6 Equivalence in Imagination

Francesco Berto

6.1 Imagination as ROMS

In their introduction to the beautiful *Knowledge Through Imagination*, Amy Kind and Peter Kung state what they call 'the puzzle of imaginative use' (Kind and Kung 2016, 1): imagination seems to be arbitrary escape from reality; how can it give knowledge of reality? The short version of their answer, shared by a number of authors in that collection (Kind (2016), Langland-Hassan (2016), Williamson (2016)), is: imagination, of the kind that can give us knowledge, is *constrained*. It deviates from reality, but in a regimented way. This chapter is about logical regimentation. Our question will be: given that one imagines that *A*, what *B*s is one to imagine, as a matter of logical necessity?

As 'imagining' is highly ambiguous (daydreaming, hallucinating, mental wandering, entertaining), we need to pin down the sort of imagination whose logic we are after. For our question to make sense, it must be propositional: one imagines that one jumps a stream, that John is tall and thin, that there are functional zombies devoid of phenomenal consciousness, that Stauffenberg puts the bomb on the other side of the table.

A number of authors (e.g. Currie and Ravenscroft (2002), Goldman (2006)) agree on a simulationist account. Imagination in general re-creates counterpart non-imaginative mental states. In perceptual imagination, one simulates perception: one hears the riff of *Smoke on the Water* in one's head and it's relevantly like the real thing, except that one's auditive apparatus is left offline. Propositional imagination, the kind thereof which seems to have a chance, if any does, at helping with knowledge, is the one in which one simulates belief (Arcangeli 2019 calls it 'cognitive imagination'). On the other hand, imagination is often contrasted with belief: the former is voluntary in a way the latter isn't, at least according to most authors (Dorsch (2012), Gendler (2000), Mulligan (1999), Walton (1990)): one can imagine that all of Edinburgh has been painted yellow but, having overwhelming evidence of Edinburgh's greyness, one cannot make oneself believe it. What's the point of a simulation that differs from what it simulates in such a core feature?

The link between imagination as mental simulation and belief is, I think, that in mental simulation we imagine in order to assess what will happen *if* something is the case, or what would have happened *if* something had been the case. What if Stauffenberg had put the bomb on the other side of the table? How would the course of WWII have changed? What if I jump the stream? Will I make it to the other side, or will I fall in the water and drawn?

The kind of belief mental simulation typically connects to for epistemic purposes, then, must be *conditional* belief (which may be quite different from belief in the corresponding conditional: more on this below). I don't really believe, say, that Brexit will cause a recession. I explore the consequences: what if Brexit causes a recession? If I conclude, given the supposition that Brexit causes a recession and what I know or believe, that I will lose my job in the hypothetical scenario, I form the relevant conditional belief, representing a subjective conditional probability. This gives us guidance on the kind of logical form we should be after, when we ask about how logical closure works for imagination as mental simulation. The form must be related to conditionals – 'Given input A, one imagines that B'; or, less tersely: 'In an exercise of mental simulation starting from suppositional input A, one imagines that B'.

This obviously relates to the Ramsey test for conditionals: one evaluates a conditional 'If A, then B' by supposing the antecedent, A, minimally adjusting one's belief or knowledge system, and seeing whether the consequent B turns out in the imagined scenario. The difference between supposing that A in an exercise of mental simulation, and supposing that one believes that A, is big. And it's the first one that goes on, in general, in suppositional thinking, as per the Ramsey test. (Of course, there will be special cases in which one supposes that one believes something.) Supposing I am prime minister of the UK, I may conclude, e.g., that I will hire a bunch of friends in key governmental roles. Supposing I believe I am prime minister of the UK, I may conclude very different things, e.g., that I need to be hospitalized for I am seriously deluded.

The Ramsey test ties our evaluation of conditionals to our updating our priors in the light of new information. It's around noon and one sees that John has moved into the kitchen. One then minimally revises one's beliefs compatibly with the news. In the updated belief system, one will have dropped, say, the previously held belief that John was in the living room; on the other hand, that John is cooking will now look likely enough for one to come to believe it. What is typical of mental simulation is that one doesn't really get the news online, e.g., by seeing that John is in the kitchen; one just pretends, in an offline mode (Williamson (2016)), that John is there, and checks what would follow. If in the hypothetical situation it turns out that John is cooking, one does not believe that John is cooking; but one can acquire a conditional belief: one believes, so to speak, that John is cooking conditional on John being in the kitchen.

(Does one come to believe the corresponding *conditional*, too, 'If John is in the kitchen, then he's cooking'? We know from Lewis' (1976) triviality results that, however tight the connection between conditional beliefs and beliefs in the corresponding conditionals, it won't be identity. Philosophers and psychologists have tried to circumvent the problem: some keep the core of the Ramsey test by stating that conditionals don't express propositions and cannot generally be embedded, e.g., Edgington (1995), Bennett (2003). Psychologists often content themselves with a robust, empirically corroborated correspondence, short of identity, e.g., Evans and Over (2004), Oaksford and Chater (2010). On this whole issue, see the masterful Douven (2016).)

As stressed e.g. by Williamson (2016), Langland-Hassan (2016), Canavotto et al. (2020), not everything is voluntary in such an exercise. The offline suppositional input is: we are free to suppose whatever *A* we like (Arcangeli (2019) makes a forceful case). Once the supposition is in, however, what *B*s will come out true in the imagined scenario is largely not up to us: it has to do with how our prior belief system is disposed to adjust itself conditional on *A*, and how this works is largely involuntary, as beliefs are.

As many have acknowledged (e.g. Byrne (2005), Kind (2016), Williamson (2016)), the adjustment is governed by principles of relevance and minimal alteration. One who mentally simulates jumping the stream to predict whether if one jumped, one would make it won't imagine growing wings on one's back allowing one to fly to the other side. One keeps reality, as one knows or believes it to be, unaltered as much as possible, compatibly with one's jumping. Imagination is *reality-oriented*. And so we have a label for what we're after: the logic of *Reality-Oriented Mental Simulation* (ROMS). The epistemological importance of such an activity is hardly over-estimated; if we use ROMS to gain knowledge of metaphysical modality (Williamson (2016)), there are good arguments to the effect that we use it to handle modalities of much more mundane kinds, too (Strohminger and Yli-Vakkuri (2019)). And so we may want logics to reason about ROMS.

