Philosophical Studies
https://doi.org/10.1007/s11098-023-02007-4

Cognitive synonymy: a dead parrot?

Francesco Berto[1,2] · **Levin Hornischer**[3]

Accepted: 22 June 2023

Abstract
Sentences φ and ψ are *cognitive synonyms* for one when they play the same role in one's cognitive life. The notion is pervasive (Sect. 1), but elusive: it is bound to be hyperintensional (Sect. 2), but excessive fine-graining would trivialize it and there are reasons for some coarse-graining (Sect. 2.1). Conceptual limitations stand in the way of a natural algebra (Sect. 2.2), and it should be sensitive to subject matters (Sect. 2.3). A cognitively adequate individuation of content may be intransitive (Sect. 3) due to 'dead parrot' series: sequences of sentences $\varphi_1, \ldots, \varphi_n$ where adjacent φ_i and φ_{i+1} are cognitive synonyms while φ_1 and φ_n are not (Sect. 3.1). Finding an intransitive account is hard: Fregean equipollence won't do (Sect. 3.2) and a result by Leitgeb shows that it wouldn't satisfy a minimal compositionality principle (Sect. 3.3).*Sed contra*, there are reasons for transitivity, too (Sect. 3.4). In Sect. 4, we come up with a formal semantics capturing this jumble of desiderata, thereby showing that the notion is coherent. In Sect. 5, we re-assess the desiderata in its light.

Keywords Aboutness · Subject matter · Hyperintensionality · Synonymy · Defeasible reasoning · Cognitive content · Leitgeb impossibility result

1 Why it matters

Sometimes distinct sentences φ and ψ seem to play the same role in one's cognitive life. Ways to spell this out (to which we shall return) may include:

1 *Equipollence*: one cannot take either as true without taking the other as true.

✉ Francesco Berto
fb96@st-andrews.ac.uk

Levin Hornischer
Levin.Hornischer@lmu.de

[1] Arché Research Centre, University of St Andrews, St Andrews, Scotland, UK

[2] Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Amsterdam, The Netherlands

[3] Munich Center for Mathematical Philosophy (MCMP), LMU Munich, Munich, Germany

2 *Understanding*: what one understands, thinks about, or communicates when either is uttered, one does, when the other is.
3 *Inferences*: whatever one concludes—deductively, abductively, inductively, or in some other way—supposing either, one does, supposing the other.
4 *Belief update*: whatever belief change would be triggered by learning either, would also be triggered by learning the other.

Call this *cognitive synonymy*, which may be understood as two-way cognitive entailment: φ cognitively entails ψ when, supposing φ, one concludes that ψ; upon learning that φ, one would come to believe that ψ; etc. Plausibly, this should be a defeasible or non-monotonic sort of entailment: supposing or learning that φ (say, Tweety is a bird) one may conclude or come to believe that ψ (Tweety flies), but one may not conclude that, or retract one's conclusion or belief change, after supposing or learning that $\varphi \wedge \chi$ (say, Tweety is a penguin bird). Cognitive synonymy should be relative to the speaker's knowledge or belief base, and storage of concepts, but it can be generalized to groups of speakers and language-sharing communities: '*Ex contraditione quodlibet* is classically valid' and 'Pseudo-Scotus' Law is classically valid' are cognitive synonyms for most logicians; 'Groundhogs are rodent' and 'Woodchucks are rodent', for most zoologists; 'Paul is a bachelor' and 'Paul is an unmarried man', for most competent English speakers.

The notion is elusive, but it plays a role in various inquiries. Absolute synonymy, understood as substitutivity *salva veritate* in all contexts, is often taken by linguists as a merely theoretical notion for it cannot be empirically tested (Cruse, 2000, Stanojević, 2009). Linguists then employ a more tractable idea they call 'cognitive synonymy' as their working concept (Lyons, 1996; Murphy, 2003). Psychologists dealing with framing effects (Busby et al., 2018; Kahneman & Tversky, 1984) investigate how differently framed necessarily equivalent claims make one believe, desire, and choose differently. Which ones do *not* make one believe, desire, or choose differently? Plausibly, those are cognitive synonyms for one.

The notion may play a role for philosophers. One important test for accounts of propositional content is how they fare with respect to subject matters and embeddings within attitude ascriptions. Take a basic propositional language $\mathcal{L}$ with denumerably many atomic sentences $\mathcal{L}_{AT} : p, q, r, p_1, p_2, \ldots$, negation $\neg$, conjunction $\wedge$, disjunction $\vee$, the usual rules of well-formedness. Let us use $\varphi, \psi, \chi, \varphi_1, \varphi_2, \ldots$ as metavariables. The account of propositions as sets of possible worlds will assign to each atom p a truth set, $|p|$—the set of worlds where p is true—and truth sets to complex formulas recursively, forming a simple Boolean algebra: conjunction is set-theoretic intersection; disjunction is union; negation is complementation; entailment is inclusion. Qua individuation of propositional contents, this is deemed by many too coarse-grained for it conflates logically or necessary equivalents. It clashes with what Steve Yablo has called 'our sense of when sentences say the same thing' (Yablo, 2014, 2). 'Equilateral triangles are equiangular' and '$7 + 5 = 12$' seem to say different things although they are true at the same worlds: only one is about equilateral triangles and how they are like. It delivers what has been called the 'problem of logical omniscience' (Fagin & Halpern, 1988; Fagin et al., 1995; Jago, 2007) for attitude ascriptions: if mental states like knowing, believing, supposing, are understood as unmediated relations between subjects and propositions and the latter are sets of possible worlds, one cannot have different attitudes towards necessarily equivalent contents. If the standard account from epistemic logic is endorsed (Hintikka, 1962), cognitive agents are represented as knowing (believing, etc.) all logical truths and all logical consequences of what they know (ditto). Theorists looking for a cognitively more plausible account of content are likely to be in the vicinity of cognitive synonymy. Unlike the idealized Hintikkan agents, Joe Bloggs can have different attitudes towards necessarily equivalent contents. But Joe *will* have the same attitudes towards φ and ψ when they play the same role in Joe's cognitive life.

When one looks for principles governing cognitive synonymy in view of modeling it, however, one will find the task difficult. They may not even carve a coherent concept.

2 Hyperintensional troubles

Cognitive synonymy would have to be more fine-grained than standard intensional equivalence, i.e., propositional identity in the possible worlds account. How much more? One who advocates a hyperintensional account of content is bound to be asked the question of 'just "how hyper" hyperintensions are' (Jespersen and Duži, 2015, 527). As we attempt an answer for cognitive synonymy, we are pulled in different directions.

2.1 Fine-graining, coarse-graining

Surely we don't want to get as fine-grained as the syntax of the language, or there would be no point in having a semantics. Impossible worlds where $7 + 5$ differs from 12, or where the laws of logic may fail so that a conjunction can be true while one of its conjuncts isn't (Nolan, 1997, 2013), have been used to make hyperintensional distinctions. In particular, 'open worlds' (worlds not closed under any non-trivial notion of logical consequence) have been proposed to model the contents of logically anarchic mental states (Kiourti, 2010; Priest, 2016). As admitted even by some of their proponents (Berto & Jago, 2019; Jago, 2014), however, a naive approach whereby any set of open worlds can deliver contents gives excessive fine-graining. If our subject is located at world w, let $\mathcal{S}_w = \{w_1 \mid wRw_1\}$, the set of (open) worlds accessible from w via epistemic accessibility R, that is, the worlds which represent epistemic possibilities for the agent. Let $\mathcal{C} = \{\varphi \mid \text{for all } w_1 \in \mathcal{S}_w, \varphi \text{ is true at } w_1\}$, the set of formulas true at all of them. The agent's epistemic or doxastic state boils down to mere syntax: one knows or believes that φ at w just in case $\varphi \in \mathcal{C}$, and $\mathcal{C}$ can be a set of formulas lacking any (non-trivial) closure property. Once the content of epistemic states gets as structured as the syntax of the language, the worlds apparatus is pointless.

Independent reasons for coarse-graining come from the consideration of truth-functional composites. It seems that φ and $\varphi \wedge \varphi$ should be cognitive synonyms, and so should $\varphi \wedge \psi$ and $\psi \wedge \varphi$, for a speaker who understands the workings of Boolean, order-insensitive conjunction. The difference between 'Lucy went to the hospital and got sick' and 'Lucy got sick and went to the hospital' may matter when assessing medical responsibilities, but then the syntactic order encodes a temporal or causal order. When the idea of such an order is absent, we would resort to pragmatic cues to interpret a conversation. Mary says: 'Lucy is a lawyer and an expert chess player'; John replies: 'No: Lucy is an expert chess player and a lawyer'. We make sense of John's utterance pragmatically: perhaps he wanted to imply that playing chess is what matters for Lucy. We resort to pragmatics because we feel the two utterances say the same thing, we believe that's so for John, too, and we make sense of his move accordingly.

But if the purpose of cognitive synonymy and entailment is to capture what Joe Bloggs would believe given certain information, etc., won't we need extreme fine-graining anyway? Surely Joe can believe that $\varphi \wedge \psi$ without believing that φ, one may say, because he is, on occasion, cognitively incapacitated or too busy to perform a step of Simplification: insofar as inference qua psychological process happens in time, one could always get distracted and fail. Perhaps we do need impossible worlds where a conjunction can be true without its conjuncts being true.

But this is too fast. Simplification has special plausibility for attitude ascriptions, as argued by a number of authors. In his discussion of epistemic closure, Holliday (2012) takes Simplification (he focuses on knowledge operators: if one knows that $\varphi \wedge \psi$, then one knows that φ) as a *pure* (contrast *deductive*) closure principle. A deductive closure principle from $\varphi_1, \ldots, \varphi_n$ to ψ has it that if an agent comes to believe ψ starting from $\varphi_1, \ldots, \varphi_n$, by competent deduction, and all the while knowing each of $\varphi_1, \ldots, \varphi_n$, then the agent knows ψ. This can go wrong for Joe Bloggs. By contrast, Simplification, qua pure closure principle, is such that 'an agent *cannot* know $\varphi \wedge \psi$ without knowing ψ—regardless of whether the agent came to believe ψ by "competent deduction" from $\varphi \wedge \psi$' (Holliday, 2012, 15). Yablo (2014) calls this 'immanent closure'. In *Knowledge and Its Limits*, Williamson endorses Simplification as a pure or immanent closure principle for knowledge ascriptions:

> ... Knowledge of a conjunction is *already* knowledge of its conjuncts. ... There is no obstacle here to the idea that knowing a conjunction *constitutes* knowing its conjuncts, just as, in mathematics, we may *count* a proof of a conjunction as a proof of its conjuncts, so that if $p \wedge q$ is proved then p is proved, not just provable. (Williamson, 2000, 282–283)

He generalizes and conjectures that Simplification may hold for all positive attitudes (Ibid.): in believing a conjunction, one believes the conjuncts; in conceiving a conjunction, one conceives the conjuncts, etc.

