

Concept originalism, reference-shift and belief reports

Seyed N. Mousavian^{1,2} ·
Mohammad Saleh Zarepour³

Received: 15 February 2016 / Accepted: 3 September 2016 / Published online: 15 September 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Concept originalism, recently introduced and defended by Sainsbury and Tye (S&T) (Proc Aristotelian Soc Suppl Vol 85:101–124, 2011, Seven puzzles of thought and how to solve them: an originalist theory of concepts. Oxford University Press, Oxford, 2012), Tye (in Brogaard, Does perception have content? Oxford University Press, Oxford, 2014), and Sainsbury (Erkenntnis 80:195–214, 2015), holds that “atomic concepts are to be individuated by their historical origins, as opposed to

Earlier drafts of this paper have been presented at the 43rd annual meeting of the Society for Exact Philosophy (2015, McMaster University, Canada), the GRG at LOGOS: Research Group in Analytic Philosophy (2014, University of Barcelona, Spain), the research seminars in theoretical philosophy at FLOV/the Department of Philosophy, Linguistics and Theory of Science (2014, University of Gothenburg, Sweden), the 3rd annual conference of the Iranian Association for Logic (2015, University of Tehran, Iran), and the summer talk series at the School of Philosophy at IPM (2015, the Institute for Research in Fundamental Sciences, Iran). We are greatly thankful to our audience on each occasion. In particular, we thank Mahrads Almotahari, Mohammad Ardeshtir, Sten Ebbesen, Dan Giberman, Katerina Ierodiakonou, Martin Kasá, Kaveh Lajevardi, Bernard Linsky, Hossein Masoumi Hamedani, Anna-Sofia Maurin, Damian Melamedoff, Mostafa Mohajeri, Mahmoud Morvarid, Zia Movahed, Amir Saemi, Sajed Tayebi, Hamid Vahid, and Mohsen Zamani for their helpful suggestions. We are indebted to two anonymous referees of this journal for their instructive and insightful comments. The first author’s research is partly done within the research programme Representation and Reality: Historical and Contemporary Perspectives on the Aristotelian Tradition (funded by Riksbankens Jubileumsfond).

✉ Seyed N. Mousavian
seyed.mousavian@gu.se
Mohammad Saleh Zarepour
msz26@cam.ac.uk

¹ Department of Philosophy, Linguistics, and Theory of Science, University of Gothenburg, Box 200, 405 30 Gothenburg, Sweden

² School of Philosophy Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

³ University of Cambridge, Clare Hall, Herschel Road, Cambridge CB3 9AL, UK

their semantic or epistemic properties” (S&T 2012, p. 40). The view is immune to Gareth Evans’s “Madagascar” objection to the Causal Theory of Reference since it allows a concept (and thus a name) to change its reference over time without losing its identity. The possibility of (inadvertent) reference-shift, however, raises the problem of misleading belief reports. S&T try to tackle the problem by strengthening the sufficient condition for a truthful belief report. We will argue that, first, their solution fails, second, and more importantly, their diagnosis of the root of the problem is misguided, third, two initially appealing ways out of the problem fail, and fourth, the prospect of finding a solution to the problem within CO is dim. The view opens the Pandora’s box of reference-shift, in a wide range of cases, without providing the necessary semantic means to take care of them.

Keywords Originalism · Concept · Thought · Content · Belief report · Reference · Madagascar

1 Concept originalism (CO)

According to Sainsbury and Tye (S&T) (2011, 2012), Tye (2014), and Sainsbury (2015), “atomic concepts are to be individuated by their historical origins, as opposed to their semantic or epistemic properties” (S&T 2012, p. 40).¹ What makes a concept that very concept is its origin, more particularly its “originating use” (ibid, 42). Each concept, as a contingent abstract object, has many contingent concrete “correlates”, namely its tokens (ibid, 41). To use a concept is to use a token of it. A *non-originating use* of a concept has two characteristics: (1) it involves “deference” to other uses of the concept (namely, to intend to use the concept as it has been used already) and (2) it makes the accumulation of information from the other uses possible. An *originating use* of a concept, in contrast, contains neither a “deferential intention” nor “informational accumulation” (ibid, 43); this is a use in which the concept is originally introduced.² Thus, the main principle of Concept Originalism (CO) is (ibid, 44):

(O) Necessarily: concept C1 = concept C2 iff the originating use of C1 = the originating use of C2.³

According to CO, concept transmission is nothing but “standing in the ancestral of a deference involving relation R to the originating use” (S&T 2012, p. 44). *Deference*, the key notion, is not to be understood semantically: “[it] can be modeled (rather over-intellectually) as the recognition that others already use a concept, together with desire to use the very concept they use, [...]”. It does not require the notion of an expert, nor does one need to think of “correct use” as use to form a true thought [...]” (ibid, 70). Therefore, although using the same concept involves the intention to use it in accordance with some previous uses, it does not involve the actual satisfaction

¹ Concept Originalism, in structure, is apparently similar to the Common Currency Conception of words, introduced and defended by Kaplan (1990, 2011). According to the latter view, *words* are individuated by their “life histories” (Kaplan 1990, p. 108). The two views, however, are different in significant respects.

² Following S&T (2012, p. 66), we put cases of language-less users of concepts aside.

³ CO needs to explain the crucial role that concepts play in the identification of thoughts and semantics of attitude reports. We will discuss these matters in Sects. 2 and 4.

of this intention. CO holds that the learner of a concept needs to intend to refer to its reference, but she may acquire the concept even if this intention fails.⁴

2 CO's terminology

Well-formed expressions of natural language have/express (are semantically associated with)⁵ concepts or structures of concepts: Names and other referring expressions have *nominative* concepts (e.g. 'Fido' has FIDO as its concept), predicates have *predicative* concepts (e.g. 'bark' has BARK as its concept), logical operators have *logical* concepts (e.g. 'and' has AND as its concept), and so on (ibid, 21).⁶ Declarative sentences have properly structured concepts or thoughts (ibid, 47). For example,

1. Fido barks

expresses a thought that may be represented by:

1T. < FIDO, BARK >

(1T) is a unary atomic thought constituted of a nominative concept, FIDO, and a predicative one, BARK, properly linked together.⁷ We may have negative thoughts, conjunctive thoughts, and the like.

