Social Cognition, Vol. 36, No. 2, 2018, pp. 167–198

THE POWER OF THE TRUTH BIAS: FALSE INFORMATION AFFECTS MEMORY AND JUDGMENT EVEN IN THE ABSENCE OF DISTRACTION

Myrto Pantazi, Mikhail Kissine, and Olivier Klein
Université Libre de Bruxelles

Truth bias is the tendency to believe information whether or not it is true. According to a prominent account, this tendency results from limited cognitive resources. We presented participants true and false statements organized in coherent narratives, and distracted half of the participants while they processed the statements. Our findings suggest that explicitly false statements are misremembered as true and affect participants' judgments regardless of cognitive load (Experiments 1 & 2). Experiment 3 replicates a distraction-independent truth bias in a paradigm with an equal number of true and false statements, suggesting that the truth bias does not depend on the frequency of true versus false statements. Experiment 4 suggests that when the statements are presented in lists, as it often happens in the relevant literature, the truth bias is significantly underscored. Taken together, our results strongly support that the truth bias may be stronger than suggested by previous studies.

Keywords: truth bias, distraction, misinformation, judgment, memory

The *Oxford Dictionaries* proclaimed "post-truth" to be the 2016 word of the year. Two major political events of 2016 have abundantly shown how media and the public space are full of fake news and misinformation. Given the high number of false allegations made during the Brexit referendum campaigns (Andrew, 2017), or the 71% inaccuracy rate of the current U.S. president, Donald Trump (PolitiFact, 2016), one might hope that people routinely monitor information they encounter

We are indebted to Klaus Fiedler, Cameron Brick, George Farmer, our editor and two anonymous reviewers for providing comments and suggestions that substantially improved this article. We are also grateful to Daniel Gilbert and his colleagues for making available the material of their original study, and to Huub van den Bergh for his help with mixed models. Special thanks to Katia Kissine and Eric Breton le Veel for being the speakers in our materials. This research was supported by the Mini-ARC "Project" grant, *At the Sources of Faith*, from the Université libre de Bruxelles.

Correspondence concerning this article should be addressed to Myrto Pantazi, Department of Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB. E-mail: mp852@cam.ac.uk.

before believing it. However, ample social psychological research suggests that people are "truth biased": they tend to accommodate new information as if it was true (Levine & Bond, 2014; Levine, Park, & McCornack, 1999; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Street & Masip, 2015).

The propensity to believe information we receive is evolutionarily efficient in a context where the information is mostly accurate (Kissine & Klein, 2013; Levine, 2014; Levine et al., 1999; Reber & Unkelbach, 2010). However, people tend to believe information even when informed that it is inaccurate. Multiple studies have shown that participants base judgments and conclusions about themselves and others on false information, even in the presence of meta-information explicitly signaling falsity (Anderson, 1983; Anderson, Lepper, & Ross, 1980; Guenther & Alicke, 2008; Schul & Burnstein, 1985; Thorson, 2015). Similarly, participants consistently show a tendency to misclassify more false information as true than true information as false, which is a clear indication of a truth bias (Peter & Koch, 2016).

A prominent account of truth bias in information processing is Gilbert and colleagues' model of statement comprehension and validation (Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993). The central tenet of this model is that comprehending and accepting statements is one and the same process such that people automatically believe any statement they happen to understand. Rejecting a statement as false necessarily occurs at a second stage, after the statement has been believed, and only if sufficient cognitive resources are available. To test this model, Gilbert and colleagues (1990, 1993) conducted a series of studies where they asked participants to read statements that were explicitly tagged as true or false (accompanied by the tags "True" or "False," or presented in different font colors). Crucially, the authors placed half of their participants under extra cognitive load during reading, by asking them to simultaneously accomplish an additional task (a sound or digit-detection task). Their hypothesis was that participants in cognitively taxing conditions would be less capable of proceeding to the second stage of disbelieving false statements, thus remaining at the first stage of believing them. In line with this prediction, distracted participants misremembered more false statements as true and were more influenced by the false statements on subsequent judgments they made.

According to Gilbert and colleagues, thus, cognitive load plays a very central role in the operation of truth bias. This has two important theoretical implications. The first implication concerns the mechanism of the truth bias. If available cognitive resources during information processing determine the extent to which people will eventually (dis)believe encountered information, then the effect will, at least partly, be located at the stage of processing or understanding the information (see Gilbert, 1991).

The second implication relates to the pervasiveness of the truth bias. The setup of Gilbert and colleagues (1990, 1993) stresses the operation of truth bias under high cognitive load, but these authors were less emphatic with regard to the potential operation of the truth bias when cognitive load is low. Actually, the statistical models in many of their studies did not directly test whether undistracted participants were truth biased. Whether people are capable of rejecting incorrect information

if they have sufficient cognitive resources is of primary importance for determining the extent and pervasiveness of the truth bias. This in turn, is crucial for the development of strategies that will seek to shield people against misinformation.

In view of these two implications of Gilbert and colleagues' account, it is crucial that some scholars did not find an effect of cognitive resource availability on the susceptibility to incorrect information (Fiedler, Armbruster, Nickel, Walther, & Asbeck, 1996; Fiedler, Walther, Armbruster, Fay, & Naumann, 1996). In one of these studies, distracted and undistracted participants viewed the interior of a flat and were then presented true and false statements about the depicted scene. In a later recognition test, participants misremembered seeing things contained in the false statements they read, regardless of whether they were distracted (Fiedler, Walther, et al., 1996). In the same vein, Marsh and Fazio (2006) found that participants are susceptible to misinformation contained in fictional stories, even when the stories are fairly easy to process, and hence should not particularly tax cognitive resources.

These studies differ from Gilbert and colleagues (1990, 1993) in that participants could not rely on external meta-information to monitor the truth value of the information they received. In order to accurately reject false information participants had to rely on their own memories of videos they previously saw (Fiedler, Walther, et al., 1996) or on general knowledge stored in their long-term memory (Marsh & Fazio, 2006). So, while these studies challenge the idea that low cognitive demands shield people from false information, it remains unclear whether explicit meta-information, coupled with the absence of high cognitive load, suffices to cancel the truth bias. The main aim of this article is to test whether cognitive load is a necessary condition for truth-bias effects, in contexts where participants have to rely on external meta-information to reject or accept information they encounter (Experiments 1–3).

The present article also addresses a methodological issue. Many studies assessing the truth bias, especially those that oppose its existence, rely on unrealistic situations. Participants are typically presented with a list of true or false unrelated statements, and are then asked to recall the statements' truth value or make judgments related to the statements (Fazio, Brashier, Payne, & Marsh, 2015; Gilbert et al., 1990; Hasson, Simmons, & Todorov, 2005; Nadarevic & Erdfelder, 2013; Richter, Schroeder, & Wöhrmann, 2009). This presentation mode is unlikely to reflect real-life processes, where statements are generally integrated and embedded within wider discourses, and are difficult to validate in a piecemeal fashion. Experiment 4 tested the hypothesis that, when the information participants are presented comes in the form of discrete individual statements, the truth bias is significantly reduced.

STUDY OVERVIEW

In Experiments 1–2, we tested the extent to which truth bias depends on cognitive load by adapting Gilbert and colleagues' paradigm (1993, Experiment 1). In the original study participants were instructed to act as trial judges, and read two

crime reports containing true and false statements, as indicated by their font color (black = true, red = false). False statements aggravated the severity of one crime and attenuated the severity of the other. Half of participants were cognitively loaded while they read the false statements through a digit-search visual task. Participants were then asked to judge the perpetrators and fill in a truth-value memory test. Gilbert and colleagues found that judgments by distracted participants were more influenced by the false aggravating and attenuating statements than judgments by control participants. At the memory level, distracted participants tended to misremember more false statements as true, but did not differ from the undistracted group as for their memory of true statements.

Gilbert and colleagues' (1993) paradigm has the advantage of presenting statements as a coherent narrative. This natural mode of information presentation is more suited to investigate potential real-life operation of a truth bias than the discrete statement presentation often adopted in other studies. Furthermore, in many countries lay citizens may be jury members, a role in which they are required to disregard information based on explicit meta-information. This element renders Gilbert and colleagues' paradigm ecologically valid. In Experiment 1, participants listened to reports similar to those of Gilbert and colleagues while the statements' truth value (true vs. false) was manipulated using two different voices (male vs. female). Information about the speakers' voice is integrated early in the statement interpretation process (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008), which renders this truth-value manipulation more naturalistic than the manipulation used in the original study. Using the judgment and memory measures of the original study, we tested both whether distraction may increase the truth bias, and whether the truth bias emerges among undistracted participants despite the presence of explicit truth-value meta-information. In Experiment 2, we tested the same hypotheses in a visual paradigm almost identical to that used by Gilbert and colleagues. This experiment also tested source memory for the statements rather than memory for their truth value, thus providing an alternative, indirect test of participants' truth-value memory.

Experiment 3 was designed to rule out the possibility that the truth bias detected in Experiments 1 and 2 would be triggered by the relative higher frequency of true compared to false statements, an element inherent in the original paradigm. Finally, Experiment 4 had two objectives. First, it tested our hypothesis that the often-used discrete statement presentation weakens the truth bias. Second, by counterbalancing the true and false statements across the truth-value factor, it additionally tested whether our stimuli were inherently believable and showed that the truth bias in these specific studies is not due to the particularities of our material. Materials, data, and analyses for all studies can be found here: osf.io/g8bjk. In all our studies we report power analyses, all eventual data exclusions (if any), and all manipulations and measures used.

EXPERIMENT 1

METHOD AND MATERIAL

The Reports. We pretested adapted translations of Gilbert and colleagues' (1993) crime reports in French to verify that each of the reports we used contained seven false statements that either aggravated or attenuated the severity of the crime.[1] The reports of two targets, Etienne and Dimitri, were, thus, crossed with version (false statements: aggravating vs. attenuating). The four reports were prerecorded in a neutral tone of voice by two native French speakers, a male, and a female, both having professional acting experience. In each recording, the two speakers took turns, one reading the neutral true statements (the core of the story) and the other the seven false (aggravating or attenuating) statements, in order to form a coherent narrative, with intertwined true and false statements. Participants listened to one of eight report pairs, resulting from the orthogonal combination of report × version × truthful speaker (male vs. female). Each participant listened to the aggravating version of one report and the attenuating version of the other report, while the truthful speaker was constant within, and counterbalanced across participants.

