The informative value of type of repetition: perceptual and conceptual fluency influences
on judgments of truth

Rita R. Silva¹

William James Center for Research, ISPA – Instituto Universitário, Lisbon

Teresa Garcia-Marques

William James Center for Research, ISPA – Instituto Universitário, Lisbon

Rolf Reber

University of Oslo

Word count: 11702 (including main text and notes; excluding tables and abstract).

References: 47

Correspondence regarding this paper should be addressed to Rita R. Silva, Social Cognition Center Cologne, University of Cologne, Richard-Strauss-Str. 2, 50931 Köln, Germany; email: rita.silva@uni-koeln.de.

¹ Present Address: Social Cognition Center Cologne, University of Cologne, Richard-Strauß-Strasse 2, 50931 Cologne, Germany.

Abstract

We contrast the effects of conceptual and perceptual fluency resulting from repetition in the truth effect. In Experiment 1, participants judged either verbatim or paraphrased repetitions, which reduce perceptual similarity to original statements. Judgments were made either immediately after the first exposure to the statements or after one week. Illusions of truth emerged for both types of repetition, with delay reducing both effects. In Experiment 2, participants judged verbatim and paraphrased repetitions with either the same or a contradictory meaning of original statements. In immediate judgments, illusions of truth emerged for repetitions with the same meaning and illusions of falseness for contradictory repetitions. In the delayed session, the illusion of falseness disappeared for contradictory statements. Results are discussed in terms of the contributions of recollection of stimulus details and of perceptual and conceptual fluency to illusions of truth at different time intervals and judgmental context conditions.

146 words

Keywords: truth effect, repetition, perceptual fluency, conceptual fluency, recognition

1. Introduction

The “illusion of truth effect” refers to the observation that statements that are repeated are considered more probably true and more valid than new statements. Hasher, Goldstein and Toppino (1977) were the first to empirically demonstrate this effect, which has been repeatedly replicated in different settings and with different materials (for a meta-analysis see Dechêne, Stahl, Hansen, & Wänke, 2009).

Four kinds of repetitions can be used in conducting studies on the illusion of truth. (1) The majority of experiments that illustrate this effect have used verbatim repetition of the statements. That is, participants evaluate the truth of exactly the same items they were previously exposed to. However, to capture the processes underlying evaluations of truth, some empirical approaches made changes to the statements in the judgment phase. The following changes are of particular relevance for that purpose, yielding three further kinds of repeated statements: (2) changes to the wording of the statements while maintaining their meaning (i.e., paraphrases), (3) changes to the meaning of the statements while maintaining most of their wording and topic (i.e., contradictory statements), and (4) changes in both the wording and the meaning of the statements, while maintaining their topic (i.e., contradictory paraphrases)

In this paper we present two experiments that compare truth judgments of repeated verbatim statements with these kinds of changes to better understand how the conceptual and perceptual fluency associated with the different kinds of repetition contribute and interact to create repetition-based illusions of truth. Throughout the manuscript we make the distinction between repetition of the *meaning* of a statement, which is the specific and exact details of a statement’s semantic content, and repetition of

its *topic*, which includes the general theme but not its exact details (e.g., Begg, Armour & Kerr, 1985).

In addition to type of repetition, we manipulated the delay between the first exposure to the statements and the truth evaluations. With this manipulation we attempted to test how conceptual and perceptual fluency associated with the different types of repetition contribute to the effect when: (1) truth judgments are more likely to be based on the conscious recollection of the details of the original statements, versus when (2) truth judgments are more likely to be based on the familiarity associated with fluent processing of repeated stimuli (e.g., Begg, Anas, & Farinacci, 1992; see also Jacoby & Dallas, 1981; Tulving, Schacter, & Stark, 1982; Whittlesea, 1993, for effects of familiarity).

2. Processes underlying repetition effects on truth judgments

2.1 Familiarity

The literature suggests that the illusion of truth effect is closely related to the feeling of familiarity elicited by repeated statements in comparison to new ones. That is, simply remembering (even if falsely) that we have heard or read a statement beforehand will influence how we judge its validity. One reason for why this may occur is that when we recognize a statement we may also recall information about its epistemic value (whether it was presented as true or false) and confer referential validity on it (e.g., Brown & Nix, 1996). The fact that we remember having heard a statement before may also signal that different sources agree on it, conveying convergent validity on it (Arkes, Hackett, & Boehm, 1989, and Arkes, Boehm, & Xu, 1991). Accordingly, Begg and Armour (1991)

showed that statements judged as old were evaluated as true whether they were repetitions that participants correctly remembered or new statements that they mistook for repetitions. Furthermore, old statements that were forgotten were evaluated just as true as new statements correctly identified as such. This evidence suggests that perceived familiarity is more important for illusions of truth than the real repetition status of the statements (see also Bacon, 1979; Boehm, 1994).

2.2 Processing fluency

The impact of familiarity on judgments of truth seems to be anchored in the fluent processing of repeated items (e.g., Begg et al., 1992). Repeated stimuli increase processing fluency (e.g., Jacoby & Dallas, 1981), and just as fluency influences judgments that one has seen a stimulus before (e.g., Jacoby & Whitehouse, 1989), it also influences judgments that a stimulus is true. In line with this argument, individuals evaluate statements as more probably true simply because they are presented in high as compared to low visual contrast (e.g., R. Reber & Schwarz, 1999) or in an easy as compared to difficult letter font (e.g., Parks & Toth, 2006). Similarly, aphorisms are perceived as more probably true when presented in a rhyming (e.g., “Birds of a feather flock together”) rather than a non-rhyming form (e.g., “Birds of a feather flock conjointly”, McGlone & Tofigbakhsh, 2000). Relatedly, semantically primed words (i.e., words that are preceded by related concepts) are judged as better answers to trivia questions than words that are not primed (Kelley & Lindsay, 1993).

The fluency involved in the processing of a repeated statement can come from the repetition of different types of its perceptual features – perceptual fluency, and/or the

repetition of the different types of its conceptual characteristics – conceptual fluency (for a classification of different types of fluency, see Alter & Oppenheimer, 2009). Previous exposure induces perceptual fluency due to the creation of a feature-based representation of the stimulus (visual, auditory, pictorial), which supposedly facilitates the encoding and processing of the statement when encountered again (Jacoby & Dallas, 1981; Shapiro, 1999). But through repetition individuals also learn about the structural properties of statements (e.g., grammars), which facilitate their subsequent processing (Buchner, 1994; A.S. Reber, 1967). Thus, even if a statement differs from a previous one but has a similar structure, individuals may experience increased fluency (for a similar argument, see R. Reber & Unkelbach, 2010), which does not necessarily stem from conceptual aspects. Conceptual fluency occurs because previous exposure creates a meaning-based representation that facilitates encoding, processing, and retrieval in future encounters (Shapiro, 1999; Whittlesea, 1993). This type of fluency can come from the activation of the meaning of the statement, or of a more general knowledge structure, such as its theme or topic. Given the convergence of effects that fluency exerts in a variety of judgments (see Alter & Oppenheimer, 2009), the power of repetition in consistently promoting illusions of truth is not surprising – repetition might simply aggregate different sources of perceptual and conceptual fluency.

3. Types of repetition

Research has focused on different types of repetition in order to better understand the mechanisms underlying the truth effect. Although illusions of truth have mostly been associated with verbatim repetition, this is not a necessary condition to observe the effect.

Illusions of truth are expected to occur if a target statement is sufficiently similar (perceptually or conceptually) to a previously presented one, facilitating its processing.

3.1 Topic repetition

Besides verbatim repetition, illusions of truth can also emerge by repeating the topic of previous statements or part of the verbatim statement. To activate the topic of statements that participants were going to evaluate for truth, Begg and colleagues (1985) exposed participants to parts of the target statements in a previous phase of the experiment (e.g., “A hen’s body temperature” was presented as part of the statement “The temperature of a hen’s body is about 104 degrees Fahrenheit”), leading to an increase in those items’ perceived validity. These results were replicated in a second experiment in which the topic of the statements was activated by previously presenting participants with questions that overlapped either more or less with the target statements’ details (e.g., for the target statement “The right arm of the Statue of Liberty is 42 feet long”, the question “Do you have any idea how long the extended right arm of the Statue of Liberty is?” repeats more relevant information than the question “Do you have any idea how long the Statue of Liberty has been in New York?”). Both conditions increased perceived truth relative to statements that had not been primed by any question, with the questions that had greater overlap with the statements resulting in the strongest effect.