A concept or a thought has a content. The content of a nominative concept, at a world, is its actual referent, if any, (which is the same as the actual referent of the corresponding name) (ibid, 48); 'Fido' has FIDO as its nominative concept and the latter has Fido, the individual, as its content. The content of a predicative concept, at a world, is its actual referent, if any, (which is the same as the actual referent of the corresponding predicate); 'barks' has BARK as its concept and the latter has the property of barking as its content. Hence, properties are not concepts; in fact, concepts represent (or refer to) properties (ibid, 2). The content of a unary atomic thought is the set of all possible worlds *w* in which each constituent concept of the thought has a content *and* "the content of the nominative concept at *w* possesses the property that is the content of the predicative concept at *w*" (ibid, 48). The content of a negative

⁴ Two points: First, Sainsbury (2015) develops CO into Name Originalism (NO) according to which "The same-name relation among uses is determined by (a) unique origination and (b) deferential historical chains" (Sainsbury 2015, p. 196). Sainsbury's (2015) main semantic thesis is that "The same-reference relation among uses of the same name is affected by distinct deferential historical chains" (ibid). NO furnishes the ontology of names, as devices of direct reference, according to CO. NO and Kaplan's Common Currency Conception of Words are not identical. For one thing, the origin of a name is different from its history. For another, Kaplan does not introduce concepts as what words express. Second, Tye (2014) builds on S&T's (2012, pp. 152–157) argument against gappy content and attempts to develop a non-Meinongian non-adverbial account of the content of hallucination based on the originalist conception of 'content'. We consider this as an application of the view, not as an essential part of it.

⁵ S&T do not explicitly *discuss* the nature of this association; however it is consistent with their view to assume that this relation has a semantic character, as they sometimes get very close to this formulation, for example: "Concepts expressed in a public language have the semantic properties of the words that express them" (S&T 2012, p. 45).

⁶ S&T (2012) use small-caps notation, e.g. 'MEAT', for terms that refer to concepts (i.e. concept terms). The first letter upper case format applies to *nominative* concept terms, e.g. 'HESPERUS'.

⁷ We fudge on the issue of proper representation of thoughts; for example, we do not represent the distinction between the concept of *bark(s)* and that of *barking*. We hope that this does not have any serious effect on our presentation of CO, as S&T have not dealt with the issue either.

thought or a conjunctive thought, in turn, can recursively be defined. Therefore, “a thought is true iff the actual world is a member of the associated set [of worlds] and is otherwise false” (ibid, 49). In addition to concepts, thoughts, and contents, S&T (2012) introduce extensions: “The extension of a nominative concept is its referent (just as for the naïve Millianism), of a predicative concept the things of which it is true, and of a thought, its truth-value. This notion can be relativized to worlds” (S&T 2012, p. 32). Now, to (inadvertent) reference-shift.

3 CO and the ‘Madagascar’ case

Gareth Evans (1973), announcing the ‘Madagascar’ case as “decisive against the Causal Theory of Names”, describes it as follows:

Change of denotation is similarly decisive against the Causal Theory of Names. Not only are changes of denotation imaginable, but it appears that they actually occur. We learn from Isaac Taylor’s book: *Names and their History*, 1898: “In the case of ‘Madagascar’ a hearsay report of Malay or Arab sailors misunderstood by Marco Polo... has had the effect of transferring a corrupt form of the name of a portion of the African mainland to the great African Island.” (Evans 1973, pp. 195–196)

Kripke (1980, p. 163) leaves “the problem for further work” and as John P. Burgess (2014, p. 195) rightly points out, he “has not yet provided any such extended discussion in print.” CO, however, is immune to the ‘Madagascar’ case. According to CO, different tokens of a proper name may refer to different objects since the nominative concept expressed by the name, may refer to different objects at different times:

Marco Polo heard some locals using the word ‘Madagascar’ (or some similar word) and was thereby introduced to a concept. He had appropriate deferential intentions: his plan was to use the concept in the way the locals did. He was confused about its referent: the locals used it to refer to Mogadishu, the town on the mainland, whereas he thought they used it to refer to the island Madagascar (hundreds of miles away). Marco Polo’s mistaken usage of the concept became standardized, and the usage of the locals had negligible later effects [...]. In this case, it seems more appropriate, to say that the original concept MADAGASCAR changed its referent.

(S&T 2012, p. 71)

CO can easily explain the ‘Madagascar’ case away. Given CO, concept transmission is pretty much like reference borrowing with regard to names, with no significant semantic or epistemic restriction on the deference relation involved. In fact, CO generalizes the possibility of (inadvertent) reference shift to all concepts, and thus all terms:

The concept MEAT used in the fifteenth century had as its reference anything edible; in our terms, its content was the property of *being edible*. The concept MEAT we use now has as its reference only flesh; in our terms, its content is the property of *being flesh*. One option is to say that earlier concept is the same as the current concept, but that its content has changed. Another option is to say

that a new concept, expressed by a word spelled and pronounced the same way, was introduced at some point, and each of the two concepts have retained their original and distinct contents. We prefer the first story. It is a case of gradual drift, with no event that seems a good candidate for the introduction of a new concept.