6.2 The Logic of ROMS

In a couple of papers on ROMS (Berto (2017, 2018)), I have proposed to use the simple notation '[*A*]*B*' for 'Given input *A*, one imagines that *B*', or 'In an exercise of mental simulation starting from suppositional input *A*, one imagines that *B*'. Logical tradition suggests that it be interpreted as some modality. We have learned since Hintikka (1962) that we can treat intentional notions like *knows*, *believes*, *is informed that*, using modal logic. We take them as normal modals: quantifiers over possible worlds, restricted by an epistemic accessibility relation. What one knows is what holds at all worlds one looks at, which represent open epistemic possibilities.

The Hintikkan approach notoriously gives logical omniscience: if *X* is the relevant mental state, given that the agent *X*s that *A*, the agent *X*s any *B* that follows logically from *A* (*A* is true at all worlds one looks at; they all make true *B* as well: one *X*s that *B*). As a special case, one *X*s all logical truths. Now we ordinary humans are not like that: we sometimes fail to see even obvious logical consequences of what we know or believe, because we are tired, cognitively impaired, or otherwise busy. For various *X*s, logical omniscience is implausible even for perfect reasoners. It has been questioned for knowledge, even for ideal knowers, on the grounds of sceptical paradoxes (Hawke (2016), Nozick (1981)) and the presence of defeaters (Brown (2018)). While these stances are controversial (see e.g. Hawthorne (2004) for criticisms), other propositional attitudes *obviously* fail to be closed under logical consequence: one desires that one's headache goes away, and this implies that one has a headache; but one doesn't desire that one has a headache, even as a perfect reasoner.

What about imagination as ROMS? Even if it's not subject to full logical closure, it should be closed under some logical operations. Try to imagine that John is tall and thin without imagining at the same time that he's thin and tall, and without imagining that he's tall; or that Mary is an expert chess player, without imagining that she's a chess player. ROMS must have some conjunctive structure: one can't imagine that $A \land B$ without simultaneously imagining that $B \land A$, and without imagining that *A*. Imagining that John is tall and thin without imagining that he is tall would be a bit like imagining that John is tall without imagining that John is tall, wouldn't it? And it might also be that one cannot imagine that *A* and that *B*, without imagining that $A \land B$. You can't imagine that John is tall and that John is thin without imagining that John is tall and thin, can you? (This may sound more controversial; I'll get back to it.)

Besides, the aforementioned reality-oriented nature of ROMS must induce plausibility and relevance constraints for our logic. When the suppositional input is that Stauffenberg puts the bomb on the other side of the table, we don't imagine that Hitler suddenly grows an armour protecting him from the explosion: this would be an implausible departure from reality. We imagine historical reality largely as we know or believe it to be, save that Stauffenberg puts the bomb on the other side. As for relevance, in ROMS we focus on what is on-topic given an input: when the input is that Stauffenberg puts the bomb on the other side of the table and we conclude that, then, Hitler gets killed, we don't imagine that either Hitler gets killed or there's life on Kepler-442b, although *A or B* is a logical consequence of *A*. And we don't imagine each irrelevant logical or necessary truth or consequence given that suppositional input, although logical truths are implied, in classical logic, by any input, and necessary truths hold at all possible worlds, thus at all worlds, where the input is true.

To comply with the plausibility requirement, we should be looking at what goes on at worlds where the suppositional input, *A*, is true, and which are, for the rest, the most plausible worlds given one's knowledge or beliefs. Readers familiar with conditional logics will hear the workings of the Lewis (1973) semantics for counterfactuals. One evaluates 'If it were/had been the case that *A*, then it would be/have been the case that *B*' by looking at the closest worlds where *A* is true. Similarly, our [*A*] will be a sententially indexed universal modality (Chellas (1975)), a restricted universal quantifier over worlds, where the restriction is determined by the *A* at issue: the suppositional input.

Relevance tells us that [*A*] cannot be a normal modality, i.e. *just* a restricted quantifier over possible worlds. Any possible world where *B* is true will also be one where *B* ∨ *C* is, but we don't want our agents to automatically imagine the irrelevant *either Hitler gets killed or there's life on Kepler-442b* because they imagine that Hitler gets killed. Nor do we want our agents busy in working out what would have been the case, had Stauffenberg put the bomb on the other side of the table, to imagine unrelated necessary truths, e.g., that either Melbourne is in Sweden or not, or that Fermat's Last Theorem is true.

In Berto (2017, 2018) I have explored two semantic approaches that make of the [*A*]-operators non-normal modals: one – *Semantics 1*, let's say – sticks to the traditional possible worlds apparatus but adds a topicality constraint that filters out irrelevant logical or necessary truths and consequences. Another one, *Semantics 2*, drops 'possible' and takes [*A*] as a restricted quantifier over worlds, some of which may be non-normal or impossible: worlds where logical or necessary truths may fail (Berto and Jago (2019)). One reason for considering both here is that they can embed an idea which is, finally, the main topic of this chapter – and that they do it in slightly but interestingly different ways. This is the idea of equivalence in imagination.

6.3 Equivalence in Imagination

The logically untamed nature of intentional states shows up in the fact that the operators expressing them seem to be hyperintensional: they can fail substitution *salva veritate* of intensional (necessary or logical) equivalents in their scope. One *X*s (knows, believes, hopes, etc.) that *A*, but not that *B*, although *A* and *B* are such equivalents – in standard possible worlds semantics: true at the same possible worlds (of all the relevant models). This may happen because one is computationally limited (*A* is an obvious tautology, say $p \rightarrow p$; *B* is a long and complicated logical truth), or because one is not on top of some concept (*A* is equivalent to $A \vee (A \wedge B)$ in classical and many non-classical logics; but one has an attitude only towards the former, for one lacks *B*-involved concepts); or because of the defeasibility of knowledge; etc.

