The Theory of Topic-Sensitive Intentional Models

Berto, F.

Publication date
2019

Document Version
Author accepted manuscript

Published in
The Logica Yearbook 2018

Preprint of a paper appeared in The Logica Yearbook 2018
Eds. Igor Sedlár, Martin Blicha

The Theory of Topic-Sensitive Intentional Modals

FRANCESCO BERTO[1]

Abstract: A Topic-Sensitive Intentional Modal (TSIM) is a two-place, variably strict modal with an aboutness or topicality constraint, of the form '$X^{\phi}\psi$' (read: 'Given ϕ, the agent X's that ψ', X being some mental state or act). TSIMs do nice things for mainstream and formal epistemology, belief revision theory, and mental simulation theory. I present a basic formal semantics for TSIMs and explore three readings of '$X^{\phi}\psi$' one gets by imposing different constraints on their truth conditions: (1) as expressing knowability relative to information ('Given total information ϕ, one is in the position to know that ψ'), inspired by Dretske's view that what one can know depends on the available (empirical) information; (2) as a mental simulation operator ('In mental simulation starting with input ϕ, one imagines that ψ') capturing features of mainstream mental simulation theories, like that of Nichols and Stich; (3) as a hyperintensional belief revision operator ('After (statically) revising by ϕ, one believes that ψ'), reducing the idealization of cognitive agents one finds in standard doxastic logics and AGM. I close by mentioning developments of TSIM theory currently in progress.

Keywords: intentionality, epistemic and doxastic logic, aboutness theory, hyperintensionality, mental simulation, belief revision, knowabilty, information

1 Introduction

We have learned since (Hintikka, 1962) how to treat notions like *knows, believes, is informed that* using modal logic: we interpret them as quanti-

[1]This paper is published within the project 'The Logic of Conceivability', funded by the European Research Council (ERC CoG), Grant Number 681404. It summarizes material published or forthcoming in (Berto, 2017a, 2018; Berto & Hawke, 2018), all of which are Open Access publications falling under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). Several people are thanked in those papers. The Logic of Conceivability gang deserves special rehearsed thanks – in particular Chris Badura, Ilaria Canavotto, Jorge Ferreira, Peter Hawke, Karolina Krzyzanowska, Aybüke Özgün, Tom Schoonen, Anthi Solaki. Thanks a lot to the organizers of the *Logica* conference for inviting me to present this material, and to the audience at *Logica 2018* for helpful comments and remarks.

fiers over possible worlds, restricted from the viewpoint of a given world by an accessibility relation (hopefully) endowed with some intuitive meaning. '$X\phi$' ('The agent Xs that ϕ') is true at w just in case ϕ is true at a bunch of worlds accessible via the relation R from w. By imposing simple conditions on R, we then validate various principles characteristic of different modal systems. Some conditions are more contentious than others. We all agree that R should be reflexive for 'X' to be read as 'knows', it shouldn't for it to be read as 'believes'. But we debate on whether R should be transitive, for we disagree on whether Positive Introspection should hold for knowledge or belief: does Xing that ϕ entail that one Xs that one Xs that ϕ?

All of this is well known. The rehearsal just provided is in order to highlight the three main ways in which the very general framework for intentional operators I want to sketch in *this* paper differs from the mainstream tradition. Call such framework the theory of Topic-Sensitive Intentional Modals (TSIM – read it as '*ZIMM!*'):

(1) The Hintikkan operators are one-place modals. The TSIMs are two-place modals: things of the form '$X^{\phi}\psi$', to be generically read as 'Given ϕ, the agent Xs that ψ', where X is some mental state or act.

(2) The $X^{\phi}\psi$'s are *variably strict* modals. Variability represents the contextual selection of information the agent imports into the Xed content on the basis of ϕ. The operators turn out to be nonmonotonic: epistemic logic, in TSIM clothing, becomes a kind of conditional logic.

(3) The $X^{\phi}\psi$'s encompass a *topicality* or *aboutness* filter capturing their standing for *intentional* mental states: states which are directed towards, or are about, a certain content or topic represented in the mind.

Ideas (1) and (2) are in the literature: two-place epistemic or doxastic operators expressing conditional belief, or static and dynamic belief revision ('$B^{\phi}\psi$': 'Conditional on ϕ, one believes ψ'; '$[\phi]\psi$': 'After revising one's beliefs by ϕ, it is the case that ψ') have been explored, e.g., in Dynamic Epistemic Logic and in modal recaptures of AGM (Spohn, 1988; Segerberg, 1995; Lindström & Rabinowicz, 1999; Board, 2004; van Ditmarsch, 2005; Asheim & Sövik, 2005; Leitgeb & Segerberg, 2005; van Benthem, 2007; Baltag & Smets, 2008; van Ditmarsch, van der Hoek, & Kooi, 2008; van Benthem, 2011; Girard & Rott, 2014; etc.).

Idea (3) is relatively new, though variously related to work on tautological or analytic entailment (Parry, 1933; van Fraassen, 1969; Angell, 1977;

Fine, 1986; Correia, 2004; Ferguson, 2014) and awareness logic (Fagin & Halpern, 1988; Schipper, 2015). So let me comment on (3) a bit.

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). Research on aboutness is burgeoning: (Fine, 2014, 2015; Hawke, 2017). Such works have clarified that what a sentence is about can be (properly) included in what another one is about. Thus, to the extent that aboutness is a component of content, contents should be capable of standing in mereological relations (Yablo, 2014, Section 2.3), (Fine, 2015, Sections 3-5). They should be capable of being fused into wholes which inherit the proper features from the parts (Yablo, 2014, Section 3.2).

Yablo and Fine address aboutness mainly as a feature of linguistic representations, but Chapter 7 of (Yablo, 2014) gets into the aboutness of epistemic states. And rightly so, because another kind of representation bears aboutness, too: mental representation. Maybe Brentano was wrong when he said that all mental states bear intentionality, but some do, and 'every intentional state or episode has an object – something it is about or directed on' (Crane, 2013, p. 4).

The insight behind TSIM theory is that we should take at face value the view of belief, knowledge, (cognitive) information, but also of other notions less explored in formal logic, like imagination and mental simulation, as (propositional) representational mental states bearing intentionality, that is, being about states of affairs, issues, situations, or circumstances which make for their contents. I will generically call these things *topics*, and provide a simple formal mereology for them. The semantics for our TSIMs will be given in a kind of conditional logic framework, with an added mereology of topics.

Besides being nonmonotonic thanks to (2), our $X^{\phi}\psi$'s will turn out to be *hyperintensional*, differentiating between necessarily or logically equivalent contents, thanks to their topicality or aboutness filter (3). And rightly so, because *thought* is hyperintensional. Our mental states – believing, supposing, desiring, hoping, fearing – can treat logically or necessarily equivalent contents differently: Lois Lane can wish that Superman is in love with her without wishing that Clark Kent is in love with her, although (if Barcan Marcus and Kripke are right) it is metaphysically impossible for Superman to be other than Clark Kent. We can think that $75 \times 12 = 900$ without thinking that Fermat's Last Theorem is true. But given the necessity of mathematical truths, the two make for the same content or proposition in possible worlds semantics: the total set of worlds.

One further feature of TSIM theory brings back continuity with the standard Hintikkan framework: starting from a basic semantics for our $X^{\phi}\psi$'s, which I will present in Section 2 below, one can add constraints on the accessibility relations (or, as we will see, functions) used in their truth conditions. Such constraints validate different logical consequences or principles involving the operators, and come with different interpretations for them. (Just as, starting from **K** as our basic normal modal logic, we get stronger systems by adding constraints on accessibility, and the different principles characteristic of **B**, **S4**, etc., come with different interpretations of the relevant modals.)