Illusions of truth can occur even when the topic of the statements is activated without much overlap with the verbatim meaning of the statements. In a context of simple topic repetition, Arkes and colleagues (1991) presented participants with statements either about China or about other topics (Week 1). One week later (Week 2)

they presented participants with text passages either about the topics covered by the China statements of Week 1, other topics about China not covered by the China statements of Week 1, or topics unrelated to China. After another week passed (Week 3), participants had to evaluate the truth of a group of statements pertaining to China, including those first presented on Week 1. Results showed that reading any of the China text passages on Week 2 led participants to consider the original China-related statements of Week 1 as more valid, independently of having been exposed to those statements on Week 1 or not.

Hawkins, Hoch, and Meyers (2001) offer further evidence of the impact of non-verbatim repetition on perceived truth. These authors tested how the perceived truth of a claim was influenced both by its exact repetition and by the repetition of associated claims that shared only some words with the original statement. They found little evidence of truth effects with the associated claims. However, both repetition conditions increased the perceived truth of a superordinate general claim to which all others were conceptually related. Mediation analyses suggested that the effect was mediated both by an increase in perceived familiarity and in perceived coherence of the claim (both presumably related with the ease of processing its content).

3.2 Contradictions

One other type of repetition includes repeating all the perceptual characteristics (words, phrase structure) of the original statements, except for small details that produce a contradiction to its original meaning. Using this kind of stimuli, Bacon (1979) found that contradictions intermixed with original statements were rated falser than new

statements, creating an illusion of falseness. This is an illusory falseness effect because there is no reason to consider the contradictory statements to be the false ones. The illusion of falseness effect was replicated in a similar experiment when participants had to judge the truth of contradictions and new statements. (Garcia-Marques, Silva, R. Reber, & Unkelbach, 2015). This observation suggests that the direct contrast between contradictions and verbatim repetitions is not necessary to produce illusions of falseness. This evidence suggests that simple perceptual overlap does not produce illusions of truth when the meaning of the original statements is significantly altered.

However, Garcia-Marques and colleagues (2015) found that illusions of falseness occurred only when truth judgments were made within a brief interval after the exposure phase. When the delay was extended to one week, the illusion of falseness reversed into an illusion of truth. This pattern of findings suggests that accurate recollection of specific content details of the statements fades with time and judgments of truth are driven by alternative mechanisms, such as fluency-driven familiarity (see Begg et al., 1992; Brown & Nix, 1996). Specifically, illusions of falseness associated with contradictions seem to depend on conscious recollection of the details associated with the meaning of the original statements while illusions of truth to a greater degree depend on the fluency that originated from the repetition of most of the exact wording of the statement. However, the Garcia-Marques et al. study left unclear whether participants in the delayed condition relied their judgments on the fluency generated by the repetition of the topic of the statements, by the repetition of most of their perceptual features, or both.

Providing convergent evidence on the importance of controlled memory mechanisms such as recollection for the rejection of contradictions as true statements,

Johar and Roggeveen's (2007) investigated the effectiveness of refutations of an original statement depending on whether or not the refutations were perceptually similar with the original. Their experiments showed that refutations that were direct contradictions of the original claim, and therefore more perceptually similar, were less believed than indirect refutations (i.e., statements not seen before but that could be inferred from old statements). Mediation analysis suggested that the similar refutations were stronger recall cues for the original claims, and that accurate recall decreased belief in the contradiction itself.

The importance of recollection for the rejection of statements thought to be false was also shown by Begg and colleagues (1992). The authors presented statements aurally through a tape recording and told participants that all statements presented by a female voice were true and all statements presented by a male voice were false (or vice-versa). When afterwards asked to evaluate the truth of old and new statements, participants evaluated old statements presented by the "false" voice during the first exposure as about equally true as new statements. In contrast, old statements that had been presented by the "true" voice were not only rated more probably true than new ones but also more probably true than "false" voice old statements (Experiment 4). However, in another experiment (Experiment 3) in which the truth-status of the source of the statements was less memorable (participants merely read statements paired with either female or male names and were told to believe one of the sexes and doubt the other), old statements associated with the "false" source were judged as being more probably true than new statements.

4. The Present Experiments

To understand the role that conceptual and perceptual fluency play in repetition-based illusions of truth, we introduce in Experiment 1 one type of repetition that has not been clearly and fully explored in previous research: paraphrases, that is, the repetition of the exact details and meaning of a statement independent of its perceptual features. This manipulation maintains the meaning of the original statements (not only their topic), while reducing the perceptual overlap usually associated with repetition. Bhagat and Hovy (2013) described 25 operations that can be implemented to transform original statements into paraphrases. These operations include, for example, synonym substitution (e.g., The infection with the highest prevalence *in the world* is Malaria → The infection with the highest prevalence *on the planet* is malaria), converse substitution (e.g., The pigeon has a life time *superior* to that of a rabbit → A rabbit has a life time *inferior* to that of a pigeon) or change of voice (e.g., The first traffic light *was installed* in the USA and it was only red and green. → The USA *installed* the first traffic light and it was only red and green). The use of these transformations ensures that paraphrases widely overlap in the semantic features of the original statements but differ in their perceptual features. When we write about “the same meaning of statements” in this article, we refer to the encompassing semantic overlap between the two statements due to these transformations. This stimulus category is essential when we want to gauge the relative contributions of perceptual and conceptual fluency to judgments of truth. With this experimental setup, we can directly contrast truth ratings given to verbatim repetitions and to paraphrases. This specific contrast, which has not been addressed in previous studies, allows to test

whether the perceptual fluency originated by the verbatim repetition of the statements adds a significant contribution to the effect produced by the repetition of their meaning.

In addition to paraphrases, participants in Experiment 2 judged contradictions of the items presented in the exposure phase. Contradictory statements maintain both the topic and the perceptual features of the original items, while differing in their meaning (e.g., Bacon, 1979; Garcia-Marques et al., 2015). Furthermore, the second experiment also introduced paraphrases of the contradictory statements, that is, items that changed both the meaning of the original statements and their perceptual features. The feature in common with the original statements is then merely their topic. Thus, the contrast between the truth ratings given to contradictory statements and those given to contradictory paraphrases allows us to test yet once more how much individuals rely on the perceptual overlap when they cannot use conceptual similarity given that the two kinds of items contradict the specific meaning of the original statements (and given that the overlap with the topic of the original statements is a factor kept constant in the two types of repetition). Additionally, the contrast between truth ratings given to paraphrases and contradictory paraphrases provides a test for another theoretically and empirically important question that has not yet been addressed. It allows the direct comparison of the contribution of the fluency elicited by the repetition of two different (even if related) conceptual aspects of the original items when perceptual overlap with the original statements is reduced to a great extent: repetition of the topic, which is present in both paraphrases and contradictory paraphrases, and repetition of the same meaning as in the original statements that is only present in paraphrases without contradiction.

According to previous studies, experimental conditions that undermined recollection of the details of the original statements (e.g., divided attention during the first exposure to the statements or increased delay between exposure and truth judgments) led to increased illusions of truth for statements initially presented as false (Begg et al., 1992; Brown & Nix, 1996) or statements that contradicted previous ones (Garcia-Marques et al., 2015). These findings suggest that reducing recollection of the details of the statements led participants to base their truth judgments on the familiarity associated with the fluency of processing repeated statements (Begg et al., 1992). Therefore, we wished to create conditions that fostered vs. hindered optimal recollection of the original statements, as a way to examine the extent to which individuals base their truth judgments on fluency stemming from the different types of repetition. For this purpose, we manipulated the delay between the first exposure to the statements and the truth judgments session, asking participants to make their judgments either immediately after the first exposure to the statements or one week later.

5. Experiment 1

The goal of Experiment 1 was to compare paraphrases and verbatim repetition effects on perceived truth, both for judgments made within a few minutes after exposure and after a delay of one week. Paraphrases differ from verbatim repetitions in that the former share fewer perceptual features with the original items (because the wording and syntactic structure are changed) than the latter do, while both share the same semantic features (because the meaning of the original statements is preserved in both).

In both delay conditions, we expected participants to rate verbatim repetitions as more probably true than new statements, replicating previous studies in which illusions of truth emerged both at short and long intervals of time (e.g., Bacon, 1979; Garcia-Marques et al., 2015; Hasher et al., 1977). Given previous results in the literature showing that mere topic repetition is enough to produce illusions of truth also after several weeks of delay between the first exposure and the truth judgments (e.g., Arkes et al., 1991), we also expected the truth ratings given to paraphrases to be higher than those given to new statements, as paraphrases maintain the conceptual features of the original items, including the same meaning. Additionally, reading time measures show that paraphrases of text passages are processed faster after having read a text with the same semantic content (e.g., Levy, Barnes & Martin, 1993; Levy et al., 1995; Tardif & Craik, 1989). This further suggests that the fluency resulting from the overlap in perceptual features (which paraphrases have to a much lesser extent than verbatim repetitions) might not be a necessary condition for illusions of truth to occur. Yet the most important result in this experiment is the comparison between the truth ratings given to paraphrases and verbatim statements. Because the most relevant difference between the two types of repetition is the perceptual similarity to the original statements, the differences that may emerge between the two will inform about the specific contribution of perceptual overlap to the truth effect.

