The Truth About the Truth: A Meta-Analytic Review of the Truth Effect

Alice Dechêne, Christoph Stahl, Jochim Hansen and Michaela Wänke

Pers Soc Psychol Rev 2010 14: 238 originally published online 18 December 2009
DOI: 10.1177/1088868309352251

Personality and Social Psychology Review
14(2) 238–257

DOI: 10.1177/1088868309352251
http://pspr.sagepub.com
SAGE

The Truth About the Truth: A Meta-Analytic Review of the Truth Effect

Alice Dechêne[1], Christoph Stahl[2], Jochim Hansen[1], and Michaela Wänke[1]

Abstract

Repetition has been shown to increase subjective truth ratings of trivia statements. This truth effect can be measured in two ways: (a) as the increase in subjective truth from the first to the second encounter (*within-items criterion*) and (b) as the difference in truth ratings between repeated and other new statements (*between-items criterion*). Qualitative differences are assumed between the processes underlying both criteria. A meta-analysis of the truth effect was conducted that compared the two criteria. In all, 51 studies of the repetition-induced truth effect were included in the analysis. Results indicate that the between-items effect is larger than the within-items effect. Moderator analyses reveal that several moderators affect both effects differentially. This lends support to the notion that different psychological comparison processes may underlie the two effects. The results are discussed within the processing fluency account of the truth effect.

Keywords

truth effect, meta-analysis, processing fluency

When presented with an unfamiliar statement, such as "The zipper was invented in Norway," most people do not know whether it is actually true or not (i.e., the statement is *ambiguous*).[1] To nevertheless judge a statement's truth, people tend to use heuristic cues. Such heuristics make use of attributes of the statement's source (e.g., the source's level of expertise on the subject matter), attributes of the context in which it is presented (e.g., at a scientific conference), and attributes of the statement itself. In particular, it has been shown that people tend to trust a statement more if it had been encountered before: After reading "The zipper was invented in Norway" for a second time, judgments of the statement's truth or validity typically increase. This so-called *truth effect* has been the subject of extended study in psychological (e.g., Arkes, Boehm, & Xu, 1991; Bacon, 1979; Begg, Anas, & Farinacci, 1992; Hasher, Goldstein, & Toppino, 1977) and consumer research (e.g., Hawkins & Hoch, 1992; Law, Hawkins, & Craik, 1998; Roggeveen & Johar, 2002, 2007). It is important to understand this effect because people's trust in statements' truth may affect behavior related to those statements (e.g., marketing claims, health beliefs).

The present article presents a meta-analytic review of research on the truth effect that aims at integrating past research and providing a common basis for systematic future research. In addition, it seeks to open new avenues for research. It attempts to do so by focusing on two theoretically important factors that have rarely been investigated so far, namely, the distinction between two components of the truth effect (i.e., the *within-items* component and the *between-items* component) and the context in which statements are judged. The theoretical analysis in this review is based on the current consensus that the truth effect is mediated by the metacognitive experience of *processing fluency* (Begg et al., 1992; Reber & Schwarz, 1999; Unkelbach, 2007; Whittlesea, 1993; for a review, see Alter & Oppenheimer, 2009). In short, processing a repeated statement is experienced as unexpectedly fluent—in other words, the processing fluency is experienced as discrepant from a comparison standard—and this discrepancy subsequently affects judgments of truth (e.g., Hansen, Dechêne, & Wänke, 2008; Whittlesea & Williams, 1998, 2000, 2001a, 2001b). Here we focus on the prediction that the comparison standard—and, thereby, the truth effect—should vary with the context in which judgments are made (e.g., Dechêne, Stahl, Hansen, & Wänke, 2009; Whittlesea & Leboe, 2003).