ROMS is no exception. Sometimes [*A*]*C* will be true while [*B*]*C* won't, for equivalent *A* and *B*: one may be subject to framing effects (Kahneman et al. (1982)), whereby one imagines different things starting from differently presented, necessarily equivalent suppositions ('If you apply for the job, you have 60% chances of making it' vs '… 40% chances of failing'). One may imagine something given *A*, nothing given an equivalent *B*, for one lacks some *B*-involved concept but no *A*-involved concept. Importantly, not *all* aspects of logical de-idealization are represented by the frameworks below. The suppositional agents are conceptually non-idealized: they may lack some concept needed to grasp a proposition; but they are syntactically idealized, in that they cannot fail to parse the syntax of certain sentences in the correct way. For example, whenever, given suppositional input *A*, they imagine that $B \wedge C$, they also imagine that $C \wedge B$ when '$\wedge$' is Boolean, order-insensitive conjunction. We know that real humans are, instead, susceptible to order-of-presentation effects in their attitudes.

The anarchy of imagination may be heuristically fruitful (see Stuart (2020)). But an operator failing substitution of intensional equivalents is likely to have a rather weak logic. When we find operators of the form 'Given *A*, *C*' (e.g. in probability logic – '*C* is likely given *A*': Adams (1998); in dynamic belief revision – 'After revision of one's beliefs by *A*, *C* is the case': van Benthem (2011); in conditional logic: 'If *A* is/was/ had been the case, then *C* is/would be/would have been the case'), they nearly always satisfy the principle whereby 'Given *A*, *C*' entails 'Given *B*, *C*' when *A* and *B* are intensional equivalents. This is so even when such operators are non-monotonic, i.e. 'Given *A*, *C*' does not entail 'Given $A \wedge D$, *C*' (adding information in the antecedent or premise is not guaranteed to preserve the consequent or conclusion: given that Tweety is a bird, it flies. Given that it's a penguin bird, it doesn't). For example, we find the principle that 'Given *A*, *C*' entails 'Given *B*, *C*' when *A* and *B* are intensional equivalents, under the label of Left Logical Equivalence, in the mainstream system P, often taken as a set of minimal principles any non-monotonic logic should obey (Kraus et al. (1990)).

However, ROMS may satisfy some *restricted* substitutivity principle. This is the idea of equivalence in imagination: although it is not the case that, given two intensionally equivalent suppositional inputs, *A* and *B*, one will imagine the same things, this will happen when *A* and *B* are at least equivalent in one's imagination.

What is equivalence in imagination? We can understand it as cognitive equivalence or synonymy, which is different from synonymy *tout-court*. *A* and *B* are cognitive equivalents or cognitive synonyms when they play the same role in one's cognitive life (Hornischer (2017)): whatever one concludes – deductively, abductively, inductively – supposing either, one does, supposing the other. Whatever one understands when either is uttered, one does, when the other is. One cannot take either as

true, without taking the other as true. Whatever one thinks about given either, one does, given the other.

Cognitive synonymy is a respectable notion in linguistics. Absolute synonymy, understood as substitutivity *salva veritate* in all contexts, has notoriously raised eyebrows among philosophers. Linguists sometimes take it as a purely theoretical, limit notion, for it cannot be empirically tested (Cruse (2000), Stanojecvić (2009)). However, cognitive equivalence or synonymy is the working concept for a number of works in linguistics (Lyons (1996), Murphy (2003)).

Cognitive equivalence should be relative to the thinker's available knowledge base and storage of concepts – thus, what's in this or that specific agent's knowledge base, and thus what is cognitively equivalent for this or that agent, is not for a general logic of suppositional thinking to settle. *John is a bachelor* and *John is an unmarried man* are equivalent for nearly any competent speaker of English: if one takes either as true but the other as false, this is likely to generate suspicion on the level of English proficiency of the speaker. *Ex falso is classically valid* and *Pseudo-Scotus' Law is classically valid* are equivalent for most logicians. *Groundhogs are rodent* and *Woodchucks are rodent* for most zoologists. One way to represent such cognitive interchangeability in AI is via pairs of defeasible, non-monotonic conditionals, 'If *A*, then *B*', 'If *B* then *A*', stored in the agent's knowledge base, e.g., in Logic Programming (Stenning and van Lambalgen (2008)). Our formalism will be in the same ballpark.

6.4 Getting Formal

We use a simple propositional language L with an indefinitely large set L_{AT} of atoms p, q, r ($p_1, p_2,...$), negation $\neg$, conjunction $\wedge$, disjunction $\vee$, a strict conditional $\prec$, square brackets [,] for our imaginative modals, and round parentheses (,) as auxiliary symbols. The well-formed formulas of L are the items in L_{AT}, and if A and B are formulas:

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

(We normally omit outermost brackets.) Read '$[A]B$' as 'In an act of ROMS with suppositional input A, one imagines that B', or 'Given input A, one imagines B', or so.

6.4.1 Semantics 1

Semantics 1 starts with a frame F = (W, $\{R_A | A \in L\}$, T, +, t) with five items:

1 W is the usual set of possible worlds;
2 $\{R_A | A \in L\}$ is a set of binary accessibilities between worlds: each formula A of L has its own $R_A \subseteq W \times W$;

3   T is a set of possible *topics* or *subject matters*: what the formulas of
    L can be *about* (more on these in a minute);

4   + is an idempotent, commutative, associative operation on T: un-
    restricted *topic fusion*; it takes two topics and returns their sum or
    fusion; *topic parthood* can be defined from it the usual way: topic x
    is part of topic y ($x \leq y$, a partial ordering) when $x + y = y$, the fusion
    of x and y is just y.

5   is a function assigning topics to all formulas of L via the following
    recursion: it first assigns topics to all items in L_{AT}; then if At(A) =
    $\{p_1,..., p_n\}$ is the set of atoms in A, t(A) is the fusion over the atoms
    of A, +At(A) = t(p_1) +... + t(p_n); intuitively: what a formula is about
    is what its atoms, taken together, are about.

Items (1) and (2) need little comment. Items (3)–(5) do: what are topics?
Why should they allow fusion into further topics, and stand in parthood
relations? Why are topics assigned to formulas in that way?