Our audio adaptation of Gilbert and colleagues' (1993) paradigm involved four additional departures from the original paradigm. First, we designed an auditory cognitive-load manipulation, whereby participants had to count the forms of the verb *avoir* (*to have*) appearing in each report. Each report contained 21 *avoir* forms in total, 13 in true statements and 8 in false statements. Second, while Gilbert and colleagues only distracted the reading of false statements, our distraction-task trials occurred simultaneously with the presentation of both false and true statements, rendering memory for the two kinds of statements more comparable. Third, while the original study only included one type of new statements in the memory test, we distinguished between aggravating, attenuating, and neutral new statements, in order to test whether the attenuating or aggravating character of the false statements could per se affect their credibility. Fourth, a limitation of Gilbert and colleagues' studies was that while they assumed cognitive load to affect truth bias, they did not test the correlation between the truth-bias effects and cognitive-load task performance. We measured participants' performance in the cognitive-load task, in order to fully assess this relationship.

DEPENDENT MEASURES[2]

Judgments. For each perpetrator participants were asked to propose a prison term (0–5 years) and to report, on an 11-point scale: a general index of punishment severity (extremely light to extremely heavy), dangerousness (slightly dangerous to extremely dangerous), potentially beneficial contribution of psychological

1. For a detailed description of the pretest and the reports used in the experiment see the online supplementary material (osf.io/g8bjk).

2. In this experiment, we also measured belief in conspiracy theories using Brotherton, French, and Pickering's (2013) scale. The scale was added at the end of the main questionnaires. As conspiracy endorsement was not related to people's tendency to believe false statements and is not central to our main hypotheses, we do not report this measure in detail.

counseling (not at all to extremely), probability of recidivism (not likely at all to extremely likely), and their feelings toward him (total indifference to total aversion).

Memory Test. All participants were presented with the same two lists of 24 statements, one for each report, and were instructed to decide whether each statement was present in the report (old) or not (new) and, if it appeared, whether it was true or false. Twenty-four statements were presented in each list: 4 old true, 4 old false aggravating, 4 old false attenuating, 4 new aggravating, 4 new attenuating, and 4 new neutral statements. For the participants exposed to the aggravated version of each report, the four old attenuating statements of the list counted as attenuating new; conversely, for the participants exposed to the attenuated version of each report, the four old aggravating statements of the list counted as aggravating new (see Table 1 in the online supplementary materials).

Cognitive-Load Task Performance. Performance in the task was measured as the absolute deviation of the number of *avoir* forms detected relative to the actual number. We tested correlations between cognitive-load task performance and the truth-bias judgment and memory measures.

PARTICIPANTS AND PROCEDURE

Gilbert and colleagues' (1993) *group* × *version* critical interaction on participants' prison terms (operationalized as the mean difference of the terms for the aggravated and the attenuated perpetrators between the cognitive-load vs. control groups, $n = 35$) corresponds to an effect size of $f = .68$. To replicate such a within-between interaction given two groups, two measurements and a correlation between repeated measurements of $r = .30$ (estimated from our pretest) with .95 power, 12 participants were needed (GPower; Faul, Erdfelder, Lang, & Buchner, 2007). Gilbert et al. (1993) did not report the necessary information for a power analysis for the cognitive-load effect on memory. GPower suggests 40 participants for an (arbitrarily assumed) medium within-between interaction effect ($f = .25$) in the generalized linear mixed model that we used for memory responses, with an assumed repeated measures correlation of $r = .5$, given two groups, six measurements (corresponding to the 2 × 3 within-subject interaction in our model) and a power of .95.

We tested 82 first year undergraduate psychology students at the Université libre de Bruxelles, compensated with course credits for their participation. We excluded 8 participants from subsequent analyses for one of the following reasons: They had participated in similar studies in the past ($n = 3$), failed to understand or to correctly follow the instructions ($n = 2$), were non-native speakers of French ($n = 1$), or the computer crashed during testing ($n = 2$). Of the remaining 74 participants (37 in the cognitive-load and 37 in the control condition, resulting in a .99 power for both the judgment and memory analyses),[3] 25 were male and 49 female.

3. Our a priori power analyses refer to ANOVAs used by Gilbert et al. (1993). While we preferred to analyze our data using hierarchical mixed modeling, which takes random variance into account, and is less prone to Type I errors compared to traditional ANOVAS (Judd, Westfall, & Kenny, 2012) our results did not change when traditional ANOVAs were used.

Participants were tested in individual computer booths in groups of maximum 8. After signing an informed consent, they were given instructions describing their task. Participants were informed that the one speaker (e.g., male) provided truthful information, while the other, (e.g., female) provided false information, which was taken from other reports that are unrelated to the present cases (see online supplementary materials for verbatim instructions). Participants in the cognitive-load condition were also instructed to count the number of every *avoir* form occurring in each report. All participants then listened first to Dimitri's and then to Etienne's report through headphones. The version order—aggravated-then-attenuated versus attenuated-then-aggravated—was counterbalanced across participants. Finally, participants filled in an online questionnaire, whereby they were asked to judge the two perpetrators and to respond to the memory test. In the memory test, participants were first reminded which speaker provided the truthful and untruthful information, and were then presented with the two 24-statement lists, related to Dimitri's and Etienne's reports. The statement order in each list was randomized and kept constant across participants. Cognitively loaded participants were asked, at the beginning of the questionnaire, to report the number of *avoir* forms in each report.

RESULTS

In all our experiments, outliers for the judgment analyses were excluded based on the median absolute deviation with a constant of 3 (Leys, Ley, Klein, Bernard, & Licata, 2013).[4] As we analyzed memory responses by means of a Generalized Linear Mixed Model for binomial data, we did not exclude outliers by subjects for the memory analyses. For correlational analyses, we used the whole sample. Effect sizes are reported for all pairwise comparisons based on Hedges' formula of Cohen's *d*, in order to avoid overestimates due to the correlated repeated measures of our mixed designs (Dechêne, Stahl, Hansen, & Wänke, 2010). The mixed model used in each analysis was determined based on the design we used, following Judd, Westfall, and Kenny's (2017) recommendations.

Judgments

We treated the five judgments for each perpetrator as separate items (the value of the prison term was multiplied by two to be analogous to the other scales) and analyzed them using mixed-effects modeling, with group (cognitive-load vs. control), and version (attenuated vs. aggravated) as fixed factors. Items, participants, and their interaction were included as random effects. Of the total responses, 3.5% were excluded as outliers. Figure 1 displays the mean judgments per group. Participants judged the perpetrator in the aggravated report more severely than the perpetrator in the attenuated report, $F(1, 93.56) = 26.04$, $p < .001$, $d = .48$, 95% CI

4. In all studies, the exclusion of outliers did not change the conclusions of our analyses.

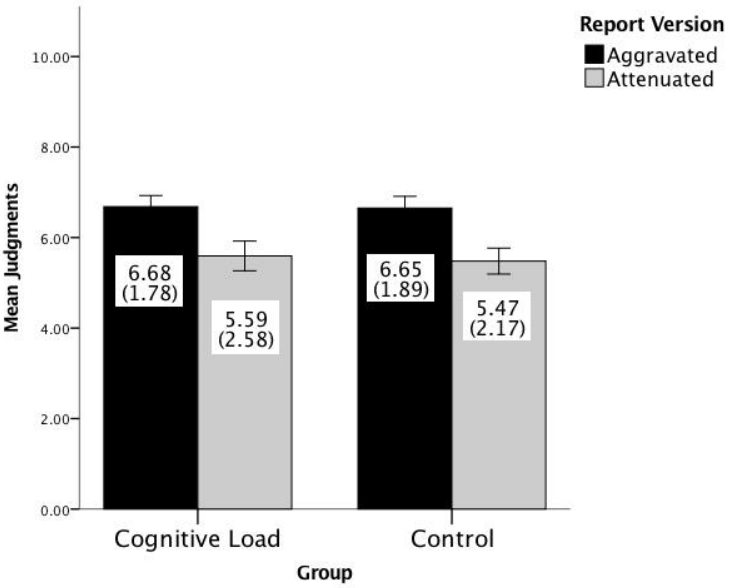

FIGURE 1. Mean judgments per perpetrator for the two groups in Experiment 1. Means and *SD*s are given numerically. Error bars represent 95% CIs.

[.29, .68] for the cognitive-load condition; d = .57 for the control, CI [.38, .77]. We found no main effect of group, $F(1, 74.70)$ = .02, p = .867, d = .02 95% CI [-.16, .10], or a *group* × *version* interaction, $F(1, 46.49)$ = .15, p = .698.

Memory

It is possible that aggravating and/or attenuating statements are inherently more likely to be believed, independently of their accompanying meta-information. We analyzed old and new statements separately to test this possibility.

Old Statements. Participants could either correctly identify each statement as true or false, confound it with the opposite type of statements (as false or true respectively), or misidentify it as new. We used a Generalized Linear Mixed Model for binomial data adapted to these three possible responses (GENLINMIXED procedure in SPSS; see Quené & van den Bergh, 2008). We treated the three possible responses as repeated measures at three different levels of the fixed factor of *classification* (correct response vs. confounded response vs. *new item* response), on a target binomial variable. The *target* variable was thus repeated three times for each of the 48 items participants answered in the memory test. *Target* was coded *1* at the level of the *classification* factor that represented participants' actual response, and *0* at the other two levels. The target variable was thus binominal, and we ran our analysis after it was logit transformed (LOGIT link Target_Option in SPSS; see Quené and van den Bergh, 2008). Besides *classification*, we included two additional fixed factors: *statement* reflected whether each statement was presented as true or false in the reports; and *group* (cognitive-load vs. control). All two- and three-way

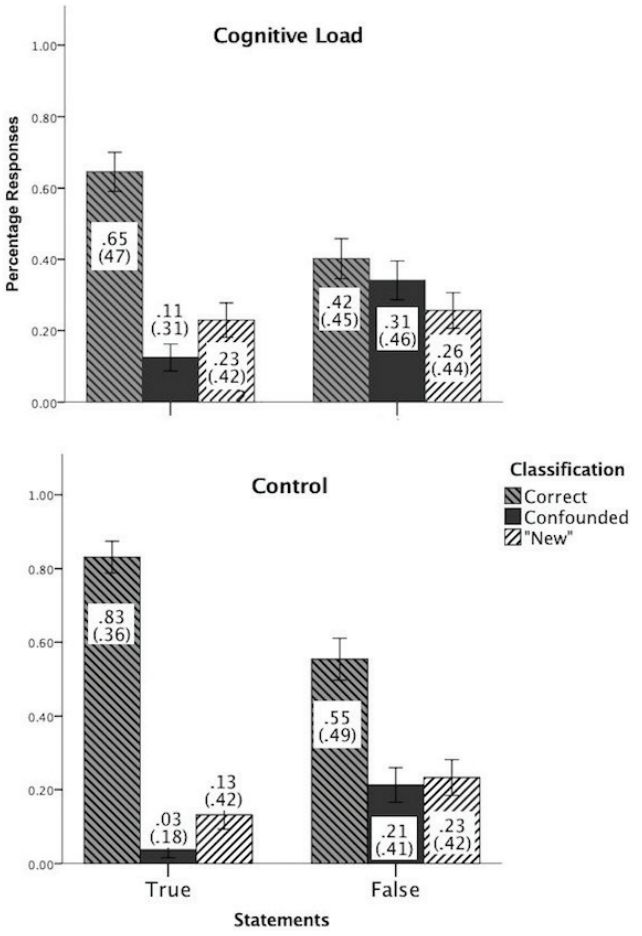

FIGURE 2. Response percentages for the old statements per statements and classification for the cognitive-load and control group in Experiment 1. Mean percentages and *SD*s are given numerically. Error bars represent 95% CIs.

interactions of the three fixed factors were tested. This model, thus, enabled us to test both between-groups differences in terms of memory performance, as well as within-subject differences in the memory for true versus false statements. In the random part, we included intercepts for participants and items.