We distinguish between two types of judgment contexts that differ with regard to the variability of the stimuli with which participants are confronted: First, truth judgments can be

[1]University of Basel, Basel, Switzerland
[2]University of Freiburg, Freiburg, Germany

Corresponding Author:
Alice Dechêne, University of Basel, Department of Psychology, Missionsstrasse 60/62, CH-4055 Basel, Switzerland
Email: alice.dechene@unibas.ch

collected in a *homogeneous* context in which a homogeneous set of repeated items is judged. To assess the effect of repetition, these judgments are compared to those collected for the same set of statements at their first encounter; we call this the *within-items* criterion. Second, truth judgments can be collected in a *heterogeneous* context in which a mixture of repeated and nonrepeated statements is judged. To assess the effect of repetition, judgments of a set of repeated statements are compared to judgments of a different set of nonrepeated statements; we call this the *between-items* criterion. To shed light on contextual influences on the truth effect, the meta-analysis separately evaluates the between- and within-items criteria, both with regard to the magnitude of the effect and with regard to influences of potential moderating variables discussed in the literature.

In the remainder of this introduction, the literature on the truth effect is briefly reviewed. The discussion of mediating variables focuses on processing fluency and the important role of the comparison standard. The distinction between contexts is elaborated, and hypotheses for a comparison of the two criteria are derived. Before turning to the meta-analysis, potential moderating variables are addressed and the focus of the present review is summarized.

The Truth Effect

In a typical study on the truth effect, a set of ambiguous statements (i.e., statements whose truth status is unknown to participants, as established by pretests) is presented to participants in a first session. In a second session, some of the statements are presented and judged again, often interspersed with new statements that have not been presented before. Repeated presentation of an ambiguous statement increases the probability that it will be judged as true (e.g., Bacon, 1979; Hasher et al., 1977; Schwartz, 1982). This effect belongs to a family of well-known effects of repetition in psychology, such as the mere exposure effect (Bornstein, 1989; Zajonc, 1968) and the false-fame effect (Jacoby, Kelley, Brown, & Jasechko, 1989).

The finding that repeated statements are believed more than new ones (e.g., Hasher et al., 1977) has been obtained under many conditions. The effect has been demonstrated for trivia statements (Bacon, 1979), product-related claims (Hawkins & Hoch, 1992; Law et al., 1998; Roggeveen & Johar, 2002, 2007), and even opinion statements (Arkes, Hackett, & Boehm, 1989). It occurs for orally presented statements (Gigerenzer, 1984; Hasher et al., 1977) and written statements (e.g., Arkes et al., 1989; Schwartz, 1982) as well as with a delay of minutes (Arkes et al., 1989; Begg & Armour, 1991; Begg, Armour, & Kerr, 1985; Begg et al., 1992; Schwartz, 1982) or weeks (Bacon, 1979; Gigerenzer, 1984; Hasher et al., 1977) between the repetitions of the statements. The effect has been shown with different presentation times of each statement (5 s to 12 s; Gigerenzer, 1984), and it occurs when the truth of each statement is rated after each repetition (Hasher et al., 1977) or when the rating takes place only after the final repetition (Schwartz, 1982). The effect generalizes to nonlaboratory settings (Gigerenzer, 1984), it has been shown that one exposure to a statement is sufficient to produce the effect (Arkes et al., 1991), and it seems to be robust against feedback after a longer delay (Brown & Nix, 1996; Skurnik, Yoon, Park, & Schwarz, 2005). It occurs equally for actually true and actually false statements (Brown & Nix, 1996). Moreover, even repeated statements from a noncredible source are judged as more probably true as compared to new statements (Begg et al., 1992). Overall, the truth effect appears to be very robust. The only constraint seems to be that the statements have to be ambiguous, that is, participants have to be uncertain about their truth status because otherwise the statements' truthfulness will be judged on the basis of their knowledge and not on the basis of fluency.

Mediating Variables: The Role of Memory Processes. By definition, the repetition-based truth effect is mediated by memory processes. Different memory processes have been discussed, for example, memory for stimulus frequency (Hasher et al., 1977), explicit recognition (Bacon, 1979; Hawkins & Hoch, 1992), familiarity (Begg et al., 1992; Schwartz, 1982), and, more recently, processing fluency (e.g., Begg et al., 1992; Reber & Schwarz, 1999; Unkelbach, 2006, 2007). Although these suggested mediators are all based on memory, they differ with regard to important qualities. A first important distinction is that between source memory and item memory: On one hand, if we remember hearing a statement from a particular source (e.g., from a distinguished expert), this information provides *referential validity* and likely increases our judgment of the statement's truth—given the source is trustful (e.g., Brown & Nix, 1996). The role of referential validity is limited in most studies because statements are usually pretested to be unfamiliar to participants. On the other hand, just remembering having heard the statement itself before (i.e., without source information) can increase truth judgments, too (Arkes et al., 1991): Remembering previous occurrences of a statement conveys what can be called *convergent validity*.[2]