It is important to note here that previous literature in the domain of memory research suggests that over longer intervals between learning of an item and the test episode, both conscious recollection and familiarity show marked forgetting effects (for a review see Yonelinas, 2002). Thus, it is possible that the magnitude of the illusions of

truth is somewhat smaller when truth judgments are delayed by one week (but see Nadarevic & Erdfelder, 2014). However, in light of all the evidence showing illusions of truth even after several weeks have passed, we predict to find truth effects in the delayed judgment condition. Finally, the difference between verbatim statements and paraphrases in the delayed condition will inform us about the role that perceptual and conceptual overlap, respectively, play in truth judgments.

5.1 Method

5.1.1 Participants and Design.

A total of 105 undergraduate students (95 women; $M_{age} = 22.00$, $SD = 7.03$) participated in the experiment for course credit. Participants were randomly assigned to one of the four conditions in a two-way between-participants design: statement repetition (old vs. new) x repetition type (verbatim vs. paraphrase) x truth judgment session (same-session vs. 1-week delay). All participants evaluated repeated and new items.

5.1.2 Stimulus Development.

We developed paraphrases of 52 statements (half factually true and half factually false) previously pre-tested as neutral regarding truth status (i.e., equally likely to be perceived as true or false; Garcia-Marques et al., 2015),² ensuring that they: (a) had approximately the same length (number of words) as the original statements, (b) repeated

² Following Bacon's (1979) methodology, a large sample of true statements about different topics were gathered. A false version of each statement was created by changing a detail of the original. Neutral statements are those that in a pre-test showed that both versions were rated as equally likely to be true (Garcia-Marques et al., 2015).

as few words as possible from the original statements (by means of synonym substitutions), (c) had a different phrase structure (by means of a reordering of the elements of the phrase), and (d) shared the same meaning of the original statements (see Bhagat & Hovy, 2013). For example, the paraphrase of the statement “The infection with the highest prevalence in the world is Malaria” is “Malaria is the most predominant infectious disease on the planet”, encompassing synonym substitutions (e.g., “*in the world*” $\rightarrow$ “*on the planet*”) and a reordering of the elements of the sentence (“*The infection (...) is Malaria*” $\rightarrow$ “*Malaria is (...) infectious disease*”).

From the original-paraphrases pairs, we selected the 42 (21 factually true and 21 factually false) that scored highest in a pre-test regarding meaning similarity (all $M_{\text{similarity}} \geq 5$ in a rating scale from 1 to 7; overall mean of the selected statements was $M_{\text{similarity}} = 6.21$, $SD = 1.37$). From those, 10 factually true and 10 factually false pairs were randomly selected to compose the list of target (i.e., repeated) stimuli of the experiment. Twenty further statements (10 original and 10 paraphrases, and in each of these, 5 were factually true and five were factually false statements) were randomly selected from the remaining highly similar pairs to serve as the new statements in the truth evaluation task. In order to prevent statement effects, we created two sets of statements (A and B) that balanced factually True/False status and repetition type (verbatim vs. paraphrase) across conditions. New statements were the same for both sets.

5.1.3 Procedure

Participants arrived in the lab and were seated at individual workstations. They were told they were going to participate in a session composed of different studies, some

of which were pre-tests of material for future experiments. As in previous studies, the first exposure to the statements was disguised as a pre-test in which participants had to evaluate 24 statements on their level of interest using a rating scale ranging from 1 - Not Interesting at all to 7 - Very Interesting (see Begg et al., 1992; Garcia-Marques et al., 2015). Instructions stated that statements would be presented through headphones, at a rhythm of 10 seconds per statement. Therefore, their judgments had to be relatively fast without much time to think about the response. Participants were also instructed that the first four statements were used as practice trials, and that they should not interrupt the hearing of the statements until the last judgment was made. Statements were then presented one at a time with an inter-stimulus interval of 3 seconds. Half of the participants listened to Set A and the remaining to Set B. During the presentation of the statements, the interest rating scale remained in the center of the screen.

After the exposure phase, participants in the 1-week delay group participated in another, unrelated experiment. Participants in the same-session group were presented with a second, “unrelated” material pre-test. Now they had to evaluate how true they considered a new group of 40 statements, using a 7-point scale (1 - Certainly False; 4 - Uncertain; 7 - Certainly True). Participants were told that half of the statements were true and half were false, and they were instructed to answer rapidly, with the first spontaneous response that came to their minds. Half of the statements were new and half were repeated from the interest ratings task, and they were randomly ordered for each participant. Depending on the experimental condition, participants evaluated either 20 verbatim repetitions or 20 paraphrases of the original statements together with 20

completely new items. Statements were presented one by one, remaining at the center of the screen with the rating scale below them until a response was given.

In total, the two tasks lasted between 10-15 minutes, and upon completion participants were thanked and debriefed. All participants were instructed to come back to the lab one week later, and whereas those in the 1-week delay group completed the truth evaluations at that time, the same-session group participated in a different, unrelated study.

5.2 Results

We expected participants to assign higher truth-values to repeated items than to new ones, showing evidence of illusions of truth. If perceptual overlap contributed to the illusion of truth, we expected a difference between verbatim repetitions and paraphrases. By contrast, if perceptual overlap in this between-participants design did not affect truth judgments because its role is overshadowed by other mechanisms, such as recollection (in the same-session) and conceptual fluency (in the 1-week delay) stemming from the repetition of the meaning of the original statements, we would not expect any a difference between verbatim repetitions and paraphrases.

Mean ratings of truth (Table 1) were analyzed in a 2 (old vs. new statements) x 2 (verbatim vs. paraphrased repetition) x 2 (same-session vs. 1-week delay) ANOVA, with the last factor as repeated measures. As expected, we obtained a main effect of repetition, $F(1, 101) = 105.54, p < .001, \eta_p^2 = .51$, evidencing a truth effect associated with old items ($M_{Old} = 5.01, SD = .82; M_{New} = 4.26, SD = .46$). A significant interaction between repetition and the evaluation session factor was also found, $F(1, 101) = 13.81 p < .001$,

$\eta_p^2 = .12$. As Table 1 shows, this effect is associated with lower differences between repeated and new statements in the 1-week delay condition than in the same-session condition, which might be explained by forgetting effects on both recollection and familiarity mechanisms over long delays (see Yonelinas, 2002; see also Roediger & Blaxton, 1987, for a decrease in repetition priming effects on fragment completion tasks after one week). No main effects and interactions associated with type of repetition were found ($F_s < 1$), suggesting that verbatim and paraphrased repetitions did not differ and induced the same effects on truth judgments in both judgment sessions.

Direct tests of our hypothesis were performed with planned comparisons testing the presence of the truth effect (repeated statements: 1; new statements: -1) in all experimental conditions. For the verbatim repetition condition, a truth effect emerged in both the same-session, $t(101) = 6.84, p < .001, d_z = 1.34$, and the 1-week delay condition, $t(101) = 4.08, p < .001, d_z = .76$. The same pattern was found for paraphrases: same-session, $t(101) = 6.59, p < .001, d_z = 1.40$; 1-week delay, $t(101) = 2.80, p = .006, d_z = .53$.

5.3 Discussion

Experiment 1 showed that paraphrases were evaluated as true as verbatim repetitions, both when truth judgments were made immediately after exposure or after a 1-week delay. This suggests that processing fluency induced mainly at the conceptual level mimics processing fluency associated with complete perceptual overlap involved in verbatim repetition. That is, the fluency elicited by the greater perceptual overlap incorporated in verbatim repetition does not seem to yield any additional effect to the

fluency elicited by the overlap with the meaning of the original statements that is present in paraphrases.

Yet, although the fluency originating from perceptual overlap does not seem to be a necessary condition for the truth effect to occur, we already know from previous research that perceptual fluency is sufficient to elicit the effect (e.g., Parks & Toth, 2006; R. Reber & Schwarz, 1999). Thus, perceptual fluency may still contribute to the repetition of the meaning of the original statements, but the between-participants design in which participants only judged either paraphrases or verbatim repetitions but never the two types of repetitions together may have prevented the detection of such effects (for a similar finding, see Dechêne, Stahl, Hansen, & Wänke, 2009).