In Section 3, I will give an overview of three such interpretations; not because they are especially good, or because they are the only ones available, but just because, as a matter of fact, these are the ones I have explored in various works, alone or with friends:

Section 3.1 : '$X^{\phi}\psi$', relabeled as '$K^{\phi}\psi$', as expressing a notion of *knowability relative to information* ('Given total information ϕ, one is in the position to know that ψ'), inspired by Dretske's (1999) view that what an agent can know is dependent on the available (empirical) information. Peter Hawke and I have developed this in a paper forthcoming in *Mind* (Berto & Hawke, 2018).

Section 3.2: '$X^{\phi}\psi$', relabeled as '$I^{\phi}\psi$', as expressing an *imagination* or *mental simulation* operator ('In an act of imagination starting with input ϕ, one imagines that ψ'), capturing ideas found, e.g., in mental simulation theories from cognitive science like (Nichols & Stich, 2003), and in Williamson's (2007) imagination-based modal epistemology. I have presented this in a paper that has come out in *Philosophical Studies* (Berto, 2017a).

Section 3.3: '$X^{\phi}\psi$', relabeled as '$B^{\phi}\psi$', as expressing a hyperintensional *conditional belief*, or (static) *belief revision* operator ('Conditional on ϕ, one believes ψ', or: 'After revising by ϕ, one believes ψ'), which reduces the logical idealization of cognitive agents affecting similar operators in standard doxastic logics as well as in AGM. I have presented this in a paper that has come out in *Erkenntnis* (Berto, 2018).

All of the above are mere initial explorations of the TSIM world. In the conclusive Section 4, I briefly speak of possible further work and of how others are currently developing some TSIM ideas.

2 The basic semantics

Take a propositional language $\mathcal{L}$ with an indefinitely large set $\mathcal{L}_{AT}$ of atomic formulas, p, q, r $(p_1, p_2, \ldots)$, negation $\neg$, conjunction $\wedge$, disjunction $\vee$, a strict conditional $\prec$, X standing for a generic TSIM, round parentheses as auxiliary symbols (,). I use ϕ, ψ, χ, $\ldots$, as metavariables for formulas of $\mathcal{L}$. The well-formed formulas are items in $\mathcal{L}_{AT}$ and, if ϕ and ψ are formulas:

$$\neg\phi \mid (\phi \wedge \psi) \mid (\phi \vee \psi) \mid (\phi \prec \psi) \mid X^{\phi}\psi$$

Outermost brackets are usually omitted. We identify $\mathcal{L}$ with the set of its well-formed formulas. In the metalanguage I use variables $w, w_1, w_2, ...$, ranging over worlds, x, y, z $(x_1, x_2, ...)$, ranging over *topics* (I'll say more on these in a minute), and the symbols $\Rightarrow$, $\Leftrightarrow$, &, *or*, $\sim$, $\forall$, $\exists$, read the usual way. A *frame* for $\mathcal{L}$ is a tuple $\mathfrak{F} = \langle W, \{R_{\phi} \mid \phi \in \mathcal{L}\}, \mathcal{T}, \oplus, t\rangle$, understood as follows:

- W is a non-empty set of possible worlds.
- $\{R_{\phi} \mid \phi \in \mathcal{L}\}$ is a set of accessibilities between worlds, where each $\phi \in \mathcal{L}$ has its own $R_{\phi} \subseteq W \times W$. These may satisfy a number of different conditions, to which I come in a minute.
- $\mathcal{T}$ is a set of *topics*. We may understand topics as the abstract or concrete situations (the configurations of objects and properties), or issues, or Yablovian or Finean subject matters the formulas of $\mathcal{L}$ involved in intentional ascriptions are *about*. (We need no more for our propositional logic purposes. In particular, we can stay silent on how they may be interpreted: as certain divisions of the set of worlds, as Finean truthmakers, structured entities, or else. We only ask them to obey the mereological constraints coming next.)
- $\oplus$ is *topic fusion*, a binary operation on $\mathcal{T}$ making of topics part of larger topics and satisfying, for all $x, y, z \in \mathcal{T}$:
  - (Idempotence) $x \oplus x = x$
  - (Commutativity) $x \oplus y = y \oplus x$
  - (Associativity) $(x \oplus y) \oplus z = x \oplus (y \oplus z)$

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

- (Reflexivity) $x \leq x$
- (Antisymmetry) $x \leq y \;\&\; y \leq x \Rightarrow x = y$
- (Transitivity) $x \leq y \;\&\; y \leq z \Rightarrow x \leq z$

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

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

  - $t(\phi) = \oplus \mathfrak{A}\mathfrak{t}\phi = t(p_1) \oplus \ldots \oplus t(p_n)$.

  A formula is about what its atoms, taken together, are about.

This mereology of topics (of which a more refined version is Peter Hawke's *issue-based theory*: see Hawke, 2017) will allow our TSIMs to make hyperintensional distinctions. However, we don't get as fine-grained as the syntax of $\mathcal{L}$. By induction on the construction of formulas, $t(\phi) = t(\neg\neg\phi)$ (recall Frege on the *Sinn*-preservation of Double Negation). Also, $t(\phi) = t(\neg\phi)$: a formula is about what its negation is about (no matter how we understand the topic of the whiteness of snow, 'Snow is white' is about that, and that is what 'Snow is not white' is also about). And not only $t(\phi \wedge \psi) = t(\phi \wedge \psi)$, but also, e.g., $t(\phi \wedge \psi) = t(\phi) \oplus t(\psi) = t(\phi \vee \psi)$. In the literature, these are often taken as key requirements for a good recursive account of aboutness or subject matter (see Yablo, 2014, p. 42; Fine, 2015, p. 1).

A *model* $\mathfrak{M} = \langle W, \{R_\phi \mid \phi \in \mathcal{L}\}, \mathcal{T}, \oplus, t, \Vdash\rangle$ is a frame with an interpretation $\Vdash \subseteq W \times \mathcal{L}_{AT}$, relating worlds to atoms: we read '$w \Vdash p$' as meaning that p is true at w, '$w \nVdash p$' as $\sim w \Vdash p$. $\Vdash$ is extended to all formulas of $\mathcal{L}$ thus:

- (S$\neg$) $w \Vdash \neg\phi \Leftrightarrow w \nVdash \phi$
- (S$\wedge$) $w \Vdash \phi \wedge \psi \Leftrightarrow w \Vdash \phi$ & $w \Vdash \psi$
- (S$\vee$) $w \Vdash \phi \vee \psi \Leftrightarrow w \Vdash \phi$ *or* $w \Vdash \psi$
- (S$\prec$) $w \Vdash \phi \prec \psi \Leftrightarrow \forall w_1(w_1 \Vdash \phi \Rightarrow w_1 \Vdash \psi)$
- (SX) $w \Vdash X^\phi\psi \Leftrightarrow$ **(1)** $\forall w_1(wR_\phi w_1 \Rightarrow w_1 \Vdash \psi)$ & **(2)** $t(\psi) \leq t(\phi)$

For '$X^\phi\psi$' to come out true at w we ask, thus, for two things to happen:

> **(1)** ψ must be true at all worlds w_1 one looks at, via the accessibility determined by ϕ (more specific readings of '$wR_\phi w_1$' will come in Section 3: these depend on the conditions we add). This is the truth-conditional component making of $X^\phi\psi$ a variably strict quantifier over worlds.
>
> **(2)** ψ must be fully on topic with respect to ϕ. This is the aboutness-preservation component.