Figure 2 displays the mean *classification* percentage per *statements*, separately for each group. There was a significant main effect of *classification*, $F(2, 3540) = 268.94$, $p < .001$. Old statements were more likely to be correctly classified than confounded, $t(3540) = 24.31$, $p < .001$; $d = .99$, 95% CI [.90, 1.07], or misclassified as new, $t(3540) = 21.43$, $p < .001$; $d = .88$, 95% CI [.79, .96]. Additionally, there was a *classification* × *statements* interaction, $F(2, 3540) = 77.13$, $p < .001$: Overall, true statements were more correctly classified than false statements, $t(3540) = 9.76$, $p < .001$, $d = .55$, 95% CI [.44, .67]; and false statements were significantly more confounded, $t(3540) = 8.72$, $p < .001$, $d = .52$, 95% CI [.40, .63]. Moreover, there were significant interac-

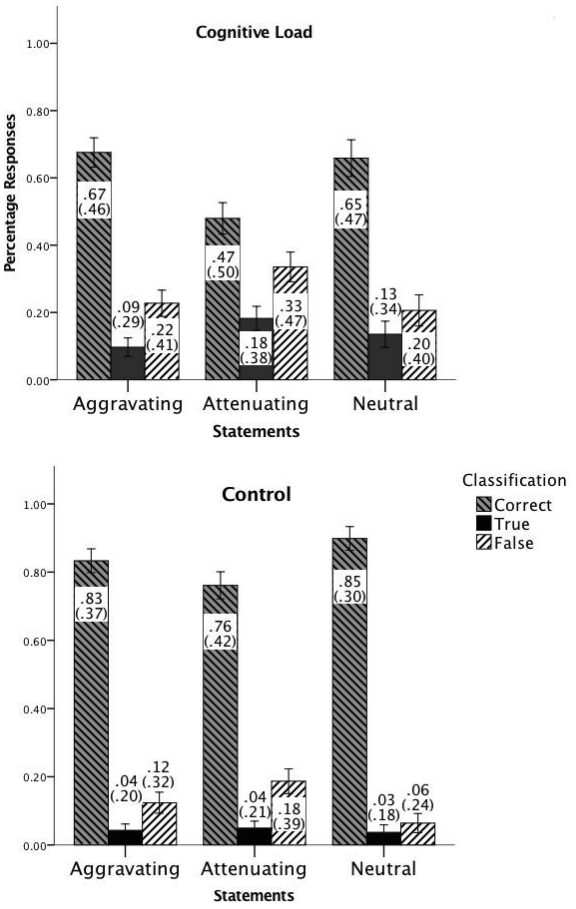

FIGURE 3. Response percentages for the new statements per statements and classification for the cognitive-load and control group in Experiment 1. Mean percentages and *SD*s are given numerically. Error bars represent 95% CIs.

tions between *classification* × *group*, $F(2, 3540) = 33.00$, $p < .001$, and *classification* × *statements* × *group*, $F(2, 3540) = 3.13$, $p = .044$. Compared to controls, cognitively loaded participants classified fewer statements correctly, $t(3540) = -6.43$, $p < .001$, $d = .35$, 95% CI [.24, .47], confounded one truth value with another, $t(3540) = 4.63$, $p < .001$, $d = .29$, 95% CI [.17, .40], and misclassified old statements as new, $t(3540) = 2.78$, $p = .005$, $d = .15$, 95% CI [.03, .26], more often. Importantly, the three-way interaction signaled that despite these differences between the *cognitive-load* and *control* group, the former did not seem to classify more false statements as new, $t(3540 = .67$, $p = .501$, $d = .04$, 95% CI [-.11, .20].

New Statements. We ran a similar model on new statements with *statements* (aggravating vs. attenuating vs. neutral), *classification* (correctly classified vs. misclassified as "true" vs. misclassified as "false"), and *group* (see Figure 3). *Classification* had a significant effect, $F(2, 7086) = 853.90$, $p < .001$. New statements were more correctly than incorrectly classified, $t(7086) = 54.64$, $p < .001$, $d = 1.64$, 95% CI [1.72,

1.58] for correctly versus misclassified as "true"; $t(7086) = 41.96$, $p < .001$, $d = 1.23$, 95% CI [1.17, 1.29] for correctly versus misclassified as "false." New statements were also more likely to be misclassified as false than as true, $t(7086) = 8.31$, $p < .001$, $d = .29$, 95% CI [.23, .35]. A *classification* × *statements* interaction, $F(4, 7086) = 24.72$, $p < .001$, signaled that the aggravating and neutral new statements were classified more correctly than the attenuating ones, $t(7086) = 5.86$, $p < .001$, $d = .28$, 95% CI [.19, .37]; $t(7086) = 6.97$, $p < .001$, $d = .33$, 95% CI [.22, .43], respectively, which tended more to be classified as false, $t(7086) = 4.27$, $p < .001$, $d = .22$, 95% CI [.12, .31] for the comparison with the aggravated; $t(7086) = 6.27$, $p < .001$, $d = .32$, 95% CI [.22, .43] for the comparison with the neutral. Besides, the aggravating were more misclassified as false than the neutral, $t(7086) = 2.38$, $p = .017$, $d = .11$, 95% CI [.006, .21], and less as true than the attenuating, $t(7086) = -2.26$, $p = .024$, $d = .17$, 95% CI [.08, .26]. Moreover, there was a main effect of *group*, $F(1, 7086) = 9.64$, $p = .002$, and a *classification* × *group* interaction, $F(2, 7086) = 114.55$, $p < .001$. Relative to control participants, the cognitive-load group classified the new statements less accurately, $t(7086) = -12.32$, $p < .001$, $d = .52$, 95% CI [.44, .60], and misclassified more new statements as true, $t(7086) = 6.53$, $p < .001$, $d = .32$, 95% CI [.24, .40], and as false, $t(7086) = 7.98$, $p < .001$, $d = .33$, 95% CI [.25, .41]. Finally a *classification* × *statements* × *group* interaction, $F(4, 7086) = 2.78$, $p = .025$, was due to the fact that, while both the cognitive-load and the control groups misclassified more attenuating than neutral statements as false, $t(7086) = 3.99$, $p < .001$, $d = .29$, 95% CI [.14, .44] for the cognitive load group; $t(7086) = 4.87$, $p < .001$, $d = .35$, 95% CI [.20, .50] for the control group, the control group additionally misclassified more aggravating than neutral statements as false, $t(7086) = 2.46$, $p = .014$, $d = .20$, 95% CI [.05, .35] for the control; $t(7086) = .70$, $p = .483$, $d = .04$, 95% CI [.09, .19] for the cognitive-load group.

Judgments–Memory Relationship. We tested the correlation between the difference of the judgments for the two perpetrators and the percentage of false aggravating and attenuating statements participants misremembered as true. As expected, the more participants misremembered false statements as true the larger the impact of the false statements on their judgments was, $r(74) = .36$, $p = .001$.

Cognitive Load. Cognitively loaded participants detected over 70% of the *avoir* forms ($M = 29.1$, $SD = 9.4$). Cognitive load task performance did not correlate with the difference between the judgments for the aggravated and attenuated perpetrator, $r(37) = -.18$, $p = .138$, nor with the percentage of false statements misremembered as true, $r(37) = -.15$, $p = .173$.

DISCUSSION

Unlike in Gilbert and colleagues' (1993) seminal study, distraction did not increase the impact that false statements had on participants' judgments. Additionally, although the cognitive-load task *did* tax participants' cognitive resources, impairing their identification of *all* kinds of statements—true, false, and new—it did not selectively affect their tendency to misclassify false statements as true. Thus, in our study the effects of cognitive load on participants' memory were more generalized, and not restricted to memory for the false statements.

Crucially, both groups confounded more false statements with true than true statements with false, and, in both groups, participants' judgments were influenced by false statements. These effects indicate the operation of a strong truth bias, for which cognitive load is not a necessary condition. Lack of significant correlations between performance in the cognitive-load task and memory and judgment measures corroborates the idea that the truth bias is independent of cognitive load. At the same time, as in Gilbert and colleagues (1993), memory and judgments correlated, suggesting that they constitute complementary and reliable indexes of the truth bias.

Although these results seem quite discrepant from those of the original study, in one of the experiments that Gilbert and colleagues ran (1993; Experiment 2), cognitive load increased not only participants' tendency to misremember false statements as true, but also their tendency to misremember them as new. Additionally, in those seminal studies there was also evidence that cognitive load increased participants' tendency to mistake new items for old true ones (Experiment 1), and true statements for new ones (Experiment 2). Thus, the tendency for cognitive load to more globally affect participants' performance in the memory test is not a completely novel finding.

It is worth emphasizing that our results were not driven by a general tendency to classify any item as true, as our participants misclassified more new statements as false than as true. This finding likely reflects the fact that in such memory tests participants partly judge the statements' truth value driven by the statements' familiarity or processing fluency (Dechêne et al., 2010). Under such a perspective, familiarity or fluency give an impression of truth, and as new statements are less familiar or fluent compared to both the true and the false old statements, participants may be biased toward judging new statements as false. Finally, the responses on the different types of new statements suggested that the aggravating and attenuating statements were more likely to be misclassified as false than the neutral ones, which indicates that the non-neutral content of the false statements did not render them especially believable. This last finding suggests that participants did not tend to judge the false statements as true merely because of their aggravating or attenuating content, but rather due to their high fluency or familiarity emerging from their past encounter.