   Topics or subject matters are what meaningful items are *about* (Fine
(2016), Humberstone (2008), Yablo (2014)). Aboutness is usually treated
as a feature of linguistic representations, but of course mental representa-
tions bear aboutness, too: we call it intentionality. When one thinks that
John is tall, one is thinking about *John's height*. Topics are sometimes
linked to questions. Lewis (1988) took them as partitions of the set of
worlds: when the topic is *the number of stars*, the associated question is,
'How many stars are there?'. Worlds end up in the same cell when they
agree on the answer: all zero-star worlds, all one-star worlds, etc. Others
(Fine (2016)) take topics as sets or fusions of truthmakers understood as
situations à la Barwise and Perry (1983). We don't need to take a stance
on the nature of topics. For our logical purposes, we only need them to
obey three constraints, widely agreed upon in subject matter semantics:

i   Intensionally equivalent sentences *A* and *B* can differ in content
    when they are about different things. In Yablo's (2014)'s version, the
    content of a sentence (in context) is not specified just by the set of
    worlds in which it is true, but also by what it's about. Subject matter
    semantics is, thus, hyperintensional: '2 + 2 = 4' and 'Either Jupiter
    is a planet, or not' differ in content in spite of being true at the same
    worlds (all of them), for they say different things: only one is about
    the number 2.

ii The space of topics has a mereological structure (Fine (2016), Yablo
    (2014)), hence our talk of fusions and parthood in (4) above. Top-
    ics can have proper parts; distinct topics may have common parts.
    Mathematics includes *arithmetic*. *Mathematics* and *philosophy*
    overlap, having (certain parts of) logic as a common part. The topic
    of $A \land B$ includes that of A as a (proper) part; if Mary says 'John is
    tall and thin' and Ann says 'John is tall', what Ann said has already

been said by Mary – who also said more: Ann has addressed a topic, say *John's height*, which is a proper part of the larger topic addressed by Mary, *the height and looks of John*.

iii   The Boolean operators should add no subject matter of their own: they are 'topic-transparent' (Fine (2016), Perry (1989), Yablo (2014)). The topic of ¬*A* is the same as that of *A* ('John isn't tall' is exactly about what 'John is tall' is about: *John's height*; it certainly is not about *not*). Conjunction and disjunction merge topics ('John is tall and thin' and 'John is tall or thin' are about the same topic: *the height and looks of John*). In our L above, this reduces the topic of a formula to the fusion of the topics of its atoms. Hence our recursion in (5) above.

A model M = (W, {R_A|*A*∈L}, T, +, t, v) is a frame with a valuation v, assigning to each atom *p* in L$_{AT}$ a truth set v(*p*)⊆W, the set of worlds where *p* is true. This is extended to the whole L as follows. Read '*w* ⊩ *A*' as saying that *A* is true at world *w*; we omit reference to the model as this is understood:

(Sat) *w* ⊩ *p* iff *w* ∈ v(*p*)
(S¬) *w* ⊩ ¬*A* iff *w* ⊮ *A*
(S∧) *w* ⊩ *A* ∧ *B* iff *w* ⊩ *A* and *w* ⊩ *B*
(S∨) *w* ⊩ *A* ∨ *B* iff *w* ⊩ *A* or *w* ⊩ *B*
(S≺) *w* ⊩ *A* ≺ *B* iff for all w_1: if w_1 ⊩ *A*, then w_1 ⊩ *B*
(S[*A*]) *w* ⊩ [*A*]*B* iff (i) for all w_1: if wR_Aw_1, then w_1 ⊩ *B* and (ii) t(*B*) ≤ t(*A*)

Negation, conjunction, and disjunction work as usual and the fishhook is a strict conditional. As for the last item: read 'wR_Aw_1' as saying that w_1 is one of the worlds one looks at, in an act of ROMS (carried out at *w*) whose initial suppositional input is *A*.

    (*What* worlds does one look at? The intuition behind the minimal alteration principle would be that these are the most plausible worlds given one's belief or knowledge state. This, however, is not really represented in the semantics. To do it, one can impose a total ordering of worlds by plausibility, mimicking what is done in the Lewis (1973) sphere semantics and in various epistemic logics for belief revision – see e.g. van Benthem (2011). I haven't done it here, or in Berto (2017, 2018), for Semantics 1 or 2. But I have done it, for Semantics 1, in Berto (2019). One reason not to do it in the present context is that a semantics with the plausibility ordering automatically satisfies the condition representing equivalence in imagination, which I will present below. I want to discuss, instead, the opportunity of adding it manually.)

    One can equivalently formulate (S[*A*]) using set-selection functions à la Lewis (1973): each formula *A* of L has its own function, f_A, taking as input the world where the act of mental simulation starting from *A* takes

place, and outputting the corresponding bunch of accessible worlds, $f_A(w) = \{w_1 \in W | w R_A w_1\}$. If $|A| = \{w \in W | w \Vdash A\}$, the truth set for A, one can rephrase the clause for $[A]$ as the more compact:

(S[A]) $w \in [A]B$ iff (i) $f_A(w) \subseteq |B|$ and (ii) $t(B) \leq t(A)$

The two formulations are equivalent, as $w R_A w_1$ iff $w_1 \in f_A(w)$. But either is at times handier than the other to talk about the semantics.

For $[A]B$ to be true, we ask for two things: (i) is the truth-conditional component making of $[A]$ a sententially indexed modal: B must be true throughout the worlds one looks at, given input A. (ii) is the topicality component: B must be fully on-topic with respect to what A is about. One natural Basic Constraint on the semantics is that, for all $A \in L$ and $w \in W$:

(BC) $f_A(w) \subseteq |A|$

This says that all the worlds one looks at, when one starts by supposing that A, will be worlds where A is true (plausibly enough: when we wonder what happens if Brexit causes a recession, we consider situations where Brexit *causes* a recession to begin with). From now on, we will only consider models satisfying BC.

We define logical consequence as truth preservation at all worlds of all models: where S is a set of formulas, $S \Vdash A$ when in all models $M = (W, \{R_A | A \in L\}, T, +, t, v)$ and for all worlds w in W if $w \Vdash B$ for all B in S, then $w \Vdash A$. As a special case, A is a logical truth, $\Vdash A$, when true at all worlds of all models.