The second important distinction is that between explicit and implicit memory processes: Although explicit recognition of a statement is clearly necessary for referential or convergent validity effects to occur, a *nonreferential* truth effect can consistently be observed even in the absence of explicit recollection (Bacon, 1979); it is this nonreferential, implicit part of the truth effect that is thought to be driven by processing fluency (Begg et al., 1992; Reber & Schwarz, 1999; Unkelbach, 2007). Importantly, although source and even item memory provide a rational basis for the truth effect, this is not the case for the fluency-based part of the effect, which is why it has been termed the *illusory* truth effect (Begg et al., 1992).

Processing Fluency and the Truth Effect. Processing fluency is defined as the metacognitive experience of ease during information processing; it may be elicited, for example, by linguistic ease, perceptual ease, semantic priming, or retrieval ease (Alter & Oppenheimer, 2009; Whittlesea, 1993). For instance, processing fluency can be increased by improving the visual contrast with which a stimulus is presented on a computer screen: A higher contrast makes it easier for participants to perceive the stimulus. This perceptual ease—or, more generally, processing fluency—may subsequently affect judgments; for example, it may lead individuals to rate statements more probably true when presented in high as compared to low visual contrast (Reber & Schwarz, 1999). In addition to truth judgments, the fluency experience has been shown to affect a broad variety of domains (cf. Alter & Oppenheimer, 2009), including confidence (e.g., Simmons & Nelson, 2006), familiarity judgments (e.g., Whittlesea, 1993), and attractiveness (Reber, Winkielman, & Schwarz, 1998). Taken together, repetition affects judgments of a statement's truth value by increasing processing fluency: A statement that is processed more fluently is typically more likely to be judged true (e.g., Begg et al., 1992).[3]

For fluency to inform judgments, one has to form an idea of how fluent a given experience was; this implies that a comparison must be made with a norm or standard (Whittlesea & Leboe, 2003). Depending on the standard, a given experience is then interpreted as relatively fluent (i.e., if it exceeds the standard) or as relatively disfluent (i.e., if it falls below the standard). In other words, for processing fluency to affect judgments, it needs to be experienced as discrepant from a comparison standard (Hansen & Wänke, 2008; Westerman, 2008; Whittlesea & Williams, 1998, 2000, 2001a, 2001b; Willems & Van der Linden, 2006). The experience of discrepancy has been shown to play an important role in effects of repetition on truth judgments (Dechêne et al., 2009; Hansen et al., 2008). Despite its theoretical importance, not much is known about how comparison standards are chosen or about the variables that affect this choice. What we do know is that the comparison standard appears to depend on features of the stimuli and of the context in which stimuli are encountered (e.g., Whittlesea & Leboe, 2003). Recent findings suggest that a standard may be based, for instance, on an expectation held in mind about how easily a stimulus will be processed (Hansen et al., 2008; Hansen & Wänke, 2008; Whittlesea & Leboe, 2003; Whittlesea & Williams, 1998, 2000, 2001a, 2001b) or on the average processing fluency of other stimuli in the same context (Dechêne et al., 2009; Whittlesea & Leboe, 2003).

At this point, the mediation of the repetition-based truth effect by processing fluency can be summarized as follows: In a first step, fluency is enhanced by repeated processing; in a second step, the enhanced processing fluency is experienced as discrepant from a comparison standard; in a third step, the experienced discrepancy informs truth judgments.

Contextual Influences on the Comparison Standard. In the present review, we argue that the truth effect may be affected by the judgment context via its influence on the construction of a comparison standard. We distinguish between two different contexts in which the critical truth judgments are collected: a homogeneous and a heterogeneous context. Below, it is outlined how these contexts may differentially affect the comparison standard and, thereby, the truth effect.