The results of Experiment 1, thus complement prior studies and attest to the existence of a strong truth bias that is independent of cognitive load (Fiedler, Walther, et al., 1996; Marsh & Fazio, 2006). While those studies suggested that cognitive load does not affect the truth bias when participants rely on their own knowledge to reject inaccurate information, Experiment 1 showed that the same holds when participants have to rely on explicit truth-value meta-information. Since our paradigm differed in several aspects from that of Gilbert et al. (1993), at this point it is unclear whether the same results can be replicated using the original visual paradigm. Experiment 2 tackles this issue.

EXPERIMENT 2

METHOD AND MATERIAL

The design and materials were identical to those used in Experiment 1, but now the procedure was very similar to Gilbert et al.'s (1993, Experiment 1), with true and false statements presented to the participants visually. We created one video for each *report* × *version* combination whereby the text of the reports crawled at a rate of 16 characters/second. The videos were presented using *E-prime* (2.0) and participants read aloud the two crime reports, while the text crawled on screen. In addition, to make this study as similar as possible to the original study, only the five judgments employed by Gilbert and colleagues were used (proposed prison term, dangerousness, potentially beneficial contribution of psychological counseling, probability of recidivism, and feelings toward perpetrators), while participants could now propose prison terms between 0–20 years.

Experiment 2 departed from the original study in three respects. First, while Gilbert and colleagues (1993) presented true statements in black and false in red font, we used green and red fonts respectively. *Green* and *red* are generally associated with concepts of truthfulness versus falsity, and additionally, this modification ensured that both kinds of statements were presented in non-default colors.

Second, in order to measure the correlation between the truth-bias indexes and cognitive-load task performance, we slightly modified Gilbert and colleagues' (1993) cognitive-load manipulation, based on a task by Lavie (2006). In the original study, participants under cognitive load had to press a button every time the digit "5" appeared on the screen, among a stream of digits crawling below the crawling report-text line. In our experiments, the digits now appeared in groups of 6 digits displayed in two rows (3 on the top 3 on the bottom) of an imaginary rectangular positioned below each false statement. Groups of digits appeared along with the first letter of a false statement and kept appearing every 2.500 ms until the last letter of a false statement disappeared. Participants had to press *x* every time the digit 5 was among the group of digits on the display and *n* otherwise. A pretest had shown that the two letters were similarly related to positive and negative responses in the task.

Third, the memory test now assessed participants' memory of the truth-value meta-information, that is, statements' color, rather than explicit truth value. Fragale and Heath (2004) have shown that people attribute information they believe to truthful sources. Accordingly, in this context participants are expected to assign the statements they believe the truthful tag (green color) and those they disbelieve the untruthful tag (red color). Color memory thus allowed for an indirect measure of truth-value memory that could be immune to potential acquiescence effects and demand characteristics in the response phase.

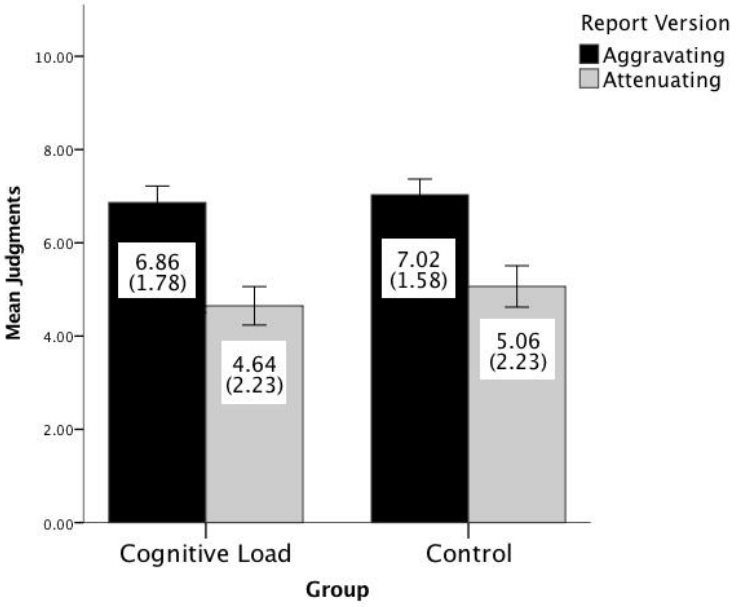

FIGURE 4. Mean judgments per perpetrator for the two groups in Experiment 2. Means and *SD*s are given numerically. Error bars represent 95% CIs.

PARTICIPANTS AND PROCEDURE

The procedure was similar to that in Experiment 1 except that participants were tested individually and read Etienne's (and not Dimitri's) report first. Following Gilbert et al. (1993) participants were asked to read the reports aloud. We tested 63 students at the Université libre de Bruxelles, rewarded with course credits. Twenty participants were excluded from the analyses because they were dyslexic (n = 3), non-native French speakers (n = 4), suspicious about the design (n = 1), incapable of reading the text aloud (n = 3) or of following the instructions (n = 5). Finally, 4 participants were excluded because *E-prime* crashed during the presentation phase. Data from 43 participants were finally analyzed (6 male, 37 female; 23 in the cognitive-load condition, 20 control).

RESULTS

Judgments

Of the total responses, 6.5% were excluded as outliers. We ran the same model as in Experiment 1, but the prison term measure was now divided by two to be comparable to the other measures. As shown in Figure 4, participants judged the perpetrator in the aggravated report more severely compared to the one in the

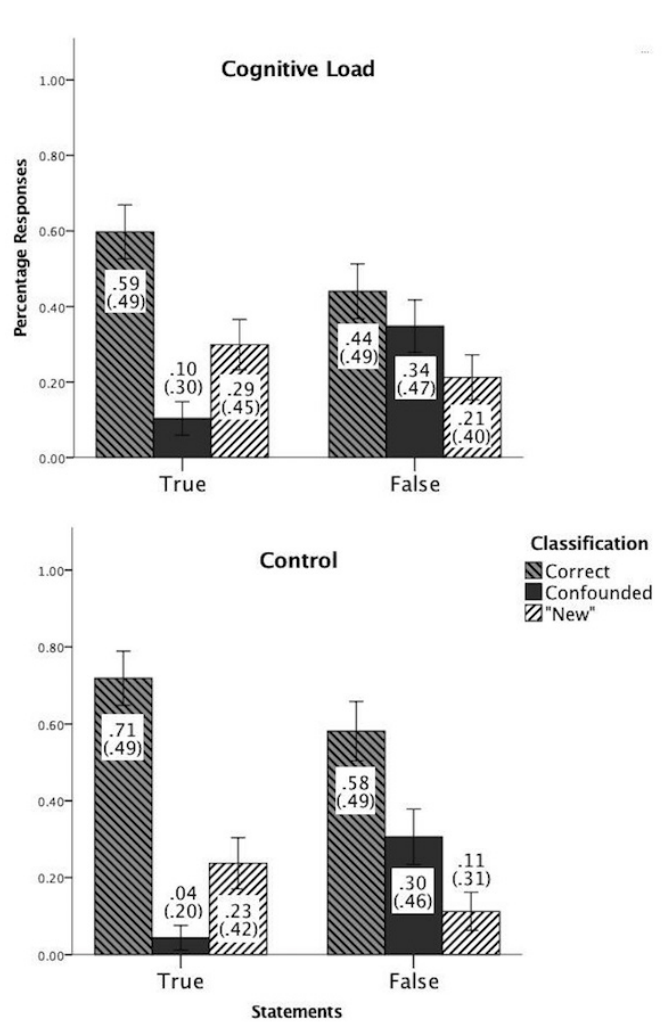

FIGURE 5. Response percentages for the old statements per statements and classification for the cognitive-load and control group in Experiment 2. Mean percentages and *SD*s are given numerically. Error bars represent 95% CIs.

attenuated report, $F(1, 16893) = 41.49$, $p < .001$, $d = 1.35$, 95% CI [1.06, 1.65] for the cognitively loaded; $d = 1.00$, 95% CI [.70, 1.31] for the control. As in Experiment 1, there was no effect of *distraction* or a *version* × *group* interaction (Fs < 1).

Memory

Old Statements. Mean *classification* (green vs. red) percentage as a function of *statements* (green vs. red) for the two groups is displayed in Figure 5. The models used both for old and for new statements were the same as in Experiment 1. There was a significant *classification* effect, $F(2, 2052) = 135.61$, $p < .001$. Statements were more correctly classified than confounded, $t(2052) = 16.47$, $p < .001$, $d = .82$, 95% CI [.71, .93], or misclassified as new, $t(2052) = 15.15$, $p < .001$, $d = .79$, 95% CI [.68, .90]. A significant *classification* × *statements* interaction, $F(2, 2052) = 39.89$, $p < .001$,

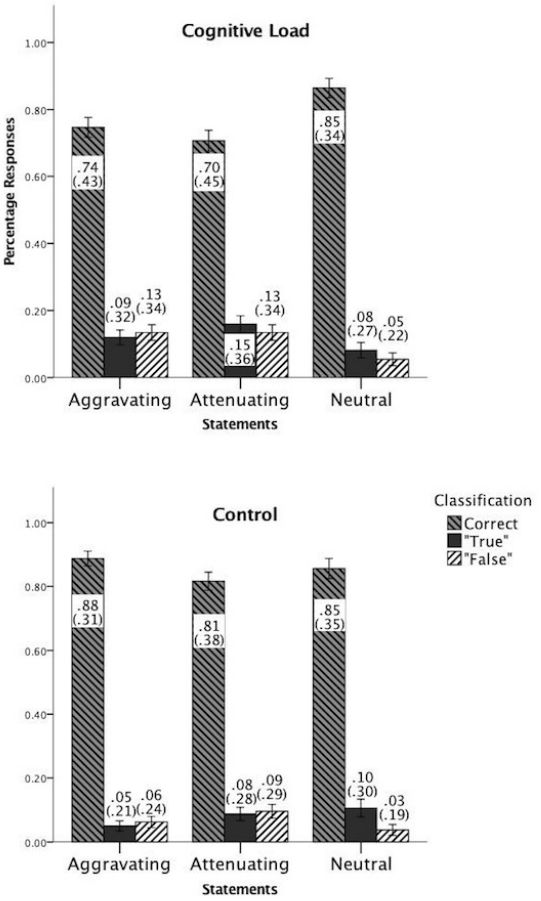

FIGURE 6. Response percentages for the new statements per statements and classification for the cognitive-load and control group in Experiment 2. Mean percentages and *SD*s are given numerically. Error bars represent 95% CIs.

signaled that red statements were less correctly classified than green statements, $t(2052) = -3.97$, $p < .001$, $d = .30$, 95% CI [.15, .45], as they were more misclassified as green than green statements as red, $t(2052) = 8.50$, $p < .001$, $d = .65$, 95% CI [.50, .81]. Additionally, green statements were significantly more misclassified as new than red statements, $t(2052) = 3.52$, $p < .001$, $d = .27$, 95% CI [.12, .42]. Finally, there was a *classification* × *group* interaction, $F(2, 2052) = 11.35$, $p < .001$. Cognitively loaded participants classified fewer statements accurately, $t(2052) = -3.53$, $p < .001$, $d = .28$, 95% CI [.13, .44], and erroneously mistook more old statements for new ones, $t(2052) = 2.75$, $p = .006$, $d = .19$, 95% CI [.04, .35]. The *group* × *identification type* × *statements* interaction was nonsignificant, $F(2, 2052) = 1.14$, $p = .318$.