(S&T 2012, p. 46)

To sum up, CO allows reference-shift over a single causal chain of reference borrowing. This is exactly the point that Evans's objection to the Causal Theory of Reference tries to make. But Evans's objection is no objection to CO since a causal chain of communication of a concept doesn't need to deliver a single reference for it; at different nodes of the chain, as it were, different tokens of the concept may have different references. In fact, this is part of the view.⁸

We claim, however, that CO's semantics of belief reports does not accommodate belief reports containing terms with reference-shift. To see the problem, we must briefly explain CO's semantics of belief reports.⁹

4 CO's semantics of belief reports

Consider:

2. S believes that P¹⁰

According to CO, $\lceil P \rceil$ is the thought put on display in (2); it is a thought of the attributer, who is not explicitly represented in (2). If the subject of the attribution (attributee), S, is appropriately related to a thought $\lceil Q \rceil$ (not necessarily identical with $\lceil P \rceil$) which itself stands in an R-relation to $\lceil P \rceil$, then (2) is true, and vice versa (S&T 2012, p. 118).¹¹ CO accepts a relational analysis of *belief states* but denies a relational analysis of *belief reports*: "Even though one who entertains a thought is in a relational state, of which a thought is a relatum, it does not follow that ascriptions of such states are semantically or syntactically relational, referring to the thought *as a relatum*" (S&T

⁸ Sainsbury (2015, p. 211) is explicit that he accepts Madagascar cases "at face value".

⁹ CO's treatment of (inadvertent) reference-shift may be criticized for being incomplete: the view needs to provide a systematic account of the relationship between concepts and their referents and how they contribute to the necessary and sufficient conditions of attitude ascriptions (as well as other complex sentences containing intensional operators). S&T (2012, pp. 69–72) briefly mention the first issue and partially discuss the second one in passing. According to S&T, "Originalism as such does not entail a view about the fixation of reference" (ibid, 69); however, obviously, the view is in serious need to adopt a theory of reference and develop a full semantics for attitude ascriptions. Having said that, we are following a different line of reasoning: We are trying to show that CO, as it stands, is inadequate by focusing on the consequences of the possibility of reference-shift for S&T's sufficient condition for truthful belief reports. Admittedly, the problem runs deeper.

¹⁰ (2) is a schema. A proper name and a declarative sentence of English, not their names, should be substituted for 'S' and 'P' respectively. Also, in what follows, we use quasi-quotation in a slightly different form from Quine (1940, pp. 33–37): instead of names of unspecified statements, we use unspecified statements themselves in quasi-quotations.

¹¹ The R-relation, here, should not be conflated with the deference relation discussed in earlier sections. The nature of this relation, as a relation between thoughts, is to some extent obscure. It "is sensitive to contextual factors" and "may differ for different verbs of attitude" (ibid, 118). S&T (2012, pp. 116–121) try to show that the R-relation is neither identity nor even an equivalence relation, necessarily. Below we will further discuss the conditions on the R-relation for truthful belief reports.

2012, p. 32). Consequently, the view rejects the so-called “Specification Assumption”, namely a true belief report must specify the thought the believer (i.e. the ascriber) has in her belief box.¹²

To illustrate, consider:

3. Mark Sainsbury believes that he is tall.

4. Michael Tye believes that he is tall.

In (3) and (4), ‘he’ refers to Mark Sainsbury. Mark and Michael believe the “same thing”,¹³ i.e. that Mark Sainsbury is tall. However, in a sense, there may be no single thought that is what they *both* believe or are belief-related to.¹⁴ They may believe the same thing in virtue of being belief-related to two different thoughts. For example, Mark may be belief-related to the thought represented by:

3M. <I, BEING TALL>¹⁵

And Michael may be belief-related to the thought represented by:

4M. <SAINSBURY, BEING TALL>

These thoughts are not identical; they have different constituents and different properties. The thought put on display in the belief reports, i.e. in (3) and (4), in turn, may be represented by:

3/4T. <HE, BEING TALL>¹⁶

This thought is in an R-relation to the thoughts represented by (3M) and (4M).

5 Belief reports containing ‘Madagascar’

Suppose that Sainsbury makes the following belief report:

5. Adam believed that Madagascar is not an island.

Adam is the local from whom Marco Polo learned ‘Madagascar’. By assumption, Adam used ‘Madagascar’ to refer to Mogadishu and Sainsbury, following Marco Polo’s usage, uses ‘Madagascar’ to refer to Madagascar. Given CO, there is one concept MADAGASCAR, once entertained by Adam and now employed by Sainsbury, that refers to different objects at different times. Given that Adam and Sainsbury use the same concept of *being an island* and *not*, we may suppose that the thought to which Adam was belief-related and the thought put on display in (5) are the same since they have the same constituents and the same structure. This thought may be represented by:

5T. <NOT < MADAGASCAR, BEING AN ISLAND>>

¹² See Kent Bach (1997, p. 222). In rejecting the Specification Assumption, Sainsbury and Tye follow Bach (1997), Taylor (2002), and Fara (2003), among others.

¹³ That Mark and Michael believe the “same thing” does not imply that they are related to the same content. By using the terminology of “believing the same thing”, we are simply following S&T who “grant that if Smith and Jones believe that the US financial situation is worsening, then they believe the same thing ... [However,] it does not follow that there exists an entity Smith and Jones believe” (S&T, 2012, p. 115).

¹⁴ Again, we are borrowing the term “belief-related” from S&T (ibid, 118–119, n. 12, 120, 128). Accordingly, S is belief-related to the thought that P iff S has the thought that P in her belief box. Nonetheless, “we need not express the very thought that is in the belief box of the subject whose beliefs we report” (S&T 2012, p. 128).

¹⁵ Note that Mark’s concept I is not the same as Michael’s, since they have different originating uses, see S&T (2012, p. 51).

¹⁶ The concept HE, in (3/4T), is attributer’s.

This suggests, independently of Specification Assumption, that (5) should be a truthful belief report. Intuitively speaking, however, this is not the case: (5) attributes to Adam the belief that Madagascar is not an island, which is now false, whereas Adam's belief, at his own time, was that Mogadishu is not an island, which is now true. Something is going wrong.¹⁷

S&T are aware of the problem, though they do not discuss it for singular terms:

On our version of originalism, there are some qualifications on the claim that if the subject of the attribution stands in the appropriate relation to the very thought displayed in the attribution, the attribution is true. This is because we allow that a concept's content can shift over time. A fifteenth-century person, Harold, might be belief-related to the thought that spinach is meat, and we might display this thought in the twenty first-century attribution "Harold believed that spinach is meat." The attribution is certainly misleading, and it is doubtful whether it is true. It's misleading, because it suggests that Harold's thought had the same content as ours, in which case he believed something obviously false, whereas what he believed, at the time he believed it, was true. Hence the best candidate for a sufficient condition is identity of thought and sameness of content.