Simplification is especially plausible if one has a view of mental states as intentional states: contentful and directed towards situations or circumstances. If thinking that $\varphi \wedge \psi$ was merely having a sentence (say, of mentalese), '$\varphi \wedge \psi$', tokened in the head, one may need to do something to move from thinking that $\varphi \wedge \psi$ to thinking that φ, that is, to having a sentence, say, of mentalese, 'φ', tokened in the head. And if one failed to apply the syntactic operation of Mental Conjunction Elimination to one's mentalese conjunctive sentence, one would have '$\varphi \wedge \psi$' tokened in the head without having 'φ' there.

But if thinking that $\varphi \wedge \psi$ is understood as having a mental state endowed with content—it's about representing a situation making $\varphi \wedge \psi$ true—then one who thinks that $\varphi \wedge \psi$, has already represented a whole situation making $\varphi \wedge \psi$ true in one's head; and (part of) that situation has already made true φ. By thinking about the whole, one has already thought about the parts: there's nothing more for one to do, such that if one failed to do it one would be thinking that $\varphi \wedge \psi$ without thinking that φ. This is bolstered by intuition. Take some putatively very anarchic mental state, such as imagining: try and *imagine* that Paul is tall and thin without imagining that Paul is tall. That seems difficult: in a non-merely-syntactic sense of 'thinking that', this would be a bit like thinking that Paul is tall without thinking that Paul is tall, wouldn't it?

2.2 Conceptual limitations

If a Boolean algebra of propositions is too coarse-grained for cognitive synonymy, can one come up with a more fine-grained but not too fine-grained algebra? One may take cognitive contents as given by sets of circumstances more fine-grained than classical possible worlds; but which, unlike anarchic open worlds, still display some degree of logical closure—like the situations of Barwise and Perry (1983). These can be thought of as parts of reality which don't take a stance on each φ: the rainy Boston situation makes true 'It's rainy in Boston', makes false 'It's sunny in Boston', but is silent on whether or not it's snowy in Oslo. Situations could also be taken as abstract representations of (parts of) reality which may, on occasion, be inconsistent. Such situations make sense of the points of evaluation in the semantics of the non-classical logic of First Degree Entailment (FDE) (Belnap, 1977; Dunn, 1976), where formulas can be both true and false, or neither. Such points are at times called 'FDE worlds' (Berto & Jago, 2019), although one will not see them as possible worlds if one does not accept situations where sentences of the form $\varphi \wedge \neg\varphi$ can be true, or sentences of the form $\varphi \vee \neg\varphi$ can fail to be, as genuine logical or metaphysical possibilities.

One way to formulate FDE is as a four-valued logic (*true only, false only, both true and false, neither true nor false*). Negation flips truth and falsity and has *both* and *neither* as fixed points; a conjunction is true iff both conjuncts are, false if either is false; dually for disjunction; the designated truth values (those preserved in valid inferences) are *true only* and *both true and false*. One recursively assigns to each formula φ of the language a pair of a truth $|\varphi|^+$ and falsity $|\varphi|^-$ set (a set of FDE worlds making φ true and, respectively, false) and sets $|\varphi| = \langle|\varphi|^+, |\varphi|^-\rangle$. This delivers a De Morgan algebra of contents: a bounded, distributive lattice where negation-complementation is an involution satisfying De Morgan's laws. The situation on the way to cognitive synonymy improves: we have that $|\varphi| = |\varphi \wedge \varphi|$ and that $|\varphi \wedge \psi| = |\psi \wedge \varphi|$ (conjunction is idempotent and commutative); $|\varphi \wedge \psi|$ will entail $|\varphi|$ and $|\psi|$ separately; (perhaps more controversially) $|\neg\neg\varphi| = |\varphi|$; etc. On the other hand, one can model locally inconsistent (but nontrivial) and incomplete belief states. So FDE has been used in epistemic logic to capture the contents of the attitudes of non-ideal agents (Fagin et al., 1995; Levesque, 1984).

But it doesn't quite work for cognitive synonymy. FDE semantics assigns the same content to φ and to $\varphi \wedge (\varphi \vee \psi)$ (and to $\varphi \vee (\varphi \wedge \psi)$). This seems wrong. It's not just that 'John is happy' and 'John is happy and either he is happy or extremally disconnectedness is no hereditary property of topological spaces' don't quite seem to say the same thing. It's also that one may just lack the concept of extremally disconnectednes, for one knows nothing about topology. Then one cannot entertain thoughts involving such concepts. One cannot suppose, believe, know, etc., that φ when one lacks some concept needed to grasp what φ's content is about. Perhaps one cannot even *see* that φ then, although one can see a situation in which φ (Barwise & Perry, 1983; Williamson, 2000): one can see a situation in which Mary is playing Go, but one cannot see that Mary plays Go if one has no idea of what Go is. FDE is not good for cognitive synonymy.

But the Absorption laws, $a = a \vee (a \wedge b)$, $a = a \wedge (a \vee b)$, are very weak algebraic identities: they hold in any lattice (any ordered set where least upper bounds $a \vee b$ and greatest lower bounds $a \wedge b$ exist for each a, b). This includes non-distributive and non-modular ones, which are even less structured than a De Morgan algebra. Finding an algebra of cognitive synonyms is not going to be easy Hornischer (2020).

Cognitive synonymy should respect *topic* or *subject matter*, in the intuitive sense of what (the content expressed by) a sentence is about. (Aboutness, as Yablo has it, 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, 1).) The idea that mental states like belief and knowledge should be sensitive to topic or subject matter is gaining popularity in recent research (Berto, 2022; Hawke, 2016; Hoek, 2022; Yalcin, 2016). One reason why φ and ψ can fail to be cognitive synonyms for one although they entail each other in classical logic as well as in non-classical logics such as FDE, is that one lacks the concepts needed to grasp what either of the two corresponding contents is about. This is an all-too-common phenomenon. To adapt an example due to (Stalnaker, 1984, 88), William III may have known (or, believed, etc.) that England could avoid war with France, without thereby knowing (or, ditto) that either England could avoid war with France, or France could develop a nuclear arsenal: he had no idea what nuclear weapons might be, so he could not entertain nuclear-weapons-involving thoughts. So conceptual limitations force a cognitive asymmetry between conjunction and disjunction. While Simplification may well hold for cognitive entailment, Addition or Disjunction Introduction—the inference from φ to $\varphi \vee \psi$—in general shouldn't.

This needn't have to do with representing synonymy and entailment for *deductively* limited or impaired agents: Addition seems to be as basic as Simplification, so a general asymmetry cannot be motivated in this way. It is motivated by the difference in topic-preservation: to think that $\varphi \wedge \psi$, one has to think both about what φ is about, and about what ψ is about. Instead, one can think that φ without thinking about what $\varphi \vee \psi$ is about. And one may not even be in a position to think about the latter, because one has no way to think about what ψ is about, even if one is an unbounded deductive reasoner: the topic of ψ may be *alien* to the agent. One is as blind to the concepts involved in grasping what ψ is about as William III was to the concept of nuclearity. Williamson again:

> $\wedge$-elimination has a special status. It may be brought out by a comparison with the equally canonical $\vee$-introduction inference to the disjunction $p \vee q$ from the disjunct p or from the disjunct q. Although the validity of $\vee$-introduction is closely tied to the meaning of $\vee$, a perfect logician who knows p may lack the empirical concepts to grasp (understand) the other disjunct q. Since knowing a proposition involves grasping it, and grasping a complex proposition involves grasping its constituents, such a logician is in no position to grasp $p \vee q$, and therefore does not know $p \vee q$. In contrast, those who know a conjunction grasp its conjunct, for they grasp the conjunction. (Williamson, 2000, 282–283)

2.3 Sensitivity to subject matters

Things improve when we move to a way of assigning contents that takes subject matters at face value (Humberstone, 2008; Plebani & Spolaore, 2021; Schipper, 2018). The aforementioned asymmetry between conjunction and disjunction is a guiding principle for influential subject-matter-sensitive accounts of content:

> A paradigm of [content] inclusion, I take it, is the relation that simple conjunctions bear to their conjuncts – the relation *Snow is white and expensive* bears, for example, to *Snow is white*. A paradigm of noninclusion is the relation disjuncts bear to disjunctions; *Snow is white* does not have *Snow is white or expensive* as a part. (Yablo, 2014, 11)

> A guiding principle behind the understanding of partial content is that the content of A and B should each be part of the content of $A \wedge B$ but that the content of $A \vee B$ should not in general be part of the content of either A or B. (Fine, 2016a, 200)

Talk of partial content and content inclusion mirrors the idea that the space of topics must display some mereological structure: topics can have proper parts; distinct topics may have common parts; one topic may be included in another in that every part of the former is also a part of the latter. *Mathematics* includes *arithmetic*. *Mathematics* and *philosophy* overlap, having (certain parts of) *logic* as a common part. Yablo (2014) proposes 'thick' or 'directed' propositional contents obtained by enriching thin propositions taken as sets of possible worlds with topics or subject matters, understood in their turn as partitions—an idea found in Lewis (1988a, 1988b)—or divisions of modal space. A subject matter, like *the number of stars*, is linked to a question, *What's the number of stars?* Worlds are split and grouped depending on how they answer the question: all zero-star worlds in one cell, all one-star worlds in another, etc. Subject matters are ways of being true, understood as ways of splitting and grouping worlds. Thick propositions allow many hyperintensional distinctions:

φ and ψ may be true in the same possible worlds, but diverge in thick content, insofar as they are about different topics.

A promising subject-matter-sensitive account comes from truthmaker semantics à la Kit Fine. His Angellic Content (AC) logic (Fine, 2016a) works with a semantics given by a set of *states*, which look like Barwise-Parry situations in that they are partial; and may also resemble FDE 'worlds', in that they can be inconsistent: there's a mereological fusion operation on states, and according to Fine (2016b) one can merge incompatible states, e.g., this table's being circular and this table's being square, thereby obtaining an impossible state where this table is a square circle. The contents of formulas are sets or fusions of truthmakers (states making the formula true) and falsemakers (states making it false), and Fine (2020) makes a convincing case that this captures subject matter adequately. If one takes AC equivalence as a candidate for cognitive synonymy, one finds that this preserves the welcome conjunction-involving FDE features, e.g., $\varphi \wedge \psi$ is still equivalent to $\psi \wedge \varphi$, and entails φ. But because Fine gives the truth and falsity conditions for conjunction and disjunction differently from FDE (following Van Fraassen (1969)'s tautological entailment), AC invalidates exactly the unwelcome Absorption principles which held in FDE: φ is not equivalent to $\varphi \wedge (\varphi \vee \psi)$ or to $\varphi \vee (\varphi \wedge \psi)$—for discussion, see Hornischer (2020).