The heterogeneous context. Here, truth judgments are collected for a heterogeneous mix of statements, consisting of a set of repeated and a different set of nonrepeated statements. The truth effect is assessed using the between-items criterion: Truth judgments for the critical set of repeated statements are compared to those for a different set of statements that were not encountered before. In the heterogeneous context, because of the mixture of repeated and nonrepeated statements, there is considerable variability in processing fluency. This variability has at least two consequences: First, participants may notice that some statements are more fluent than others, and they may be more likely to use this information as a heuristic cue to inform their judgments. Second, and most importantly, the heterogeneous context provides a situation in which a useful and precise comparison standard can be easily construed on the fly based on the statements' average fluency (Whittlesea & Leboe, 2003). With the average fluency as a comparison standard, repeated statements are likely to be perceived as above average whereas nonrepeated statements are perceived as below average. Such an average standard is the most efficient way to classify items as fluent versus disfluent, as repeated versus new, or as more versus less likely true. In addition, the contrast in fluency between fluent and disfluent stimuli may lead to a shift of judgments of nonrepeated items toward the "false" end of the scale. We call this the *negative truth effect*. If such an effect were to exist, it would inflate the between-items criterion but would not affect the within-items criterion (i.e., because the latter does not consider judgments of nonrepeated statements).

The homogeneous context. In the homogeneous context, truth judgments are collected for a homogeneous set of repeated statements only. The truth effect is assessed using the within-items criterion: Truth judgments are compared to those made to the same set of statements at the first encounter. In contrast to the heterogeneous context, there is little variability in processing fluency in the homogeneous context. As a consequence, processing fluency conveys no information and is less likely to be recruited as a cue by participants (Whittlesea & Leboe, 2003).

Furthermore, if fluency is nevertheless used, a sensible comparison standard is not easily construed. Using average fluency as a standard is obviously not useful here; it would render all statements equally nondiscrepant from the standard. Global standards or expectancies form other possible bases for a comparison standard (e.g., Hansen et al., 2008; Whittlesea & Leboe, 2003): A global expectancy may be formed on the basis of previous experience but will likely lead to a rather diffuse comparison standard. Perhaps the most useful expectancy to be recruited is based on the specific statements' processing fluency at previous encounters; it might lead to a less diffuse standard but depends on participants' ability to retrieve their processing fluency experiences from the first session. It is unclear whether and how fluency experiences are stored in memory; given that fluency is a rather subtle perceptual signal, memory for this signal is likely to decline substantially as a function of intersession delay. Other standards may also be used, but such standards are likely to be idiosyncratic and more variable across participants as they are less restrained by the context. Because the homogeneous context provides no information about the relevant range of processing fluency, the comparison standards are likely to be more variable on two levels: Within participants, they may rely on more diffuse prior expectancies; furthermore, expectancies are more likely to differ between participants. As a result, the use of fluency is rendered both less likely and less efficient.

Thus, although the truth effect is driven by fluency discrepancy in both contexts, the comparison standards may differ across contexts: In the absence of relevant contextual information in the homogeneous context, the truth effect in this context may, for example, be based on an individual's rather diffuse global fluency expectancies. In contrast, in the heterogeneous context, the truth effect is likely to be based on a more specific standard computed on the fly from the information provided by that context. From these differences, three predictions can be derived that will be the focus of the meta-analysis. First, the truth effects in the two contexts should differ in magnitude (i.e., the effect should be larger in the heterogeneous than in the homogeneous context); second, they might be affected differently by moderating variables (e.g., by moderators that affect the construction of the comparison standard). A third prediction is that a negative truth effect should be observed in the heterogeneous context: Nonrepeated statements should be less likely to be judged true when judged in the context of repeated statements as compared to when judged in the context of other nonrepeated statements. These predictions are evaluated in the meta-analysis by comparing the between-items and within-items criteria.

Note that the postulated links between the within-item criterion and the homogeneous context, and between the between-item criterion and the heterogeneous context, are not perfect: Although a truth effect in a homogeneous context is necessarily measured using the within-items criterion, a truth effect obtained in a heterogeneous context can often be measured using either criterion. Specifically, the within-items criterion can also be computed for the heterogeneous context (provided that first-session truth ratings are available). In fact, a large proportion of the within-items effect sizes in the meta-analysis are from heterogeneous contexts. This implies that the meta-analytic results obtained for the within-items effect were also, to a substantial extent, driven by the heterogeneous context. Hence, for separating the effects of the different contexts, the present comparison between the two criteria must be seen as a rough initial assessment that should be followed and complemented by more targeted experimental research (e.g., Dechêne et al., 2009).