(SX) can be equivalently expressed using set-selection functions (Lewis (1973), pp. 57-60). Each $\phi \in \mathcal{L}$ comes with a function $f_\phi : W \rightarrow \mathcal{P}(W)$ outputting the set of accessible worlds, $f_\phi(w) = \{w_1 \in W | wR_\phi w_1\}$. If $|\phi| = \{w \in W | w \Vdash \phi\}$, we can rephrase the clause for X as:

- (SX) $w \Vdash X^\phi\psi \Leftrightarrow$ **(1)** $f_\phi(w) \subseteq |\psi|$ & **(2)** $t(\psi) \leq t(\phi)$

The two formulations are equivalent as $wR_\phi w_1 \Leftrightarrow w_1 \in f_\phi(w)$. However, either formulation is at times handier than the other. In particular, we will phrase the additional conditions on the semantics of our TSIMs in Section 3 using the f's.

Finally, logical consequence is truth preservation at all worlds of all models. With Σ a set of formulas:

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

For single-premise entailment, I write $\phi \vDash \psi$ for $\{\phi\} \vDash \psi$. Logical validity, $\vDash \phi$, truth at all worlds of all models, is $\emptyset \vDash \phi$, entailment by the empty set of premises.

The logic induced by the semantics for the extensional operators is just classical propositional, with $\prec$ a strict S5-like conditional (i.e., one equivalent to the necessitation of a material conditional, where the relevant necessity is S5). The novelty comes with $X^\phi\psi$, whose logical behavior we are now going to unpack.

3 Adding conditions

One can impose different conditions on the f's:

(C0) $|\phi| \subseteq f_\phi(w)$

(C1) $f_\phi(w) \subseteq |\phi|$

(C2) $|\phi| \neq \emptyset \Rightarrow f_\phi(w) \neq \emptyset$

(C3) $f_\phi(w) \subseteq |\psi| \;\&\; f_\psi(w) \subseteq |\phi| \Rightarrow f_\phi(w) = f_\psi(w)$

(C4) $f_\phi(w) \cap |\psi| \neq \emptyset \Rightarrow f_{\phi\wedge\psi}(w) \subseteq f_\phi(w)$

The three interpretations of our TSIMs to be explored now come, respectively, from (1) adding (C0), (2) adding (C1) (and, tentatively, (C3)), and (3) imposing a total ordering on W (read as comparative plausibility in a belief system) that automatically validates (C1)-(C4). In each case, we restrict our attention to models that satisfy the relevant conditions. In each of the three subsections, I will only explore *some* notable validities and invalidities involving the TSIMs. There are many more, for which I refer to the source papers mentioned in Section 1 above.

3.1 Knowability relative to information

(C0) says that all the ϕ-worlds are selected, but allows for selected worlds which are not ϕ-worlds. With this one in place, we relabel our '$X^\phi\psi$' as '$K^\phi\psi$' and read it as expressing the *Knowability of ψ, Relative to Information ϕ* (KRI). This comes from Dretske (1999), who stressed the view that knowledge depends on the (empirical) information available to us, where the role of incoming information is to narrow down the set of epistemically viable alternatives.[2] Thus, we read the accessibility $wR_\phi w_1$ as: 'Relative to w, w_1 is epistemically accessible on the basis of total information that ϕ', or: 'Relative to w, w_1 is not ruled out by knowledge based on the total information that ϕ'.

Information (1) eliminates possibilities, just as the truth of a meaningful sentence is, in general, compatible with some possibilities and not others; and (2) is about something, just as a meaningful sentence has a subject matter that it addresses. Knowability of ψ is, then, determined by the available information ϕ twice over: (1) once via the worlds ϕ makes epistemically accessible (that's the truth-conditional component of TSIMs), and (2) once via the topic ϕ is concerned with (that's the aboutness component).

A first simple validity comes via (C0) (the proof is trivial) and captures the idea that knowledge is factive:

(**Factivity**) $\{K^\phi\psi, \phi\} \vDash \psi$

When ψ is knowable based on the information that ϕ, and ϕ is true, ψ must be true as well. Notice that ϕ *needn't* be true: one point of departure of KRI from Dretske's view, is that Dretske takes all information to be veridical, whereas KRI is neutral on this, as per the (trivially proved) invalidity:

$$K^\phi\psi \nvDash \phi$$

(On the debate concerning the factivity of information, see, e.g., Floridi, 2015. Floridi himself is in favour. In the literature on belief revision, however, a weaker sense of information is often adopted, whereby (declarative) information is meaningful data, not perforce truthful. This is connected to what is sometimes called 'soft information', see, e.g., van Benthem, 2011; van Benthem & Smets, 2015.)

KRI, as well as all the other TSIMs, is closed with respect to conjunction elimination:

[2] The standard Hintikkan framework already embeds the impulse to parameterize knowledge to information: it models agent a's epistemic situation as a set of possible worlds, most straightforwardly understood as a's information or knowledge. Ascriptions $K_a\varphi$ are then naturally understood as capturing what is *knowable on this basis*. Various proposed readings draw out the conditionality. Consider the preferred interpretation in (Hintikka, 1962): $K_a p$ means roughly 'Relative to her knowledge, a is permitted to infer p'. Or consider a purely descriptive interpretation raised in Sect. 2.10 of (Hintikka, 1962): 'It follows from what a knows that p'.

(**Simplification**) $K^{\phi}(\psi \wedge \chi) \vDash K^{\phi}\psi \qquad K^{\phi}(\psi \wedge \chi) \vDash K^{\phi}\chi$

Proof. We do the first one (for the second, replace ψ with χ appropriately). Let $w \Vdash K^{\phi}(\psi \wedge \chi)$. By (S$X$), for all w_1 such that $wR_{\phi}w_1$, $w_1 \Vdash \psi \wedge \chi$, thus by (S$\wedge$), $w_1 \Vdash \psi$. Also, $t(\psi \wedge \chi) = t(\psi) \oplus t(\chi) \leq t(\phi)$, thus $t(\psi) \leq t(\phi)$. Then, by (SX) again, $w \Vdash K^{\phi}\psi$. □

The 'tracking' notion of knowledge due to Nozick (1981) does not necessitate that one who knows a conjunction is positioned to know the conjuncts. According to Kripke (2011a), this is a damning defect for Nozick's approach. KRI is free from such a defect. The companion of Simplification is:

(**Adjunction**) $\{K^{\phi}\psi, K^{\phi}\chi\} \vDash K^{\phi}(\psi \wedge \chi)$

Proof. Let $w \Vdash K^{\phi}\psi$ and $w \Vdash K^{\phi}\chi$, that is, by (SX): for all w_1 such that $wR_{\phi}w_1$, $w_1 \Vdash \psi$ and $w_1 \Vdash \chi$, so by (S$\wedge$) $w_1 \Vdash \phi \wedge \psi$. Also, $t(\psi) \leq t(\phi)$ and $t(\chi) \leq t(\phi)$, thus $t(\psi) \oplus t(\chi) = t(\psi \wedge \chi) \leq t(\phi)$. Then, by (S$X$) again, $w \Vdash K^{\phi}(\psi \wedge \chi)$. □

All of the TSIMs explored in this paper share two further core features: (a) they are nonmonotonic, and (b) they display their hyperintensionality by invalidating, among other things, Closure under strict implication. Thus, in particular, for KRI:

(**Monotonicity**) $K^{\phi}\psi \nvDash K^{\phi \wedge \chi}\psi$

Countermodel. Let $W = \{w, w_1\}$, w R_p-accesses nothing, $wR_{p \wedge r}w_1$, $w_1 \nVdash q$, $t(p) = t(q) = t(r)$. Then $w \Vdash K^{p}q$, but $w \nVdash K^{p \wedge r}q$. □

(**Closure under** $\prec$) $\{K^{\phi}\psi, \psi \prec \chi\} \nvDash K^{\phi}\chi$