New Statements. Mean *classification* (correctly classified vs. misclassified as green vs. misclassified as red) per *statements* (aggravating vs. attenuating vs. neutral) for new items is displayed in Figure 6. The *classification* effect was significant, $F(2, 12366) = 2048.45$, $p < .001$. New items were more correctly than incorrectly classified, $t(12366) = 89.68$, $p < .001$, $d = 2.01$, 95% CI [1.95, 2.06] for misclassified *green*;

$t(12366) = 92.24$, $p < .001$, $d = 2.06$ for misclassified *red*, 95% CI [2.01, 2.11]. The *classification* × *statements* interaction was also significant, $F(2, 12366) = 17.83$, $p < .001$: neutral statements were more correctly classified than the aggravating, $t(12366) = 2.11$, $p = .026$, $d = .13$, 95% CI [.05, .21], and attenuating ones, $t(12366) = 6.11$, $p < .001$, $d = .28$, 95% CI [.20, .36], while the aggravating were more correctly classified than the attenuating ones, $t(12366) = 4.09$, $p < .001$, $d = .14$, 95% CI [.07, .21]. The attenuating statements on the other hand, were more misclassified as green than the aggravating, $t(12366) = 3.13$, $p = .002$, $d = .13$, 95% CI [.06, .20], or neutral ones, $t(12366) = 1.97$, $p = .048$, $d = .09$, 95% CI [.02, .16]. Finally, the neutral statements were less misclassified as red compared to both the aggravating, $t(12366) = -3.78$, $p < .001$, $d = .22$, 95% CI [.14, .30], and the attenuating ones, $t(12366) = -5.29$, $p < .001$, $d = .21$, 95% CI [.13, .29]. Besides, there were significant interactions between *classification* × *group* and *classification* × *statements* × *group*, $F(12366) = 13.35$, $p < .001$. Compared to control individuals, cognitively loaded individuals misclassified less aggravating, $t(12366) = -7.32$, $p < .001$, $d = .34$, 95% CI [.20, .48], and attenuating items correctly, $t(12366) = -5.17$, $p < .001$, $d = .25$, 95% CI [.11, .39], as they misclassified them more as green, $t(12366) = 4.19$, $p < .001$, $d = .21$, 95% CI [.11, .31] for the aggravating; $t(12366) = 4.12$, $p < .001$, $d = .21$, 95% CI [.11, .31] for the attenuating, and as red, $t(12366) = 4.25$, $p < .001$, $d = .24$, 95% CI [.13, .33] for the aggravating; $t(12366) = 2.22$, $p = .026$, $d = .16$, 95% CI [.02, .22] for the attenuating.

Judgments–Memory Relationship. The difference in the judgments of the two perpetrators correlated with the number of red statements participants misclassified red as green, $r(43) = .30$, $p = .026$.

Cognitive Load. There were 78 cognitive-load trials in each report (38 in each). Accurate responses corresponded to pressing *x* in the presence of *5* and *n* in its absence. Mean accuracy was $M = 14.61$ (18% of the trials, $SD = 7.94$) and mean *RT* was $M = 1007.91$ ms ($SD = 171.96$). As in Experiment 1, these measures did not correlate with the differences in the judgments of the perpetrators in the two versions, $r(23) = .04$, $p = .412$ for accuracy; $r(23) = -.02$, $p = .466$ for RTs, or with the amount of red statements misremembered as green, $r(23) = .06$, $p = .379$ for the accuracy; $r(23) = .02$, $p = .450$ for the RTs.

DISCUSSION

In line with Experiment 1, the judgment measures of Experiment 2 replicated a distraction-independent truth bias, in a paradigm almost identical as that of Gilbert et al. (1993). As for the memory task, we obtained analogous results to Experiment 1 even if truth-value memory was now substituted for with truth-value meta-memory. This finding provides additional evidence for a genuine truth bias at the memory level that is not driven by task demands or social norms (see Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014).

Similar to what was found in Experiment 1, relative to new neutral statements, new aggravating and new attenuating statements were misremembered more as being red. This may indicate that participants somehow realized the aggravating

and attenuating nature of the false statements. This realization may, in turn, have made them classify aggravating and attenuating statements as false. Even if such a tendency existed, however, it was not enough to counter the truth-bias effect for old aggravating and attenuating statements. Participants' responses in the memory test may, thus, have been driven by multiple processes, that to some extent opposed each other. It is possible, for example, that participants, overall, tended to classify aggravating and attenuating statements as false, while the familiarity or processing fluency of old statements may have finally overcome this tendency, leading to the truth-bias effect (Dechêne et al., 2010; Jacoby, 1991).

In sum, Experiments 1 and 2 robustly demonstrated a truth bias, that emerged regardless of participants' cognitive load and was independent of distraction. The difference between our results and those by Gilbert et al. (1993) could be due to the slightly different stimuli (different language, different color for the true statements) or the slightly different distraction tasks we used. Note, however, that our material and paradigm, especially in the second experiment, were very close to those in their original study. Another possibility is that given the pervasive use of media and video games in the young generations today, our participants, 30 years later, may be much better at multitasking compared to Gilbert et al.'s participants. Thus, despite the overall cognitive-load effects on memory accuracy, the specific effects of cognitive load on the truth bias may have been overshadowed in our studies.

In any event, the most important implication of our results for the misinformation effects in real life is not the absence of a cognitive-load effect on the truth bias, but rather the strength of the truth bias *even* in the absence of a cognitive load. Our findings show that participants misremembered statements they *knew* to be false as true, in a context where they were expected to be vigilant. This finding suggests that the truth bias may actually be stronger than what is often assumed in the literature, and is very likely to routinely operate in many not so cognitively demanding everyday situations.

There is, nevertheless, an alternative explanation of our results. Just as in Gilbert and colleagues (1993), in our reports, the number of true statements outnumbered false ones. Arguably, this is consistent with people's expectations of real-life situations, where communicated information is anticipated to be predominantly accurate (Kissine & Klein, 2013; Levine, 2014; Nadarevic & Erdfelder, 2013; Reber & Unkelbach, 2010). Yet, it remains possible that in our experimental paradigm the imbalance between true and false statements may have enhanced participants' tendency to consider old statements as true. The next experiment is designed to rule out this possibility.

EXPERIMENT 3

DESIGN AND MATERIAL

The same speakers as in Experiment 1 audio-recorded eight new versions of the reports used in Experiment 1, now with an equal number of true and false state-

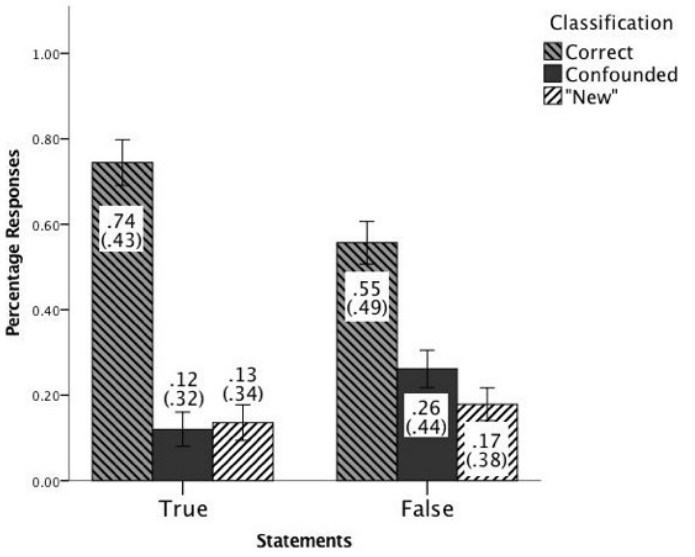

FIGURE 7. Response percentages for the old statements per statements and classification in Experiment 3. Mean percentages and *SD*s are given numerically. Error bars represent 95% CIs.

ments. The untruthful speaker uttered the false statements of the original reports, plus other statements that were either in accordance with the (aggravating or attenuating) report version or perceived as neutral according to our pretest. For reasons of coherence it was impossible to fully counterbalance the word number across true and false statements, yet, these differences, are too small to possibly bias participants (see Table 2 in the online supplementary materials).

All the measures were identical to Experiment 1, except that we excluded the *usefulness of psychological counseling* from the judgments, as we realized that perpetrators may profit from counseling regardless of their crime.

PROCEDURE AND PARTICIPANTS

The smaller effect in our previous studies was that on participants' memory. The average memory effect in the first two experiments was $d = .58$. To obtain such an effect in a within-subjects F test with the correlation between false statements misremembered as true and true statements misremembered as true being $r = .237$ with a power of .9 and an alpha level of .05 Gpower suggested 49 participants. Participants were 48 first-year psychology students at the Université libre de Bruxelles who participated in the study in exchange for course credits. The procedure was the same as in Experiment 1, except for the fact that there was only one, nondistracted group of participants. Based on the same exclusion criteria as above, 5 participants were excluded from the analyses, leaving 43 participants (36 female, 6 male, 1 unknown).

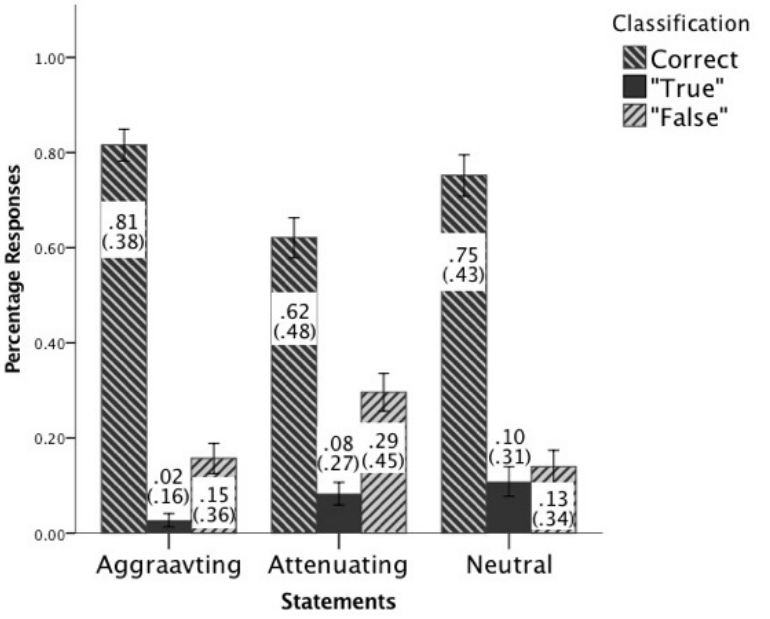

FIGURE 8. Response percentages for the new statements per statements and classification in Experiment 3. Mean percentages and *SD*s are given numerically. Error bars represent 95% CIs.