Moderating Variables

A variety of moderators of the truth effect have been investigated. First, as mentioned above, the truth effect is observed only for ambiguous statements; it disappears when the actual truth status is known. Evidence for this comes from findings that the truth effect disappears when feedback about the actual truth status is given (Brown & Nix, 1996). However, when a delay is introduced between feedback and judgment, causing memory for the actual truth status to decline, the effect reappears.

Several attributes of the statements' source also moderate the truth effect. Overall, repeated statements from credible sources are believed more, as compared to repeated statements from noncredible sources (when source credibility is remembered). However, repeated statements from noncredible sources are still believed more, as compared to new statements from those sources (Begg et al., 1992). This result indicates that source credibility moderates the size of the effect, but it does not completely eradicate the truth effect. In addition, statements are believed more when the statements' source is misattributed as originating from outside the experimental setting, which may be interpreted as an external validity cue. But even in this case, repeated statements that are attributed to the artificial laboratory situation are still believed more than nonrepeated statements (Arkes et al., 1989; Law et al., 1998). Similar to source credibility, source variability has been shown to influence the truth effect, but only when participants are aware of the variability in sources (Roggeveen & Johar, 2002). The effect appears to be more pronounced when the statements' source is not correctly remembered but the statements' content is (Law & Hawkins, 1997). Finally, older adults who tend to have impaired source memory are especially susceptible to the truth effect (Law et al., 1998; Skurnik et al., 2005; for differing results, see Parks & Toth, 2006). Thus, a disconnection of content and source memory seems to provide beneficial conditions for the occurrence of the truth effect. This can be seen as functional; for instance, it is plausible that familiar or recognized information from a source that is not recalled—and thus not evaluable—may be interpreted as a part of general semantic knowledge and by this interpreted as probably true without further evaluation (Law & Hawkins, 1997).

Taken together, research on the influence of feedback, source characteristics, and age of participants on the truth effect suggests ways in which different memory processes are involved in the truth effect, above and beyond the explicit use of recognition and the implicit use of cognitive feelings of discrepant fluency (that both are enhanced by repetition). Perhaps most importantly, imprecise source memory may support the effect. Therefore, repetition may affect truth ratings in two opposite ways: First, repetition enhances the cognitive processing fluency of statements; second, repetition also enhances the probability that both a statement's content and its source are stored in memory.

Focus of the Meta-Analysis

As discussed above, the nonreferential part of the repetition-based truth effect is the product of two processes: An increase in fluency and the subsequent experience of this increased fluency as discrepant from a comparison standard. Importantly, the second process is likely affected by the (homogeneous or heterogeneous) judgment context; this contextual effect is reflected in differences between the within- and between-items criteria. It is an inherent limitation of the meta-analytic method that these measures do not represent process-pure indicators of the above context effects, and the present meta-analysis cannot replace controlled experiments of such context effects. Nevertheless, by focusing on the differences between the within-items and the between-items components of the truth effect, the present review can provide an initial assessment of the role of the comparison standard as well as contextual moderation of the truth effect, and it hopes to stimulate experimental research in this direction.

The present review separately analyzes the between-items and within-items truth effects and investigates possible differential effects of procedural and participant variables on both components. First, we investigate the overall effect of repetition on subjective truth separately for both components. Next, we examine whether the effect of repetition on both components of subjective truth judgments is enhanced or diminished by a number of procedural and participant moderators that are of potential theoretical interest. Moderators include characteristics of stimuli, presentation, and measurement of the truth effect as well as levels of processing, the number of repetitions, and participants' age.

Method

Literature Search. We retrieved published articles, book chapters, and dissertations through an extensive search in PsycINFO, PubMed, and Web of Science, the main databases for psychological literature. Furthermore, we searched in ProQuest, the main database for doctoral dissertations, Google Scholar, and two business literature databases (Business Source Premier and EconLit) because a considerable amount of research on the truth effect has been published in consumer and marketing journals. The following keywords were used: *truth effect*, *illusory truth*, and *truth judgment(s)*. We also searched Web of Science for articles that cited the first truth effect article, by Hasher et al. (1977). We asked for unpublished data through inquiries using the mailing lists of the Society of Personality and Social Psychology and the European Association of Social Psychology (the former European Association of Experimental Social Psychology).