Countermodel. Let $W = \{w, w_1\}$, $wR_p w_1$, $w \nVdash q$, $w_1 \Vdash q$, $w_1 \Vdash r$, $t(p) = t(q) \neq t(r)$. Then $f_p(w) \subseteq |q|$ and $t(q) \leq t(p)$, thus by (SX), $w \Vdash K^{p}q$. Also, $|q| \subseteq |r|$, thus by (S$\prec$), $w \Vdash q \prec r$. But although $f_p(w) \subseteq |r|$, $t(r) \nleq t(p)$, thus $w \nVdash K^{p}r$. □

(a) Failure of Monotonicity comes from the TSIMs' variable strictness: $f_{\phi}(w)$ can differ from $f_{\phi \wedge \chi}(w)$. In particular, for KRI: the addition of new information may reduce one's knowledge. (b) Failure of Closure comes from the fact that $\prec$ can take one off-topic, whereas information is topic-sensitive: although all the ψ-worlds are χ-worlds, thus all the ϕ-selected

ψ-worlds are χ-worlds, ϕ may be information about the topic of ψ yet not be information about the topic of χ. Thus, given ϕ, one can come to know ψ but not χ even if there just is no way for ψ to be true while χ is not.

In the *Mind* paper with Peter, I argued that *both* the failure of Monotonicity *and* that of Closure help with the Kripke-Harman Dogmatism Paradox (Harman, 1973; Kripke, 2011b), whereby knowing agents seem to be immune to rational persuasion with new evidence. Suppose that agent x knows φ on the basis of information I_1. Suppose that evidence e, when received, disconfirms φ. Now, φ entails that $\neg(e \wedge \neg\varphi)$: hence φ entails that e, if true, is misleading evidence. So Closure yields that x knows that e (if true) is misleading on the question of φ. Now suppose that x receives new information I_2, positioning her to know e. In this case, by Monotonicity, x knows that e is misleading. In general, it seems that x can always ignore new countervailing evidence! Central strategies on the market deal with this paradox by targeting precisely Monotonicity (e.g., Harman, 1973) or Closure (e.g., Sharon & Spectre, 2010, 2017).

What of the venerable Platonic insight that knowledge as *epistéme* must, in some sense, be stable? It's captured by KRI's validating Transitivity. This is invalid for the other TSIMs explored below due to their variable strictness, but it holds for KRI thanks to (C0):

(**Transitivity**) $\{K^{\phi}\psi, K^{\psi}\chi\} \vDash K^{\phi}\chi$

Proof. Assume that $w \Vdash K^{\phi}\psi$ and $w \Vdash K^{\psi}\chi$. Thus: $\forall w_1(wR_{\phi}w_1 \Rightarrow w_1 \Vdash \psi)$ & $t(\psi) \leq t(\phi)$ and $\forall w_2(wR_{\psi}w_2 \Rightarrow w_2 \Vdash \chi)$ & $t(\chi) \leq t(\psi)$. Then $t(\chi) \leq t(\psi) \leq t(\phi)$. Further: by (C0), we have that $|\psi| \subseteq f_{\psi}(w)$ and, by (SX), that $f_{\psi}(w) \subseteq |\chi|$. Thus, $|\psi| \subseteq |\chi|$. Now, by (SX) again, we have that $f_{\phi}(w) \subseteq |\psi|$. Hence, $f_{\phi}(w) \subseteq |\chi|$. $\square$

Knowledge is stable in that old knowledge cannot be lost as new one is accumulated. The intuitive case for Monotonicity is that it captures the core idea of the stability of knowledge. KRI suggests a different hypothesis: knowledge is stable in that it respects Transitivity. Suppose χ is known on the basis of information ψ. And suppose that one's information is refined insofar as new information χ is received upon which knowledge of ψ can be based. Transitivity says that χ is still knowable: no knowledge is lost in the update from ψ to ϕ.

Failure of Closure helps with Cartesian skepticism (Dretske, 1970). One's ordinary empirical information, delivered via sensory perception, positions one to know mundane facts, e.g., that one has hands. Now, having hands is incompatible with being a bodiless brain-in-vat whose phenomenal experience is systematically misleading. Yet it seems implausible that ordinary empirical information eliminates the possibility of radical sensory deception.

What of the intuition that knowability is closed under deduction? KRI validates *Closure Over Known Implication and Topic*:

(**COOKIT**) $\{K_A B, K_A(B \prec C)\} \vDash K_A C$

Proof. Let $w \Vdash K_A B$ and $w \Vdash K_A(B \prec C)$. By the former and (SX), for all w_1 such that $wR_A w_1$, $w_1 \Vdash B$, and $t(B) \leq t(A)$. By the latter and (SK) again, for all w_1 such that $wR_A w_1$, $w_1 \Vdash B \prec C$. Thus for all w_1 such that $wR_A w_1$, $w_1 \Vdash C$. Also, $t(B \prec C) = t(B) \oplus t(C) \leq t(A)$, thus $t(C) \leq t(A)$. Thus by (SX), $w \Vdash K_A C$. $\square$

The Cartesian threat is reduced for COOKIT: it allows that, on pain of circularity, deductions from one's mundane empirical knowledge cannot yield knowledge that the senses are reliable. But it assures that mundane information positions one to know every *mundane* consequence of that information.

3.2 Imagination as mental simulation

I used (C1), and tentatively added (C3), in (Berto, 2017a), a paper on imagination as mental simulation. (C1) has it that all the ϕ-selected worlds will be ϕ-worlds – worlds making ϕ true. With this constraint in place, we relabel our '$X^\phi \psi$' as '$I^\phi \psi$' and read it as 'Given input ϕ, one imagines ψ', or, less tersely, 'In an act of imagination starting with input ϕ, one imagines ψ'. We now read the accessibility $wR_\phi w_1$ as: 'w_1 is one of the worlds where things are as imagined (at w), starting with input ϕ'.

'Imagination' is highly ambiguous: we use it to refer to all sorts of intentional activities, from free mental wandering to daydreaming and hallucinating. What we want to model here, though, is the kind of imaginative exercise we engage in when we want to anticipate what will happen if such-and-so turns out to be the case ('What will I do if I can't pay my mortgage tomorrow?'), or when, counterfactually, we want to ascertain responsibilities ('Would he have been hit by the car, had the driver respected the speed limit?'). We simulate alternatives to reality in our mind, to explore what would and would not happen if they were realized. It is widely agreed in cognitive psychology as well as philosophy (Byrne, 2005; Kind & Kung, 2016; Markman, Klein, & Surh, 2009) that imagination as mental simulation is of epistemic value, not only to improve future performance (that's on the know-how side), but also to make contingency plans by successfully anticipating future outcomes, or by learning from mistakes via the consideration of alternative courses of action (that's on the know-that side).

Acts of imagination as mental simulation have a deliberate starting point, given by the initial input: we set out to target an explicit content, *that* ϕ. It has been generally acknowledged (Langland-Hassan, 2016; Van Leeuwen, 2016; Wansing, 2015; Williamson, 2016) that imagination, so understood, is episodic and voluntary in ways belief is not: you can imagine that all of London has been painted blue (and try to guess how Londoners would react to that), but, having overwhelming evidence of the contrary, you cannot make yourself believe it.

In their well-known cognitive model of mental simulation, Nichols and Stich (2003) have 'an initial premiss or set of premisses, which are the basic assumptions about what is to be pretended' (p. 24). This may be made up by the conceiver ('Now let us imagine what would happen if . . . '), or it may be given as an external instruction (think of going through a novel and take the sentences you read as your sequential input). But also, we integrate the explicit input ϕ with background information we import, contextually, depending on ϕ and what we know or believe: once the initial input is in, Nichols and Stich (2003) claim, 'children and adults elaborate the pretend scenarios in ways that are not inferential at all', filling in the explicit instruction with 'an increasingly detailed description of what the world would be like if the initiating representation were true' (pp. 26-28).

