\tau]\psi \to [im_\alpha^\tau](\varphi \wedge \psi)$$

is not a valid principle. As we can see, in this case exploiting a reference to a topic is crucial to provide a correct interpretation of the kind of act of imagination we are performing.

Case 2. *The possibility of selective imagination.*

The second case concerns an aspect of the connection between imagination and knowledge: our acts of ROMS are constrained both by what we know or believe, and by the goal for which they are pursued. In Williamson (2016), p. 114, the following situation is discussed, concerning a group of our ancestors supposing that there are wolves in the forest they are about to enter:

> To serve that purpose well, the imagination must be both selective and reality-oriented. They could imagine the wolves bringing them food to eat, but doing so would be a waste of time, and a distraction from more practically relevant possibilities. An imagination that clutters up the mind with a bewildering plethora of wildly unlikely scenarios is almost as bad as no imagination at all. It is better to have an imagination that concentrates on fewer and more likely scenarios. One's imagination should not be completely independent of one's knowledge of what the world is like.

We can represent how imagination is selective and reality-oriented. On the one hand, oriented selectivity has to do with the fact that the selected scenarios are to be close to what we take the actual world to be like: function C_α is introduced to do this job. On the other hand, oriented selectivity has to do with the fact that the selected scenarios are to be consistent with a specific issue, or question addressed in the exercise: this is captured by the selection of the topic operated by C_α^τ.

Our language also allows us to express several features of imagination acts, as:

- which input and topic are actually selected:

  $\langle im_\alpha \rangle \mathbf{var}(\alpha)$ says that an act based on input α is entertained;

  $\langle im_\alpha^\tau \rangle \mathbf{var}(\alpha)$ says that input α is processed in light of topic τ.

- which inputs and topics the agent had the option to select:

  $\neg \langle im_\alpha \rangle \mathbf{var}(\alpha) \wedge \Diamond \langle im_\alpha \rangle \mathbf{var}(\alpha)$ says that an act based on input α could have been entertained;

  $\neg \langle im_\alpha^\tau \rangle \mathbf{var}(\alpha) \wedge \langle ag \rangle \langle im_\alpha^\tau \rangle \mathbf{var}(\alpha)$ says that input α could have been processed in light of topic τ.

- that an agent voluntarily entertains an imagination act based on certain input and topic:

  $\langle im_\alpha^\tau \rangle \mathbf{var}(\alpha) \wedge \langle ag \rangle \neg \langle im_\alpha^\tau \rangle \mathbf{var}(\alpha)$ says that the agent entertains an imagination act based on input α and processed in light of topic τ, but could have chosen not to select that topic.

- that an agent voluntarily entertains an imagination act based on certain input, topic, and traits specification:

  $\langle im_\alpha^\tau \rangle \mathbf{var}(\alpha) \wedge [im_\alpha^\tau]\varphi \wedge \langle ag \rangle (\langle im_\alpha^\tau \rangle \mathbf{var}(\alpha) \wedge \langle im_\alpha^\tau \rangle \neg\varphi)$ says that the agent entertains an imagination act based on input α, processed in light of topic τ, and a traits specification that necessitates φ, but could have chosen a specification not necessitating φ.

Finally, if we assume that the set of topics is finite, we can use the formula

$$\langle im_\alpha \rangle \mathbf{var}(\alpha) \wedge \bigwedge_{\tau \in Var_T} [im_\alpha^\tau] \neg \mathbf{var}(\alpha)$$

to say that the agent is entertaining an imagination act based on input α, but chooses not to process it in light of any topic.

Case 3. *The possibility of learning from imagination.*

The final case concerns a second aspect of the relation between imagination and knowledge: how we are able to learn through imagination, via a suitable combination of voluntary and involuntary acts in virtue of which possible scenarios are increasingly reduced. In Williamson (2016), p. 116, the following situation is discussed:

> When the hunter makes himself imagine trying to jump the stream, his imagination operates in voluntary mode. But he neither makes himself imagine succeeding nor makes himself imagine failing. Rather, having forced the initial conditions, he lets the rest of the imaginative exercise unfold without further interference. For that remainder, his imagination operates in involuntary mode. He imagines the antecedent of the conditional voluntarily, the consequent involuntarily. Left to itself, the imagination develops the scenario in a reality-oriented way, by default.

In our framework we are able to represent that through propositions like:

Learning 1: $\neg[im_\alpha]\varphi \wedge [im_\alpha^\tau]\varphi$
Learning 2: $\neg[im_\alpha]\varphi \wedge [im_\alpha^\tau]\varphi \wedge [ag]([im_\alpha^\tau]\varphi \vee [im_\alpha^\tau]\neg\varphi)$.