Criteria for Inclusion. We applied the following criteria to determine the eligibility of each study for inclusion in the meta-analysis:

1. Studies that investigated the effect of repetition on subjective ratings of truth were included.
2. Studies that reported means or proportions of subjective truth ratings at least of new and repeated statements in Session 2 (between-items criterion) or of repeated statements at their first and second (or more) presentation or judgment (within-items criterion) were included. Studies that did not report means or proportions of subjective truth judgments were excluded ($k = 13$).

Furthermore, we avoided duplication by excluding data that were reported in previously published work and that were thus already included in the meta-analysis ($k = 2$). All studies had an experimental or at least a quasi-experimental design. After application of the exclusion criteria, 51 independent studies matched the criteria and therefore were included in the analysis.

Coding of Study Characteristics. Appropriate studies were coded by the first and the third author, using a data coding form. Both authors were familiar with the truth effect literature and discussed various preliminary versions of the coding form before starting the coding. The topics of the coded study characteristics can be grouped to the following areas: (a) study and sample descriptors (e.g., year and type of publication, type of sample, and proportion of male participants), (b) research design descriptors (e.g., the moderators investigated), (c) procedure descriptors (e.g., delay between assessments, characteristics of data collection), and (d) data-level information (e.g., type of data reported). The relevant variables are described in more detail below. The interrater agreement was high: Of 456 ratings, 6 were divergent. These differences were resolved through discussion.

Study and sample descriptors. Although we had no hypotheses about influences of publication year, publication type, gender, and sample size on the truth effect, we coded these variables for descriptive reasons. Studies in journal articles ($k = 42$) and book chapters ($k = 1$) were published between 1977 and 2006. Unpublished studies (doctoral dissertations, $k = 4$; master's theses, $k = 1$; and other manuscripts, $k = 3$) were from 1998 to 2007. The sample sizes ranged from $n = 12$ to $n = 145$ participants ($M = 46.78$, $Mdn = 40$, $SD = 25.59$). The gender of participants was rarely reported.

Furthermore, we recorded the type of sample used in the studies (2 representative samples, 43 convenience samples consisting of university students, and 6 quasi-experimental studies that compared older and younger adults). Participants' mean age was reported only in studies that investigated the effect of age (M_{old} = 72.86, SD = 5.09; M_{young} = 22.54, SD = 4.11). Given the large proportion of student samples, it can be assumed that participants' mean age was similar to that of the young participants in studies on age effects.

Research design descriptors. Categorizing the study designs involved coding the type of moderator (e.g., level of processing, age) that was investigated in a study. If the information given did not allow for a definite coding judgment or the study did not involve a moderator, the data were coded as missing. In addition, we coded the type of dependent variable used to measure the truth judgments (a 7-point Likert-type scale was most frequently used; anchored either with 1 = *true* and 7 = *false* or with 1 = *false* and 7 = *true*).

Procedure descriptors. Technical characteristics of the study procedures were coded. These included the number of measurements, the time of the measurements (e.g., at Sessions 1 and 2, only Session 2), the number of presentations of the items, the modality of data collection (e.g., paper and pencil, computer), the presentation mode of the items (e.g., visual or auditory), and the response mode (e.g., written form, orally). Furthermore, the delay between two sessions was coded (e.g., < 30 min, 1 day, 1 week) as well as the duration of presentation of one item, if this information was available. There was always a category to indicate if there was no information available or a different procedure was used.

In general, the coded variables can be categorized as either procedural (e.g., stimulus variables, measurement variables) or participant (e.g., age of participants, level of processing) variables. At least one effect was coded from each study (within-item criterion and/or between-item criterion). If a study systematically investigated a variable of interest in this meta-analysis, one effect for each level was coded if the manipulation was between participants. We did not include effects from manipulations that were implemented within participants because they violate the assumption of independence of the effect sizes. In this case, we collapsed over manipulations (see Hedges & Olkin, 1985; Rosenthal & Rubin, 1986).