Subject-matter-sensitive semantics score many points. But a general problem may affect them if proposed as accounts of cognitive synonymy.

3 The intransitivity of cognitive content

All the accounts we've met partition the totality of sentences of the target language into equivalence classes of same-sayers: φ and ψ end up in the same box when they are assigned the same content. The partition can be coarse-grained, as in the Boolean algebra of contents delivered by standard, 'merely intensional' possible worlds semantics; or it can be more fine-grained, because we use circumstances more fine-grained than possible worlds, like situations, FDE worlds, or Finean states. At the extreme of fine-graining (where we don't want do end up, as we have seen), we find the naive open worlds setting where each syntactically distinct sentence can be assigned a different content taken as a set of open worlds. But any such account will partition into equivalence classes. Among the features that make for an equivalence relation, reflexivity seems unobjectionable for cognitive synonymy: surely φ will play exactly the same role as itself (qua type) in one's cognitive life. And symmetry seems fine, too: if φ is cognitively synonymous with ψ, then ψ will be cognitively synonymous with φ. The problem is transitivity.

3.1 Dead parrots

There is some fuzziness (context-sensitivity, non-monotonicity, vagueness, etc.) in what is cognitively synonymous to what for someone, and this threatens same-saying. Finally coming to why our paper got its title, take the Monty Python dead parrot sequence:

> It passed on! This parrot is no more! He has ceased to be! It's expired and gone to meet its maker! ... Bereft of life, it rests in peace! ... Its metabolic processes are now history! ... It's kicked the bucket, it's shuffled off its mortal coil, run down the curtain and joined the choir invisible! *This is an ex-parrot*!

A competent speaker may have no trouble taking '[This parrot] passed on' as immediately entailing 'This parrot is no more' or as equivalent to it, and similarly for any other adjacent pair in the sequence, while having troubles when looking at the first and last item. The example may not be persuasive (it was just funny to use), but its recipe has persuasive instances: take a sequence $\varphi_1, \ldots, \varphi_n$, such that any φ_i cognitively entails, or is synonymous to, φ_{i+1}, but φ_1 and φ_n are not. Bjerring and Schwarz (2017), who formulated the transitivity objection against various hyperintensional ways of individuating contents, have $\varphi_1, \ldots, \varphi_n$ as a sequence of algebraic equations where each neighboring pair makes for a trivial transformation, but the transformation between φ_1 and φ_n is not trivial in the least. Or take a sequence of (two-way) logical entailments: one moves to the next item via an elementary inference rule application, whereas deducing the last item from the first (or *vice versa*) is highly nontrivial.

Cognitive synonymy or (two-way) cognitive entailment seems to be intransitive. We may keep calling it 'synonymy', but it starts looking more like similarity (reflexive and symmetric, but not transitive) than sameness of meaning.

3.2 Fregean equipollence

Are there non-transitive accounts of content on the market? One candidate would be Fregean equipollence, which Frege (at times) advances as a criterion for sameness of sense ('at times', for Frege seems to say things on equipollence that don't hang together well: Schellenberg (2012) gives a masterful reconstruction). Roughly, φ and ψ are cognitively equipollent when one could not rationally take either of the two as true and the other as untrue (e.g. Frege (1891, 14) and Frege (1979, 197)): compare item (1) in our initial list of features of cognitive synonymy.

The fuzziness of what counts as equipollent makes Fregean equipollence intransitive (Bjerring and Schwarz, 2017, 33). But it, too, can't easily capture cognitive synonymy as such. Equipollence was taken by Frege (sometimes, at least) as a criterion for sameness of sense, but it's not very clear what Fregean senses are and so what precise intransitive account could be given for them. E.g., if one takes senses as standard intensions, as Carnap (1947) basically did, or as 'primary' intensions, as in two-dimensional semantics (Schroeter, 2021), thus as sets of epistemically possible scenarios (Lewisian centered worlds; see e.g. Chalmers (2011)), such individuations of content are still transitive. One hyperintensional account of content using a notion akin to Fregean equipollence, which is intransitive and formally quite precise, is Skipper and Bjerring (2020)'s Frege-inspired view. One worry about Fregean equipollence in general, even when made intransitive, however, is that it may not give a good account of subject matter. Most competent speakers of English cannot rationally take one of '2 is a number' and 'Triangles have three angles' as true and the other as not true, but they say different things: only one is about triangles. Fregean equipollence violates our Yablovian 'sense of when sentences say the same thing'. It cannot give a good account of same-saying, and so it cannot give a good account of cognitive synonymy to the extent that the latter is topic-sensitive.

Next, there are worries for *any* account that tries to replace synonymy as sameness or equivalence of content with content similarity.

3.3 Is content similarity even coherent?

Goodman (1949), Mates (1952), Churchland (1993), etc., have claimed that strictly speaking there is no synonymy, but only degrees of similarity in meaning. The view has been criticized, e.g., by Fodor and Lepore (1999). But the most relevant critical point is due to Hannes Leitgeb (2008). In a neat result which got less attention than it should have, Leitgeb has showed that a set of plausible assumptions governing content similarity is inconsistent. The assumptions on our $\mathcal{L}$, on which a similarity relation $\approx$ is defined, are:

(*Similarity*) $\approx$ is reflexive and symmetric.
(*Connectedness*) $\approx$ is connected: any two sentences φ and ψ, however dissimilar, will be linked by some (perhaps long) similarity chain. There will be a sequence $\varphi_1 \approx \varphi_2 \approx \ldots \approx \varphi_n$ with $\varphi = \varphi_1$ and $\varphi_n = \psi$.
(*Non-triviality*) No classical tautology in $\mathcal{L}$ can be $\approx$-similar to a classical contradiction.
(*Closure*) The sentences in $\mathcal{L}$ are closed under negation, conjunction and disjunction.
(*Compositionality*) Similarity on $\mathcal{L}$ is compositional, that is, it's preserved by the logical operators: if $\varphi \approx \psi$ and $\chi \approx \theta$, then $\neg\varphi \approx \neg\psi$, $\varphi \wedge \chi \approx \psi \wedge \theta$ and $\varphi \vee \chi \approx \psi \vee \theta$.

All are reasonable. *Similarity*: surely $\approx$ should be reflexive and symmetric. *Connected*: if $\approx$ wasn't connected, there would be φ and ψ that are entirely unrelated in that no chain of associations could ever link the two. (And, in any case, we could then focus on the connected subspaces of our similarity space.) *Non-triviality*: a tautology and a contradiction are the most content-dissimilar. So they, at the very least, shouldn't be $\approx$-related, or $\approx$ would trivialize basic dissimilarity intuitions. *Closure*: we can always form the negation, conjunction, and disjunction of given sentences.[1] *Compositionality*: if $\approx$ were meaning identity, this would be an obvious way of stating compositionality: if φ and ψ mean the same, then also $\neg\varphi$ and $\neg\psi$ do, etc., that is: the meaning of a negation, conjunction, or disjunction is a function of the

[1] Philosophically, this is highly plausible. However, when it comes to managing the computational tractability of a logic, limiting sentence-formation is a strategy: see, e.g., guarded fragments of first-order logic (Andréka et al., 1998)

meanings of the negated, conjoined, or disjoined sentences. This is just transferred to similarity.

Leitgeb has proved that these are jointly inconsistent.[2] Leitgeb's diagnosis: either meaning similarity is doomed, or 'our common views on compositionality have to be changed significantly in order to make compositionality compatible with semantic resemblance' (Leitgeb, 2008, 295).

There are intuitions for the transitivity of cognitive synonymy as well, pulling us around again. Here follow three arguments.

3.4 Reasons for transitivity

Argument 1: let's write '$\varphi \leftrightsquigarrow \psi$' for cognitive synonymy and '$\varphi \rightsquigarrow \psi$' for cognitive entailment. We have seen that any attempt at spelling out cognitive synonymy as sameness of content (being same-sayers, being identical in meaning) will, qua equivalence relation, render cognitive synonymy transitive. Are there other ways of spelling out cognitive synonymy that do not presuppose contents of sentences? A famous suggestion of Quine (1951) is as substitutability *salva veritate*:

> (*Substitution salva veritate*) $\varphi \leftrightsquigarrow \psi$ if and only if, for all allowed sentential contexts $\chi[\ldots]$, we have $\chi[\varphi]$ iff $\chi[\psi]$.[3]

(We already saw compositionality for synonymy also demanding some substitutability: substituting synonyms preserves synonymy. But this principle is subtly different: substituting synonyms preserves equivalence.) For example, φ = 'Paul is a bachelor' and ψ = 'Paul is an unmarried man' are cognitive synonyms because, for any allowed context, like $\chi[\ldots]$ = 'Paul doesn't have a wedding ring because ...', if, as in this case, we obtain a true sentence by inserting φ, we also obtain a true sentence by inserting ψ and *vice versa*. Substitutability *salva veritate* can be motivated as capturing the idea that φ and ψ are cognitively synonymous for competent speakers of English: the sentences play the same role in any allowed context. Compare this to item (2) in our initial list: cognitive synonymy as same-understanding.

The qualifier 'allowed' is badly needed. If quotation contexts like 'Jane literally said that ...' were allowed, the principle would trivialize cognitive synonymy to syntactic identity. But there should not be too few contexts either: as (Quine,

[2] The idea of the proof is: use Closure to get a tautology φ and a contradiction ψ. Use Connectedness to find a sequence $\varphi = \varphi_1 \approx \varphi_2 \approx \varphi_3 \approx \ldots \approx \varphi_n = \psi$. Use Similarity, Closure, and Compositionality for a clever lemma saying that for any such sequence, we can find a shorter sequence $\varphi'_1 \approx \varphi'_3 \approx \ldots \approx \varphi'_n$ where φ_k is logically equivalent to φ'_k for $k = 1, 3, \ldots, n$. Apply this repeatedly to shorten the original sequence to one of length two where the first item is logically equivalent to φ_1, a tautology, and the second is logically equivalent to φ_n, a contradiction, thereby violating Non-triviality.

[3] Here and elsewhere we will allow us some imprecision in our notation. Strictly speaking, we should use Quine corners and write '$\ulcorner\chi[\varphi]\urcorner$ iff $\ulcorner\chi[\psi]\urcorner$'. This is because we want to say that the sentence obtained by inserting the sentence denoted by the variable φ into the context denoted by χ is equivalent to the sentence that we obtain doing the same thing but using ψ. We do not want to say that the names '$\chi[\varphi]$' and '$\chi[\psi]$' for these two sentences are equivalent—which would not make sense.