RESULTS

We used a mixed model with *version* as fixed factor, and intercepts and slopes of participants and items as well as intercepts of participants per item as random factors.

Judgments

According to Leys et al.'s method (2013) there were no outliers. As before, participants' judgments were more severe for the aggravated (M = 5.74, SD = 2.43) than for the attenuated perpetrator (M = 4.69, SD = 2.53), $F(1, 180.04) = 5.22$, $p = .023$, $d = .42$ 95% CI [.23, .61].

Memory

Old Statements. We used a Generalized Linear Mixed Model for binomial data with *classification*, *statements*, and their interaction as fixed factors. Figure 7 displays mean *classification* percentage per *statements*. The *classification* effect was significant, $F(2, 1926) = 98.93$, $p < .001$. Again, old statements were more correctly classified than confounded, $t(1926) = 13.80$ $p < .001$, $d = .97$, 95% CI [.85, 1.08], or mistaken for new, $t(1926) = 14.91$, $p < .001$, $d = 1.10$, 95% CI [.99, 1.22]. Additionally, there was a *classification* × *statements* interaction, $F(2, 1926) = 11.12$, $p < .001$. False statements were classified significantly less accurately, $t(1926) = -3.50$, $p < .001$, $d =$

.40, 95% CI [.24, .56], and were more likely to be confounded than true statements, $t(1926) = 3.14$, $p = .002$, $d = .35$, 95% CI [.19, .51].

New Statements. A similar model as for the old statements was used for the analysis of new statements. Figure 8 displays mean *classification* percentage per *statements*. A main effect of classification, $F(2, 4251) = 319.50$, $p < .001$, revealed that new statements were more correctly classified than misclassified as true, $t(4251) = 35.39$, $p < .001$, $d = 1.84$, 95% CI [1.75, 1.93], or as false, $t(4251) = 24.57$, $p < .001$, $d = 1.31$, 95% CI [1.15, 1.31]. Besides, once more, the new statements tended more to be misclassified as false than as true, $t(4251) = 7.07$, $p < .001$, $d = .46$, 95% CI [.34, .59]. A classification × statements interaction, $F(2, 4252) = 15.36$, $p < .001$, signaled that the attenuating statements were less correctly classified than both the aggravating, $t(4252) = -4.76$, $p < .001$, $d = .43$, 95% CI [.31, .56], and the neutral statements, $t(4252) = -2.83$, $p < .001$, $d = .28$, 95% CI [.15, .41]. This reflected that the attenuating statements tended more to be classified as false, $t(4252) = 3.74$, $p < .001$, $d = .34$, 95% CI [.22, .46] for the comparison with the aggravated; $t(4252) = 4.12$, $p < .001$, $d = .39$, 95% CI [.52, .26] for the comparison with the neutral. The aggravating statements on the other hand were less misclassified as true compared to both the attenuating, $t(4252) = -3.13$, $p = .002$, $d = -.27$, 95% CI [-.14, -.39], and the neutral statements, $t(4252) = -3.83$, $p < .001$, $d = -.33$, 95% CI [-.20, -.47].

Judgments–Memory Relationship. Again, participants' judgments correlated with their tendency to misclassify false statements as true, $r(43) = .42$, $p = .003$.

DISCUSSION

The truth bias effect found in Experiments 1 and 2 was replicated in an orthogonal paradigm, with an equal number of true and false statements. This study, therefore, invalidates alternative explanations for our findings in these experiments. Given the results of Experiment 3, it is unlikely that participants tended to believe false statements to be true based on an implicit learning mechanism or a response strategy, triggered by a higher frequency of true statements in our previous paradigms.

As regards the memory pattern on new statements, the present experiment largely replicates the finding of the previous ones. Once more, new statements tended more to be misclassified as false than as true. This finding provides additional evidence for the potential impact of processing fluency or familiarity on participants' responses in the memory test. On the other hand, in this experiment only the attenuating new statements, but not the aggravating ones, tended more to be misclassified as false compared to the neutral ones. At the same time, the aggravating statements were less misclassified as true than the neutral ones. These two findings combined, suggest once more that participants may have been somehow sensitive to the link in our material between aggravating and attenuating information on the one hand, and falsity on the other hand. Future studies could specifically test how judgments of truth for unlearned information may be affected by the known truth or falsity of previously learned information.

Taken together, Experiments 1–3 support the existence of a strong truth bias that occurs even in the absence of cognitive load and operates when statements are ex-

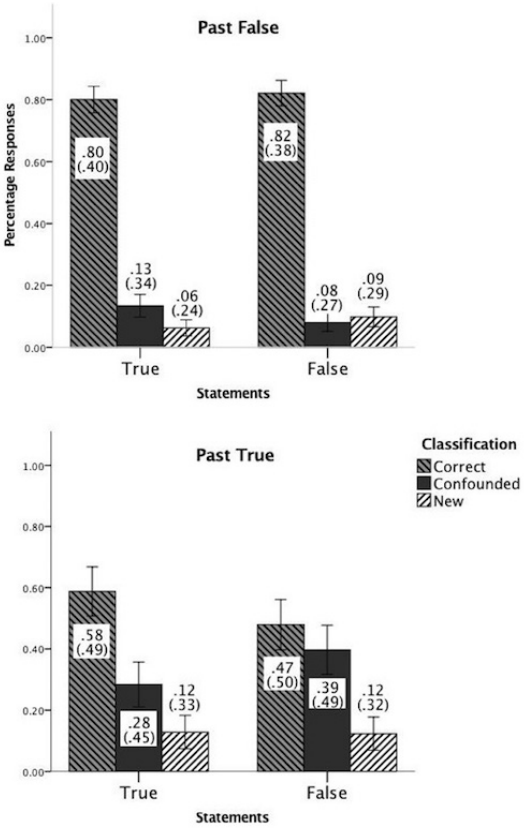

FIGURE 9. Response percentages for the old statements per statements and classification for the past false and past true statements in Experiment 4. Mean percentages and *SD*s are given numerically. Error bars represent 95% CIs.

plicitly tagged as false. These studies, thus undermine the hypothesis that cognitive load or distraction constitute *necessary* conditions for the operation of the truth bias. We leave the relevant implications to the general discussion. Before that, we report a final study, designed to determine whether the truth bias critically depends on the way the true and false information is presented to participants.

EXPERIMENT 4

DESIGN AND MATERIAL

The primary goal of Experiment 4 was to test whether the truth bias is sensitive to the discrete statement presentation mode. Along the way, this experiment also tested whether some random characteristics, inherent in the false statements we used rendered them particularly believable. To this end, while in Experiments 1–3 each of the old statements was presented as either true or as false, in this experi-

ment all statements appeared both as true and as false across participants. Assessing error rates for each statement as a function of its truth value would now reveal any such random characteristics.

We created a short summary of the reports used in Experiment 1, based on a subset of the true statements of these reports. Each summary was accompanied by a list of the 12 old (the four true and eight false, four aggravating and four attenuating) statements included in the memory test of Experiment 1. As in many studies assessing truth-value judgments, these statements now appeared individually and sequentially, rather than in the form of a narrative. Crucially, two lists were created for each report, and each of the statements was presented as true in the one list and false in the other list, while in each list half of the statements were presented as true and half as false. The orthogonal combination of the two lists for each report resulted in four final lists. Each participant was randomly assigned to one of these four lists (see Table 3 in the online supplementary materials).

The memory test items were identical to those used in the previous experiments. Each report now included both aggravating and attenuating information, so that no judgment differences between report should be expected. However, we still included the prison term measure (0–10 years) and the general index of punishment severity (0–10 scale) to double-check that the reports of Etienne and Dimitri were equivalent.

PARTICIPANTS AND PROCEDURE

The same power analysis as in Experiment 3 applies for the detection of a memory-based truth bias. Participants (N = 43; 34 female, 9 male) were recruited and tested as in the previous experiments. They received similar instructions, but now they were informed that they would read two short crime reports and several additional true and false statements concerning each of them. We used *E-Prime* (2.0) for the presentation of the report summaries and the accompanying statements. Participants first read the two summaries presented on the upper part of a computer screen, at their own reading pace. Next, they launched the presentation of the 12 additional true and false statements by pressing the spacebar. Statements appeared individually, below the summary in a random order, accompanied by their truth-value tag (*true* or *false*), which was printed below them. In order to eliminate variation due to differences in reading styles or strategies, the duration of the presentation for each statement varied according to its length, and corresponded to a relatively slow reading rate of 16 characters/second (see Just, Carpenter, & Woolley, 1982). Participants first read the report and statements related to Dimitri's crime and then those related to Etienne's crime. Once they had finished reading the material, the two judgment questions appeared on the screen, followed by the memory task, in which statements appeared on the screen one at a time, in a random order, until participants gave their response (*true*, *false*, or new).

RESULTS

We used a mixed model to analyze participants' memory for the old statements, including the fixed factors *classification* and *statements* as in Experiment 1, and a third fixed factor, *past presentation*, that reflected each statements' truth value in the previous experiments. Slopes of participants and items were included for each level of the classification factor.