The additional details come from our information base (Van Leeuwen, 2016, p. 95): we imagine the last meeting between Heathcliff and Catherine in *Wuthering Heights*. We represent Heathcliff dressed as an an Eighteenth-Century country gentleman, not as a NASA astronaut. The text of the novel never says this explicitly, nor do we infer this from the text via sheer logic. Rather, we import such information into the represented situation, based of what we know: we know that the story is temporally located in the Eighteenth Century, and we assume, lacking information to the contrary from the text, that Heathcliff is dressed as we know country gentlemen were dressed at the time. The variability of strictness of our TSIMs now accounts for the contextual selection of the information we import in an act of imagination when we integrate its explicit input.

Also, in reality-oriented mental simulation we do not indiscriminately import unrelated contents into the conceived scenarios: '[We require] that the world be imagined as it is *in all relevant respects*' (Kind, 2016, p. 153). What is imported is constrained by what is *on-topic* with respect to the input. This is, again, the job of our topic-preservation filter. Topicality is a distinguishing feature of reality-oriented mental simulation, as opposed to free-floating mental wandering: you know that Nuku'alofa is the capital of Tonga, but this is immaterial to your imagining Catherine and Heathcliff's adventures as per Brontë's book, in so far as such adventures do not involve Tonga at all. The story is not *about* that. So you will not, in general, import such irrelevant content in your scenario.

In this setting, a reflexivity principle holds, thanks to (C1) (the proof is trivial), securing that the initial input is always imagined:

(**Success**) $\vDash I^{\phi}\phi$

(I flag here that this may make the framework unsuitable to model cases of so-called 'imaginative resistance': see Gendler, 2000). On the other hand, lacking (C0), Factivity fails – and rightly so, for mental simulation isn't factive: I imagine Crispin Wright working in Stirling, ϕ, but I develop the scenario in my imagination by importing my background (false) belief that Stirling is in England (may the Scots forgive me). I imagine that Crispin works in an English city, $I^{\phi}\psi$. ϕ is true, but it doesn't follow that ψ is true, Crispin works in an English city.

Two disjunction-involving features that hold for all of our TSIMs deserve some comment in the imagination reading. Yablo's 'paradigm of non-inclusion', that is, of (classically valid) entailment which is not aboutness-preserving, is the entailment from a formula to a disjunction between it and something else. This needs to fail, in particular, for the aboutness of imagination. When one imagines in an act whose explicit input is ϕ, that ψ, one does not thereby imagine a disjunction between the latter and an unrelated χ. Intuitively enough, the mental simulator need not be aware of that disconnected χ at all. Thus we need, and we get, a failure of:

(**Addition**) $I^{\phi}\psi \nvDash I^{\phi}(\psi \vee \chi)$

Countermodel. Let $W = \{w, w_1\}$, $wR_p w_1$, $w_1 \Vdash q$, $t(p) = t(q) \neq t(r)$. Then $t(q) \leq t(p)$, so by (SX), $w \Vdash I^p q$. But $t(q \vee r) = t(q) \oplus t(r) \nleq t(p)$, thus $w \nVdash I^p(q \vee r)$. □

(Notice that the inference fails for the right reason: although $\psi \vDash \psi \vee \chi$, disjunction brings in irrelevant, alien content.)

The other disjunction-involving issue has to do with the fact that imagination generally under-determines its contents (this is true, I think, for intentional states by default). We imagine things vaguely, without this entailing that we imagine vague things. An often-used example: you imagine the crowded streets of New York and you think about a complex scenario involving cabs running around, people in restaurants, skyscrapers, etc. You do not imagine all the details, but you want the details to be there, so to speak. Although you do not imagine the city building by building, New York is not a vague object in your scenario, one with an objectively indeterminate number of buildings. Either the number of buildings of New York is odd, or it is even. But you do not imagine it either way. So we need, and we get, a failure of:

(**Distribution**) $I^{\phi}(\psi \vee \chi) \nvDash I^{\phi}\psi \vee I^{\phi}\chi$

Countermodel. Let $W = \{w, w_1, w_2\}$, wR_pw_1, wR_pw_2, $w_1 \Vdash q$ but $w_1 \nVdash r$, $w_2 \Vdash r$ but $w_2 \nVdash q$, $t(p) = t(q) = t(r)$. Then by (S$\vee$), $w_1 \Vdash q \vee r$ and $w_2 \Vdash q \vee r$, so for all w_x such that wR_pw_x, $w_x \Vdash q \vee r$. Also, $t(q \vee r) = t(q) \oplus t(r) \leq t(p)$, thus by (S$X$), $w \Vdash I^p(q \vee r)$. However, $w \nVdash I^pq$ and $w \nVdash I^pr$ for both q and r fail at some R_p-accessible world. Thus by (S$\vee$), $w \nVdash I^pq \vee I^pr$. □

In the *Phil Studies* paper, I tentatively added Condition (C3) above. Let us look at it again:

(C3) $f_{\phi}(w) \subseteq |\psi| \ \&\ f_{\psi}(w) \subseteq |\phi| \Rightarrow f_{\phi}(w) = f_{\psi}(w)$

(C3) was labeled there as a 'Principle of Equivalents in Imagination': in the context of the interpretation of TSIMs as imagination operators, one reads it as saying that when all the selected ϕ-worlds make ψ true and vice versa, ϕ and ψ are equivalent in that, when we imagine either, we look at the same set of worlds. (C3) validates a nice Substitutivity principle for equivalents in imagination:

(**Substitutivity**) $\{I^{\phi}\psi, I^{\psi}\phi, I^{\phi}\chi\} \vDash I^{\psi}\chi$

Proof. Suppose $w \Vdash I^{\phi}\psi$, $w \Vdash I^{\psi}\phi$, $w \Vdash I^{\phi}\chi$. By (SX), these entail, respectively, (a) $f_{\phi}(w) \subseteq |\psi|$ and $t(\psi) \leq t(\phi)$, (b) $f_{\psi}(w) \subseteq |\phi|$ and $t(\phi) \leq t(\psi)$, (c) $f_{\phi}(w) \subseteq |\chi|$ and $t(\chi) \leq t(\phi)$. From (a) and (b) we get $f_{\phi}(w) = f_{\psi}(w)$ (by (C3)) and $t(\phi) = t(\psi)$ (by antisymmetry of topic parthood). From these and (c) we get $f_{\psi}(w) \subseteq |\chi|$ and $t(\chi) \leq t(\psi)$. Thus by (SX) again, $w \Vdash I^{\psi}\chi$. □

Substitutivity says that 'equivalents in imagination' ϕ and ψ can be replaced *salva veritate* as indexes in $I^{\cdots}$. This seems right, in spite of the many hyperintensional distinctions we may draw in our mind. For suppose that *bachelor* and *unmarried man* are for you equivalent in imagination: you are so firmly aware of their meaning the same, that you cannot imagine someone being one thing without imagining him being the other ($I^\phi\psi$ & $I^\psi\phi$ entails $t(\phi) = t(\psi)$: equivalents in imagination are always about the same thing for the conceiving subject). Thus, $I^\phi\psi$, when you imagine that John is unmarried, you imagine that he is a bachelor, and $I^\psi\phi$, when you imagine that John is a bachelor, you imagine that he is unmarried. Suppose $I^\phi\chi$: as you imagine that John is unmarried, you imagine that he has no marriage allowance. Then the same happens as you imagine that he is a bachelor, $I^\psi\chi$.