To address the issue that a between-participants design may undermine the effects of fluency due to perceptual overlap, in Experiment 2 we made some modifications to the design of the experiment to make comparisons between different items possible to participants. Therefore, participants were asked to evaluate all types of repeated statements. We also introduced a factor manipulating whether repeated statements maintained or contradicted the meaning of the statements originally presented (even if maintaining their topic). Finally, because the contribution of memory mechanisms in Experiment 1 could only be inferred from the difference of the ratings in the same-session vs. 1-week delay conditions, a recognition test was added at the end of Experiment 2. The results of this memory test will possibly inform the degree to which fluency elicited by perceptual vs. conceptual similarities interacts with memory mechanisms such as accurate recollection of the details of the statements in the production of illusions of truth.

6. Experiment 2

Experiment 2 was designed to further investigate the role that conceptual and perceptual fluency exert in truth evaluations of repeated statements, by allowing participants to contrast statements with different degrees of perceptual similarity (same words and sentence structure but not the same meaning) and conceptual similarity (same topic and meaning, but not the same wording or sentence structure) with the original items. Thus, four types of repeated statements were presented: (1) verbatim repetitions, (2) paraphrases, (3) contradictory statements and (4) paraphrases of the contradictory statements (contradictory paraphrases), mixed with completely new items.

In their meta-analysis, Dechêne and colleagues (2010) contrasted the results of experiments that manipulated the repetition of topic and related issues with those that focused on verbatim repetition of the statements. Their analyses suggest that topic repetition produces small truth effects when participants are presented with all types of items (e.g., both old and new) in a within-participants design that allows comparisons between the different types of items (a heterogeneous context in Dechêne et al.'s terminology) and null effects when participants were presented with only one type of item (e.g., either old or new) in a between-participants design that made comparisons between different types of items impossible (a homogeneous context according to Dechêne et al., 2010). In order to address these differential effects elicited by different designs with which truth effects are tested, participants in Experiment 2 evaluated statements pertaining to all the different types of repetitions mixed with completely new statements. With this design we intended to create an evaluation setting that parallels

Dechêne et al.'s (2010) conception of a heterogeneous context, in contrast to Experiment 1 more homogeneous context (since participants rated only one type of repeated statements intermixed with new ones).

As mentioned earlier, the comparisons between all the different types of repetitions allow us to test the extent to which individuals rely on the different levels of perceptual and conceptual similarity when making their judgments. We expected a replication of the truth effect, evidenced by higher truth evaluations to repeated items than to new ones. In line with Experiment 1, delay of evaluation is expected to moderate the truth effect, reducing its magnitude. The more heterogeneous context may however lead to some differences regarding the contrast between verbatim and paraphrased repetitions. This context may highlight all the different nuances of repetition and make them more relevant for the truth judgment task. That is, since participants can directly compare the two types of repetitions, they may rely more on the perceptual overlap included in verbatim repetitions when making their judgments than they did in Experiment 1 that employed a between-participants design.

By opposing the meaning of original items, contradictory statements are expected to produce an illusion of falseness and to be evaluated as less probably true than new items (Bacon, 1979; Garcia-Marques et al., 2015). However, this effect should be reduced in the delayed evaluation because of the similarity of contradictory statements with the perceptual features and topic of the original items (Garcia-Marques et al., 2015). In line with contradictory statements, contradictory paraphrases are expected to produce an illusion of falseness in the same-session condition, as they also contradict the meaning of original statements. Yet, if perceptual fluency contributes to the reversal of the

falseness effect in the delayed judgment condition, we should not expect the two types of contradictory items to produce the same effects. With the passage of time, recollection of the meaning of the statements is hampered and conceptual fluency due to overlap with the topic may exert some influence instead. Reduced recollection and increased influence of conceptual fluency both reduce the illusion of falseness. However, as contradictory statements share also the perceptual features of original statements, while contradictory paraphrases have almost no perceptual overlap and share mostly the topic, we would expect that the reversal of the illusion of falseness is more pronounced for contradictory statements than for contradictory paraphrases. Finally, the comparison of the truth ratings given to paraphrases and to contradictory paraphrases (especially when judgment is delayed) allows to contrast the contribution of the fluency stemming from the repetition of the topic of the original statements and the fluency stemming from the repetition of their meaning when there is little perceptual similarity with the originals.

By analyzing truth ratings that take recognition classifications of the statements into account (i.e., conditional analysis), we can test differences between statements thought to be repeated and those thought to be new relative to actual repetitions and new statements. Therefore, a recognition test was added to the experiment. Although we expect recognition memory and truth judgments to share variability (as both judgments are influenced by the ease with which a statement is processed), we expect these not to be perfectly correlated. Additionally, any effects associated with the different types of statements in the conditional analysis will inform about the potential role that conceptual and perceptual similarity with original items play in the relation between recognition memory and truth judgments.

6.1 Method

6.1.1 Participants and Design

Forty-five undergraduate students (36 women; $M_{\text{age}} = 20.20$, $SD = 3.00$) participated in the experiment for course credit. They were randomly assigned either to the same-session group ($n = 24$) or the 1-week delay group ($n = 21$; between-participants). All participants judged five types of statements: verbatim repetitions, paraphrases, contradictory statements, contradictory paraphrases, and new statements.

6.1.2 Stimulus Development

From the 52 pairs of neutral statements and paraphrases pre-tested in Experiment 1, we selected the 40 pairs (20 true and 20 false) that scored the highest on meaning similarity to use as target stimuli. For each statement and its paraphrase, a contradictory version (also pre-tested as neutral) was introduced by changing a small detail in the statements (usually a word or two) that alters and contradicts its original meaning (these contradictory statements were selected from the same pre-test of statements' truth status as in Experiment 1, Garcia-Marques et al., 2015). For example, for the statement "Crocodiles sleep with the eyes open": the paraphrase is "When sleeping, crocodiles do not close their eyelids"; the contradictory version is "Crocodiles sleep with the eyes closed"; and the contradictory paraphrase version is "When sleeping, crocodiles close their eyelids". Twenty other items (half true and half false) served as new statements.

6.1.3 Procedure

The procedure was the same as in Experiment 1, except that in the truth evaluation task participants rated all the four types of repeated statements (a total of 40 statements, 10 of each type of repetition; statements in each of the repetition conditions were counterbalanced, so that each statement was evaluated equally often in all four versions), mixed with 20 new ones. Statements were randomly ordered for each participant. We acknowledge that the number of stimuli in each repetition category is not as large as in Experiment 1 (10 vs. 20, respectively) and the number of critical (repeated) items is also smaller than the number of new stimuli. However, in light of Dechêne et al.'s (2010) results which suggest that the proportion of critical (repeated) items does not significantly alter the truth effect, and the fact that several other authors were able to reliably test their specific hypothesis with similarly small sets of stimuli (e.g., Arkes et al., 1991, in Experiment 2 used 8 statements per repetition condition; Begg et al., 1992, in Experiments 1 and 2 used 10 repeated statements per each of the 4 source credibility $\times$ source gender combinations; or Boehm, 1994, who used between 2 and 14 repeated statements), there is good reason to assume that this will not significantly harm the reliability of our results. To make sure that our experiment had adequate power to test our hypotheses and detect the possible effects, we conducted a statistical power analysis with the online application "Power Analysis with Crossed Random Effects" designed by Westfall, Judd and Kenny (2014). The parameters included in this power analysis take into account both participant and stimuli as a source of random variation in the data. Setting effect size for repetition effects at $d = .80$ (based on the smallest effect size Garcia-Marques et al., 2015, observed for illusions of truth with the same class of stimuli as in the present experiments, $d = .83$) and entering our total number of participants (45)

and of stimuli (60), we are able to reach a power of .99. Conversely, using the same parameters to calculate how many stimuli would be necessary to reach statistical power at the recommended .80 level (with the same number of participants and effect size), the results of the power analysis show that 9 stimuli would already be sufficient.

After providing truth ratings, participants were presented with exactly the same set of statements (randomly ordered) and were asked to indicate if each item had been presented in the first pre-test of the session, the interest ratings task (for participants in the 1-week delay condition, instructions referred to the interest pre-test presented one week ago). Statements were presented one by one, and remained at the center of the screen until a response was given. The labels “S (Yes)” and “L (No)” were presented close to the lower left and right corners of the screen on each trial, respectively, and to respond participants pressed one of the keys.

The three tasks (interest ratings, truth evaluation, and memory test) lasted between 20-30 minutes, depending on individual speed. When participants completed the recognition judgments they were thanked and debriefed.

6.2 Results

6.2.1 Truth Judgments

Truth evaluations (Table 2) were analyzed with an ANOVA³ defined by 5 (type of statement) x 2 (truth evaluation session). No direct effect of the time factor was observed

³ Four participants were excluded from our analysis given their extreme values in their response times. The main findings reported here did not change when these participants were included in the analysis.