(S&T 2012, pp. 118–119, n.12)

CO's solution is to strengthen the sufficient condition for truthful belief reports. A belief report of the form 'S believes that P' is true if (a) the thought put on display, i.e. 'P', is identical to the thought that the attributee, S, is/was belief-related to and (b) the thought in question has the same content across time.¹⁸ Having this sufficient condition handy, it is not strange why (5) is not a truthful belief report: (b) does not hold. The content of (5T), namely the set of possible worlds in which the constituent concepts of (5T) have reference and (5T) is true, roughly speaking, at the time of Adam,

¹⁷ S&T and we share that (5), or a similar attribution, "is certainly misleading, and it's doubtful whether it is true" (S&T 2012, p. 119). To make room for this, S&T suggest a strong sufficient condition for a truthful belief report (see below) such that it cannot be satisfied in cases like (5). We will show that this strategy is not successful. A non-originalist account, e.g. a Russellian view, may reject reference-shift altogether and introduce two names, spelled and pronounced similarly, as 'Madagascar'. Since a Russellian view does not require a nominative concept to contribute to the thought expressed by/associated with a sentence containing a proper name, it may easily provide two readings of (5), one true and one false, depending on the intended name and its reference. CO cannot employ double-name strategy in its full generality since it requires the denial of the possibility of reference-shift (see Sect. 9). Alternatively, one may challenge the intuition that (5) is false. Here is one way to develop this idea: (5) expresses/conveys the content of:

5'. Adam believed that what he called 'Madagascar' was not an island.

Since what Adam called 'Madagascar' was Mogadishu and Mogadishu was not an island, Adam believed something true and (5') correctly represents his mental state. Thus (5') is true. If (5) expresses/conveys the content of (5'), then (5) expresses/conveys something true. Admittedly, in *some cases* belief reports like (5) may *say* something true or be taken that way. This is not a specific problem for CO. However, that a sentence like (5') evokes the intuition of something true being said is most likely explained by its indirect recourse to the *original reference* of 'Madagascar' via the expression "what he *called* 'Madagascar'". If the originalist is willing to take (5) as true, then she needs to explain how the original reference of 'Madagascar' contributes to the truth conditions of (5), via expressing/saying the content of (5') or not, without threatening the founding idea of CO (recall, an atomic concept is not individuated by its semantic properties, its reference included). S&T have not followed this strategy. Thus, at this stage, we may put it aside as well.

¹⁸ In most cases, we freely move between 'concept'/'thought', on the one hand, and 'concept-token'/'thought-token', on the other hand. The context, we hope, clarifies the intended meaning.

contains the actual world whereas now it does not. So, the thought in question does not have the same content across time. This can explain why (5)'s being misleading and false does not contradict CO.¹⁹ If the sufficient condition were not strengthened, then (5) would satisfy it and thus would become true (contradicting our intuition that the report is misinforming and incorrect). Now, the new condition cannot be met. Hence, there is room to take (5) as false. This solution, however, fails, or so we will argue next.

6 CO's solution fails

Consider a belief report whose 'that'-clause contains an atomic simple sentence containing a singular (or general) term with reference-shift such that the thought expressed by the sentence is necessarily false all along. Since the reference-shift does not change the identity of the concept, there would be no change in the identity of the thought. Thus, condition (a) holds. On the other hand, the content of a thought, according to CO, is a set of possible worlds w in which each constituent concept of the thought has a content (reference) and the thought is true. Therefore, if the thought is necessarily false, its content would be the empty set all along. Hence, there would be no change in its content. Hence, condition (b) holds too. However, intuitively speaking, the belief report is not truthful.

For instance, suppose that Adam, mistakenly, believed that Atlantis, the mythical island, exists and is the same as Mogadishu. Assume that he has been belief-related to the thought represented by:

(6T) < MADAGASCAR, =, ATLANTIS >

By assumption, Adam and we use the same concept MADAGASCAR, only with reference-shift. (Also, by assumption, Adam and we use the same empty concept ATLANTIS.) Given CO's semantics of belief reports, (6) should be true:

(6) Adam believed that Madagascar is Atlantis.

First, the thought put on display in (6) is the same as the thought entertained by Adam; they are constituted of the same concepts and have the same structure. Therefore, condition (a) holds. Second, the thought entertained by Adam has the same content as the thought put on display in (6), namely the empty set: there is no possible world in which MADAGASCAR and ATLANTIS have one and the same reference, as we use them. Recall that CO allows empty concepts (S&T 2012, pp. 139–144): an empty concept has no content at any world (ibid, 48), and all identity thoughts containing such concepts are necessarily false (ibid, 141). Therefore, there is no change in the content of the thought represented by (6T) over time. Hence, condition (b) holds. Nevertheless, (6) is misleading and it is doubtful that it is true; Adam believed that Mogadishu, not Madagascar, is Atlantis. That what Adam believed is necessarily false should not

¹⁹ Note that the conjunction of (a) and (b) is a sufficient condition for having a truthful belief report. Thus, not meeting this condition does not logically imply that the belief report in question, i.e. (5), is not truthful. The problem, however, is not only that CO does not provide the necessary and sufficient conditions for attitude ascriptions in general, but also that the sufficient condition introduced above does not hold. See Sect. 6.

conceal the point that Adam had some belief *about* Mogadishu, not Madagascar. (6) misrepresents Adam's mental state; in particular, (6) misrepresents the subject matter of Adam's belief.