Our first logical validity is secured via BC (for proofs of all the validities and invalidities listed from now on for Semantics 1, see Berto (2018)):

$\Vdash [A]A$

The suppositional input is always imagined. The next validities, taken together, characterize ROMS as 'fully conjunctive':

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

The insight is that imagined scenarios have some sort of mereological structure. Simplification has it that one who imagines the whole of a situation imagines the parts. You can't imagine that John is tall and thin without imagining that John is tall because by imagining the former – the whole – you have *already* imagined the latter – the part. Adjunction has it that one who imagines the parts imagines the whole. You can't

imagine that John is tall and John is thin without imagining that John is tall and thin because, by imagining the parts all together, you have *already* imagined the whole.

One may have more doubts about Adjunction than about Simplification. I discussed the point in Berto (2017, 2018), resorting to an old chestnut by Quine (1960) concerning same-antecedent counterfactuals allowing for conjunction of their consequents. The suppositional input in [A] has Caesar being in command of the US troops in the Korean war. We can imagine him using atomic bombs, B, if in our ROMS we keep fixed the weapons available in the Korean war, or we can imagine him using catapults, C, if we keep fixed the military apparatus available to Caesar. One would not thereby infer [A](B ∧ C), that Caesar employs both bombs and catapults. One can imagine that, too, if one wills, but it will spoil the reality-oriented nature of the exercise. It should not come as an automatic entailment mandated by the logic of mental simulation.

In those works, I tried to argue that Adjunction should be maintained, and that doubts on it may have to do with the formalism not capturing the contextual character and temporal development of ROMS episodes. These are indexed only to formulas (the suppositional inputs), but it seems clear that the same input can trigger different imaginings in different contexts, and at different times, as it unfolds. The Quinean example embeds a shift. Once one sticks to a single context and time, Adjunction will work: you can't imagine *in one go*, that is, in the same act of ROMS and at the same time, that John is tall and that John is thin without imagining that John is tall and thin, can you?

Another line of argument may start by assuming that imagination (ROMS included) essentially involves mental imagery (see Kind (2001)). Unlike propositional or language-like mental representations, pictorial mental representations, also called mental imagery, have quasi-spatial features (Paivio (1986)). This may point at some mereological structure represented in the mind: when we visually imagine our bedroom, we can zoom into one part of the scenario – where the bedside table is – then zoom into one further sub-part – the book on the bedside table – then move upwards, etc. Classic empirical work in psychology (Block (1983), Shephard and Metzler (1971)) showed that the time taken to scan between two points of a mental image is generally proportional to their subjective distance; that larger objects fill the imagined scenario sooner than smaller ones; etc. Maybe the conjunctive nature of imagination is supported by such considerations on the intuitive mereology of mental imagery.

It is, however, controversial (Gregory (2016), Van Leeuwen (2013), Williamson (2016)) that imagination essentially involves mental imagery, even if one admits that some mental representations represent pictorially, which is itself a controversial claim (see the so-called imagery debate, e.g., Pylyshyn (2002)). In mental simulation, we sometimes imagine

scenarios involving only abstract objects (think of a mathematician mentally going through a proof, without making use of quasi-spatial structures like graphs or so); and sometimes we imagine complex situations whose representation needn't involve mental imagery (think of yourself trying to predict how the markets worldwide will react to the next economic downturn). More logical reasons for dismissing such conjunctive hunches will come from considerations linked to equivalence in imagination, to which we get below.

Lots of things come out *in*valid in the semantics, making the logic of our imaginative operators correspondingly weak – and rightly so. Four examples. First, their variable strictness models a key feature of ROMS: imagination is non-monotonic.

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

Supposing Tweety is a bird, you imagine that it can fly. Supposing Tweety is a penguin bird, you don't imagine that.

Second, imagination under-determines its contents. This is guaranteed by A-accessibilities allowing one to look at a plurality of worlds:

$$[A](B \vee C) \nvDash [A]B \vee [A]C$$

When you imagine, given inputs provided by the page of *The Lord of the Rings* you are reading, that Boromir is either left-handed or right-handed (or ambidextrous) – he's a normally endowed human being, after all – you don't thereby imagine that he is left-handed, and you don't thereby imagine that he is right-handed (LOTR tells you nothing about Boromir's dominant hand). There'll be worlds compatible with what you imagine where he's left-handed, and compatible worlds where he's right-handed.

Third, imagination is not additive. This is guaranteed by topic-sensitivity:

$$[A]B \nvDash [A](B \vee C)$$

When, starting from the supposition that Stauffenberg puts the bomb on the other side of the table, you imagine that Hitler gets killed, you don't thereby imagine that either Hitler gets killed or there's life on Kepler-442b.

Fourth, imagination is hyperintensional. Even when A strictly or necessarily implies B, that is, there is just no possible way for A to be true without B being true, one needn't imagine B starting from suppositional input A, when B is off-topic:

$$A \prec B \nvDash [A]B$$

For a detailed discussion of the workings of Semantics 1, see Berto (2018). Now let's look at the alternative.

6.4.2 *Semantics 2*

Semantics 2 starts with a frame F = (W, N, $\{R_A|A \in L\}$) with three items:

1. W is a set of worlds;
2. N $\subseteq$ W is the subset of *normal* worlds (think possible worlds); the items in W – N are the non-normal worlds (think of them as logically impossible worlds);
3. $\{R_A|A \in L\}$ is as in Semantics 1.

There is no structure of topics. The truth conditions go as follows. A model M = (W, N, $\{R_A|A \in L\}$, r) is a frame with a *relation* r, relating, for each world w, each atom p in L_{AT} to truth, 'r_wp1', falsity, 'r_wp0', both, or neither. Unlike the usual valuations, a relation can connect a formula to more than one truth value at a world: formulas can be both true and false, truth value gluts. Formulas can also be truth value gaps, being related neither to truth, nor to falsity. Readers familiar with non-classical logics will recognize one way of presenting the semantics for First Degree Entailment (FDE) logic (Belnap (1977), Dunn (1976), Priest (2008)), where one drops the classical presupposition that truth and falsity be exhaustive and exclusive.