($F < 1$). A main effect of type of statement suggests differences in the evaluation of the 5 types of items, $F(4, 156) = 31.48, p < .001, \eta_p^2 = .45$. The effect of type of statement was moderated by the time of the evaluation session, $F(4, 156) = 9.33, p < .001; \eta_p^2 = .19$.

As in Experiment 1, we tested the direct hypotheses with planned comparisons addressing the presence of truth effects (repeated statements: 1; new statements: -1) with each type of repeated statement in both evaluation session conditions. We expected to replicate the truth effect both with verbatim repetitions and paraphrases, and in both judgment conditions. In line with Garcia-Marques and colleagues' (2015) previous results, we expected both types of contradictions (i.e., contradictory statements and contradictory paraphrases) to induce illusions of falseness in the immediate judgment condition. In the delayed judgment condition we expected contradictory statements to reverse into an illusion of truth (replicating Garcia-Marques et al., 2015). However, if perceptual fluency contributes to such reversal, then the same may not happen for contradictory paraphrases, as these items lack the perceptual similarities to original statements.

Replicating Experiment 1, we found clear evidence of a truth effect associated with verbatim repetitions both in the same-session, $t(39) = 6.26, p < .001, d_z = 1.40$, and in the 1-week delay condition, $t(39) = 2.51, p < .016, d_z = .55$. The same pattern was found for paraphrases: same-session condition, $t(39) = 6.23, p < .001, d_z = 1.39$, and 1-week delay condition, $t(39) = 2.88, p < .007; d_z = .63$. As in Experiment 1, paraphrases were rated to be as true as verbatim repetitions ($t < 1$). These results suggest that even when the context favored the direct comparison between different types of similarity with

the original statements, perceptual overlap did not add a significant contribution to the illusions of truth elicited by simple repetition of the meaning of the statements.

Contradictory statements showed a tendency to produce illusions of falseness in the same-session condition, presenting a marginally significant tendency to be judged as less true than new items, $t(39) = -2.01, p = .052; d_z = .45$, and therefore also less probably true than original statements. This tendency started to reverse in the 1-week delay condition, in which contradictions were judged to be as true as verbatim repetitions, $t(39) = 1.11, p = .275$, and showed a marginal tendency to be evaluated as more probably true than new statements, $t(39) = 1.52, p = .136$ (a unilateral hypothesis, marginally significant, $p = .07, d_z = .33$, in line with Garcia-Marques et al.'s, 2015 results in a homogeneous context that did not allow for comparisons between item types). Furthermore, when comparing the truth ratings given to contradictory statements in the two delay judgments conditions, it is possible to see that these items were considered significantly more probably true in the 1-week delay condition than in the same-session condition, $t(39) = 2.62, p = .013, d_s = .82$. These results suggest that when testing conditions did not favor detailed recollection for the original statements (i.e., longer delay between exposure and test), participants made almost no difference between contradictory statements and verbatim items. However, they also suggest that when contradictory statements are presented in a heterogeneous context, participants may rely less on the fluency brought about by perceptual overlap (and by topic repetition). One possible explanation for this observation is that context highlights the different sources of similarity, helping participants to discriminate between appropriate sources of

information for their judgment (e.g., Whittlesea, 1993; see also Silva, Garcia-Marques, & Mello, 2016).

Contradictory paraphrases were perceived as less true than new statements in the same-session condition, $t(39) = -2.75$; $p = .009$; $d_z = .60$ (and thus also less true than verbatim repetitions), reinforcing the idea that changing the meaning of the original statements leads to an illusion of falseness. However, unlike contradictory statements, in the 1-week delay condition contradictory paraphrases were still judged falser than verbatim repetitions, $t(39) = 2.72$, $p = .010$, $d_z = .61$, but just as true as new ones ($t < 1$). The simultaneous change in the meaning and in the perceptual features seems to have prevented participants from judging contradictory paraphrases as true as original items, even though they shared the original topic. Thus, our results suggest that in a heterogeneous context that allows comparison between the different types of statements, sharing the same topic of previously presented statements is not sufficient to induce illusions of truth, even in conditions that undermine the detailed recollection of the original statements (i.e., 1-week delay).

The comparison of contradictory statements and contradictory paraphrases allows us to assess the contribution of perceptual fluency effects (stemming from the repetition of the perceptual features of the original statements) to judgments of truth in the absence of conceptual fluency stemming from sharing the meaning of the original items (the other possible source of conceptual fluency is kept constant, as both types of statements have topic similarity with the originals statements). In the 1-week delay condition, contradictory statements showed a marginally significant tendency to be judged more probably true than contradictory paraphrases, $t(39) = 2.00$, $p = .052$, $d_z = .44$. This result

is most likely explained by the fact that contradictory statements repeat almost all the perceptual features of the original statements, which contradictory paraphrases do not, and that participants partially relied in the resulting perceptual fluency when making their judgments. The lack of differences between the two types of statements in the same-session condition ($t < 1$) suggests that in that condition the illusion of falseness associated with contradictory paraphrases was not caused only by the lack of perceptual similarities with the original statements. Rather, it seems to have been caused by the change in the meaning, which participants could probably more accurately recall in the immediate judgment condition (just as with contradictory statements).

Finally, by comparing paraphrases with contradictory paraphrases in the 1-week delay condition we contrasted the contributions of two possible sources of conceptual fluency in the absence of perceptual fluency (since both types of repetitions have no perceptual overlap with the originals): repetition of the meaning of the original statements and repetition of only their topic. In this condition, paraphrases were rated more probably true than contradictory paraphrases, $t(39) = 3.14$, $p = .003$, $d_z = .69$.

6.2.2 *Memory Judgments.*

Proportions of hits (saying yes to the verbatim repetitions of the original items) and false alarms (FA, saying yes to paraphrases, contradictory statements, contradictory paraphrases and new statements) for old/new discrimination of statements are depicted in Table 3. Mean proportions of FA to the different types of repetitions (i.e., paraphrases, contradictory statements, contradictory paraphrases) were in between $M = .50$ and $M = .72$, whereas FA to the completely new statements were $M = .16$. And while only 29.5%

of participants made FA to the new items, almost all participants (98%) made at least one FA to the other types of repetitions.

To understand the independent influence of each type of statement similarity in level of discrimination between old and new items, we computed Signal Detection Theory d' estimates (Table 3). These discrimination estimates inform us about how much participants could distinguish between verbatim repetitions (the real old items) and each of the other four categories of distractor statements. d' estimates are calculated by subtracting the z score of each of the four different FA rates (i.e., proportion of “Yes” responses to paraphrases, contradictory statements, contradictory paraphrases, and new statements) from the z score of the hit rate (i.e., proportion of “Yes” responses to verbatim repetitions). The higher a particular d' estimate is, the better participants were in discriminating verbatim repetitions from that particular type of distractor, with $d' = 0$ representing complete inability to discriminate (see Mcmillan & Creelman, 1991; Stanislaw & Todorov, 1999).

A set of hypotheses can be deduced from our previous claims and results. First, the fact that paraphrases were judged just as true as verbatim repetitions suggests that paraphrases will be more likely to be wrongly categorized as “old” than any other type of distractor item. Thus, when subtracting the FA rate of paraphrases from the hit rate of verbatim repetitions, the resulting d' scores should be very low. In fact, we expected these d' estimates to be the lowest of all discrimination scores, and this should be reflected in a main effect of type of statement (H1). Second, delaying judgments is expected to impair memory mechanisms, both recollection and familiarity (see Yonelinas, 2002) and consequently decrease the difference between the hit and FA rates.

Therefore, we expected a main effect of judgment session to be associated with lower d' values in the delayed judgment session (H2). Given the relation between truth and memory judgments, we expected an interaction between the type of statement and evaluation session factors on memory discrimination, mimicking what was observed for the truth judgments (H3). Specifically, we expected that discrimination estimates calculated with contradictory paraphrases and completely new statements were not different after a delay given that these two types of items have little to no similarities with the original statements and thus fluency should not support judgments to the same extent as is the case with the other types of statements.

Discrimination scores were analyzed with a repeated measures ANOVA, with judgment session as a between-participants factor. Results of the ANOVA support our predictions, showing the presence of significant main effects of type of statement, $F(3, 117) = 76.26, p < .001, \eta_p^2 = .66$, and judgment session, $F(1, 39) = 15.09, p < .001, \eta_p^2 = .28$, as well as a significant interaction between the two factors, $F(3, 117) = 6.69, p < .001, \eta_p^2 = .15$. In order to understand whether the pattern of these effects fit our hypotheses, we performed the same specific planned comparisons as we did for the truth judgments. Thus, main contrasts compared how good participants were in discriminating verbatim repetitions from each of the three non-verbatim types of repetition (i.e., the d' calculated with the FA given to paraphrases, contradictory statements and contradictory paraphrases; contrast weights: 1) to how good participants were in discriminating verbatim repetitions from completely new statements (i.e., the d' calculated with the FA given to new statements; contrast weight: -1), in the two evaluation session conditions.