As another example, consider a belief report whose 'that'-clause contains an atomic simple sentence containing a singular (or general) term with reference-shift such that the thought expressed by the sentence is necessarily true all along and its truth requires the existence of one set of objects.

For instance, to have an example of reference-shift in general terms, consider a mathematical predicative concept with reference-shift. Suppose that the concept BEING PERFECT in number theory has gradually changed its semantic reference from the property of *being equal to the sum of all or some of the proper divisors of a number*, in 1000 BC, to the property of *being equal to the sum of all the proper divisors of a number*, now. The two properties are similar, though distinct. Suppose that Harold, in 1000 BC, was belief-related to the thought that 6 is a perfect number. This thought may now be represented by:

7T. $\langle 6, \text{BEING PERFECT} \rangle$

By assumption, Harold and we use the same concept BEING PERFECT. Given CO's semantics of belief report, the following belief report should be true:

7. Harold believed that 6 is perfect.

The thought put on display in (7), is the same as the thought to which Harold was belief-related. Hence condition (a) holds. Moreover, this very thought is necessarily true all along: 6 has both the property of *being equal to the sum of all or some of its proper divisors* and the property of *being equal to the sum of all its proper divisors* necessarily. Finally, note that 6 is a necessary object and exists in all possible worlds. Hence the content of the thought represented by (7T) is the set of all possible worlds all along. Therefore, there is no change in its content. Hence, condition (b) holds. Nevertheless, (7) is misleading and it is doubtful that it is true: what Harold believed of 6, as a perfect number, is significantly different from what one believes of 6, as a perfect number, now. In some sense, it is by mere coincidence that the thought put on display in (7) has the same content across time. If instead of 6, Harold had thought of 12, the content of his thought, as well as its extension (truth-value), would have changed (12 only possesses the property of *being equal to the sum of all or some of its proper divisors*, and not the property of *being equal to the sum of all its proper divisors*). (7) is misleading and it is doubtful that it is true since it misrepresents the conceptual content of Harold's belief state.

7 Objections and replies

First, it might be thought that the above counterexamples are hypothetical and thus have no counterparts in natural language. CO, the objection continues, provides a theory of the concepts expressed in natural language. Thus, remote possibilities of reference-shift shall not count as actual evidence against the view. We believe that this objection fails. Admittedly, the counterexamples are hypothetical; they may not be, historically speaking, true stories. This, however, shall not, and does not, affect

their theoretical value. CO is a theory of the concepts *expressible* (and not necessarily expressed) in natural language.

Second, one might object that the belief reports in question appear misleading only if one reads them *de re*. It is only when one characterizes (6) as reporting a belief about, or of, Mogadishu, that (6) might appear misleading. If one reads them *de dicto*, the contrary intuitions should fade away. This objection fails for two reasons: first, the originalist formulation of pure *de re* belief reports does not support this claim and second the problem is not restricted to *de re* belief reports. Here is S&T's characterization of pure *de re* belief reports:²⁰

Setting aside hybrid cases, Sally thinks of Hesperus that it is visible iff for some *de re* concept C, whose referent is Hesperus, Sally thinks a thought whose first element is C and whose second element is R-related to the concept IS VISIBLE. As before, the R-relation will sometimes but not always be identity.

(S&T 2012, p. 122)

According to the above characterization, (6) should count as a pure *de re* belief report. Thus, if its *de re* reading is misleading and false, intuitively speaking, appeal to *de re/de dicto* distinction does not block our counterexamples.

More importantly, a belief report like (7), in which reference-shift occurs in a predicative concept (not in a nominative one), has nothing to do with S&T's treatment of *de re* belief reports, or in general with the *de re/de dicto* distinction. What makes such a belief report misleading is not that the subject-matter of the attributee's belief state is misrepresented; rather, the distortion occurs in the representation of a different part of the belief state, i.e. the predicative concept. In general, given CO, concepts of other kinds, e.g. logical or adverbial, may suffer the same kind of misrepresentation. Hence, the *de re/de dicto* distinction is not the key to the problem.

8 CO's diagnosis of the root of the problem is misguided

S&T maintain that the belief report in their spinach example, namely:

8. Harold believed that spinach is meat.

is "misleading, because it suggests that Harold's thought had the same *content* as ours, in which case he believed something obviously false, whereas what he believed, at the time he believed it, was true" (S&T 2012, p. 71), emphasis is ours. We agree that (8) is misleading. Nonetheless, S&T's diagnosis of the root of this problem is misguided.

Recall that 'content', given CO, gives birth to two different kinds of entity: on the one hand, the content of a nominative concept or that of a predicative one is an individual or a property, and on the other hand, the content of a unary atomic thought (expressed by a declarative atomic simple sentence) is the set of possible worlds w in which each constituent concept of the thought has a content (reference) and the thought is true, roughly speaking. Thus, the content of a thought, in a sense, does not represent the contents of its constituents. (8) is misleading, and perhaps false, not

²⁰ Following S&T (2012, p. 122), we put hybrid cases, namely *de re* reports of *de dicto* attitudes and *de dicto* reports of *de re* attitudes, aside.

because it suggests that Harold's thought had the same content, in the originalist sense of 'content', as ours. Rather, it is misleading and perhaps false since it suggests that Harold believed a structured content, or a Russellian proposition, he could not have. Consider the following two Russellian propositions:

9P. <Spinach, *Being edible*>

10P. <Spinach, *Being flesh*>.²¹

What makes the belief report in question incorrect is that it suggests that Harold's belief box contained the proposition represented by (10P), rather than the one represented by (9P). Note that, according to CO, Russellian propositions are not legitimate entities; they are neither thoughts, nor their contents, and not even their extensions (truth-values). They differ from thoughts in that they are constituted of individuals and properties, not concepts. They differ from the contents of thoughts in that they are structured objects, not sets of possible worlds. Neither the originalist 'thoughts' nor their 'contents' do what Russellian propositions do. The problem, we believe, is rooted in the fact that S&T do not avail themselves of a theoretical apparatus equivalent to Russellian propositions.