This is extended to the whole L via the following truth-and-falsity conditions at *normal* worlds (we need to give falsity conditions, too, as the semantics is not bivalent):

(S1¬) $r_w(\neg A)1$ iff r_wA0
(S2¬) $r_w(\neg A)0$ iff r_wA1
(S1∧) $r_w(A \wedge B)1$ iff r_wA1 and r_wB1
(S2∧) $r_w(A \wedge B)0$ iff r_wA0 or r_wB0
(S1∨) $r_w(A \vee B)1$ iff r_wA1 or r_wB1
(S2∨) $r_w(A \vee B)0$ iff r_wA0 and r_wB0
(S1≺) $r_w(A \prec B)1$ iff for all $w_1 \in$ N: if $r_{w1}A1$, then $r_{w1}B1$
(S2≺) $r_w(A \prec B)0$ iff for some $w_1 \in$ N: $r_{w1}A1$ and $r_{w1}B0$
(S1[A]) $r_w([A]B)1$ iff for all $w_1 \in$ W such that wR_Aw_1, $r_{w1}B1$
(S2[A]) $r_w([A]B)0$ iff for some $w_1 \in$ W such that wR_Aw_1, $r_{w1}B0$

Negation, conjunction, and disjunction get the relational semantics for FDE. The fishhook is, again, a strict conditional: true at a normal world when all normal worlds where the antecedent is true also make the consequent true. [A] works, again, with accessibilities indexed by formulas. We can equivalently rephrase its truth and falsity conditions in a more compact way using set-selection functions:

(S1[A]) $r_w([A]B)1$ iff $f_A(w) \subseteq |B|$
(S2[A]) $r_w([A]B)0$ iff $f_A(w) \cap |\neg B| \neq \varnothing$

At points in W − N, r relates complex formulas to truth values directly, irrespective of their syntax. At non-normal worlds, a disjunction can be true although both disjuncts are false, etc. – this is the logically anarchic nature of such worlds. (The trick of assigning arbitrary truth values to complex formulas at non-normal worlds is due to Rantala (1982). Priest (2005, 2008) has put it to work in logic and metaphysics; I've done the same in Berto and Jago (2019).)

The Basic Constraint goes thus for all $A \in$ L:

(BC) If $w \in$ N, then $f_A(w) \subseteq |A|$

(We restrict to normal worlds, as non-normal worlds can do what they like.) Logical consequence is truth preservation at all *normal* worlds of all models: S ⊩ A when in all models M = (W, N, {$R_A|A \in$ L}, r)) and for all worlds w in N, if $r_w B1$ for all B in S, then $r_w A1$. Logical truth, ⊩ A, is truth at all normal worlds of all models. Non-normal worlds are taken as logically impossible worlds, where logic can fail. In the definitions of logical truth and consequence, we only look at worlds where logic does *not* fail.

Like Semantics 1, Semantics 2 makes of [A]A a validity, thanks to the BC. Unlike Semantics 1, Semantics 2 doesn't make ROMS operators adjunctive as such: one can imagine, given input A, that $B \wedge C$ is the case without imagining that B is the case and that C is the case, and vice versa. If we want to rule this out, we need the following two conditions. For Simplification:

(C1) For all $w \in$ N: if wR_Aw_1, and $r_{w1}(B \wedge C)1$, then $r_{w1}B$ 1 and $r_{w1}C1$

For Adjunction:

(C2) For all $w \in$ N: if wR_Aw_1, $r_{w1}B$ 1 and $r_{w1}C1$, then $r_{w1}(B \wedge C)1$

(For proofs that C1 and C2 validate, respectively, Simplification and Adjunction, and for proofs of all the validities and invalidities listed from now on for Semantics 2, see Berto (2017).) One may want to retain C1 and Simplification but make imagination non-adjunctive by not having C2, e.g., if one is persuaded by the quasi-Quinean worries discussed above, and unpersuaded by my replies. Semantics 2 wins in flexibility over Semantics 1. One may say that it loses in naturalness, as C1 and C2 may look contrived, or ad hoc.

Otherwise, Semantics 2 proceeds hand in hand with Semantics 1 on a number of invalidities. It agrees with the latter in modelling the

non-monotonic nature of ROMS ([A]B does not entail [A ∧ C]B) and its under-determinacy ([A](B ∨ C) does not entail [A]B ∨ [A]C), thanks to [A]'s being a variably strict modal and to the accessibility of a plurality of worlds. But it can also mimic hyperintensionality (A ≺ B does not entail [A]B) and the failure of Addition ([A]B does not entail [A](B ∨ C)), with no need for a topicality filter, thanks to the accessibility of non-normal worlds which can break necessary truths and consequences. The resulting logic is rather weak in both cases, but we can make it stronger by adding a constraint that captures the idea of equivalence in imagination.

6.5 Adding Equivalence in Imagination

The constraint – let's call it the Principle of Imaginative Equivalents (PIE) – is the same for both kinds of semantics. For all *A, B* in L:

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

If all the selected *A*-worlds make *B* true and vice versa, then *A* and *B* are equivalent in imagination. Intuitively, when we take either as our suppositional input in an act of ROMS, we look at the same circumstances. PIE validates, both in Semantics 1 and in Semantics 2, a Substitutivity principle that greatly strengthens the logic:

(Substitutivity) [A]B, [B]A, [A]C �ⱶ [B]C

Substitutivity says that equivalents in imagination, *A* and *B*, can be replaced *salva veritate* as modal indexes inside [ ]. It may be the case that we imagine different things in exercises of ROMS that start with intensionally equivalent suppositional inputs. But when we start from inputs that play the same role in our cognitive life, we *will* imagine the same things. (Again, which *A*s and *B*s count as cognitively equivalent depends on the specific imaginative agent, and it's not an issue for the logic to settle: *x is a woodchuck* and *x is a groundhog* may be equivalent for your fellow zoologist but not for you, etc.)

  Given the naturalness of the notion of imaginative equivalence – given how it connects to the plausible ideas of cognitive equivalence and cognitive synonymy – Substitutivity looks like a good principle to have in a logic of imagination as ROMS. Say that *bachelor* and *unmarried man* are cognitively equivalent for you qua competent speaker of English: when you suppose that John is unmarried, you imagine that he's a bachelor ([A]B) and vice versa ([B]A). Supposing that John is unmarried, you imagine that he has no marriage allowance, [A]C. Then the same happens when you suppose that John is a bachelor, [B]C. Your imagination won't work that smoothly with *Groundhogs are rodent* and *Woodchucks are rodent* – unless you're a zoologist; in this case, it probably will.