Discrimination of paraphrases was worse than discrimination of new items in both the same-session, $t(39) = -11.01, p < .001, d_z = 2.46$, and the 1-week delay condition, $t(39) = -6.13, p < .001, d_z = 1.34$. In fact, supporting our first prediction (H1), discrimination of paraphrases was worse than all other types of statements in the two judgment conditions (same-session, $t(39) = -9.70, p < .001, d_z = 2.17$, and delayed-session, $t(39) = -4.15, p < .001, d_z = .91$). The relevance of sharing the same meaning with the original items is such that the level of discrimination of paraphrased statements did not differ from a pure chance level ($d' = 0$), either in the same-session ($t < 1$) or the 1-week delay conditions ($t(20) = 1.61, p = .123$). Thus, once again repetition of statements' meaning led participants to accept paraphrases as unchanged repetitions.

Similarly, contradictory statements showed worse discrimination than new statements in the two judgment session conditions (same-session, $t(39) = -6.74, p < .001, d_z = 1.51$, and 1-week delay, $t(39) = -4.91, p < .001, d_z = 1.07$). The overlap of perceptual features is likely to have harmed individuals' ability to discriminate between original items and statements that contradict their meaning.

Sharing the original statements' topic also led to poorer discrimination than totally new statements. This is clearly shown in the contrast between the discrimination scores of contradictory paraphrases and new items, in both evaluation sessions (same-session, $t(39) = -7.86, p < .001, d_z = 1.76$, and 1-week delay, $t(39) = -6.14, p < .001, d_z = 1.34$). Thus, even without sharing the meaning and the perceptual features of original items, just sharing the same topic induced some level of false recognitions (i.e., FA).

To address our prediction that discrimination should decrease in the 1-week delay condition for all types of repetitions with exception of contradictory paraphrases and new

statements (H3), we compared d' estimates observed in the same-session condition to those in the 1-week delay condition individually for each type of statements. Results of these specific comparisons reveal that our prediction was not supported. In fact, discrimination in the delayed-session was significantly worse for all types of statements (H2, main effect of evaluation session) except for paraphrases (paraphrases, $t < 1$; contradictory items, $t(39) = 3.19$, $p = .003$, $d_s = 1.00$; contradictory paraphrases, $t(39) = 3.75$, $p = .001$, $d_s = 1.17$ new items, $t(39) = 3.70$, $p = .001$, $d_s = 1.16$). As Table 3 shows, delaying judgments has affected participants' willingness to categorize statements as "old", reducing greatly the proportion of hits and bringing it closer to the proportions of FA, which also decreased for all types of repetitions (although not to the same extent as the proportion of hits), and consequently all d' estimates were lowered. This was less noticeable in the case of paraphrases because discrimination with this type of statement was already low in the immediate judgment condition.

6.2.3 Conditional Analysis of Truth Ratings

The preceding analyses revealed differences in both rated truth and recognition judgments for the different types of statements. To understand whether perceived familiarity expressed as recognition judgments influenced truth judgments, we calculated the mean ratings of truth of only those items that were considered "old" in the recognition test, for each type of repetition; new statements were excluded. This was necessary because although all types of items produced FAs, mean truth estimates of the perceived to be "old" items have a different reliability than those of the items perceived as "new", for two main reasons. First, as results of the recognition test revealed (Table 3), only a

small proportion of participants (29.5%) committed FAs with new statements and only a subset of new statements were falsely recognized ($M = .16$). This makes mean truth ratings of the new statements falsely recognized as “old” less reliable than the estimates of the other types of items. Second, exactly the opposite happens with the estimation of means for all the other statements when perceived to be “new”, as most participants (98%) showed false recognitions for at least 50% of items belonging to each type of repetition.

Truth ratings of the perceived to be “old” items (Table 4) were analyzed with an ANOVA with the 4 types of repeated statements as a within-participants factor and evaluation session as a between-participants factor. Results revealed a main effect of type of statement, $F(3, 114) = 7.21, p < .001, \eta_p^2 = .159$, showing that contradictory paraphrases were considered less probably true than all the other types of repetitions ($M_{\text{Contradictory-paraphrases}} = 4.23, SD = 1.16$ vs. $M_{\text{Contradictory}} = 4.68, M_{\text{Original}} = 4.97, SD = 0.93, M_{\text{Paraphrases}} = 5.11, SD = 0.86$; contrast, $t(38) = 3.34; p = .002; d = 1.08$). Thus, statements that lacked similarity with both the meaning and the perceptual features of original items were not perceived to be as true as any of the statements with more conceptual and/or perceptual fluency, even when they were falsely recognized. Time of the judgment session did not affect truth ratings of the perceived to be “old” items ($F < 1$), and it also did not interact significantly with the type of statement factor, $F(3, 114) = 1.21, p = .310$.

6.3 Discussion

Experiment 2 replicated and extended the results of Experiment 1 in several ways. First, the effects observed in Experiment 1 were replicated in a heterogeneous context in

which participants judged all different types of statements. Even though this context presumably accentuated the differences between statements, participants reacted to paraphrases as if they were verbatim repetitions. The simultaneous presentation of verbatim repetitions and paraphrases did not lead to a greater differentiation between these two types of items in their perceived truth-value ($M_{verbatim} = 4.87$, $SD = .74$, and $M_{paraphrases} = 4.88$, $SD = .64$), suggesting that the fluency stemming from perceptual similarity does not bring an additional contribution to the fluency stemming from repetition of the semantic content of statements (meaning and topic). The results regarding recognition judgments also suggest that paraphrases were treated like verbatim repetitions in both judgment session conditions (overall similar proportions of “old” responses: $M_{verbatim} = .93$; $M_{paraphrases} = .89$). Presumably, when instructions for the memory task do not explicitly stress that only verbatim repetitions should be categorized as old items, individuals use conceptual similarity as the default criteria to make the decision. Of course it is possible that with more exclusive instructions participants would also attend to the perceptual features of the statements when making their recognition judgments. But the fact is that when faced with a general instruction participants assumed paraphrases should be taken as repetitions (even in a context allowing participants to see that some of the statements looked exactly as they had appeared before while others changed).

Second, Experiment 2 adds specificity and a deeper understanding to the illusion of falseness observed for contradictory statements when judgments were made after a short delay, and which reversed into an illusion of truth after a 1-week delay (Garcia-Marques et al., 2015). The heterogeneous nature of our experiment, with type of

repetition manipulated within-participants (note that in Garcia-Marques et al.'s experiment the contradictions vs. verbatim repetitions factor was manipulated between-participants), reduced both the illusion of falseness after a short delay and its reversal after a 1-week delay. How could this difference in the results of the two experiments be explained? Likely, participants got access to more sources of information for their judgments in the heterogeneous context than in a homogeneous context where only one type of repetition is presented. Thus, the overlap in perceptual features or the similarity with the topic of the original items may no longer be experienced as an appropriate source for truth judgments, especially if a stronger conceptual fluency source like the repetition of the meaning of the statements is available. This possibility is consistent with Whittlesea's (1993) notion that fluency cues will only affect judgments as long as they are interpreted as appropriate sources of information.

Third, Experiment 2 extends previous results by showing that, at least in a heterogeneous context, sharing only the topic of an original statement (contradictory paraphrases) was not enough to induce illusions of truth, neither under an immediate nor a delayed judgment condition. However, one should not interpret the latter finding as evidence that topic similarity never affects perceived truth, as again this may depend on the context and the availability of more appropriate fluency sources on which to base the judgments.

The different patterns of results obtained separately for truth and recognition measures suggest that not all features that produce false alarms in recognition produce illusions of truth. For example, statements that share only the topic of their original version (contradictory paraphrases) and statements that also share their perceptual

features (contradictory statements) affect truth ratings independently from having been judged as “old” or “new”. Presumably, the contradictory status of these items prevents their impact on recognition judgments (reducing false alarms), but not the influence of the remaining similarities with the original statements (i.e., perceptual features and topic) on perceived truth.