1951, 30) had it, if only extensional contexts were allowed, substitutability *salva veritate* wouldn't capture agreement of expressions due to meaning ('bachelor' and 'unmarried man') but merely due to accidental facts, turning 'creature with a heart' and 'creature with a kidney' into synonyms. So we also need intensional contexts:

> If a language contains an intensional adverb 'necessarily' ... or other particles to the same effect, then interchangeability *salva veritate* in such a language does afford a sufficient condition of cognitive synonymy (Quine, 1951, 30).

Notoriously, Quine objected that such an intensional language is 'intelligible only if the notion of analyticity is already clearly understood in advance' (30). Since the rise of modal logics, this is less of a worry.

But now the argument is this: If the substitution *salva veritate* principle is correct, then, no matter which set C of allowed contexts is chosen, cognitive synonymy is transitive. Because if $\varphi_1 \leftrightsquigarrow \varphi_2$ and $\varphi_2 \leftrightsquigarrow \varphi_3$, then $\forall \chi \in C : \chi[\varphi_1]$ iff $\chi[\varphi_2]$ and $\forall \chi \in C : \chi[\varphi_2]$ iff $\chi[\varphi_3]$, so, by transitivity of 'iff', $\forall \chi \in C : \chi[\varphi_1]$ iff $\chi[\varphi_3]$, hence $\varphi_1 \leftrightsquigarrow \varphi_3$ (see [*reference omitted*]).

One might object: What if 'allowed' is itself context-sensitive? Then the choice of C is not uniform but depends (at least) on φ and ψ—indicated by writing $C(\varphi, \psi)$. And it could well be that $C(\varphi, \psi)$ and $C(\psi, \chi)$ contain few enough contexts to render $\varphi \leftrightsquigarrow \psi$ and $\psi \leftrightsquigarrow \chi$, but $C(\varphi, \chi)$ is large enough to contain a context in which φ and χ aren't substitutable. We come back to this idea in Sect. 5.

Argument 2: as another approach to spelling out cognitive synonymy, consider item (3) in our initial list: sameness of inferential role. We could consider many types of inference (deductive, inductive, abductive, non-monotonic, etc.), so as a placeholder, let's say that χ is *cognitively inferrable* from φ when χ is (easily) inferred, in the considered sense, from φ. There is some fuzziness in what is easily derivable from what for a cognitive agent with bounded resources, and cognitive inferrability between sentences may well fail to be transitive. But regardless of its exact nature, sameness of inferential role inspires the following principle:

> (*Cognitive role*) $\varphi \leftrightsquigarrow \psi$ if and only if, for all χ, we have that χ is cognitively inferrable from φ iff χ is cognitively inferrable from ψ.

Now the argument is this: even if the relation $\varphi R \psi$ of cognitive inferrability is *not* transitive, the cognitive role principle implies that $\leftrightsquigarrow$ *is* transitive: if $\varphi_1 \leftrightsquigarrow \varphi_2$ and $\varphi_2 \leftrightsquigarrow \varphi_3$, then $\forall \chi : \varphi_1 R \chi$ iff $\varphi_2 R \chi$ and $\forall \chi : \varphi_2 R \chi$ iff $\varphi_3 R \chi$, so, by transitivity of 'iff', $\forall \chi : \varphi_1 R \chi$ iff $\varphi_3 R \chi$, hence $\varphi_1 \leftrightsquigarrow \varphi_3$. In other words, regardless of what type of inference we consider, taking cognitive synonymy to be sameness of inferential role renders it transitive.

Again, a context-sensitive retort may be: not any χ is relevant to be considered as potentially being cognitively inferrable, but only some χ from a candidate set C; and this C may depend, at least, on φ and ψ. We pick this up again in Sect. 5.

Argument 3: as mentioned at the outset of section 1, cognitive entailment appears to have non-monotonic features. The idea is that φ cognitively entails ψ for one when one has something like a defeasible conditional 'If φ, then (normally) ψ' in one's knowledge or belief base.[4] Further information can lead one to retract one's conclusions based on it. One has 'If x is a bird, then x flies' in one's mental repository. Supposing Tweety is a bird, one would conclude that Tweety flies; if one were to learn that Tweety is a bird, one would believe that Tweety flies; etc.—see item (4) from our initial list. But if one then learned that Tweety is a penguin bird, then one would retract one's conclusion.

Then principles of conditional/non-monotonic logics (Kraus et al., 1990) may plausibly be taken as applying to $\rightsquigarrow$:

(*Cautious monotonicity*) If $\varphi \rightsquigarrow \chi$ and $\varphi \rightsquigarrow \psi$, then $\varphi \wedge \psi \rightsquigarrow \chi$.
(*Cut*) If $\varphi \wedge \psi \rightsquigarrow \chi$ and $\varphi \rightsquigarrow \psi$, then $\varphi \rightsquigarrow \chi$.
(*Commutativity*) If $\varphi \wedge \psi \rightsquigarrow \chi$, then $\psi \wedge \varphi \rightsquigarrow \chi$.

All are reasonable. *Cautious monotonicity*: while new assumptions may defeat previously derived consequences, if we get a new one ψ from a given premise φ, this shouldn't defeat the earlier derivation of χ from φ. *Cut*: we can establish $\varphi \rightsquigarrow \chi$ by first making the additional assumption ψ in concluding χ from φ and then showing that ψ could already be obtained from φ. *Commutativity* is just a weaker version of non-monotonic logic's Left Logical Equivalence principle: if φ and ψ are classically equivalent and $\varphi \rightsquigarrow \chi$, then $\psi \rightsquigarrow \chi$. This is implausible for cognitive entailment: it would imply that classically equivalent sentences have the same cognitive role—which, we saw in Sect. 2, should be avoided. But the weaker Commutativity principle only requires that both orders of a conjunction play the same cognitive role, which, as already discussed, is highly plausible for Boolean conjunction and sufficient for our argument. (It would even work without Commutativity if Cut is rephrased as: if $\psi \wedge \varphi \rightsquigarrow \chi$ and $\varphi \rightsquigarrow \psi$, then $\varphi \rightsquigarrow \chi$. But that feels more like hiding the Commutativity rule, than simplifying the argument.)

So here is the argument: those three principles imply

(*Equivalence*) If $\varphi \rightsquigarrow \psi$, $\psi \rightsquigarrow \varphi$, and $\psi \rightsquigarrow \chi$, then $\varphi \rightsquigarrow \chi$.

From $\psi \rightsquigarrow \chi$ and $\psi \rightsquigarrow \varphi$, Cautious monotonicity yields $\psi \wedge \varphi \rightsquigarrow \chi$. Commutativity yields $\varphi \wedge \psi \rightsquigarrow \chi$. Together with $\varphi \rightsquigarrow \psi$, Cut implies $\varphi \rightsquigarrow \chi$ (cf. Kraus et al., 1990, lem. 3.3).

Equivalence is also plausible on its own: if two sentences φ and ψ are cognitively synonymous, they should play the same cognitive role in that any cognitive entailment χ from one should also be obtainable from the other. While Left Logical

4 Reiterating footnote 3, here we also should say '$\ulcorner\varphi\urcorner$ cognitively entails $\ulcorner\psi\urcorner$ for one' (or 'that φ cognitively entails that ψ for one'). Only then do we get a sentential operator (and we intend cognitive entailment to be a conditional connective) and not a predicate applying to names of sentences. Thanks to an anonymous referee for pointing this out.

Equivalence implausibly demanded that to hold already for classically equivalent sentences, Equivalence reasonably demands it only for cognitively synonymous ones. But taking $\varphi \leftrightsquigarrow \psi$ as two-way cognitive entailment (i.e., $\varphi \rightsquigarrow \psi$ and $\psi \rightsquigarrow \varphi$), $\varphi \leftrightsquigarrow \chi$ follows from $\varphi \leftrightsquigarrow \psi$ and $\psi \leftrightsquigarrow \chi$ using Equivalence twice. So the much motivated Equivalence principle yields transitivity again.

So, does cognitive synonymy encode irreducibly inconsistent intuitions? In the next section, we come up with a model capturing all the features of cognitive entailment and synonymy singled out so far. The model expands on ideas presented by one of us (FB) in Berto (2022) and other works (Berto, 2018a, 2018b), but never applied so far specifically to model the notion of cognitive synonymy. This will go some way towards showing that the notion is coherent. In the section after that, we use the semantics to further analyze the issue of transitivity.

4 Cognitive synonymy, redux

4.1 Language

We add to our $\mathcal{L}$ an operator allowing us to talk of cognitive entailment and synonymy inside the language. So now we have: a set $\mathcal{L}_{AT}$ of atomic formulas $p, q, r, \ldots, p_1, p_2, \ldots$, negation $\neg$, conjunction $\wedge$, disjunction $\vee$, but also an arrow $\rightsquigarrow$ (and, round parentheses as auxiliary symbols). The well-formed formulas are the items in $\mathcal{L}_{AT}$ and, if φ and ψ are formulas:

$$\neg\varphi \mid (\varphi \wedge \psi) \mid (\varphi \vee \psi) \mid (\varphi \rightsquigarrow \psi)$$

(We normally omit outermost brackets.) We take $\mathcal{L}$ as the set of its well-formed formulas. We read '$(\varphi \rightsquigarrow \psi)$' as '$\varphi$ cognitively entails ψ (for one)', so cognitive synonymy as two-way entailment is $\varphi \leftrightsquigarrow \psi := (\varphi \rightsquigarrow \psi) \wedge (\psi \rightsquigarrow \varphi)$.

Cognitive entailment may work as a sort of defeasible conditional. Then $\varphi \rightsquigarrow \psi$ should be a non-monotonic operator that fails Antecedent Strengthening: $\varphi \rightsquigarrow \psi$ should not entail $(\varphi \wedge \chi) \rightsquigarrow \psi$. There are well-known possible worlds semantics for conditionals of this kind, e.g., the semantics for counterfactuals by Stalnaker (1968) and Lewis (1973)—for overviews, see Nute (1984), Priest (2008), Egré and Rott (2021). As it happens, they make their conditionals intransitive. We will resort to a similar framework, but with a twist. Remember that cognitive synonymy is to preserve subject matter: we don't want $\varphi \leftrightsquigarrow \psi$ to generally hold for one when φ and ψ are logically or necessarily equivalent, for one may just lack the concepts to grasp what either of φ and ψ is about. So we don't want a $\varphi \rightsquigarrow \psi$ to be a validity in our logic of cognitive synonymy just because φ logically or necessarily entails ψ. And although $\varphi \leftrightsquigarrow \varphi$ should be valid, we don't want $\varphi \leftrightsquigarrow \varphi \wedge (\varphi \vee \psi)$ to be: Absorption must go. This will be achieved in our semantics by making $\rightsquigarrow$ subject-matter-sensitive: φ will cognitively entail ψ for one, only when φ and ψ are suitably connected in what they are about, so that the concepts one needs to be on top of to grasp the former will be enough for one to be on top of the concepts needed to grasp the latter.