I said that the addition of (C3) was 'tentative'. That's because it also validates a kind of Special Transitivity principle which has good instances, but, in the context of imagination, may face counterexamples:

(**Special Transitivity**) $\{I^\phi\psi, I^{\phi\wedge\psi}\chi\} \vDash I^\phi\chi$

Proof. Suppose (a) $w \Vdash I^\phi\psi$ and (b) $w \Vdash I^{\phi\wedge\psi}\chi$. From (a), Success, and Adjunction we get $w \Vdash I^\phi(\phi \wedge \psi)$, thus, by (S$X$), $f_\phi(w) \subseteq |\phi \wedge \psi|$ and $t(\phi \wedge \psi) \leq t(\phi)$. Also, $w \Vdash I^{\phi\wedge\psi}\phi$ (from Success $\vDash I^{\phi\wedge\psi}(\phi \wedge \psi)$ and Simplification). By (SX) again,$f_{\phi\wedge\psi}(w) \subseteq |\phi|$ and (of course) $t(\phi) \leq t(\phi \wedge \psi)$. Thus, by (C3) $f_\phi(w) = f_{\phi\wedge\psi}(w)$, and $t(\phi \wedge \psi) = t(\phi)$ (by antisymmetry of content parthood). Next, from (b) and (SX) again, $f_{\phi\wedge\psi}(w) \subseteq |\chi|$ and $t(\chi) \leq t(\phi \wedge \psi)$. Therefore, $f_{\phi\wedge\psi}(w) = f_\phi(w) \subseteq |\chi|$ and $t(\chi) \leq t(\phi) = t(\phi \wedge \psi)$. Thus by (S$X$) again, $w \Vdash I^\phi\chi$. □

Special Transitivity has good instances. $I^\phi\psi$: as you imagine that John has won the lottery, you imagine that he has a lot of money. $I^{\phi\wedge\psi}\chi$: as you imagine that John has won the lottery and has a lot of money, you imagine that he is to pay substantive amounts of taxes. Thus, $I^\phi\chi$: as you imagine that John has won the lottery, you imagine that he is to pay substantive amounts of taxes.

Special Transitivity for the imagination operator may face counterexamples. Here's a situation suggested by Claudio Calosi, that may do. $I^\phi\psi$: given the input that I am wearing a red shirt in Pamplona, I imagine that I am being chased by bulls. $I^{\phi\wedge\psi}\chi$: given the input that I am being chased by bulls on the streets of Pamplona while wearing a red shirt, I imagine that I die on the street. But it's not the case that $I^\phi\chi$: Given that I am wearing a red shirt in Pamplona, I don't imagine that I die on its streets. So it might be that (C3) has to go for imagination, in spite of its usefulness.[3]

3.3 Hyperintensional belief revision

One automatically gets all of (C1)-(C4) if one imposes a plausibility ordering on W, thereby getting a system of spheres in the style of (Lewis, 1973). One adds to the semantics a function, \$, assigning to each w a finite set of nested subsets of W (the spheres): $\$(w) = \{S_0^w, S_1^w, ..., S_n^w\}$, with $n \in \mathbb{N}$, such that $S_0^w \subseteq S_1^w \subseteq ... \subseteq S_n^w = W$. Next, for each $\phi \in \mathcal{L}$ and $w \in W$, $f_\phi(w)$ goes thus: if $|\phi| = \emptyset$, then $f_\phi(w) = \emptyset$. Otherwise, $f_\phi(w) = S_i^w \cap |\phi|$, where $S_i^w \in \$(w)$ is the smallest sphere such that $S_i^w \cap |\phi| \neq \emptyset$.

In the *Erkenntnis* paper (Berto, 2018), I used this set-up to deal with AGM belief revision theory (Alchourrón, Gärdenfors, & Makinson, 1985). The feature of TSIMs under the spotlight then, is hyperintensionality. We relabel '$X^\phi\psi$' as '$B^\phi\psi$' and read it as 'Conditional on ϕ, one believes that ψ', or: 'After revising by ϕ, one believes that ψ'.

The accessibility $wR_\phi w_1$ now has us look at the *most plausible* worlds w_1 where ϕ holds, given the system of beliefs of the agent located at w, as modeled by the spheres. As in the (Grove, 1988) reformulation of the Lewisian insight, we don't demand that $w \in S_0^w$, that is, the relevant world be in the innermost sphere: in Lewis' terminology, we have a system of spheres which is not even weakly centered. That's because our spheres do not express objective world similarity, but subjective world plausibility, or belief entrenchment. The innermost sphere at the core, S_0^w, gives the most plausible worlds for the agent located at w; w itself need not be among the innermost worlds, for the agent may have false beliefs.

The relevant TSIM reduces the logical idealization of cognitive agents affecting similar operators in doxastic and epistemic logics, as well as in AGM. The first postulate for belief revision in (Alchourrón et al., 1985), (K*1), has it that $K * \phi$ (belief set K after revision by ϕ) is closed under the full strength of classical logical consequence. Postulate (K*5) trivializes belief sets revised in the light of inconsistent information: if ϕ is a logical inconsistency, then $K * \phi = K_\perp$, the trivial belief set; agents who revise

[3]In (Berto, 2017a), I suggested that if one resorts to an extended semantics that uses impossible worlds (of a non-adjunctive kind) besides possible ones, one can have (C3) and its welcome child, Substitutivity, without having Special Transitivity because Simplification and/or Adjunction can fail in such a framework. The proof of Special Transitivity essentially uses both, whereas the one of Substitutivity doesn't.

via inconsistent inputs trivially believe everything. And postulate (K*6) requires that, if ϕ and ψ are logically equivalent, then $K * \phi = K * \psi$, that is, revising by either gives the same belief set.

These principles are rather implausible for agents like us all. Against (K*1), our belief states need not be closed under classical logical consequence (perhaps under any kind of monotonic logical consequence: see, e.g., Jago, 2014 for extended discussion). Against (K*5), we do not trivially believe everything just because we occasionally hold inconsistent beliefs, and we should not be modeled as undergoing a trivialization of our belief system just because we can be, as we occasionally are, exposed to inconsistent information (given that information is not factive). Against (K*6), it is well known that how we revise our beliefs, as well as our preferences, is subject to what psychologists call *framing effects* (Kahneman & Tversky, 1984): logically or necessarily equivalent contents can trigger different revisions depending on how they are presented. Agents may revise their beliefs in one way when told they have 60% chances of succeeding in a task, in another way when told they have 40% chances of failing.

Our TSIM now takes care of all of these. Success (guaranteed again by (C1)) mirrors the Success postulate of AGM – After revising by ϕ, one does believe ϕ:

(**Success**) $\vDash B^{\phi}\phi$

(Notice that there is no problem with this holding unrestrictely, as what we are modeling is static belief revision. Things may go differently for dynamic belief revision when, e.g., Moore formulas are concerned: see van Benthem & Smets, 2015.)