The first conjunct of the previous formulas states that φ is not necessarily imagined given input α, i.e., that the truth value of φ can vary based on the choice of a topic. The agent is free to opt for a significant topic and a specific point of view on that topic, i.e., one or more cells in the corresponding partition. Still, *once* that point of view is selected, the consequences of the selection, that is, all the propositions that come out to be true given the selection, are no longer in control of the agent: the antecedent and the cell in the partition corresponding to the topic can be voluntarily chosen, but the consequent, what is implied by that cell, is involuntarily settled. The second conjunct of the previous formulas states that φ is necessarily imagined given the selection of the cells in the partition corresponding to the topic. The final conjunct in the second formula states that the truth value of φ is determined by the cells of the partition corresponding to topic τ, i.e., that φ is on topic (Lewis, 1988). Hence, *Learning 1* captures the idea that φ can be learned, or discovered, in virtue of the selected topic specification, while *Learning 2* captures the idea that the learned proposition is also on that very topic.[8]

---

[8] Strictly speaking, both *Learning 1* and *Learning 2* only represent what we *could learn* when we entertain an imaginative act. In fact, in order to represent what we *actually learn*, we should include, beside a proposition stating that φ holds at all the scenarios whose selection is based on input α with topic τ, a proposition stating that φ holds in the α-worlds that are closest to the actual world. In other terms, we should supplement the epistemic conditional with a corresponding ontic conditional, thus grounding the truthfulness of our epistemic act.

§7. Conclusion.   We have claimed that imagination as ROMS has a number of features: it's agentive and episodic; it starts from a deliberate input, an initial supposition concerning what is to be mentally simulated; it integrates such input on the basis of the agent's background knowledge and beliefs, without mobilizing all of them—rather, only those deemed relevant to the topic of the imaginative act; and it has a purpose or goal, which is to address some question, the answer to which drives the agent's interests. We have argued that some of these components of an exercise of ROMS (e.g., picking an initial input, choosing the goal) involve voluntary choices on the side of the agent, while others are involuntary and automatic (e.g., retrieving information from one's background knowledge and beliefs to integrate the input). We have then presented a formal language with a sound and complete logic of Voluntary Imagination (**VI**), involving modal operators expressing agency, imagination given an input, a topic, and indiscernibility with respect to input and topic. The logic combines techniques of epistemic and action logics with a mereology of topics from aboutness or subject matter semantics.

VI allows us to express and address issues concerning (what we claim to be) three main, interconnected questions concerning imagination as ROMS: (1) *What is its logic?* We have shown that the imagination operators of **VI** have some noteworthy closure properties, in spite of their being non-normal and nonmonotonic modals (in particular, given an input, or an input and a topic, one doesn't imagine all the logical consequences of the input itself). (2) *How does imagination as ROMS relate to knowledge?* We have shown how one can express in **VI** the conditions under which an agent can learn something new via an act of ROMS; this relates rather strictly to question (3), *What is voluntary and what isn't in ROMS?* **VI** can express the distinction between voluntary and involuntary components, and thus help to make formally precise the idea, entertained by various authors, that imagination can allow us to gain new knowledge because some aspects of mental simulation are not arbitrary, but governed by the automatic and generally reliable mechanisms that govern the administration and revision of our beliefs in the light of new information.

One main direction of further investigation within the proposed framework, flagged since §2, involves the temporal dynamics of ROMS: how episodes of mental simulation develop temporally, while representing actions and events that themselves unfold in time. We have left this issue for future work, while conjecturing that the techniques of Dynamic Epistemic Logic may help with it.[9]

BIBLIOGRAPHY

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

Baltag, A., Moss, L. S., and Solecki, S. (1998). The logic of public announcements, common knowledge, and private suspicions. In Gilboa, I., editor. *Proceedings of TARK 98*. Evanston, IL: Morgan and Kaufmann, pp. 43–56.

Baltag, A., and Smets, S. (2008). A qualitative theory of dynamic interactive belief revision. In Bonanno, G., van der Hoek, W., and Wooldridge, M., editors. *Logic and the Foundations of Game and Decision Theory*. Amsterdam: Amsterdam University Press, pp. 9–58.

Belnap, N., Perloff, M., and Xu, M. (2001). *Facing the Future: Agents and Choices in Our Indeterministic World*. Oxford: Oxford University Press.

Berto, F. (2018a). Aboutness in imagination. *Philosophical Studies*, **175**, 1871–1886.