4.2 Semantics

In the metalanguage we use variables $w, v, w_1, w_2, \ldots$, ranging over worlds, $x, y, z, x_1, x_2, \ldots$, ranging over topics or subject matters (more on these in a moment), and the symbols $\Rightarrow, \Leftrightarrow, \&, \text{or}, \sim, \forall, \exists$, read the usual way. A *frame* for $\mathcal{L}$ is a tuple $\mathfrak{F} = \langle W, \{R_\varphi \mid \varphi \in \mathcal{L}\}, \mathcal{T}, \oplus, t \rangle$ where:

- W is a non-empty set of possible worlds.
- $\{R_\varphi \mid \varphi \in \mathcal{L}\}$ is a set of accessibilities between worlds: each $\varphi \in \mathcal{L}$ has its own $R_\varphi \subseteq W \times W$. Notice that these are indexed to *formulas*: this will matter soon.
- $\mathcal{T}$ is a non-empty set of *topics* or *subject matters* that formulas of $\mathcal{L}$ can be about. It doesn't matter what we take them to be: sets or fusions of truthmakers and falsemakers (Fine, 2016a), ways of dividing modal space (Lewis, 1988b; Yablo, 2014), or else. For our logical purposes, we only ask them to obey the mereological constraints coming next.
- $\oplus$ is *topic fusion*: a binary operation on $\mathcal{T}$ fusing two topics into a new one, satisfying for all $x, y, z \in \mathcal{T}$:

  – (*Idempotence*) $x \oplus x = x$
  – (*Commutativity*) $x \oplus y = y \oplus x$
  – (*Associativity*) $(x \oplus y) \oplus z = x \oplus (y \oplus z)$

  (To keep things simple, fusion shall be unrestricted: $\oplus$ is always defined on $\mathcal{T}$.) Thus, $\langle \mathcal{T}, \oplus \rangle$ is what's known as a join semilattice and *topic parthood*, $\leq$, can then be defined the usual way: $\forall xy \in \mathcal{T}(x \leq y \Leftrightarrow x \oplus y = y)$. Then $\leq$ is a partial order, i.e., for all $x, y, z \in \mathcal{T}$:

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

- $t : \mathcal{L}_{AT} \rightarrow \mathcal{T}$ is a topic-assignment function, assigning an item in $\mathcal{T}$ to each atomic formula. It is extended to the whole of $\mathcal{L}$ as follows: if the set of atoms in φ is $\{p_1, \ldots, p_n\}$, then $t(\varphi) = t(p_1) \oplus \ldots \oplus t(p_n)$. So a formula is about what its atoms, taken together, are about. (See Hawke, 2016, for the atom-based approach to subject matters.)

We'll see that this mereology of subject matters makes our individuation of cognitive contents hyperintensional. However, as shown more extensively in Berto (2018b, 2022) we don't risk ending up as fine-grained as the syntax of $\mathcal{L}$. E.g., by induction on the construction of formulas, $t(\varphi) = t(\neg\neg\varphi)$. Also, $t(\varphi) = t(\neg\varphi)$: a formula is about what its negation is about ('Grass isn't green' is exactly about what 'Grass is green' is about; one who thinks that grass isn't green, thinks about the same topic as one who thinks that grass is green). And not only $t(\varphi \wedge \psi) = t(\varphi \wedge \psi)$, but also, $t(\varphi \wedge \psi) = t(\varphi) \oplus t(\psi) = t(\varphi \vee \psi)$: conjunction and disjunction merge topics ('John is tall and thin' and 'John is tall or thin' are about the height and looks of John). Fine (2020), Hawke (2018), and others, forcefully defend the view that the truth-functional connectives should be 'topic-transparent', contributing no subject matter of their own to the sentences where they show up.

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

(S$\neg$) $w \Vdash \neg\varphi \Leftrightarrow w \nVdash \varphi$ (i.e., it is not the case that $w \Vdash \varphi$)
(S$\wedge$) $w \Vdash \varphi \wedge \psi \Leftrightarrow w \Vdash \varphi \,\&\, w \Vdash \psi$
(S$\vee$) $w \Vdash \varphi \vee \psi \Leftrightarrow w \Vdash \varphi \; or \; w \Vdash \psi$
(S$\rightsquigarrow$) $w \Vdash \varphi \rightsquigarrow \psi \Leftrightarrow (1)\ \forall w_1(wR_\varphi w_1 \Rightarrow w_1 \Vdash \psi) \,\&\, (2)\ t(\psi) \leq t(\varphi)$

The clause (S$\rightsquigarrow$) is reminiscent of the Parry 'analytic containment' conditionals (Deutsch, 1984; Epstein, 1993; Ferguson, 2014; Parry, 1933; Perry, 1989). For $\varphi \rightsquigarrow \psi$ to come out true at w we ask for two things to happen:

(1) ψ must be true at all worlds w_1 accessible via the relation indexed to φ. We may think of these as the set of possible scenarios or situations one looks at in one's thought, given input φ. This makes of $\varphi \rightsquigarrow \psi$ a variably strict quantifier over worlds and a defeasible conditional-like operator. (The idea of taking non-monotonic conditionals as formula-indexed necessitations goes back to Chellas (1975).)
(2) ψ must be fully on topic with respect to φ. This is the aboutness-preservation component, ensuring that the concepts one needs to grasp the latter will be enough for one to possess the concepts needed to grasp the former,[5]

[5] A helpful referee remarks: isn't full topic inclusion between ψ and φ too restrictive? Surely cognitive entailment is to account for 'ampliative' inferences one may want to make, e.g., inductive ones. But if the inference from φ to ψ is ampliative, $t(\psi)$ may well outstrip the boundaries of $t(\varphi)$. Great point! We have a long reply to this in chapter 5 of Berto (2022). We can only propose a short summary here, for reasons of space. First, we think it's a mistake to read topic-inclusion as some kind of 'Kantian analytic containment' (as it was in the original Parry setting inspiring us). It's not the case that when $t(\varphi) \leq x$, that is, φ is entirely about x, one (ideal reasoner) should always be able to extract the former from the latter a priori, via conceptual analysis, whatever this amounts to. The topic of 'Maine experiences cold winters' can be included in the topic of what New England is like (Goodman, 1961) but one cannot extract the former a priori via analysis of the concept of New England. The topic of 'Caesar crosses the Rubicon' can be included in whatever topic is suitably associated to Caesar, but one cannot extract the former a priori via analysis of the concept of Caesar (Leibnizian hopes notwithstanding). Next, if one still wants to better model the idea that, in a good cognitive entailment $\varphi \rightsquigarrow \psi$, $t(\psi)$ must sometimes exceed $t(\varphi)$—this is represented in chapter 5 of Berto (2022) by changing condition (2) above a bit: instead of requiring that the topic of ψ be included in that of φ, we require that the topic of ψ be included in the *topological closure* of that of φ. That is, rather than asking that $t(\psi) \leq t(\varphi)$, we ask that $t(\psi) \leq f(t(\varphi))$, where f is a Kuratowski closure operator, i.e., one satisfying:

(i) $x \leq f(x)$ [Inclusion]
(ii) $f(x) = f(f(x))$ [Idempotence]
(iii) $f(x \oplus y) = f(x) \oplus f(y)$ [Additivity]

The idea is that the topic of φ can be expanded to other, distinct topics. The expansion, however, is constrained: only topics suitably connected to $t(\varphi)$ will be considered, where 'connectedness' is regimented as topological connectedness. In Berto (2022), it is shown that this makes intuitive sense of ampliative inferences. Besides, importantly, in Özgün and Cotnoir (2021) it is proved that the change in clause (2) makes little difference for the *logic* of the operator: the (in)validities involving the operator with the topological clause remain the same as those of the simple setting presented in this paper.

We also impose a natural constraint on our models:

(*Success*) If $wR_\varphi w_1$, then $w_1 \Vdash \varphi$.

All the worlds one looks at given φ are φ-worlds. This makes sense for cognitive entailments: when one supposes that φ in order to see what follows, one looks, to begin with, only at circumstances where φ is true; if one were to learn that φ, one would take φ as true and, thus, update one's beliefs by kicking out worlds where it's not; etc.

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

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

For single-premise entailment, we write $\varphi \vDash \psi$ for $\{\varphi\} \vDash \psi$. Logical validity, $\vDash \varphi$, truth at all worlds of all models, is $\emptyset \vDash \varphi$, entailment by the empty set of premises. We write $|\varphi| = \{w \in W | w \Vdash \varphi\}$ for the truth-conditional content of φ and $[\varphi] = \langle |\varphi|, t(\varphi)\rangle$ for the 'truth and topic' content of φ (a 'thick proposition', to speak Yablovian).

Clause (S$\rightsquigarrow$) can be equivalently expressed using formula-indexed selection functions (Lewis, 1973, 57–60). Each $\varphi \in \mathcal{L}$ comes with a function $f_\varphi : W \to \mathcal{P}(W)$ outputting the set of accessible worlds, $f_\varphi(w) = \{w_1 \in W | wR_\varphi w_1\}$. We can then rephrase the clause:

$$(\text{S}\rightsquigarrow)\ w \Vdash \varphi \rightsquigarrow \psi \Leftrightarrow (1)\, f_\varphi(w) \subseteq |\psi| \ \&\ (2)\ t(\psi) \leq t(\varphi)$$

The two formulations are equivalent as $wR_\varphi w_1 \Leftrightarrow w_1 \in f_\varphi(w)$. Notice what the semantics delivers: φ and ψ are cognitive synonyms for one exactly when:

(1) Both $f_\varphi(w) \subseteq |\psi|$ and $f_\psi(w) \subseteq |\varphi|$ hold: all the worlds one looks at supposing (or learning, etc.) φ also make true ψ, and *vice versa*.
(2) Both $t(\psi) \leq t(\varphi)$ and $t(\varphi) \leq t(\psi)$ hols, which, by Antisymmetry, means $t(\varphi) = t(\psi)$: cognitive synonyms coincide in subject matter.

We went for the most general formula-indexed relational/functional approach, but one may not like the intrusion of syntax in the semantics. So let's discuss a restriction that avoids this. In Sect. 5 below, we discuss whether we should go for the restricted or general semantics. The answer will be that this depends ultimately on whether we want cognitive synonymy to be transitive or not, respectively. Both are viable but mutually exclusive options. We leave open this choice, because we want our semantic model to be able to accommodate both options. More precisely, will identify one concept—namely, *uniformity*—that, when required, yields the restrictive transitive semantics for cognitive synonymy, and when not required yields the general non-transitive semantics.