Now for the glitch. In Semantics 1, PIE validates a further inference we may call, following literature on non-monotonic logics, Special or Restricted Transitivity:

(RT) $[A]B, [A \land B]C \Vdash [A]C$

Is RT good? It's easy to give plausible instances. The issue has to do with cases where C easily pops to mind given B alone, but is only dimly related to A. 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, which may work as a counterexample. $[A]B$: supposing that I am wearing a red shirt in Pamplona, I imagine that I am being chased by bulls. $[A \land B]C$: supposing 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 $[A]C$: supposing that I am wearing a red shirt in Pamplona, I don't imagine that I die on its streets.

We here mark a crucial difference with Semantics 2. There, the proof that RT is valid given PIE needs to make use of the Simplification- and Adjunction-validating conditions C1 and C2 above. It goes thus: suppose normal world w verifies (i) $[A]B$ and (ii) $[A \land B]C$. Given the BC, w also verifies $[A]A$ and so, by (i) and C2, it verifies $[A](A \land B)$. From BC again, w also verifies $[A \land B](A \land B)$ and hence, by C1, it verifies $[A \land B]A$. Then by (S1[A]), $f_A(w) \subseteq |A \land B|$ and $f_{A \land B}(w) \subseteq |A|$ and so, by PIE, $f_A(w) = f_{A \land B}(w)$. Since (ii) w also verifies $[A \land B]C$, by (S1[A]) again we have $f_{A \land B}(w) \subseteq |C|$, hence $f_A(w) \subseteq |C|$, and so w verifies $[A]C$.

Now both C1 and C2 are added to Semantics 2 manually: if one drops either, one can retain Substitutivity as desired, without having RT. For the proof of Substitutivity doesn't need them. It goes thus: suppose normal world w verifies $[A]B, [B]A, [A]C$. By (S1[A]), we have $f_A(w) \subseteq |B|$, $f_B(w) \subseteq |A|$, $f_A(w) \subseteq |C|$. PIE then gives us $f_A(w) = f_B(w)$, and hence $f_B(w) \subseteq |C|$. So by (S1[A]), w verifies $[B]C$. Semantics 1 cannot afford such flexibility: it is inherently adjunctive, for there is no addition of alien topic in the move from a conjunction to its conjuncts or vice versa. It may be that intuitive counterexamples to RT are forceful enough to have us reconsider the idea that ROMS is fully conjunctive after all!

Here's another glitch involving PIE. Pierre Saint-Germier (forthcoming) has recently pointed out that if A and A' are logically equivalent and the topic of A' is included in that of A, then $[A]B$ and $[A']B$ turn out to be logically equivalent. The problem is generated by the equivalence of instances like $[p]q$ and $[p \lor (p \land r)]q$. I reply that the only case in which the difference between p and $p \lor (p \land r)$ matters, for they should not be cognitively equivalent qua suppositional inputs, is when the agent does not grasp the proposition expressed by r, due to some conceptual deficit. But I grant that it *is* indeed a problem for my approach as it is now, that it doesn't represent the difference. I should think about what to do!

6.6 Summary

I have proposed to understand one kind of epistemically useful imagination as reality-oriented mental simulation (ROMS): we suppose that *A* takes on board relevant background beliefs and knowledge, and we wonder whether *B* turns out in the imagined scenario. The scenario we focus on is constrained by its relevant connections with the suppositional input *A* and governed by a principle of minimal alteration. ROMS allows us to enlarge our stock of conditional beliefs: if *B* turns out in the imagined scenario, we add to our belief stock the belief in *B* conditional on *A*.

I have then discussed the idea that, in spite of its hyperintensional anarchy, ROMS obeys a principle of equivalence in imagination: when *A* and *B* play interchangeable roles in one's cognitive life, they play interchangeable roles, in particular, as suppositional inputs in ROMS. I have proposed to use the notation '[*A*]*B*' for 'Given input *A*, one imagines that *B*', or 'In an exercise of mental simulation starting from supposition *A*, one imagines that *B*'. I have interpreted [*A*] as a variably strict, non-normal modal and given two alternative semantics for it: Semantics 1 uses possible worlds plus an added structure of topics. Semantics 2 uses non-normal or impossible worlds. I have shown how a constraint representing equivalence in imagination can be added to both semantics, and how it works within them in slightly but interestingly different ways. Specifically, Semantics 2 has the advantage of validating Substitutivity, which seems to capture something important about likeness in cognitive role, without validating Special/Restricted Transitivity, which may have clear counterexamples. But it does so via manually added constraints that may look artificial, or ad hoc. Semantics 1 cannot keep Substitutivity and Special/Restricted transitivity apart. It is, however, based on adding a topicality filter to an otherwise very classical, normal modal logical framework; this may make Semantics 1 more appealing to conservative logicians.[1]

Note

References

Adams, E. 1998. *A Primer of Probability Logic*. Stanford, CA: CSLI.
Arcangeli, M. 2019. *Supposition and the Imaginative Realm*. London and New York: Routledge.

Barwise, J. and J. Perry. 1983. *Situations and Attitudes*. Cambridge, MA: MIT Press.

Belnap, N. 1977. "A Useful Four-valued Logic". In *Modern Uses of Multiple-valued Logic*, ed. by J. Dunn and G. Epstein. Dordrecht: Reidel.

Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford: Oxford UP.

van Benthem, J. 2011. *Logical Dynamics of Information and Interaction*. Cambridge: Cambridge UP.

Berto, F. 2017. "Impossible Worlds and the Logic of Imagination". *Erkenntnis* 82: 1277–97.

Berto, F. 2018. "Aboutness in Imagination". *Philosophical Studies* 175: 1871–86.

Berto, F. 2019. "Taming the Runabout Imagination Ticket". *Synthese*, On Line First https://doi.org/10.1007/s11229-018-1751-6.

Berto, F. and M. Jago. 2019. *Impossible Worlds*. Oxford: Oxford UP.

Block, N. 1983. "Mental Pictures and Cognitive Science". *The Philosophical Review* 92: 499–541.

Brown, J. 2018. *Fallibilism: Evidence and Knowledge*. Oxford: Oxford UP.