9 Two initially appealing ways out of the problem fail

In a recent critical notice of S&T's book, [Horwich \(2014\)](#) suggests an alternative solution to S&T's meat example:

I suspect these difficulties could have been avoided had the authors applied their theory to 'meat' in a rather different way. It might be simpler and better for them to say: (i) that some person at a specific time introduced the new use of 'meat' with an unconstrained, nondeferential utterance of 'Meat is flesh'; (ii) that this new 'concept' spread gradually through the community over the next few hundred years; (iii) that during this period both 'concepts' were in play; and (iv) that by now the new one has completely supplanted the old one.

([Horwich 2014](#), p. 1131)

Admittedly, Horwich's proposal describes a possible scenario in which the problem of misleading belief reports may not arise in the form discussed by S&T. Nonetheless, as we believe S&T are well aware, Horwich's proposal is not a solution to the problem. The problem we are concerned with is rooted in the *possibility* of reference-shift in concepts.²² CO is committed to the view that a single concept *may* change its reference

²¹ We are ignoring the problems raised by mass nouns, like 'spinach', here. Though significant, we assume that these problems should not occupy our attention now. Spinach, *Being edible* and *Being flesh* represent stuff and properties out there that ultimately determine the truth-value of sentences like 'Spinach is edible' and 'Spinach is flesh'.

²² Here we are not concerned with the 'Madagascar' case in general, or as an objection to the Causal Theory of Reference. CO allows (inadvertent) reference-shift and acknowledges that this may lead to misleading and false belief reports. The problem we are concerned with is whether CO's sufficient condition for a truthful belief report holds, and if not, how the view can supply such a condition. By focusing on this problem, we overlook some significant and relevant questions, e.g. "Does CO need an overall coherent account of reference determination?" and "how to fit (inadvertent) reference-shift into a decent *general* account of de dicto belief attribution?" At this stage, we leave these as open questions.

over time and preserves its identity. Horwich's proposal may block the problem in some cases but it does not provide a genuine solution to it *in general*.

Horwich's proposal *can* be a solution if it is brought to its extreme by claiming that there is no possibility of reference-shift and all cases of misleading belief reports are cases of twin concepts whose expressions are spelled and pronounced similarly when one may prevail over the other, given enough time. Denying the possibility of reference-shift, however, has two implausible consequences for CO: First, Gareth Evans's objection to the causal theory of reference will be back: There are cases in which, intuitively speaking, reference-shift is admissible when the causal-chain of reference borrowing does not allow it. Second, and worse, if reference-shift is not permitted, it means that there is no single concept with two references over time. Otherwise put, it says that a change in a concept's reference is enough to change the identity of the concept. A concept's reference, after all, is a semantic property of the concept. And this implies that some semantic properties, e.g. reference, contribute to the individuation of concepts. This violates the founding idea of CO, namely "atomic concepts are to be individuated by their historical origins, as opposed to their semantic or epistemic properties" (S&T 2012, p. 40).

Alternatively, one might try to revise CO by introducing Russellian propositions as the contents of thoughts. If so, then the misleading belief reports do not satisfy S&T's *second* condition for a truthful belief report (namely the thought put on display in a belief report should have the same content as the thought to which the attributee is/was belief-related) and thus, given CO, it does not follow that they are true (in other words, CO can accommodate such belief reports as false). Accordingly, concepts may change their references but since the content of a thought is a Russellian proposition, all changes of the contents of the constituent concepts of the thought would be visible, so to speak, through the content of the thought. For example, there is no necessity to take a belief report like (7), namely 'Harold believed that 6 is perfect', to be true. This is so because the content of the thought put on display in (7), i.e. the Russellian proposition represented by:

7P. *<6, Being equal to the sum of all the proper divisors of a number>*

is distinct from the content of the thought to which Harold was belief-related, i.e. the Russellian proposition represented by:

7'P. *<6, Being equal to the sum of all or some of the proper divisors of a number>*.

This proposal faces three serious problems. First, CO's resistance to accept Russellian propositions as useful entities:

Originalism differs from Millianism in finding no use for propositions, conceived as sequences of objects and properties. It accounts for many Fregean data in terms of sameness and difference among concepts, rather than in terms of guises, notions or modes of presentation. (S&T 2012, p. 30)

Second, the problem of empty concepts will be back. If the contents of concepts, nominative or predicative, may be empty, namely there may be no corresponding individuals or properties, and the content of a thought is a structured entity constituted of individuals and properties, it is not clear what the content of a thought containing an empty concept, e.g. the thought associated with 'Vulcan exists', will be. It is hard to accept that such a thought has no content, if its content is the ultimate truth-value

bearer. After all the thought seems to be false and S&T need to explain this. Moreover, introducing an unfilled proposition or a gappy content flies in the face of both S&T's originalist approach²³ and Tye's solution to the problem of the content of a hallucinatory experience. Last but not least, there are independent reasons against gappy contents.²⁴ Third, the proposal pushes CO toward a version of Russellianism and, thus, makes it less attractive. The Russellian would wonder why she should buy all the apparatus of concepts, if at the end of the day Russellian propositions should do the most important theoretical tasks.

10 Does adding the *Reference* to the picture solve the problem?

One might further strengthen the sufficient condition for a truthful belief report by requiring the sameness of the references of the corresponding constituent concepts of the thought entertained by the attributee and the one put on display in the belief report. It might be suggested that this solves the problem of misleading belief reports for CO without deviating from its principal elements. This is not the case, we will explain.

The suggestion may be developed in different ways; we use S&T's notion of *isomorphic* thoughts:

Isomorphism: Thoughts are isomorphic iff they share a complete tree structure, and their corresponding terminal nodes are concepts with the same content [i.e. the same referent].

Isomorphism is not necessary for sameness of content [in the originalist sense] (though it is sufficient in extensional contexts).