The straightforward way to avoid the intrusion of syntax is by indexing to contents, not formulas. Usually contents are propositions regarded as sets of worlds, but here we more naturally consider contents as 'thick'. We then restrict ourselves to models which satisfy:

(*Syntax-insensitivity*) If $[\varphi] = [\psi]$, then $R_\varphi = R_\psi$.

Now content-identical sentences can be substituted in antecedents:

(*Antecedent substitution*) If $[\varphi] = [\psi]$, then $[\varphi \rightsquigarrow \chi] = [\psi \rightsquigarrow \chi]$.[6]

Without assuming Syntax-insensitivity, this can fail (see 'non-introspective conjunction' in Sect. 4.3 below). This rule is very similar to Left Logical Equivalence (LLE): however, here we require not just truth-conditional (or logical) equivalence for substitutability, but also topic identity. Failing LLE may be regarded as a form of hyperintensionality and several authors argued against LLE (Egré and Rott, 2021, sec. 3).

Another, less general approach to avoid the intrusion of syntax employs comparative similarity (Burgess, 1981; Kraus et al., 1990; Lewis, 1973; Veltman, 1985) described as a single ternary relation R on worlds: $Rww'w''$ (or, $w' \leq_w w''$) says that w' is at least as similar to w as w'', hence $\leq_w$ is reflexive and transitive. Roughly, $\varphi \rightsquigarrow \psi$ is then true at w if the $\leq_w$-minimal φ-worlds are also ψ-worlds. On this approach, Cautious Monotonicity, Cut, and Commutativity come out valid making $\leftrightsquigarrow$ transitive.

4.3 (In)Validities

We now show that the semantics does indeed give the (in)validities we want based on the preceding discussion. To start, it's readily seen that cognitive synonymy is reflexive (by Success) and symmetric:

(*Reflexivity*) $\vDash \varphi \leftrightsquigarrow \varphi$
(*Symmetry*) $\varphi \leftrightsquigarrow \psi \vDash \psi \leftrightsquigarrow \varphi$

So cognitive synonymy is indeed a similarity relation. As (*prima facie*) desired, it also is not transitive:

[6] *Proof*: We have $t(\varphi \rightsquigarrow \chi) = t(\varphi) \oplus t(\chi) = t(\psi) \oplus t(\chi) = t(\varphi \rightsquigarrow \chi)$, so we need to show $|\varphi \rightsquigarrow \chi| = |\psi \rightsquigarrow \chi|$. Let $w \Vdash \varphi \rightsquigarrow \chi$ and show $w \Vdash \psi \rightsquigarrow \chi$ (the other direction is analogous). First, if $wR_\psi w_1$, then, since, by Syntax-insensitivity, $R_\varphi = R_\psi$, we have $wR_\varphi w_1$, so the assumption implies $w_1 \Vdash \chi$. Second, by the assumption, $t(\chi) \leq t(\varphi) = t(\psi)$.

(*No 2-way transitivity*) $\{p \leftrightsquigarrow q, q \leftrightsquigarrow r\} \not\models p \leftrightsquigarrow r$[7]

Cognitive entailment also has the desired non-monotonicity and simplification features:

(*Non-monotonic*) $p \rightsquigarrow r \not\models p \wedge q \rightsquigarrow r$ (and in fact not even cautiously monotonic $\{p \rightsquigarrow r, p \rightsquigarrow q\} \not\models p \wedge q \rightsquigarrow r$)[8]

(*Simplification*) $\varphi \rightsquigarrow (\psi \wedge \chi) \models \varphi \rightsquigarrow \psi$[9]

We have commutativity of conjunction on the meta-level: $\varphi \wedge \psi \models \psi \wedge \varphi$. In fact, $[\varphi \wedge \psi] = [\psi \wedge \varphi]$. But that doesn't mean that the content-identity is 'introspective':

(*Non-introspective conjunction*) $q \wedge p \rightsquigarrow r \not\models p \wedge q \rightsquigarrow r$[10]

In particular, as already discussed, Antecedent substitution fails: $[\varphi \wedge \psi] = [\psi \wedge \varphi]$ doesn't entail $[\varphi \wedge \psi \rightsquigarrow \chi] = [\psi \wedge \varphi \rightsquigarrow \chi]$. This is a failure of compositionality: the content of $\varphi \rightsquigarrow \chi$ is not just determined by the connective $\rightsquigarrow$ and the contents $[\varphi]$ and $[\chi]$. That's in line with Leitgeb's impossibility result:

(*Non-compositionality*) $\{p \leftrightsquigarrow p', q \leftrightsquigarrow q'\} \not\models p \wedge q \leftrightsquigarrow p' \wedge q'$[11]

Again as desired, Absorption fails:

[7] *Countermodel*: $W = \{w\}$ consists of just one world; R_φ is empty if $\varphi \neq p$ and otherwise wR_pw; $\mathcal{T} = \{x\}$ consists of just one topic; $x \oplus x := x$, which clearly satisfies idempotence, commutativity, and associativity; $t(p) := x$ for all p; and w makes true p and q and no other propositional atom. Success is satisfied since the only non-vacuous relation is wR_pw and $w \Vdash p$. Then $w \Vdash p \rightsquigarrow q$ since $\forall w_1 : wR_pw_1 \Rightarrow w_1 \Vdash q$ (because $w \Vdash q$) and $t(q) = x \leq x = t(p)$. And $w \Vdash q \rightsquigarrow p$ since $\forall w_1 : wR_pw_1 \Rightarrow w_1 \Vdash q$ (because R_q is empty) and $t(p) = x \leq x = t(q)$. Similarly, $w \Vdash q \rightsquigarrow r$ and $w \Vdash r \rightsquigarrow q$. However, $w \nVdash p \rightsquigarrow r$ because wR_pw but $w \nVdash r$. So w makes true the sentences of $\{p \leftrightsquigarrow q, q \leftrightsquigarrow r\}$ but doesn't make true $p \leftrightsquigarrow r$.

[8] *Countermodel*: $W = \{w\}$; R_φ is empty if $\varphi \neq p \wedge q$ and otherwise $wR_{p\wedge q}w$; $\mathcal{T} = \{x\}$ with $x \oplus x := x$ and $t(p) := x$ for all p; and w makes true p and q but no other propositional atom. Success is satisfied since the only non-vacuous relation is $wR_{p\wedge q}w$ and $w \models p \wedge q$. Then $w \Vdash p \rightsquigarrow r$ since $\forall w_1 : wR_pw_1 \Rightarrow w_1 \Vdash q$ (because R_p is empty) and $t(r) = x \leq x = t(p)$. Similarly, $w \Vdash p \rightsquigarrow q$. However, $w \nVdash p \wedge q \rightsquigarrow r$ because $wR_{p\wedge q}w$ but $w \nVdash r$.

[9] *Proof*: Let $\mathfrak{M}$ be a model and w a world with $w \Vdash \varphi \rightsquigarrow (\psi \wedge \chi)$. To show $w \Vdash \varphi \rightsquigarrow \psi$, let $wR_\varphi w_1$, hence, by the assumption, $w_1 \models \psi \wedge \chi$, which implies $w_1 \Vdash \psi$, as needed.

[10] *Countermodel* from footnote 8. Then $w \Vdash q \wedge p \rightsquigarrow r$ since $\forall w_1 : wR_{q\wedge p}w_1 \Rightarrow w_1 \Vdash r$ (because $R_{q\wedge p}$ is empty) and $t(r) = x \leq x = t(p \wedge q)$. However, $w \nVdash p \wedge q \rightsquigarrow r$ because $wR_{p\wedge q}w$ but $w \nVdash r$.

[11] *Countermodel* from footnote 8. Then $w \Vdash p \rightsquigarrow p'$ since $\forall w_1 : wR_pw_1 \Rightarrow w_1 \Vdash r$ (because R_p is empty) and $t(p') = x \leq x = t(p)$. Similarly, $w \Vdash p' \rightsquigarrow p$, $w \Vdash q \rightsquigarrow q'$, and $w \Vdash q' \rightsquigarrow q$. However, $w \nVdash p \wedge q \rightsquigarrow p' \wedge q'$ because $wR_{p\wedge q}w$ but $w \nVdash p' \wedge q'$.

(*No absorption*) $p \rightsquigarrow q \not\models p \rightsquigarrow q \vee (q \wedge r)$[12]

In particular, Disjunction Introduction fails: $\varphi \rightsquigarrow \psi$ does not entail $\varphi \rightsquigarrow (\psi \vee \chi)$.

Finally, as suggested by an anonymous referee, let's consider the relationship between content identity (i.e., $[\varphi] = [\psi]$) and cognitive synonymy (i.e., $\varphi \leftrightsquigarrow \psi$ being true at every world). In the general setup, neither implies the other.[13] This is welcome given the problems that we've seen of aligning cognitive synonymy with some sort of sameness of content. If we do move to the restricted setup where syntax-insensitivity is required, we do at least get that sameness of content implies cognitive synonymy.[14]

5 Two ways to look at it

Our modeling gives evidence that there is a coherent notion of cognitive synonymy, delivering the expected (in)validities and vindicating intuitions for intransitivity. The take-home of 3.3 and 3.4, however, is that the non-transitive notion is not very stable. Let us discuss this in light of our semantics.

Say we accept transitivity: in Sect. 3.4, we saw three arguments for it, based on the ideas of substitution *salva veritate*, cognitive role, and non-monotonic reasoning respectively. The semantics adds another: if we used comparative similarity in the clause for cognitive entailment, we would get a transitive cognitive synonymy. As with the substitution *salva veritate* and cognitive role arguments, this, too, may be interpreted as saying that enough uniformity entails transitivity. Comparative similarity provides a more uniform structure to interpret the conditional than a selection function: in the former, it is always the minimal worlds that are selected; in the latter, there is hardly any constraint (except Success) on which worlds can be selected.

However, semanticists working in the selection function tradition also have considered a uniformity condition on models (Starr, 2019, appendix B):

(*Uniformity*) $f_\varphi(w) \subseteq |\psi| \ \& \ f_\psi(w) \subseteq |\varphi| \Rightarrow f_\varphi(w) = f_\psi(w)$

[12] *Countermodel*: $W = \{w\}$; R_φ is empty for all φ; $\mathcal{T} = \big\{\{0\}, \{1\}, \{0, 1\}\big\}$; $\oplus$ is set-union (which clearly is idempotent, commutative, and associative); and for all propositional atoms p, $t(p) := \{0\}$ if $p \neq r$ and $t(r) := \{1\}$; and w makes true no propositional atom. Success is satisfied since all relations are empty. Then $w \Vdash p \rightsquigarrow q$ since $\forall w_1 : wR_p w_1 \Rightarrow w_1 \Vdash r$ (because R_p is empty) and $t(q) = \{0\} \leq \{0\} = t(p)$. However, $w \nVdash p \rightsquigarrow q \vee (q \wedge r)$ because $t(q \vee (q \wedge r)) = \{0\} \cup \{0\} \cup \{1\} = \{0, 1\} \nleq \{0\} = t(p)$.