Berto, F. (2018b). Taming the runabout imagination ticket. *Synthese*, online first, doi: 10.1007/s11229–018–1751–6

Byrne, R. (2005). *The Rational Imagination. How People Create Alternatives to Reality*. Cambridge, MA: MIT Press.

Byrne, R., and Girotto, V. (2009). Cognitive processes in counterfactual thinking. In Markman, K. D., Klein, W., and Suhr, J. A., editors. *Handbook of Imagination and Mental Simulation*. New York: Taylor and Francis, pp. 151–160.

Chalmers, D. (2002). Does conceivability entail possibility? In Gendler, T. S., and Hawthorne, J., editors. *Conceivability and Possibility*. Oxford: Oxford University Press, pp. 145–199.

Currie, G. (1990). *The Nature of Fiction*. Cambridge: Cambridge University Press.

Evans, J., and Over, D. (2004). *If*. Oxford: Oxford University Press.

Fine, K. (2016). Angellic content. *Journal of Philosophical Logic*, **45**(2), 199–226.

Fraude-Koivisto, T. S., Wuerz, D., and Gollwitzer, P. M. (2009). Implementation intentions. In Markman, K. D., Klein, W., and Suhr, J. A., editors. *Handbook of Imagination and Mental Simulation*. New York: Taylor and Francis, pp. 69–87.

Gendler, T. S. (2000). The puzzle of imaginative resistance. *Journal of Philosophy*, **97**, 55–81.

Giordani, A. (2018). Ability and responsibility in general action logic. In Broersen, J. M., Condoravdi, C., Shyam, N., and Pigozzi, G., editors. *Deontic Logic and Normative Systems – 14th International Conference (DEON 2018)*. London: College Publications, pp. 121–138.

Giordani, A. (2019). Axiomatizing the logic of iagination. *Studia Logica*, **107**, 639–657.

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.

Horty, J. F. (2001). *Agency and Deontic Logic*. Oxford and New York: Oxford University Press.

Kahneman, D., and Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, **39**, 341–350.

Kind, A. (2016). Imagining under constraints. In Kind, A., and Kung, P., editors. *Knowledge Through Imagination*. Oxford: Oxford University Press, pp. 145–159.

Kind, A., and Kung, P. (2016). *Knowledge through Imagination*. Oxford: Oxford University Press.

Kosslyn, S., and Moulton, S. T. (2009). Mental imagery and implicit memory. In Markman, K. D., Klein, W., and Suhr, J. A., editors. *Handbook of Imagination and Mental Simulation*. New York: Taylor and Francis, pp. 35–52.

Langland-Hassan, P. (2016). On choosing what to imagine. In Kind, A., and Kung, P., editors. *Knowledge Through Imagination*. Oxford: Oxford University Press, pp. 61–84.

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

Lewis, D. (1988). Relevant implication. *Theoria*, **54**(3), 161–174.

Markman, K. D., Klein, W., and Surh, J. A. (2009). *Handbook of Imagination and Mental Simulation*. New York: Taylor and Francis.

Nichols, S., and Stich, S. P. (2003). *Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.

Oaksford, M., and Chater, N. (2010). *Cognition and Conditionals*. Oxford: Oxford University Press.

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

Stanovich, K. E., and West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, **23**(5), 645–665.

van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logic*, **17**, 129–155.

Van Leeuwen, N. (2016). The imaginative agent. In Kind, A., and Kung, P., editors. *Knowledge Through Imagination*. Oxford: Oxford University Press, pp. 85–111.

Walton, K. (1990). *Mimesis as Make-Believe: On the Foundations of Representational Arts*. Cambridge, MA: Harvard University Press.

Wansing, H. (2017). Remarks on the logic of imagination. A step towards understanding doxastic control through imagination. *Synthese*, **194**(8), 2843–2861.

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

Williamson, T. (2016). Knowing by imagining. In Kind, A., and Kung, P., editors. *Knowledge Through Imagination*. Oxford: Oxford University Press, pp. 113–123.

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

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

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION
UNIVERSITY OF AMSTERDAM
1098 XG AMSTERDAM, THE NETHERLANDS

E-mail: i.canavotto@uva.nl

ARCHÉ
UNIVERSITY OF ST. ANDREWS
ST. ANDREWS KY16 9AJ, UNITED KINGDOM
and
INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION
UNIVERSITY OF AMSTERDAM
1098 XG AMSTERDAM, THE NETHERLANDS

E-mail: fb96@st-andrews.ac.uk

DEPARTMENT OF PHILOSOPHY
CATHOLIC UNIVERSITY OF MILAN
LARGO AGOSTINO GEMELLI 1, 20123 MILANO, ITALY

E-mail: alessandro.giordani@unicatt.it