Memory. The first hypothesis was whether the statements now presented as false would tend to be classified as true, regardless of their truth value in the previous experiments (*classification* × *statements* interaction). Second, we were interested in whether the past false statements generally tended to be misclassified compared to the past true statements (*classification* × *past presentation* interaction). Mean classification percentage per statements is displayed in Figure 9. There was a *classification* effect, $F(2, 2886) = 46.12$, $p < .001$, and a *classification* × *past presentation* interaction, $F(2, 2886) = 11.33$, $p < .001$. Overall, statements were more correctly classified than confounded, $t(2886) = 8.43$, $p < .001$, $d = 1.33$, 95% CI [1.23, 1.43], or misclassified as new, $t(2886) = 11.29$, $p < .001$, $d = 1.69$, 95% CI [1.58, 1.70], while they were also more likely to be confounded than misclassified as new, $t(2886) = 2.27$, $p = .023$, $d = .23$, 95% CI [.14, .32]. However, past false statements were more correctly classified, $t(2886) = 3.11$, $p = .002$, $d = .60$, 95% CI [.46, .74], and less likely to be confounded, $t(2886) = -2.52$, $p = .012$, $d = .39$, 95% CI [.25, .52], than the past true statements. There was no *classification* × *statements* interaction, $F(2, 2886) = 1.36$, $p = .255$, which means that the statements presented as false in this study were not more confounded than those presented as true, $t(2886) = .38$, $p = .699$, $d = 0$, 95% CI [-.14, .14]. However, there was a significant *classification* × *statements* × *past presentation* interaction, $F(2, 2886) = 8.56$, $p = .001$. When the past true statements were now presented as false, they were less likely to be correctly identified, $t(2886) = -2.16$, $p = .031$, $d = .22$, 95% CI [-.007, .45], and more likely to be confounded, $t(2886) = 2.34$, $p = .019$, $d = .23$, 95% CI [.004, .46], than when they were presented as true. By contrast, past false statements were more likely to be confounded when presented as true than as false, $t(2886) = -2.07$, $p = .039$, $d = -.16$, 95% CI [-.01, -.31].

Judgments. Six of the responses were excluded as outliers. Although there was a small difference in the judgments of the two perpetrators, this difference did not reach conventionally accepted levels of significance, $F(1, 43.38) = 2.90$, $p = .096$, $M = 5.02$, $SD = 1.8$ for Dimitri; $M = 5.5$, $SD = 2.01$ for Etienne; $d = .25$, 95% CI [.10, .40].

DISCUSSION

When explicitly false statements were presented individually and discretely, the truth bias was significantly reduced. There are various possible reasons for this. First, in our studies, the truth bias was measured via participants' memory about the statements' truth value. List-wise presentation may have facilitated the piecemeal storing of each statement in memory along with its truth-value meta-information, leading to fewer dissociations between the two. This in turn, may account for participants' better performance in the memory test. Another explanation of

our findings can be drawn from the perspective of the referential theory of the illusory-truth effect (Unkelbach & Rom, 2017). According to this theory, coherence is a crucial factor in judgments of truth, so that statements that seem coherent to other knowledge one has, are judged as having a higher truth value. As the statements in Experiments 1–3 were organized in a narrative frame they may have seemed more coherent to participants, compared to the statements in the present experiment, thus increasing their perceived truthfulness. Future studies could directly test these two alternative explanations of the truth-bias reduction when information is discretely provided.

Note, that in the present experiment the truth-bias pattern was partly replicated for past true statements, which displayed different memory patterns than past false statements and were more confounded when presented as false than as true. Yet, the effect was now much smaller (d = .23) compared to that of Experiment 3 (d = .48), which used a sample of similar size. Actually, counterbalancing the statements now resulted to 24 true and false statements in the memory test, as opposed to 16 in the previous studies, which should render this experiment statistically more powerful (see Westfall, Kenny, & Judd, 2014). The fact that the memory-based truth bias, if anything, was much smaller supports the idea that a discrete presentation moderates truth bias.

Interestingly, the critical items of Experiments 1–3, viz. past false statements, were more accurately classified than past true statements. Past false statements were also more accurately classified when presented as false than as true. The enhanced memorability of the past false statements compared to the past true statements may be due to their extra valence (aggravating or attenuating). That new aggravating and attenuating statements in Experiments 1 and 2 displayed different classification patterns compared to neutral statements is also compatible with the idea that information with extra valence may trigger differential processing (see also Alves et al., 2015; Fessler, Pisor, & Navarrete, 2014 on positive vs. negative information processing). This differential processing may explain the enhanced memorability of the past false statements in the case of the present study. In any event, both memory patterns revealed for past false statements go against an alternative explanation of the truth bias based on stimulus properties in the previous studies. If anything, in Experiments 1–3 the truth bias was sufficiently strong to overshadow these tendencies.

GENERAL DISCUSSION

The chief objective of this article was to test if cognitive load is a necessary condition for the operation of truth bias in a context where participants have to rely on external meta-information in order to show disbelief. Experiments 1 and 2 revealed a truth bias operating in default situations, despite participants knowing that the statements they encountered were false, and being able to fully focus on their assessment. Our results, thus, corroborate Fiedler, Armbruster, et al.'s (1996) and Fiedler, Walther, et al.'s (1996) claim that the truth bias is independent of cognitive

load and show that this holds even when participants have available truth-value meta-information upon comprehension, and are explicitly urged to be vigilant.

These results strongly support the idea that people's cognitive resources during information encoding do not significantly determine the extent to which they will be truth biased. As such, they undermine the "automatic belief" model advanced by Gilbert and colleagues (1990, 1993) according to which the truth bias occurs at the very moment people understand a piece of information. The independence of truth bias from cognitive load suggests that the effect results from processes operating at a later stage, once participants have understood and stored the information. That is, the truth bias is more likely due to dissociations of information from truth-value or source meta-information in memory, or constructive biases (see Fiedler, Armbruster, et al., 1996; Rapp, 2016), rather than due to automatic belief upon comprehension (cf. Gilbert, 1991).

This view of the truth bias can actually compromise the extensive documentation of this effect with a strand of literature suggesting that people can automatically reject inaccurate information. Studies using implicit measures have shown that reading false statements interferes with positive responses in a Stroop-like or a lexical decision task (Hasson et al., 2005; Isberner & Richter, 2014; Richter et al., 2009). Such interference effects suggest that people can, to some extent, automatically reject incoming information that they consider inaccurate. The possibility of automatic rejection can only be compromised with the operation of truth bias, if the latter effect results from processes occurring after comprehension. Under this perspective, it would be perfectly possible for an individual to end up believing a statement she initially rejects upon comprehension. Future studies could test this hypothesis, by using a combination of measures of information assessment upon comprehension (e.g., a Stroop-like task), and post-comprehension measures of truth bias.

A limitation of our studies is that they did not directly test the mechanism that underlies the truth bias. We assessed the truth bias via two inter-correlated measures, judgments and memory, each of which is likely to reflect a separate mechanism. The judgment-based truth bias seems to be a clear instance of *metacognitive myopia* (Fiedler, 2012; Fiedler et al., 2015), understood as the tendency to use large amounts of stimulus information, while being "naive and almost blind regarding the history and validity of the stimulus data" (Fiedler, 2012, p. 2). In this line Fiedler, Armbruster, et al. (1996) tested whether participants' judgments about a person are affected by questions they are asked, independently of the answers they give. In support of this hypothesis, if participants were asked whether a person is aggressive, they later tended to rate the person as aggressive, even though they initially denied that this was the case. These findings suggest that people are affected by the semantic content of messages they receive, while they disregard the speech act within which the messages are embedded (e.g., a question, a negation, etc.). Our judgment results likely demonstrate similar effects of meta-cognitive myopia: participants ignored the meta-information signaling that the semantic

content of the false statements was inaccurate, and used this semantic content to judge the perpetrators.

The second truth-bias index, memory, suggests that participants explicitly misremembered false information they encountered as true (Experiments 1 & 3) or as being related to meta-information signaling it is true (Experiment 2). This finding could be explained by a dual-process account like the one proposed for a closely related phenomenon—the illusory truth effect. The illusory truth effect reflects people's tendency to judge previously encountered information as more true than new information (Begg, Anas, & Farinacci, 1992). In a typical illusory truth effect study, participants first receive statements that are either true or false, and then complete a memory test containing old statements along with new statements (Begg et al., 1992; Garcia-Marques, Silva, Reber, & Unkelbach, 2015; Henkel & Mattson, 2011). In such a memory test, participants' responses may be driven by two processes: on the one hand, by explicit memory for the source or truth value of the encountered statements; and, on the other hand, by the statements' perceived familiarity or processing fluency (Reber & Schwarz, 1999; Unkelbach, 2007; Unkelbach & Stahl, 2009). A crucial aspect of this account is that old statements feel more familiar or are processed more fluently than new statements, and that this automatic feeling makes participants judge the former as truer than the latter (Dechêne, Stahl, Hansen, & Wänke, 2009; Dechêne et al., 2010). In a similar vein, our participants' tendency to classify false statements as true might be driven by feelings of familiarity or fluency. Since familiarity or fluency are relative measures, this interpretation of our results entails that participants to some extent compared the fluency or familiarity of old statements against that of new statements (Dechêne et al., 2009; Dechêne et al., 2010). The finding in Experiment 1 that new statements were more likely to be misclassified as false than as true supports the possibility that participants to some extent compared old and new statements. In fact, this latter finding could be seen as the inverse of the truth-bias effect: not only are old statements more familiar or fluent than new ones, leading to the truth bias; new statements are *less* familiar and fluent than old ones, and are thus, judged as false. In fact, Dechêne et al. (2009) provide support for this hypothesis by showing that when memory tests only contain old statements, the illusory truth effect is moderated. It seems then that false statements might lack the fluency or familiarity advantage that new statements offer, and fail to be misclassified as true.

In any event, a dual-process account of our memory data is compatible with the more general meta-cognitive myopia framework (Fiedler, 2012). Participants' responses could be affected by an automatic feeling of familiarity to the extent that they disregard the meta-information signaling that false statements are false. Note that the operation of the automatic component of the illusory truth effect is partly independent of whether participants can explicitly recall the meta-information signaling that a false statement is false (Henkel & Mattson, 2011). Hence, it is likely that our participants misclassified false statements as true, even if the truth-value meta-information was not actually lost from their memory. Even if their memory

was not impaired, however, our participants must have been myopic with respect to the meta-information signaling that false statements are false. A combination of measures of meta-information memory (cf. Experiment 2) and judgments of truth (cf. Experiment 1) could disentangle whether participants misremember false statements as true because the meta-information is lost in their memory or because they fail to rely on it. Understanding whether the truth bias is due to poor memory capacities or rather due to a generalized incapacity to rely on meta-information is particularly important if we are to develop successful interventions against fake news and misinformation.

Experiment 3 suggests that the truth bias persists in contexts with equal numbers oftrue and false statements. It is a common assumption that in real life true information is more common than false information, which to some extent provides an evolutionary and ecological foundation for the truth bias (Kissine & Klein, 2013; Nadarevic & Erdfelder, 2013). On these grounds, processing fluency has been argued to be a default ecologically valid cue for judgments of truth, one that can actually be used in a controlled manner (Reber & Unkelbach, 2010). For example, in experimental contexts where processing fluency is positively linked to falsity rather than to truthfulness, fluently processed statements are judged as more false rather than more true (Unkelbach & Stahl, 2009). We believe that more nuance is needed in such claims. To the extent that our memory results were driven by a familiarity of processing fluency, participants in Experiment 3 have not been able to use such cues in an ecologically valid way. Experiment 3 thus, challenges the idea that people can easily adapt to a given context and switch the interpretation of familiarity and processing fluency in a controlled manner (cf. Unkelbach & Stahl, 2009).