7. General Discussion

The goal of two experiments was to understand how the fluency stemming from the repetition of different conceptual and perceptual features of statements influences their subjective truth-value. Additionally, we wished to explore how the effects of the different sources of fluency interact with conditions in which truth judgments are likely to be based on the recollection of the details of a statement versus conditions when the judgments are more likely to be based on familiarity. Experiment 1 showed that paraphrases were judged virtually as true as verbatim repetitions of the original statements, both in immediate evaluations and after a 1-week delay. This result suggests that perceptual overlap is not necessary for the emergence of illusions of truth. However, the relevance of perceptual fluency is shown in Experiment 2. In that experiment, marginally significant illusions of falseness emerged for statements that contradicted the meaning of the original items when judgments immediately followed the exposure phase. This was true independently of whether repetitions maintained the perceptual characteristics of the previous items (contradictory statements) or not (contradictory paraphrases). However, when truth evaluations were delayed, the fluency associated with repeating the perceptual features of original statements, such as words and sentence parts,

led to the disappearance of the illusions of falseness and contradictory statements were perceived to be as true as verbatim repetitions. In contrast, contradictory paraphrases continued to be perceived as less true than original items, an observation that can be explained by the absence of perceptual overlap with original statements. The present results are thus a relevant extension to the work by Garcia-Marques et al. (2015) because the reversal the authors found for contradictory items could also have been caused by the fluency associated with the repetition of the topic of the original statements and not necessarily by the perceptual fluency originating from the nearly identical perceptual form of the sentence. The findings on the contradictory-paraphrases in the current study more convincingly suggest that the reversal was driven by perceptual similarity (as the effect for contradictory-paraphrases was not reversed).

Previous studies in the literature show that while variations in conceptual fluency seem to consistently and robustly affect conceptual judgments or tasks, variations in perceptual fluency are less prone to do so (Whittlesea, 1993; see also Roediger, 1990, for the transfer-appropriate procedures approach, showing that memory performance is improved when the cognitive operations enacted at the test phase match those enacted during the learning episode). Such evidence suggests that there is a certain domain-specificity of fluency as a source for judgments. Our results are in line with this assumption. For illusions of truth to occur, variations of conceptual fluency elicited by repetition (vs. contradiction) of the semantic meaning of the original statements are probably more relevant than the variations on perceptual fluency elicited by repeating (vs. changing) their perceptual features. Specifically, paraphrases were treated in the same way as verbatim repetitions while contradictions could never fully elicit illusions of

truth (both contradictory statements and contradictory paraphrases were not considered significantly more probably true than new statements, even in the delayed judgment session). These results are also in line with previous research on differential fluency effects in judgments of truth, as the study by Parks and Toth (2006) illustrates. These authors found that a manipulation of conceptual fluency (e.g., presenting claims in a meaningful vs. meaningless context) had a much greater impact on judgments of truth than a manipulation of perceptual fluency (easy vs. difficult to read letter font) (see also Silva et al., 2016 for a comparison of the effect sizes of illusions of truth obtained with conceptual and perceptual fluency manipulations).

Our experiments suggest that time of the truth judgment session (immediate versus delayed) affected the likelihood of conscious recollection of the details of a statement. In the immediate session condition, recollection is assumed to support both illusions of truth for items that match the meaning of what has been encountered before, and illusions of falseness for items that differed in meaning. In the immediate judgment condition, conscious recollection may have prevented participants from being influenced by the effects of perceptual fluency associated with the perceptual overlap present in contradictory statements. A different pattern of findings could be observed in the delayed judgment condition. With the passage of time recollection of the details of the stimuli is likely to be reduced and judgments might then be based on the familiarity induced by the fluency of processing repeated items (e.g., Jacoby & Dallas, 1981). Thus, due to reduced recollection, truth judgments are likely to have been partially influenced by the perceptual fluency associated with contradictory statements, because they overlap with the perceptual features of original statements. This would explain why in the delayed

judgment condition contradictory statements were judged just as true as verbatim repetitions.

These findings are akin to those of Johar and Roggeveen's (2007) study reported in the Introduction section. The authors found that refutations with high perceptual similarity to the original claim were believed less than the less similar ones, because they contained stronger recall cues for the original statement and this recall decreased belief in the contradiction itself. However, when the authors tried to overshadow the accuracy of memory for the details by raising perceptual similarity between the original claim and the refutations to a degree that participants no longer recognized refutations as such, then highly similar contradictions were believed more. A similar pattern of results can be found in our Experiment 2.

In immediate truth evaluations, there were no differences in the truth-value attributed to contradictory statements and contradictory paraphrases, which were both believed less than original items. However, when recollection of the details of the statements was hindered by delaying the truth judgments, participants tended to consider contradictory statements as more probably true than contradictory paraphrases and equally true as original statements. This probably happened because contradictory statements share the original perceptual features to a great extent (while contradictory paraphrases do not) and participants no longer recognized them as contradictions. This result is in line with a recent finding that warnings about the effect of repetition and instructions to try to control for it do not completely eliminate illusion of truth, presumably because participants cannot discount the influence of fluency on judgments of truth (Nadarevic & Aßfalg, 2016).

It is important to note again that previous studies in the memory literature have found strong forgetting effects both for recollection and familiarity components across long delays between learning and test episodes (Yonelinas, 2002). However, the results of our experiments suggest that, in the context of truth judgments, the familiarity associated with the fluency stemming from perceptual and conceptual similarity with the original statements still exerts its effects (while recollection is affected more strongly). This is in line with findings in the implicit memory literature that showed that explicit memory fades faster than implicit memory (Tulving et al., 1982); however, the mechanisms underlying these differential decay rates remained unclear. One future avenue for research is thus to dissociate the contributions of recollection and familiarity components of memory to judgments of truth under short and long delays after exposure to the statements, as Begg and colleagues (1992) did when using Jacoby's (1991; Jacoby, Toth, & Yonelinas, 1993) process dissociation procedure together with manipulations of attention (full versus divided). Such a study would clarify if the familiarity component involved in truth judgments is less affected by long delays than recollection is.

The context in which individuals make their truth evaluations considerably moderates fluency effects (Dechêne et al., 2009; Wänke & Hansen, 2015), and the present experiments contribute further relevant evidence. Conceptual fluency effects were found in both a homogenous (Experiment 1, in which participants evaluated only one type of repeated statements) and a heterogeneous context (Experiment 2, in which participants evaluated all types of repeated statements). However, the marginally significant illusion of falseness found in Experiment 2 for contradictory statements in the immediate judgment condition was not completely reversed into an illusion of truth in the

delayed judgment condition, as it was in the study by Garcia-Marques and collaborators (2015) who presented items in a homogeneous evaluation context (i.e., participants judged only one type of repetition, either contradictory statements or verbatim repetitions). This difference between the earlier experiment and the current study might be due to heterogeneous contexts (like the one in Experiment 2) facilitating comparisons between the fluency experiences generated by the different types of repetition, increasing discrimination between the different types of items (see Dechêne et al., 2009).

Interestingly, contradictory paraphrases were considered less probably true than all other types of repetitions even if they were falsely recognized as statements presented before. Since items that are perceived as familiar are considered more valid than items perceived as new (e.g., Bacon, 1979; Begg & Armour, 1991), this result suggests that the objective dissimilarities regarding the meaning and perceptual features of original statements reduced the influence of subjective familiarity on truth evaluations. Thus, the impact of subjective familiarity on judgments of truth depends on the level at which that feeling originated, with the fluency that comes from simple topic repetition not being sufficient to increase the truth-value of a statement (at least in a heterogeneous evaluation context). This dissociation between recognition and truth judgments opens the door for a better understanding of the mechanism underlying the truth effect. Rajaram's (1996, 1998) distinctiveness–fluency hypothesis might become relevant here. This framework assumes that processing the salience or distinctiveness of the conceptual or perceptual features of a stimulus boosts the experience of *remembering*, while leaving unaffected the “*know*” responses – which are enhanced by factors increasing the fluency with which stimuli are processed (e.g., Rajaram, 1996). Thus, in future studies it may prove useful to

introduce remember/know decisions following each recognition judgment, as this may shed more light on which characteristics of the stimuli participants relied on when they evaluated each type of statement, and how this may change with a longer delay after the first exposure.

Our experiments show the flexible use of fluency as information for judgments of truth. R. Reber, Wurtz, and Zimmermann (2004) and Wurtz, R. Reber, and Zimmermann (2008) presented evidence that different manipulations of perceptual fluency have different objective effects (measured by response times) at different stages of the perceptual process but they all seem to feed into one and the same subjective experience of ease. Our data suggest that absence of conceptual fluency and reduction of recollection enables perceptual fluency to take over as the relevant source of information for judgments, producing convergent effects in subjective judgments of truth.

As Oscar Wilde wrote, “The truth is never pure and rarely simple”. By examining the effects of different sources of fluency originating from repetition of conceptual and perceptual features of information, their interplay with memory mechanisms such as conscious recollection and familiarity (Begg et al., 1992; Jacoby, 1991), and the influence of the heterogeneous versus homogeneous nature of the judgment context, this study yields new insights into the multitude of processes that determine judgments of truth.