Belief revision is not automatically trivialized by incoming inconsistent information. Against principles such as AGM's (K*5), the following ensures that we do not come to believe arbitrary, irrelevant things just because we have taken on board explicitly inconsistent information:

(**Explosion**) $\nvDash B^{\phi\wedge\neg\phi}\psi$

Countermodel. Let $W = \{w\}$, $t(p) \neq t(q)$. $|p \wedge \neg p| = \emptyset$, thus $f_{p\wedge\neg p}(w) = \emptyset \leq |q|$. However, $t(q) \nleq t(p \wedge \neg p) = t(p) \oplus t(\neg p) = t(p)$. Thus, by (S$X$), $w \nVdash B^{p\wedge\neg p}q$. □

Although there is no possible world where a contradiction is true, inconsistent information may still be about *something*. In general $\phi \wedge \neg\phi$ is not contentless: its topic is whatever ϕ is about, and this may not include the topic of ψ (*Snow is white and not white* is about snow's being white, not about grass' being purple).[4]

The counterpart of AGM's (K*6) fails (where $\phi \equiv \psi$ abbreviates $\phi \prec \psi \wedge \psi \prec \phi$), thereby modelling:

(**Framing**) $\{B^{\phi}\chi, \phi \equiv \psi\} \nvDash B^{\psi}\chi$

Countermodel. Let $W = \{w, w_1\}$, $wR_p w_1$, $wR_q w_1$, $w \nVdash p$, $w \nVdash q$, $w_1 \Vdash r$, $t(p) = t(r) \neq t(q)$. Then $f_p(w) \subseteq |r|$ and $t(r) \leq t(p)$, thus by (SX), $w \Vdash B^p r$. Also, by (C1), $w_1 \Vdash p$ and $w_1 \Vdash q$, thus $|p| \subseteq |q|$ and $|q| \subseteq |p|$. Then by (S$\prec$) and (S$\wedge$), $w \Vdash p \equiv q$. But although $f_q(w) \subseteq |r|$, $t(r) \nleq t(q)$, thus $w \nVdash B^q r$. $\square$

Again, failure of topic-preservation does the hyperintensional trick of differentiating necessarily equivalent contents. This invalidity allows a proper appreciation of framing effects: after being informed that one's probability of making it to the short list is $1/3$, one believes that one should apply for the job. But after being informed that one's probability of failing the short list is $2/3$, one does not believe that it's worth applying. There is no way that the chances of making it are $1/3$ without the chances of failing being $2/3$ and vice versa, but one has been caught into Framing.

However, by having (C3) (relabeled, for obvious reasons, as 'Principle of Equivalents in Plausibility' in the context of $B^{\phi}\psi$) the system of spheres allows a *limited recovery* of the idea encoded in principles like AGM's (K*6), thanks to Substitutivity – looking at it again:

(**Substitutivity**) $\{B^{\phi}\psi, B^{\psi}\phi, B^{\phi}\chi\} \vDash B^{\psi}\chi$

'Equivalents in plausibility' are now formulas ϕ and ψ such that, when we revise by either, we come to believe the other. Substitutivity now says that such equivalents can be replaced *salva veritate* as inputs for belief revision: when we revise by either, we come to have the same beliefs.[5] The

[4]Here's where non-classical frameworks get a revenge. Even if our TSIMs are technically not explosive, they do satisfy 'small explosion' principles like $\vDash B^{\phi \wedge \neg\phi \wedge \psi}\neg\psi$ (for, trivially, $\phi \wedge \neg\phi \wedge \psi$ is true nowhere, *and* topicality is preserved here). A framework expanded to include non-normal or impossible worlds where a contradiction can be true would help against such small detonations. I have used such a framework to model intentional operators in (Berto, 2014, 2017b).

[5]The behavior in our framework of (counterparts of) AGM principles other than (K*1),

hyperintensional belief revision operator is, overall, well-behaved qua non-monotonic operator (it is very much unlike the KRI, in this respect). Neither Factivity nor Transitivity holds, whereas it validates Success, Restricted Transitivity, and also the following (thanks to (C3) again):

(**Cautious Monotonicity**) $\{B^{\phi}\psi, B^{\phi}\chi\} \vDash B^{\phi\wedge\psi}\chi$

Proof. Suppose (a) $w \Vdash B^{\phi}\psi$ and (b) $w \Vdash B^{\phi}\chi$. From (a), Success ($\vDash B^{\phi}\phi$), and Adjunction, we get $w \Vdash B^{\phi}(\phi\wedge\psi)$, thus by (S$X$), $f_{\phi}(w) \subseteq |\phi\wedge\psi|$. Also, $w \Vdash B^{\phi\wedge\psi}\phi$ (from Success $\vDash B^{\phi\wedge\psi}(\phi\wedge\psi)$ and Simplification), so by (SX) again, $f_{\phi\wedge\psi}(w) \subseteq |\phi|$. Then, by (C3), $f_{\phi}(w) = f_{\phi\wedge\psi}(w)$. From (b) and (S$X$) again, we get $f_{\phi}(w) \subseteq |\chi|$, thus $f_{\phi\wedge\psi}(w) \subseteq |\chi|$. Also, $t(\chi) \leq t(\phi) \oplus t(\psi) = t(\phi\wedge\psi)$. Thus, $w \Vdash B^{\phi\wedge\psi}\chi$. □

Special Transitivity and Cautious Monotonicity are defended in (Board, 2004, p. 56), as 'principles of informational economy'. They correspond to the Cut and Cautious Monotonicity principles for non-monotonic logic considered in (Kraus, Lehmann, & Magidor, 1990). By satisfying these as well as Success, our $B^{\phi}\psi$ complies, thus, with Gabbay's (1985) minimal conditions for nonmonotonic entailments.

(K*2), (K*5) and (K*6) is somewhat less interesting, for that's not where the original features of the theory emerge. But in (Berto, 2018) I mentioned a peculiar asymmetry related to the AGM principles (K*7) and (K*8). A natural counterpart of (K*7) (see Board, 2004, p. 55) fails in our semantics:

$$\{\neg B^{\phi}\neg\psi, B^{\phi\wedge\psi}\chi\} \nvDash B^{\phi}(\psi \prec \chi)$$

Countermodel. Let $W = \{w\}, f_p(w) = \emptyset, f_{p\wedge q} = \emptyset, t(p) \neq t(q) = t(r)$. Then by (S$X$), $w \nVdash B^{p}\neg q$ because $t(\neg q) = t(q) \nleq t(p)$, so $w \Vdash \neg B^{p}\neg q$; and $w \Vdash B^{p\wedge q}r$, because (trivially) $f_{p\wedge q}(w) \subseteq |r|$, and $t(r) = t(q) \leq t(p) \oplus t(q) = t(p \wedge q)$. However, $w \nVdash B^{p}(q \prec r)$, because $t(q \prec r) = t(q) \oplus t(r) = t(q) \nleq t(p)$. □

On the other hand, a natural counterpart of (K*8) (see (Board, 2004), Ibid), obtained by flipping premise and conclusion in the former, holds:

$$\{\neg B^{\phi}\neg\psi, B^{\phi}(\psi \prec \chi)\} \vDash B^{\phi\wedge\psi}\chi$$

Proof. Suppose (a) $w \Vdash \neg B^{\phi}\neg\psi$ and (b) $w \Vdash B^{\phi}(\psi \prec \chi)$. By (a) and (S$\neg$), $w \nVdash B^{\phi}\neg\psi$, that is: either $f_{\phi}(w) \nsubseteq |\neg\psi|$, that is, $f_{\phi}(w) \cap |\psi| \neq \emptyset$, or $t(\neg\psi) = t(\psi) \nleq t(\phi)$. But it can't be the latter, because by (b) and (SX), $t(\psi \prec \chi) = t(\psi) \oplus t(\chi) \leq t(\phi)$, thus in particular $t(\psi) \leq t(\phi)$; so it must be the former. Applying Condition (C4) to it, $f_{\phi\wedge\psi}(w) \subseteq f_{\phi}(w)$. By (C1), $f_{\phi\wedge\psi}(w) \subseteq |\phi\wedge\psi|$, so by (S$\wedge$), $f_{\phi\wedge\psi}(w) \subseteq |\psi|$. By (b) and (S$X$) again, $f_{\phi}(w) \subseteq |\psi \prec \chi|$. Putting things together: $f_{\phi\wedge\psi}(w) \subseteq f_{\phi}(w) \subseteq |\psi \prec \chi|$, so $f_{\phi\wedge\psi}(w) \subseteq |\psi \prec \chi|$; and since $f_{\phi\wedge\psi}(w) \subseteq |\psi|$, then by modus ponens $f_{\phi\wedge\psi}(w) \subseteq |\chi|$. Also, by (b) again, $t(\psi) \oplus t(\chi) \leq t(\phi) \leq t(\phi\wedge\psi)$, thus $t(\chi) \leq t(\phi\wedge\psi)$. Thus, by (S$X$), $w \Vdash B^{\phi\wedge\psi}\chi$. □