(S&T 2012, p. 50)

Thus, on behalf of S&T, one might suggest that a belief report of the form 'S believes that P' is true if (a) the thought put on display, i.e. 'P', is both (a) isomorphic and (b) identical to the thought that the attributee, S, is/was belief-related to (we suppose that isomorphism guarantees the sameness of the truth value, or the content, of the thought in question over time). Even this is not a strong enough sufficient condition, given CO, as we will shortly show. Before doing so, let's quickly see how one might use the idea, in a slightly different form, to explain the *falsehood* of the misleading belief reports discussed above, S&T's meat example included. If isomorphism is introduced as a *necessary* condition for a truthful belief report, then the misleading belief reports are false (and thus misleading) because they violate this necessary condition. Reference-shift in a concept brings about difference in its content and difference in its content undermines isomorphism. To illustrate, consider (6) again, namely 'Adam believed that Madagascar is Atlantis'. Even though the thought put on display in the belief report and the thought to which Adam was belief-related are identical (and have the same truth value), they are not isomorphic. Our concept MADAGASCAR and Adam's concept MADAGASCAR, though identical, have different contents: one refers to Madagascar and the other to Mogadishu, by assumption. Therefore, the two thoughts having these

²³ See S&T (2012, vol. 140, pp. 152–157) where they explicitly argue against the idea of gappy propositions.

²⁴ See Everett (2003) and Mousavian (2011).

concepts as their constituents are not isomorphic. Thus, the necessary condition is not met and this explains why the belief report is not truthful.

Unfortunately, even this seemingly promising strategy cannot solve the general problem of misleading belief reports. CO acknowledges cases of apparent reference-shift in *empty* concepts; let's call them "purported-reference-shift". These cases make the task of furnishing a semantics for belief reports puzzling unless one spells out the nature of "purported reference" and get them integrated into the truth conditions of belief reports, something that has not yet been done nor is it clear how to do it. Here is an example:

Joke: A small group, as a kind of "in" joke, decides to use "Holmes" for Watson and "Watson" for Holmes. Then all the texts are destroyed in some cataclysm. People learning from the small group don't realize there has been a jocular inversion. These benighted souls become the only users of the specific names. Although the purported referent of "Watson" used to be Watson (the side-kick), it has become Holmes (the detective), and the purported referent of "Holmes" is now Watson. We again apply the principle that semantic reference is conventionalized speaker's reference. In the example, the purported reference of empty names has changed.

(Sainsbury 2015, p. 213)

Note that 'Holmes' and 'Watson', by assumption, are genuinely empty, they have no semantic reference; they only *purport* to refer. Nonetheless, as Sainsbury claims, they may undergo purported-reference-shift. Suppose that before the shift, Harold mistakenly believed that Holmes and Watson *really* exist and have some properties. He believed, for example, that Holmes is smart. More precisely, assume that before the purported-reference-shift Harold was belief-related to the thought represented by (9T):

9T. <HOLMES, BEING SMART>

Now, consider the following belief report made after the shift:

9. Harold believed that Holmes is smart.

For the sake of simplicity, we can assume that the post-shift attributer (who is not explicitly represented in (9)) does not know anything about the purported-reference-shift and mistakenly believes, like Harold, that Holmes and Watson really exist.²⁵ (9) is intended to describe Harold's belief state before the shift.²⁶ Therefore, the epistemic

²⁵ Here are two points in order. First, the attributer *has no reasons* to report Harold's belief by a quasi-quotational device since she does not know about the purported-reference-shift. In the 'Madagascar' case, likewise, the attributer has no reasons to appeal to a quasi-quotational device since she does not know about the reference-shift. The difference is that the concept under discussion here, i.e. HOLMES, is empty, unbeknownst to Harold and his belief attributer, whereas MADAGASCAR is not. (See note 17 as well.) Second, the attributer has no reasons to use a fictional operator (e.g. 'according to the story') since neither she nor Harold believes that Holmes and Watson are fictional.

²⁶ As a follow up to the previous note, it might be thought that the attributer, in this particular case, *should* report Harold's belief state by a quasi-quotational device or a fictional operator. This concern is not well-placed. How Harold's belief state should be reported is not our immediate problem; rather, how it should *not* be reported is at stake: (9) is not a reliable way to report Harold's belief state but the new proposal cannot block it. In other words, for whatever reason one may use (9) to describe Harold's belief state, (9) creates a problematic case for CO: it passes the further strengthened sufficient condition for a truthful belief report but it is not truthful (see below). Moreover, we believe that if Harold and the attributer both know

states of Harold and his post-shift belief attributer are similar to those of Adam and his post-shift belief attributer in Sect. 6. Recall:

(6) Adam believed that Madagascar is Atlantis.

In the case of (6), Adam believes something false, namely that Mogadishu is Atlantis. In the case of (9), similarly, Harold believes something false, namely Holmes (as a real person) is smart. In the case of (6), the problem is how to correctly ascribe (after the shift) the false belief Adam has (before the shift) to him, given CO. We argued that S&T's strengthened sufficient condition for a truthful belief report is inappropriate because (6) satisfies this condition but it ascribes a *different* false belief to Adam, one *about* Madagascar (due to the reference-shift in 'Madagascar'). In the case of (9), again similarly, the problem is how to correctly ascribe (after the shift) the false belief Harold has (before the shift) to him, given CO. We submit that the *further* strengthened sufficient condition for a truthful belief report (by adding isomorphism) is inappropriate because (9) satisfies this condition but it ascribes a *different* false belief to Harold, one *purportedly about* Watson (due to the purported-reference-shift in 'Holmes'). Let us explain.