[13] Proof: In one direction, consider the model from footnote 8 except that the one world w doesn't make true any atom. Then $p \wedge q$ and $q \wedge p$ have the same content (as always), but w doesn't make true $p \wedge q \rightsquigarrow q \wedge p$, because w can $R_{p \wedge q}$-access a world, namely itself, where $p \wedge q$ is not true. In the other direction, again consider the same model but now with all relations being empty. Then trivially $p \wedge \neg p \leftrightsquigarrow p \vee \neg p$ is true at every world (since they have the same topic), but the two sentences don't have the same content since the former is false at w while w is true.

[14] Proof: Assume $[\varphi] = [\psi]$. By antecedent substitution, $[\varphi \rightsquigarrow \varphi] = [\psi \rightsquigarrow \varphi]$ and $[\psi \rightsquigarrow \psi] = [\varphi \rightsquigarrow \psi]$. By reflexivity, $\varphi \rightsquigarrow \varphi$ and $\psi \rightsquigarrow \psi$ are true at every world, so $\varphi \leftrightsquigarrow \psi = \varphi \rightsquigarrow \psi \wedge \psi \rightsquigarrow \varphi$ is true at every world, as desired.

If all worlds one looks at given φ are ψ-worlds and *vice versa*, then the φ-selected worlds are the same as the ψ-selected worlds. The motivation usually given is that Uniformity entails Equivalence—restated formally:

$(Equivalence)\ \{\varphi \leftrightsquigarrow \psi, \varphi \rightsquigarrow \chi\} \vDash \psi \rightsquigarrow \chi$[15]

Uniformity together with Success entails Syntax-insensitivity, too.[16] And these make cognitive synonymy transitive: 2-way transitivity is validated.

One who accepts one of the independent justifications for (the principles entailing) transitivity should explain the intuition against transitivity conveyed by dead parrot sequences. Here is a conjecture that might be empirically testable (though it is not us who will test it). Perhaps what goes on in a dead parrot series $\varphi_1, \varphi_2, \ldots, \varphi_n$ is that, instead of cognitive synonymy, one actually only acknowledges one-way cognitive entailments, $\varphi_1 \rightsquigarrow \varphi_2 \rightsquigarrow \ldots \rightsquigarrow \varphi_n$. Humans parse the syntax of sentences sequentially. In a dead parrot series, one may just move forward following the flow of syntax, finding that each item obviously entails the successor, without also checking the reverse entailment at each step. But then no problem remains because even assuming Uniformity, cognitive entailment still fails transitivity:

$(1\text{-}way\ transitivity)\ \{p \rightsquigarrow q, q \rightsquigarrow r\} \nvDash p \rightsquigarrow r$[17]

So the agent can still consistently fail to be on top of the transitive closure of a dead parrot series.

Say we keep stubbornly rejecting transitivity for cognitive synonymy. We have to explain how to resist the transitivity arguments' plausible assumptions. Here are some ideas. Why is it that the formal semantics can violate the

---

[15] *Proof:* Let $\mathfrak{M}$ be a model and w a world with (a) $w \Vdash \varphi \leftrightsquigarrow \psi$ and (b) $w \Vdash \varphi \rightsquigarrow \chi$. By (a), $f_\varphi(w) \subseteq |\psi|$ and $f_\psi(w) \subseteq |\varphi|$. So Uniformity implies $f_\varphi(w) = f_\psi(w)$. Now, to show $w \Vdash \psi \rightsquigarrow \chi$, first let $wR_\psi w_1$ and show $w_1 \Vdash \chi$. Since $f_\varphi(w) = f_\psi(w)$, also $wR_\varphi w_1$. So, by (b), we have $w_1 \Vdash \chi$. Second, (a) also implies $t(\varphi) = t(\psi)$, so we have, by (b), that $t(\chi) \leq t(\varphi) = t(\psi)$.

[16] *Proof*: It's enough to assume $|\varphi| = |\psi|$. We need to show, for w and w_1, that $wR_\varphi w_1$ iff $wR_\psi w_1$. By Success, $f_\varphi(w) \subseteq |\varphi| = |\psi|$ and $f_\psi(w) \subseteq |\psi| = |\varphi|$. So Uniformity implies $f_\varphi(w) = f_\psi(w)$, and the conclusion follows.

[17] *Countermodel*: $W = \{w, w'\}$; $\mathcal{T} = \{x\}$; $x \oplus x = x$; $t(p) = x$ for all p; $\Vdash$ is defined by p being true at w, q being true at w and w', and r being true at w, and nothing more. Write $P = \{w'\}, Q = \{w, w'\}, R = \{w\}$. Define $wR_\varphi w'$ if $|\varphi| = P$ and $wR_\varphi w$ if $|\varphi| = Q$ and $|\varphi| = R$ if $wR_\varphi w$, and no further relations. (Technically, we need to define this by induction on φ: if φ is atomic, $|\varphi|$ is defined by $\Vdash$ above; if φ is Boolean, $|\varphi|$ is defined with the usual clauses; if φ is of the form $\psi \rightsquigarrow \chi$ and R_ψ is already defined, then $|\varphi| = \{w : (1)\ \forall v : wR_\psi v \Rightarrow v \Vdash \chi \text{ and } (2)\ t(\chi) \leq t(\psi)\}$ is defined, too.) Success is satisfied: If $w_1 R_\varphi w_2$, then $w_1 = w$ and either $|\varphi| = P$ and $w_2 = w' \in P = |\varphi|$ or $|\varphi| = Q$ and $w_2 = w \in Q = |\varphi|$ or $|\varphi| = R$ and $w_2 = w \in R = |\varphi|$. Uniformity is satisfied: Assume $f_\varphi(v) \subseteq |\psi|$ and $f_\psi(v) \subseteq |\varphi|$, and show $f_\varphi(v) = f_\psi(v)$. If $v \neq w$, then $f_\varphi(v) = \emptyset = f_\psi(v)$. So let $v = w$. Note that $|\varphi| \in \mathcal{P}(W) = \{P, Q, R, \emptyset\}$. If $|\varphi| = P$, then $P = f_\varphi(w) \subseteq |\psi|$, so $|\psi| = P$ or $|\psi| = Q$. The latter cannot be, since otherwise $\{w\} = f_\psi(w) \subseteq |\varphi| = P$. So $|\psi| = P$, hence $f_\varphi(w) = f_\psi(w)$. If $|\varphi| = Q$, then $\{w\} = f_\varphi(w) \subseteq |\psi|$, so $|\psi| = R$ or $|\psi| = Q$. In either case, $f_\psi(w) = \{w\} = f_\varphi(w)$. If $|\varphi| = R$, then $\{w\} = f_\varphi(w) \subseteq |\psi|$, so again $|\psi| = R$ or $|\psi| = Q$, and the claim follows. If $|\varphi| = \emptyset$, then, since $f_\psi(w) \subseteq |\varphi|$, we have $f_\psi(w) = \emptyset$, so $|\psi| = \emptyset$, hence $f_\varphi(w) = f_\psi(w)$. Finally, $w \Vdash p \rightsquigarrow q$ since $f_p(w) = \{w'\} \subseteq Q$; $w \Vdash q \rightsquigarrow r$ since $f_q(w) = \{w\} \subseteq R$; but $w \nVdash p \rightsquigarrow r$ since $f_p(w) = \{w'\} \nsubseteq R$.

Equivalence principle even though it supposedly is so fundamental? The accessibility relations R_φ or functions f_φ are indexed to formulas. Thus, two sentences like p and $p \wedge p$ that obviously coincide in subject matter and truth set can cognitively entail different things insofar as they prompt the agent to consider different worlds, i.e., $f_p \neq f_{p \wedge p}$. So the agent is *syntax sensitive*. The failure of compositionality that then is entailed by Leitgeb's impossibility result may then be taken as the agent being on top of the relevant cognitive contents only to some degree.

To motivate this, one may take cognitive entailment $\varphi \rightsquigarrow \psi$ as the agent having in their knowledge base a (defeasible) rule 'If φ, then usually ψ'. Such rules would be syntax-sensitive for even replacing φ with $\varphi \wedge \varphi$ takes cognitive effort: conscious reasoning, learning, etc. Cognitive entailment and synonymy are relative to a knowledge base which is a collection of syntactic rules and facts. Cognitive synonymy may fail to be transitive because the agent has in the knowledge base the rules 'If φ, then ψ, and *vice versa*' and 'If ψ, then χ, and *vice versa*', but hasn't (yet) made the cognitive effort to also add to the knowledge base the transitive closure 'If φ, then χ, and *vice versa*'. For a worked out version of the idea, see Hornischer (2017).

Also, one may invoke context-sensitivity: cognitive entailment and synonymy as being judged relative to a knowledge base provided by context. This may be the knowledge base of a particular agent, of a generic biologist, or common knowledge. In Sect. 3.4, we saw context-sensitivity as a means to keep the spirit of the substitution *salva veritate* and cognitive role principles without giving them the strength to imply transitivity. Context-sensitivity matches our semantics in its generality, without Uniformity: take a world w as involving a pair of a possible cognitive state of the agent and a context with respect to which cognitive synonymy is assessed. Then even if φ and ψ are truth- and topic-equivalent, at w the φ-selected worlds and the ψ-selected worlds may differ because the agent hasn't yet processed their equivalence.

Finally, one may remark that a dead parrot series is similar to a soritical series $\varphi_1, \dots, \varphi_n$, with φ_1 true and φ_n false, while, for each $k < n$, it seems that if φ_k, then φ_{k+1} (cf. cognitive entailment in our dead parrot series). As an example, take φ_k = 'A man with k-many hairs is bald' and n = 100,000. Sorites sequences breed paradox since the premises are plausible but repeated application of the inductive principle yields the contradiction that φ_n is true. There are contextualist replies (see e.g. Shapiro, 2006): they invoke different contexts with respect to which the involved sentences are evaluated. To adapt an idea of Raffman (1994) to our setting, assume we move along the dead parrot series. At some point φ_k, we wouldn't judge φ_k to be cognitively synonymous with φ_1 anymore. This means: we shifted context. We entered a new cognitive state, e.g., with different verbal dispositions, in which φ_k is taken as the new paradigm intended formulation. According to the former context in which φ_1 was the paradigm way of expressing things, φ_{k-1} was still an okay version, but φ_k wasn't anymore. Now, according to the new context, φ_{k-1} still is okay, but φ_1 isn't anymore.