Experiment 4 raises an important methodological issue, as it suggests that the paradigm commonly used to test truth-bias effects may underestimate their actual magnitude. When information is presented in chunks independent from one another it is more easily discounted. Although Fazio, Dolan, and Marsh (2014) report that suggestibility to incorrect information increases when information is presented in lists rather than coherently, participants in those studies were given a vague warning about the existence of misinformation. In this context, participants may have better memorized incorrect statements, which in the absence of concrete meta-information may have been more believable. To the contrary, our participants had specific meta-information at their disposal informing them which statement was true and which statement was false. In the context of our studies, it may have been easier to store or retrieve the meta-information when the statements were presented individually, which in turn reduced the magnitude of the truth bias. In any event, Experiment 4 suggests that presentation mode is an important factor to take into account when assessing truth-bias effects.

To conclude, our finding that the truth bias is independent of cognitive load and appears in its absence undermines the "automatic belief" model advanced by Gilbert and colleagues (Gilbert, 1991; Gilbert et al., 1990; 1993). More than 20 years after these seminal studies, many scholars seem to assume that cognitive load is an important component of the truth-bias effect (e.g., Bergstrom, Moehlmann, &

Boyer, 2006; Knowles & Condon, 1999; Prentice, Gerrig, & Bailis, 1997; Shrum, Wyer, & O'Guinn, 1998). The present studies both disentangle the operation of the truth bias from cognitive load, and provide even stronger evidence for Gilbert and colleagues' assumption that people are truth biased toward incoming information. In fact, the results of our studies are more troubling than those of the original studies, as far as the impact of misinformation and fake news is concerned. While our materials specifically pertain to judicial contexts, the contemporary socio-political scene confirms that truth-bias effects operate across several other contexts, including the media and politics. Psychological research can help mitigate the impact of these effects in real-world settings by advancing knowledge of the cognitive processes that underpin them. Assessing alternative explanations of the truth bias is a significant step toward resisting "alternative" facts.

REFERENCES

Alves, H., Unkelbach, C., Burghardt, J., Koch, A., Krüger, T., & Becker, V. (2015). A density explanation of valence asymmetries in recognition memory. *Memory and Cognition, 43*(6), 896-909. https://doi.org/10.3758/s13421-015-0515-5

Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology, 19*(2), 93-108. https://doi.org/10.1016/0022-1031(83)90031-8

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*(6), 1037-1049. https://doi.org/10.1037/h0077720

Andrew, G. (2017, January 18). Fake news handed Brexiteers the referendum—And now they have no idea what they're doing. *The Independent*. Retrieved from http://www.independent.co.uk/voices/michael-gove-boris-johnson-brexit-eurosceptic-press-theresa-may-a7533806.html

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*(4), 446-458. https://doi.org/10.1037//0096-3445.121.4.446

Bergstrom, B., Moehlmann, B., & Boyer, P. (2006). Extending the testimony problem: Evaluating the truth, scope, and source of cultural information. *Child Development, 77*(3), 531-538. https://doi.org/10.1111/j.1467-8624.2006.00888.x

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology, 4*, 1-15. https://doi.org/10.3389/fpsyg.2013.00279

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*(5), 1117-1122. https://doi.org/10.1016/j.jesp.2009.06.019

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*(2), 238-257. https://doi.org/10.1177/1088868309352251

Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107*(1), 122-141. https://doi.org/10.1037/a0036673

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G * Power 3: A flexible statistical power analysis program for the

social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Fazio, L. K., Brashier, N. M., Payne, K. B., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology, 144*(5), 993-1002.

Fazio, L. K., Dolan, P. O., & Marsh, E. J. (2014). Learning misinformation from fictional sources: Understanding the contributions of transportation and item-specific processing. *Memory,* 1-11. https://doi.org/10.1080/09658211.2013.877146

Fessler, D. M. T., Pisor, A. C., & Navarrete, C. D. (2014). Negatively-biased credulity and the cultural evolution of beliefs. *PloS One, 9*(4), e95167. https://doi.org/10.1371/journal.pone.0095167

Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. *Psychology of Learning and Motivation, 57*, 1-55. https://doi.org/10.1016/B978-0-12-394293-7.00001-7

Fiedler, K., Armbruster, T., Nickel, S., Walther, E., & Asbeck, J. (1996). Constructive biases in social judgment: Experiments on the self-verification of question contents. *Journal of Personality and Social Psychology, 71*(5), 861-873. https://doi.org/10.1037/0022-3514.71.5.861

Fiedler, K., Kareev, Y., Avrahami, J., Beier, S., Kutzner, F., & Hütter, M. (2015). Anomalies in the detection of change: When changes in sample size are mistaken for changes in proportions. *Memory and Cognition, 43*(5). https://doi.org/10.3758/s13421-015-0537-z

Fiedler, K., Walther, E., Armbruster, T., Fay, D., & Naumann, U. (1996). Do you really know what you have seen? Intrusion errors and presuppositions effects on constructive memory. *Journal of Experimental Social Psychology, 32*(32), 484-511. https://doi.org/10.1006/jesp.1996.0022

Fragale, A. R., & Heath, C. (2004). Evolving informational credentials: The (mis)attribution of believable facts to credible sources. *Personality and Social Psychology Bulletin, 30*(2), 225-236. https://doi.org/10.1177/0146167203259933

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*(January), 126-129. https://doi.org/10.1016/j.jesp.2014.09.015

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*(2), 107-119.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*(4), 601-613. https://doi.org/10.1037//0022-3514.59.4.601

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65*(2), 221-233.

Guenther, C. L., & Alicke, M. D. (2008). Self-enhancement and belief perseverance. *Journal of Experimental Social Psychology, 44*(3), 706-712. https://doi.org/10.1016/j.jesp.2007.04.010

Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: On the possibility of suspending belief. *Psychological Science, 16*(7), 566-571. https://doi.org/10.1111/j.0956-7976.2005.01576.x

Henkel, L., & Mattson, M. E. (2011). Reading is believing: The truth effect and source credibility. *Consciousness and Cognition, 20*, 1705-1721. https://doi.org/10.1016/j.concog.2011.08.018

Isberner, M.-B., & Richter, T. (2014). Does validation during language comprehension depend on an evaluative mindset? *Discourse Processes, 51*(1-2), 7-25. https://doi.org/10.1080/0163853X.2013.855867

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513-541. https://doi.org/10.1016/0749-596X(91)90025-F

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. J*ournal of Personality and Social Psychology, 103*(1), 54–69. https://doi.org/10.1037/a0028347

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*(1), 601-625. https://doi.org/10.1146/annurev-psych-122414-033702

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General General, 111*(2), 228-238. https://doi.org/10.1037/0096-3445.111.2.228

Kissine, M., & Klein, O. (2013). Models of communicatoin, epistemic trust and espistemic vigilance. In J. Laszlo, J. Forgas, & O. Vincze (Eds.), *Social cognition and communication*. New York: Psychology Press.

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77*(2), 379-386. https://doi.org/10.1037/0022-3514.77.2.379

Lavie, N. (2006). The role of perceptual load in visual awareness. *Brain Research, 1080*(1), 91-100. https://doi.org/10.1016/j.brainres.2005.10.023

Levine, T. R. (2014). Truth-Default Theory (TDT). *Journal of Language and Social Psychology, 33*(4), 378-392. https://doi.org/10.1177/0261927X14535916

Levine, T. R., & Bond, C. F., Jr. (2014). Direct and indirect measures of lie detection tell the same story: A reply to ten Brinke, Stimson, and Carney. *Psychological Science, 25*(10), 1960-1961. https://doi.org/10.1177/0956797614536740

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect." *Communication Monographs, 66*(2), 125-144. https://doi.org/10.1080/03637759909376468

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106-131. https://doi.org/10.1177/1529100612451018

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviations around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology,* 764-766. https://doi.org/http://dx.doi.org/10.1016/j.jesp.2013.03.013

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory and Cognition, 34*(5), 1140-1149.

Nadarevic, L., & Erdfelder, E. (2013). Spinoza's error: Memory for truth and falsity. *Memory and Cognition, 41*(2), 176-186. https://doi.org/10.3758/s13421-012-0251-z

Peter, C., & Koch, T. (2016). When debunking scientific myths fails (and when it does not): The backfire effect in the context of journalistic coverage and immediate judgments as prevention strategy. *Science Communication, 38*(1), 1-23. https://doi.org/10.1177/1075547015613523

PolitiFact. (2016). Retrieved October 27, 2016, from http://www.politifact.com/personalities/donald-trump/

Prentice, D., Gerrig, R., & Bailis, D. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin and Review, 4*(3), 416-420. Retrieved from http://link.springer.com/article/10.3758/BF03210803

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*(4), 413-425. https://doi.org/10.1016/j.jml.2008.02.002

Rapp, D. N. (2016). The consequences of reading inaccurate information. *Psychological Science, 25*(4), 281-285. https://doi.org/10.1177/0963721416649347

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and 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*(4), 563-581. https://doi.org/10.1007/s13164-010-0039-7

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge

permits fast and efficient validation of information. *Journal of Personality and Social Psychology, 96*(3), 538-558. https://doi.org/10.1037/a0014038

Schul, Y., & Burnstein, E. (1985). When discounting fails: Conditions under which individuals use discredited information in making a judgment. *Journal of Personality and Social Psychology, 49*(4), 894-903. https://doi.org/10.1037//0022-3514.49.4.894

Shrum, L. J., Wyer, R. S., & O'Guinn, T. C. (1998). The effects of television consumption on social perceptions: The use of priming procedures to investigate psychological processes. *Journal of Consumer Research, 24*(4), 447-458.

Street, C. N. H., & Masip, J. (2015). The source of the truth bias: Heuristic processing? *Scandinavian Journal of Psychology, 56*(3), 254-263. https://doi.org/10.1111/sjop.12204

Thorson, E. (2015). Belief echoes: The persistent effects of corrected misinformation. *Political Communication, 4609*(November), 1-21. https://doi.org/10.1080/10584609.2015.1102187

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 219-230. https://doi.org/10.1037/0278-7393.33.1.219

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition, 160*, 110-126. https://doi.org/10.1016/j.cognition.2016.12.016

Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition, 18*(1), 22-38. https://doi.org/10.1016/j.concog.2008.09.006

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580-591. https://doi.org/10.1162/jocn.2008.20054

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*(5), 2020-2045. https://doi.org/10.1037/xge0000014