4 Further work

Some work in TSIM theory, beyond what I have summarized in this paper, is already being carried out. Sound and complete axiomatizations of the semantics proposed above are being developed by Alessandro Giordani (forthcoming) and Aybüke Ozgün. Heinrich Wansing is working on how to combine the semantics of imagination from Section 3.2 above with his agentive STIT logic of imagination (Olkhovikov & Wansing, 2017; Wansing, 2015). And there are also developments in the direction of dynamic epistemic and doxastic logic (Ozgün again, and Peter Hawke): the hyperintensional belief revision operator is static, but we are exploring some ideas on how we make the framework dynamic.

Further possible areas of research include, e.g., moving to a first-order language, a nice question then being how we want topicality to work there. Perhaps the biggest open issue concerning topicality is the following. All the ways of playing with the TSIMs explored so far tamper only with accessibilities. None tampers with their topic-sensitivity. All our $X^{\phi}\psi$'s embed a rather draconian topicality or aboutness constraint: ψ must be fully on-topic with respect to ϕ, or what ψ is about must be fully included in what ϕ is about. I haven't explored how to play with the mereology of topics yet, but there are reasons to relax such a constraint, allowing, e.g., partial overlap of topics rather than full inclusion, for various purposes. If we allow ψ to only be partly on topic with respect to ϕ, this brings '$X_{\phi}\psi$' in the vicinity of a variably strict relevant conditional. And there are surely other options for more complicated topic-embeddings. It's a nice territory and I hope more people get interested in exploring it.

References

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

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

Asheim, G., & Sövik, Y. (2005). Preference-based belief operators. *Mathematical Social Sciences*, *50*, 61–82.

Baltag, A., & Smets, S. (2008). A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, & M. Wooldridge

(Eds.), *Logic and the Foundations of Game and Decision Theory* (pp. 9–58). Amsterdam: Amsterdam University Press.

Berto, F. (2014). On conceiving the inconsistent. *Proceedings of the Aristotelian Society*, *114*, 21–103.

Berto, F. (2017a). Aboutness in imagination. *Philosophical Studies*, *175*, 1871-86.

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

Berto, F. (2018). Simple hyperintensional belief revision. *Erkenntnis*.

Berto, F., & Hawke, P. (2018). Knowability relative to information. *Mind*.

Board, O. (2004). Dynamic interactive epistemology. *Games and Economic Behaviour*, *49*, 49–80.

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

Correia, F. (2004). Semantics for analytic containment. *Studia Logica*, *77*, 87–104.

Crane, T. (2013). *The Objects of Thought*. Oxford: Oxford University Press.

Dretske, F. (1970). Epistemic operators. *Journal of Philosophy*, *67*(24), 1007–1023.

Dretske, F. (1999). *Knowledge and the Flow of Information*. CLSI Publications.

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

Ferguson, T. (2014). A computational interpretation of conceptivism. *Journal of Applied Non-Classical Logics*, *24*(4), 333–367.

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

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

Fine, K. (2015). Angellic content. *Journal of Philosophical Logic*, *On line first*.

Floridi, L. (2015). Semantic conceptions of information. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/spr2015/entries/information-semantic`.

Gabbay, D. (1985). *Theoretical Foundations for Non-Monotonic Reasoning*. Berlin: Springer.

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

Girard, P., & Rott, H. (2014). Belief revision and dynamic logic. In A. Baltag & S. Smets (Eds.), *Johan van Benthem on Logic and Information Dynamics* (pp. 203–33). Dordrecht: Springer.

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

Harman, G. (1973). *Thought*. Princeton University Press.

Hawke, P. (2017). Theories of aboutness. *Australasian Journal of Philosophy*, 1-27.

Hintikka, J. (1962). *Knowledge and Belief. An Introduction to the Logic of the Two Notions*. Ithaca, NY: Cornell University Press.

Jago, M. (2014). *The Impossible. An Essay on Hyperintensionality*. Oxford: Oxford University Press.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–50.

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

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

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, *44*, 167-207.

Kripke, S. (2011a). Nozick on knowledge. In *Philosophical Troubles: Collected Papers, Volume 1*. Oxford University Press.

Kripke, S. (2011b). On two paradoxes of knowledge. In *Philosophical Troubles: Collected Papers, Volume 1*. Oxford University Press.

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

Leitgeb, H., & Segerberg, K. (2005). Dynamic doxastic logic: Why, how, and where to? *Synthèse*, *155*, 167–90.

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

Lindström, S., & Rabinowicz, W. (1999). DDL unlimited: Dynamic doxastic logic for introspective agents. *Erkenntnis*, *50*, 353–85.

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

Nichols, S., & Stich, S. (2003). *Mindreading. An Integrated Account of*

Pretence, Self-Awareness, and Understanding Other Minds. Oxford: Oxford University Press.

Nozick, R. (1981). *Philosophical Explanations*. Harvard University Press.

Olkhovikov, G., & Wansing, H. (2017). An axiomatic system and a tableau calculus for STIT imagination logic. *Journal of Philosophical Logic*, *On line first*.

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

Schipper, B. (2015). Awareness. In H. van Ditmarsch, J. Halpern, W. van der Hoek, & B. Kooi (Eds.), *Handbook of Epistemic Logic* (p. 79-146). London: College Publications.

Segerberg, K. (1995). Belief revision from the point of view of doxastic logic. *Bulletin of the IGPL*, *3*, 535–53.

Sharon, A., & Spectre, L. (2010). Dogmatism repuzzled. *Philosophical Studies*, *148*(2), 307–321.

Sharon, A., & Spectre, L. (2017). Evidence and the openess of knowledge. *Philosophical Studies*, *174*, 1001–1037.

Spohn, W. (1988). Ordinal conditional functions: a dynamic theory of epistemic states. In L. Hrper & B. Skyrms (Eds.), *Causation in Decision, Belief Change, and Statistics* (Vol. 2, p. 105-34). Dordrecht: Kluwer.

van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logic*, *17*, 129–55.

van Benthem, J. (2011). *Logical Dynamics of Information and Interaction*. Cambridge: Cambridge University Press.

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

van Ditmarsch, H. (2005). Prolegomena to dynamic logic for belief revision. *Synthèse*, *147*, 229–75.

van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2008). *Dynamic Epistemic Logic*. Dordrecht: Springer.

van Fraassen, B. (1969). Facts and tautological entailments. *Journal of Philosophy*, *66*, 477–87.

Van Leeuwen, N. (2016). The imaginative agent. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination* (p. 85-111). Oxford: Oxford University Press.

Wansing, H. (2015). Remarks on the logic of imagination. a step towards

understanding doxastic control through imagination. *Synthese*, *On line first*.
Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.
Williamson, T. (2016). Knowing by imagining. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination* (pp. 113–23). Oxford: Oxford University Press.
Yablo, S. (2014). *Aboutness*. Princeton: Princeton University Press.

Francesco Berto
University of St Andrews and University of Amsterdam
United Kingdom and The Netherlands
E-mail: fb96@st-andrews.ac.uk