Acknowledgments

This research was supported by the grant SFRH/BD/39153/2007 awarded to Rita R. Silva, the grant PTDC/PSI-PCO/121916/2010 awarded to Teresa Garcia-Marques, and the grant UID/PSI/04810/2013 awarded to the William James Center for Research (ISPA – Instituto Universitário), by the Portuguese National Science Foundation (FCT). The authors thank to Inês Dias and Joana Mello for the assistance provided in the preparation of the materials used in the experiments.

References

- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*, 219-235.
- Arkes, H., Boehm, L., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology, 27*, 576-605.
- Arkes, H., Hacket, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making, 2*, 81-94.
- Bacon, F. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 241-252.
- Begg, I., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*, 446-458.
- Begg, I., & Armour, V. (1991). Repetition and the ring of truth: Biasing comments. *Canadian Journal of Behavioural Science, 23*, 195-213.
- Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioral Science, 17*, 199-214.
- Bhagat, R., & Hovy, E. (2013). What is a paraphrase? *Cognitive Linguistics, 39*, 463-472.
- Boehm, L. E. (1994). The validity effect: A search for mediating variables. *Personality and Social Psychology Bulletin, 20*, 285-285.

- Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1088-1100.
- Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 550-566.
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). Mix me a list: Context moderates the truth effect and the mere-exposure effect. *Journal of Experimental Social Psychology*, 45, 1117-1122.
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, 14, 238-257.
- Garcia-Marques, T., Silva, R. R., Reber, R., Unkelbach, C. (2015). Hearing a statement now and believing the opposite later. *Journal of Experimental Social Psychology*, 56, 126-129.
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16, 107-112.
- Hawkins, S. A., Hoch, S. J., & Meyers-Levy, J. (2001). Low-involvement learning: Repetition and coherence in familiarity and belief. *Journal of Consumer Psychology*, 11, 1-11.

- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, 30, 513-541.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306-340.
- Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122, 139-154.
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118, 126-135.
- Johar, G.V., & Roggeveen, A. L. (2007). Changing false beliefs from repeated advertising: The role of claim-refutation alignment. *Journal of Consumer Psychology*, 17, 118-127.
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32, 1-24.
- Levy, B. A., Barnes, L., & Martin, L. (1993). Transfer of fluency across repetitions and across texts. *Canadian Journal of Experimental Psychology*, 47, 401-427.
- Levy, B. A., Campsall, J., Browne, J., Cooper, D., Waterhouse, C., & Wilson, C. (1995). Reading fluency: Episodic integration across texts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1169-1185.

- McGlone, M. S., & Tofigbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, 11, 424-428.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- Nadarevic, L., & Erdfelder, E. (2014). Initial judgment task and delay of the final validity-rating task moderate the truth effect. *Consciousness and Cognition*, 23, 74-84.
- Nadarevic, L., & Aßfalg, A. (2016). Unveiling the truth: warnings reduce the repetition-based truth effect. *Psychological Research*, in press.
- Parks, C. M., & Toth, J. P. (2006). Fluency, familiarity, aging, and the illusion of truth. *Aging, Neuropsychology, and Cognition*, 13, 225-253.
- Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 365-377.
- Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review*, 5, 71-78.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, 6, 855-863.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness & Cognition*, 8, 338-342.
- Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology*, 1, 563-581.

Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring “fringe” consciousness:

The subjective experience of perceptual fluency and its objective bases.

Consciousness and Cognition, 13, 47-60.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American*

Psychologist, 45, 1043-1056.

Roediger, H. L., & Blaxton, T. A. (1987). Effects of varying modality, surface features,

and retention interval on priming in word-fragment completion. *Memory &*

cognition, 15, 379-388.

Shapiro, S. (1999). When an ad's influence is beyond our conscious control: Perceptual

and conceptual fluency effects caused by incidental ad exposure. *Journal of*

Consumer Research, 26, 16-36.

Silva, R. R., Garcia-Marques, T., & Mello, J. (2016). The Differential Effects of Fluency

Due to Repetition and Fluency Due to Color Contrast on Judgments of Truth.

Psychological Research, 80, 821-837.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures.

Behavior research methods, instruments, & computers, 31, 137-149.

Tardif, T., & Craik, F. I. M. (1989). Reading a week later: Perceptual and conceptual

factors. *Journal of Memory and Language*, 28, 107–125.

Tulving, E., Schacter, D. L., Stark, H. A. (1982). Priming effects in word-fragment

completion are independent of recognition memory. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 8, 336–342. doi:10.1037/0278-

7393.8.4.336

- Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Current Directions in Psychological Science*, 24, 195-199.
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143, 2020-2045.
- Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1235-1253.
- Wurtz, P., Reber, R., & Zimmermann, T. D. (2008). The feeling of fluent perception: A single experience from multiple asynchronous sources. *Consciousness and Cognition*, 17, 171-184.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, 46, 441-517.

Table 1

Mean (SD) Ratings of Truth to Old and New Statements in Experiment 1, by Type of Repetition and Truth Evaluation Session

	Same-Session			1-Week Delay		
	Verbatim	Paraphrases	Total	Verbatim	Paraphrases	Total
Old	5.14 (1.01)	5.37 (.90)	5.24 (.96)	4.92 (.67)	4.71 (.60)	4.82 (.64)
New	4.10 (.43)	4.28 (.36)	4.18 (.41)	4.33 (.56)	4.29 (.42)	4.32 (.49)
<i>n</i>	26	22	48	29	28	57

Table 2

*Mean (SD) Ratings of Truth in Experiment 2, by Type of Item and Truth Evaluation**Session*

	Verbatim	Paraphrases	Contradictory	Contradictory Paraphrases	New
Same-Session (<i>n</i> = 20)	5.17 (.69)	5.12 (.68)	3.66 (1.02)	3.61 (.62)	3.95 (.54)
1-Week Delay (<i>n</i> = 21)	4.58 (.67)	4.64 (.52)	4.33 (.61)	4.05 (.63)	4.10 (.46)
Total (<i>n</i> = 41)	4.87 (.74)	4.88 (.64)	4.00 (.90)	3.83 (.65)	4.03 (.50)

Table 3.

Mean (SD) Proportions of Hits and False Alarms (FA), and SDT Discrimination estimates (d') in Experiment 2, by Type of Item and Truth Evaluation Session⁴

	Hits Verbatim	FA Paraphrases	FA Contradictory	FA Contradictory Paraphrases	FA New
Same-Session ($n = 20$)	.93 (.11) ---	.89 (.13) $d' = .12$ (.61)	.56 (.25) $d' = 1.17$ (.68)	.59 (.22) $d' = 1.11$ (.65)	.16 (.26) $d' = 2.60$ (.97)
1-week delay ($n = 21$)	.63 (.23) ---	.57 (.22) $d' = .20$ (.58)	.45 (.22) $d' = .53$ (.60)	.49 (.19) $d' = .42$ (.54)	.16 (.15) $d' = 1.55$ (.84)
Total ($n = 41$)	.78 (.23) ---	.72 (.24) $d' = .26$ (.59)	.50 (.24) $d' = .84$ (.71)	.53 (.24) $d' = .76$ (.69)	.16 (.21) $d' = 2.06$ (1.04)

⁴ Discrimination estimates (d') are the difference between the z score of the hit rate (i.e., proportion of “Yes” responses to verbatim repetitions) and the z score of each of the four different FA rates (i.e., proportion of “Yes” responses to paraphrases, contradictory statements, contradictory paraphrases, and new statements). Each d' score represents how good participants were in distinguishing the actually repeated items (verbatim repetitions) from each type of distractor (the higher the value, the higher the discrimination).

Table 4.

Mean (SD) Ratings of Truth of the Perceived to be “Old” Statements in Experiment 2, by Type of Item and Truth Evaluation Session⁵

	Verbatim	Paraphrases	Contradictory	Contradictory Paraphrases
Same-Session (<i>n</i> = 20)	5.13 (.93)	5.26 (.90)	4.60 (1.04)	4.08 (1.27)
1-week delay (<i>n</i> = 20)	4.81 (.92)	4.95 (.82)	4.76 (.83)	4.38 (1.04)
Total (<i>n</i> = 40)	4.97 (.93)	5.11 (.86)	4.68 (.93)	4.23 (1.16)

⁵ Table 4 omits truth ratings of the perceived to be “new” and of (actually) new statements because these estimations are neither reliable nor relevant for the current analysis (see the initial paragraphs in the section “6.2.3 Conditional Analysis of Truth Ratings”).