6 Conclusion

Cognitive synonymy is important but elusive, difficult to model, and seemingly shaped by conflicting intuitions. But we have provided a model for it, which captures the intuitions. The conflict is analyzed as the demand for and against transitivity. Our model can accommodate either by tweaking a single parameter: uniformity. It *explains* demanding transitivity in the uniformity and stability of our cognitive apparatus and rejecting transitivity in its sensitivity to context and syntax.

There is, of course, further work ahead. Here's one direction of future investigation (suggested by an anonymous referee): Fregean identity puzzles.

1 Superman is cool.
2 Clark Kent is cool.

Lois Lane may take (1) as true, not (2), and so these wouldn't play the same role in Lois' cognitive life. What does our account have to say on this? Well, it could go either way. Besides being true at the same worlds, perhaps (1) and (2) *are* about the same topic: they're both about the individual Kal-El, or whatever unique topic is suitably associated, in context, to the individual Kal-El. If so, Lois' different attitudes would have to be explained by syntax-sensitivity in the non-uniform setting, where the accessibilities or selection functions are indexed to sentences: (1) and (2) make Lois look at different scenarios in her thought, perhaps by triggering different guises or modes of presentation (Salmon, 1986) which are, themselves, extraneous to topicality.

However, there are approaches to subject matters on the market, which can make it so that 1) and (2) differ in topic, e.g., Hawke (2018)'s. Hawke has the topics of atomic sentences as structured constructions out of things akin to abstract Fregean senses. (1) and (2) can be about different things by having different Fregean senses featuring in the respective subject matters (say, one features *the super-hero wearing a red-blue costume* the other *the shy journalist wearing glasses*, etc.). To model this adequately in our setting, one would probably need a predicative language rather than our merely propositional one, and atomic sentences which get structured topics featuring as constituents the topics assigned to their subsentential components, given that 1) and (2) differ only by the substitution of (rigid) co-referents 'Superman' and 'Clark Kent'. One reason to leave this to further work is that it seems still a very open issue, among subject matter theorists, whether Frege puzzles should be dealt with by employing (distinctions concerning) subject matters, rather than devices of other kind.

Acknowledgements Versions of this paper have been presented at the Logic and Metaphysics seminar at the Arché research centre, University of St Andrews, at the LIRa seminar at ILLC, University of Amsterdam, and at the LanCog group, University of Lisbon. Thanks to the audiences and to two anonymous referees of this Journal for helpful questions, objections, and remarks. This research is published within the project 'The Logic of Conceivability', funded by the European Research Council (ERC CoG), Grant Number 681404.

References

Andréka, H., Németi, I., & van Benthem, J. (1998). Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic, 27*, 217–274.

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. CSLI Publications.

Belnap, N. (1977). A useful four-valued logic. In M. Dunn & G. Epstein (Eds.), *Modern uses of multiple-valued logics* (pp. 5–37). Reidel.

Berto, F. (2018a). Simple hyperintensional belief revision. *Erkenntnis, 84*, 559–75.

Berto, F. (2018b). Aboutness in imagination. *Philosophical Studies, 175*, 1871–86.

Berto, F. (2022). *Topics of thought. The logic of knowledge, belief, imagination*. Oxford University Press.

Berto, F., & Jago, M. (2019). *Impossible worlds*. Oxford University Press.

Bjerring, J. C., & Schwarz, W. (2017). Granularity problems. *The Philosophical Quarterly, 67*, 22–37.

Burgess, John P. (1981). Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic, 22*(1), 76–84.

Busby, E., Flynn, D. J., & Druckman, J. N. (2018). Studying framing effects on political preferences. In P. D Angelo (Ed.), *Doing news framing analysis* (Vol. II, pp. 67–90). Routledge.

Carnap, R. (1947). *Meaning and necessity*. Chicago University Press.

Chalmers, D. (2011). Propositions and attitude ascriptions: A Fregean account. *Noûs, 45*, 595–639.

Chellas, B. (1975). Basic conditional logic. *Journal of Philosophical Logic, 4*, 133–53.

Churchland, P. (1993). State-space semantics and meaning holism. *Philosophy and Phenomenological Research, 53*, 667–72.

Cruse, D. A. (2000). *Meaning in language*. Oxford University Press.

Deutsch, H. (1984). Parconsistent analytic implication. *Journal of Philosophical Logic, 13*, 1–11.

Dunn, J. M. (1976). Intuitive semantics for first-degree entailment and 'coupled trees'. *Philosophical Studies, 29*, 149–68.

Egré, P., & Rott, H. (2021). The logic of conditionals. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* Metaphysics Research Lab, Stanford University, Winter 2021 edition.

Epstein, R. L. (1993). *The semantic foundations of logic: Propositional logics*. Oxford University Press.

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

Fagin, R., Halpern, J., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. MIT Press.

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

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

Fine, K. (2016b). Constructing the impossible. In L. Walters & J. Hawthorne (Eds.), *Conditionals, probability, and paradox: Themes from the Philosophy of Dorothy Edgington.* Oxford University Press.

Fine, K. (2020). Yablo on subject matter. *Philosophical Studies, 177*, 129–71.

Fodor, J., & Lepore, E. (1999). All at sea in semantic space: Churchland on meaning similarity. *The Journal of Philosophy, 96*, 381–403.

Frege, G. (1891). *Funktion und Begriff*. Hermann Pohle.

Frege, G. (1979). A brief survey of my logical doctrines. In H. Hermes, F. Kambartel, & F. Kaulbach (Eds.), *Posthumous writings* (pp. 197–202). Blackwell.

Goodman, N. (1949). On likeness of meaning. *Analysis, 10*, 1–7.
Goodman, N. (1961). About. *Mind, 70*, 1–24.
Hawke, P. (2016). Questions, topics and restricted closure. *Philosophical Studies, 73*, 2759–84.
Hawke, P. (2018). Theories of aboutness. *Australasian Journal of Philosophy, 96*, 697–723.
Hintikka, J. (1962). *Knowledge and belief. An introduction to the logic of the two notions*. Cornell University Press.
Hoek, D. (2022). Minimal rationality and the web of questions. In D. Kindermann, P. Van Elswyk, & E. Egan (Eds.), *Unstructured content.* Oxford University Press.
Holliday, W. (2012). *Knowing what follows: Epistemic closure and epistemic logic*. PhD dissertation, Stanford.
Hornischer, L. (2017). Hyperintensionality and synonymy: A logical, philosophical, and cognitive investigation. Master's thesis. https://www.illc.uva.nl/Research/Publications/Reports/MoL-2017-07.text.pdf
Hornischer, L. (2020). Logics of synonymy. *Journal of Philosophical Logic, 49*, 767–805.
Humberstone, L. (2008). Parts and partitions. *Theoria, 66*, 41–82.
Jago, M. (2007). Hintikka and cresswell on logical omniscience. *Logic and Logical Philosophy, 15*, 325–54.
Jago, M. (2014). *The impossible. An essay on hyperintensionality*. Oxford University Press.
Jespersen, B., & Duži, M. (2015). Introduction, special issue on hyperintensionality. *Synthese, 192*, 525–34.
Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341–50.
Kiourti, I. (2010). *Real impossible worlds: The bounds of possibility*. PhD thesis, University of St Andrews.
Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence, 44*, 167–207.
Leitgeb, H. (2008). An impossibility result on semantic resemblance. *Dialectica, 62*, 293–306.
Levesque, H. (1984). A logic of implicit and explicit belief. In *National conference on AI*, AAAI-84 (pp. 198–202).
Lewis, D. (1973). *Counterfactuals*. Blackwell.
Lewis, D. (1988a). Relevant implication. *Theoria, 54*, 161–74.
Lewis, D. (1988b). Statements partly about observation. *Philosophical Papers, 17*, 1–31.
Lyons, J. (1996). *Linguistic semantics*. Cambridge University Press.
Mates, B. (1952). Synonymity. In L. Linsky (Ed.), *Semantics and philosophy of language.* University of Illinois Press.
Murphy, M. L. (2003). *Semantic relations and the lexicon*. Cambridge University Press.
Nolan, D. (1997). Impossible worlds: A modest approach. *Notre Dame Journal of Formal Logic, 38*, 535–72.
Nolan, D. (2013). Impossible worlds. *Philosophy Compass, 8*, 360–72.
Nute, D. (1984). Conditional logic. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic: Volume II: Extensions of classical logic.* Springer.
Özgün, A., & Cotnoir, A. (2021). Imagination and mereotopology. Unpublished manuscript.
Parry, W. T. (1933). Ein axiomensystem für eine neue art von implikation (analitische implikation). *Ergebnisse eines Mathematischen Kolloquiums, 4*, 5–6.
Perry, J. (1989). Possible worlds and subject matter. In *The problem of the essential indexical and other essays* (pp. 145–160). CSLI Publications.
Plebani, M., & Spolaore, G. (2021). Subject matter: A modest proposal. *The Philosophical Quarterly, 71*, 605–22.
Priest, G. (2008). *An introduction to non-classical logic. From If to Is*. Cambridge University Press.
Priest, G. (2016). *Towards non-being: The logic and metaphysics of intentionality* (2nd ed.). Oxford University Press.
Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review, 60*(1), 20–43.
Raffman, D. (1994). Vagueness without paradox. *The Philosophical Review, 103*(1), 41–74.
Salmon, N. (1986). *Frege's puzzle*. MIT Press.
Schellenberg, S. (2012). Sameness of Fregean sense. *Synthese, 189*, 163–75.
Schipper, A. (2018). Aboutness and negative truths: A modest strategy for truthmaker theorists. *Synthese, 195*, 3685–722.

Schroeter, L. (2021). Two-dimensional semantics. In Edward N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* Metaphysics Research Lab, Stanford University, Spring 2021 edition.
Shapiro, S. (2006). *Vagueness in context*. Oxford University Press.
Skipper, M., & Bjerring, J. C. (2020). Hyperintensional semantics: A Fregean approach. *Synthese, 197*, 3535–58.
Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory (American Philosophical Quarterly Monographs 2)* (pp. 98–112). Blackwell.
Stalnaker, R. (1984). *Inquiry*. MIT Press.
Stanojević, M. (2009). Cognitive synonymy: A general overview. *Facta Universitatis, 7*, 193–200.
Starr, W. (2019). Counterfactuals. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* Metaphysics Research Lab, Stanford University, Fall 2019 edition.
Van Fraassen, B. (1969). Facts and tautological entailments. *Journal of Philosophy, 66*, 477–87.
Veltman, F. (1985). *Logics for conditionals*. PhD thesis, Universiteit van Amsterdam.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.
Yablo, S. (2014). *Aboutness*. Princeton University Press.
Yalcin, S. (2016). Belief as question-sensitive. *Philosophy and Phenomenological Research, 97*, 23–47.