The thought put on display in the belief report (9) and the thought that Harold had in his belief box, i.e. (9T), are both isomorphic and identical, since they are constituted of similar concepts with the same contents.²⁷ More precisely, their first concepts are genuinely empty and their second concepts refer to the property of *being smart*. This guarantees that (9T) has one and the same truth-value, or content, over time, i.e. falsehood. Therefore, (9) satisfies the sufficient condition proposed above. Nonetheless, (9) is misleading and perhaps false because it suggests that Harold believed something that was purportedly about Watson (recall, 'Holmes' after the purported-reference-shift purports to refer to Watson), whereas Harold, by assumption, believed something that was purportedly about Holmes. Moreover, in cases of purported-reference-shift like Joke and reports like (9), since isomorphism, strictly speaking, is not violated, the falsehood of the misleading belief report may not be explained by appeal to isomorphism as a *necessary* condition of a truthful belief report either. The problem simply persists since CO extends 'reference'-shift to empty concepts.

To properly represent that Harold had a belief that was purportedly about Holmes, not Watson, one also needs to get the purported-reference of 'Holmes' involved in the semantics of belief reports. Something that has not yet been done, nor is it clear how to do it within CO without seriously deforming the view. Note that postulating nonexistent objects as the (purported) referents of the so-called empty names and then appealing to isomorphism as a sufficient or necessary condition for a truthful belief report, has three odd consequences for CO: first, it pushes the view toward a version of 'Meinongianism' and thus makes it less appealing. Second, it makes the whole apparatus of *concepts* as useful means to tackle the problem of negative existentials

Footnote 26 continued

that Holmes is fictional (not real), still a belief report like (9'), made after the purported-reference-shift, may be troublesome for CO:

(9') Harold believed that according to the story, Holmes is smart.

Nonetheless, we do not dwell on this since S&T, in the works under discussion, are silent about the semantics of fictional operators and the identity condition of stories.

²⁷ By assumption there is no change in the concept BEING SMART.

redundant: If there are nonexistent objects and they may be referred to by proper names, then one may truly attribute nonexistence to them as well. Third, this proposal threatens one of the primary motives for CO, i.e. the view acknowledges *genuinely* empty names as legitimate proper names of natural language. To repeat, CO is insistent that “empty concepts can be intelligible; even though they all have the *same* content” (S&T 2012, pp. 143).²⁸

11 Concluding remarks

We argued that CO’s semantics of belief reports is defective and no easy solution is foreseeable. Horwich’s (2014) proposal is a local solution and its generalized form undermines CO’s founding idea. CO is pretty new; it has not yet been under critical assessment from a perspective similar to ours.²⁹ Providing an identification criterion for concepts cannot be done independently of furnishing a semantics for attitude ascriptions and the latter needs to be based on the semantic properties of concepts and thoughts. Allowing (inadvertent) reference-shift makes concepts semantically unstable and extending it to purported-reference-shift makes concepts semantically untrackable. Neither thoughts nor their contents, not even the contents of the constituent concepts of thoughts, promise to help with the semantics of belief reports, as long as ‘reference’-shift is allowed for *all* concepts (referring and non-referring). CO opens the Pandora’s box of reference-shift in a wide range of cases without providing the necessary semantic means to take care of them.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bach, K. (1997). Do belief reports report belief? *Pacific Philosophical Quarterly*, 78(3), 215–241.
- Burgess, J. P. (2014). Madagascar revisited. *Analysis*, 74(2), 195–201.
- Evans, G. (1973). The causal theory of names. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 47, 187–225.
- Everett, A. (2003). Empty names and ‘Gappy’ propositions. *Philosophical Studies*, 116, 1–36.
- Fara, D. G. (2003). Desires, scope and tense. *Philosophical Perspectives 17: Philosophy of Language and Philosophical Linguistics*: 141–163 [Originally published under the name ‘Delia Graff’].

²⁸ Our emphasis.

²⁹ For example, Gasparri (2016) takes the question of ‘Is it possible “to manipulate the time of origin of a word type W without causing W to develop differences in intrinsic properties that would prove detectable [...]” (ibid, 132)’? He tries to show that the originalist and structural-functional word typing are not inconsistent. Onofri (2016), in contrast, approvingly describes CO as being “inconsistent with publicity: two concepts will be different as long as the associated terms were introduced by different linguistic communities” (ibid, 14). Onofri’s (2016) main thesis is that no theory of concept can be faithful to both *publicity* (namely, that “concepts are ‘the sort of thing that people can, and do, share’ (ibid, 3)) and *multiplicity*, as we call it (namely, that “a subject who finds herself in a so-called ‘Frege case’ appears to have different concepts for the same object” (ibid, 3)). CO apparently corroborates Onofri’s thesis.

- Gasparri, L. (2016). Originalism about word types. *Thought*, 5, 126–133.
- Horwich, P. (2014). Critical notice of seven puzzles of thought and how to solve them: An originalist theory of concepts, by R. M. Sainsbury and Michael Tye. *Mind*, 123(492), 1123–1139.
- Kaplan, D. (1990). Words. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 64, 93–119.
- Kaplan, D. (2011). Words on words. *The Journal of Philosophy*, 108(9), 504–529.
- Kripke, S. (1980). *Naming and necessity*. Oxford: Blackwell.
- Mousavian, S. N. (2011). Gappy propositions? *Canadian Journal of Philosophy*, 41, 125–158.
- Onofri, A. (2016). Two constraints on a theory of concepts. *Dialectica*, 70(1), 3–27.
- Quine, W. V. O. (1940). *Mathematical logic*. Cambridge, MA: Harvard University Press.
- Sainsbury, M. (2015). The same name. *Erkenntnis*, 80, 195–214.
- Sainsbury, R. M., & Tye, M. (2011). I—An originalist theory of concepts. *Proceedings of the Aristotelian Society, Supplementary*, 85, 101–124.
- Sainsbury, R. M., & Tye, M. (2012). *Seven puzzles of thought and how to solve them: An originalist theory of concepts*. Oxford: Oxford University Press.
- Taylor, K. A. (2002). De Re and De Dicto: Against the conventional wisdom. *Noûs*, 36(16), 225–265.
- Tye, M. (2014). What is the content of a hallucinatory experience? In B. Brogaard (Ed.), *Does perception have content?* (pp. 291–310). Oxford: Oxford University Press.