Byrne, R. 2005. *The Rational Imagination*. Cambridge, MA: MIT Press.

Canavotto, I., F. Berto and A. Giordani. 2020. "Voluntary Imagination: A Fine-Grained Analysis". *Review of Symbolic Logic*, forthcoming.

Chellas, B. 1975. "Basic Conditional Logic". *Journal of Philosophical Logic* 4: 133–53.

Cruse, D.A. 2000. *Meaning in Language*. Oxford: Oxford UP.

Currie, G. and I. Ravenscroft. 2002. *Recreative Minds*. Oxford: Oxford UP.

Dorsch, F. 2012. *The Unity of Imagining*. Frankfurt: Ontos Verlag.

Douven, I. 2016. *The Epistemology of Indicative Conditionals: Formal and Empirical Approaches*. Cambridge: Cambridge UP.

Dunn, J.M. 1976. "Intuitive Semantics for First-Degree Entailments and Coupled Trees". *Philosophical Studies* 29: 149–68.

Edgington, D. 1995. "On Conditionals". *Mind* 104: 235–329.

Evans, J. and D. Over. 2004. *If*. Oxford: Oxford UP.

Fine, K. 2016. "Angellic Content". *Journal of Philosophical Logic* 45: 199–226.

Gendler, T.S. 2000. "The Puzzle of Imaginative Resistance". *Journal of Philosophy* 97: 55–81.

Goldman, A. 2006. *Simulating Minds*. Oxford: Oxford UP.

Gregory, D. 2016. "Imagination and Mental Imagery". In *The Routledge Handbook of Philosophy of Imagination*, ed. by A. Kind. London; New York: Routledge.

Hawke, P. 2016. "Questions, Topics, and Restricted Closure". *Philosophical Studies* 173: 2759–84.

Hawthorne, J. 2004. *Knowledge and Lotteries*. Oxford: Oxford UP.

Hintikka, J. 1962. *Knowledge and Belief*. Ithaca, NY: Cornell UP.

Hornischer, L. 2017. *Hyperintensionality and Synonymy*. MSc Master of Logic, University of Amsterdam.

Humberstone, L. 2008. "Parts and Partitions". *Theoria* 66: 41–82.

Kahneman, D., P. Slovic and A. Tversky. 1982. *Judgment under Uncertainty*. Cambridge: Cambridge UP.

Kind, A. 2001. "Putting the Image Back in Imagination". *Philosophy and Phenomenological Research* 62: 85–109.

Kind, A. 2016. "Imagining under Constraints". In *Knowledge Through Imagination*, ed. by A. Kind and P. Kung. Oxford: Oxford UP: 145–59.

Kind, A. and P. Kung. 2016. *Knowledge Through Imagination*. Oxford: Oxford UP.

Kraus, S., D. Lehmann and M. Magidor. 1990. "Nonmonotonic Reasoning, Preferential Models, and Cumulative Logics". *Artificial Intelligence* 44: 167–207.

Langland-Hassan, P. 2016. "On Choosing What to Imagine". In *Knowledge Through Imagination*, ed. by A. Kind and P. Kung. Oxford: Oxford UP: 61–84.

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

Lewis, D. 1976. "Probabilities of Conditionals and Conditional Probabilities". *Philosophical Review* 95: 581–9.

Lewis, D. 1988. "Relevant Implication". *Theoria* 54: 161–74.

Lyons, J. 1996. *Linguistic Semantics*. Cambridge: Cambridge UP.

Mulligan, K. 1999. "La varietà e l'unità dell'immaginazione". *Rivista di estetica* 11: 53–67.

Murphy, M.L. 2003. *Semantic Relations and the Lexicon*. Cambridge: Cambridge UP.

Nozick, R. 1981. *Philosophical explanations*. Cambridge, MA: Harvard UP.

Oaksford, M. and N. Chater (eds.). 2010. *Cognition and Conditionals*. Oxford: Oxford UP.

Paivio, A. 1986. *Mental Representation*. Oxford: Oxford UP.

Perry, J. 1989. "Possible Worlds and Subject Matter". In *The Problem of the Essential Indexical and Other Essays*, ed. by J. Perry. Palo Alto, CA: CSLI: 145–60.

Priest, G. 2005. *Towards Non-Being*. Oxford: Oxford UP.

Priest, G. 2008. *An Introduction to Non-Classical Logic*. Cambridge: Cambridge UP.

Pylyshyn, Z. 2002. "Mental Imagery: In Search of a Theory". *Behavioral and Brain Sciences* 25: 157–82.

Quine, W.V.O. 1960. *Word and Object*. Cambridge, MA: MIT Press.

Rantala, V. 1982. "Impossible World Semantics and Logical Omniscience". *Acta Philosophica Fennica* 35: 106–15.

Saint-Germier, P. forthcoming. "Hyperintensionality in Imagination". In *Logic in High Definition, Trends in Logical Semantics*, ed. by A. Giordani and J. Malinowski. Dordrecht: Springer: 79–117.

Shephard, R. and J. Metzler. 1971. "Mental Rotation of Three-Dimensional Objects". *Science* 171: 701–3.

Stanojecvić, M. 2009. "Cognitive Synonymy: A General Overview". *Facta Universitatis* 7: 193–200.

Stenning, K. and M. van Lambalgen. 2008. *Human Reasoning and Cognitive Science*. Cambridge, MA: MIT Press.

Strohminger, M. and J. Yli-Vakkuri. 2019. "Knowledge of Objective Modality". *Philosophical Studies* 176: 1155–75.

Stuart, M. 2020. "The Productive Anarchy of Scientific Imagination". *Philosophy of Science*, forthcoming http://philsci-archive.pitt.edu/16749/.

Van Leeuwen, N. 2013. "The Meanings of 'Imagine' Part I: Constructive Imagination". *Philosophy Compass* 8: 220–30.

Walton, K. 1990. *Mimesis as Make-Believe*. Cambridge, MA: Harvard UP.

Williamson, T. 2016. "Knowing by Imagining". In *Knowledge Through Imagination*, ed. by A. Kind and P. Kung. Oxford: Oxford UP: 113–23.

Yablo, S. 2014. *Aboutness*. Princeton, NJ: Princeton UP.