Meta-Analytic Procedure.

Calculating effect sizes. We used the standardized mean difference to estimate the influence of repetition on truth judgments. The effect size d is a scale-free measure of the difference of two group means (Cohen, 1988). We calculated the effect sizes by using Hedges's formula based on means and pooled standard deviations:

$$g = (M1 - M2)/SD_{pooled}. \tag{1}$$

The pooled standard deviation was computed as,

$$SD_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n + n - 2}}. \tag{2}$$

We refrained from computing effect sizes from F tests, t tests, and p values because of the repeated measures designs typically used in truth effect experiments. The test statistics of repeated measures designs use error terms that are affected by the correlation between the measures: The larger the correlation between the measures, the smaller the error (and the larger the test statistic). If an effect size is computed from such a statistic without taking the correlation between the measures into account, the effect size will be overestimated (Dunlap, Cortina, Vaslow, & Burke, 1996). Because none of the studies reported the correlations, rendering adequate corrections impossible, we used only the reported means and standard deviations for the computation of the effect sizes. Although this may imply somewhat less precise estimates, it avoids potentially severe biases.

Twenty-one studies provided standard deviations for the reported means; seven studies reported a range of standard deviations. In the latter case, we computed the pooled standard deviations from the range. Where no standard deviations were provided, we chose to impute the pooled standard deviation from an overall estimate that was obtained from those studies in which standard deviations were reported or could be extracted.

A positive effect size indicates that truth ratings increase with repetition. For the computation of the within-items criterion effect size, we subtracted the mean of the first truth rating from the mean of the repeated truth rating. We obtained k = 32 independent effect sizes for the within-items criterion. For the between-items criterion, we subtracted the mean truth rating of new statements from the mean truth rating of repeated statements. We obtained k = 70 independent effect sizes of the between-items criterion.

We applied a small-sample correction on the effect sizes, using the following formula (Hedges & Olkin, 1985):

$$d_{unbiased} = g \times \left(1 - \frac{3}{4(N-2)-1}\right). \tag{3}$$

The resulting effect size d is unbiased by sample size and can be interpreted following the conventions of Cohen (1988) with d = .2 indicating a small effect, d = .5 indicating a medium effect, and d = .8 indicating a large effect.

Calculation of average effect sizes. In computing average effect sizes, individual study effect sizes were weighted by the inverse of their variance. This procedure lends greater weight to studies with more precise effect size measures (e.g., because of larger samples, stricter experimental control, etc.) and thereby ensures that the meta-analytic results are not easily distorted by outcomes from a single small study.

Tests for moderation. We used homogeneity analyses to test for moderation (Cooper & Hedges, 1994; Hedges & Olkin, 1985). In these analyses, the amount of variance in an observed set of effect sizes is compared to the amount of variance that would be expected by sampling error alone. The null hypothesis

Table 1. Overall Meta-Analytic Results for Within- and Between-Items Criterion

			95% Confidence Interval			
Criterion	*k*	*d*	Low Estimate	High Estimate	*z*	Q_w
Within items	32	.39*** (.39)***	0.32 (0.30)	0.47 (0.49)	10.33 (8.08)	47.21*
Between items	70	.49*** (.50)***	0.45 (0.43)	0.55 (0.57)	19.67 (13.41)	142.04***

Note: Fixed-effects estimates are presented outside parentheses, and random-effects estimates are within parentheses.
*$p < .05$. ***$p < .0001$.

of homogeneity (i.e., no variation across effect sizes) is tested using a within-class goodness-of-fit statistic, Q_w, which has an approximate chi-square distribution with $k - 1$ degrees of freedom, where k equals the number of effect sizes (Hedges & Olkin, 1985). A significant Q_w statistic indicates heterogeneity, that is, a systematic variation among effect sizes; such a finding would suggest that other variables moderate the effect (Cooper, 1998).

Homogeneity analyses can be used in a similar way to test whether groups of average effect sizes vary more than predicted by sampling error. The null hypothesis of no variation across groups can be tested by computing a between-class goodness-of-fit statistic, Q_b. It has a chi-squared distribution with $j - 1$ degrees of freedom, where j equals the number of tested groups. If the Q_b statistic is significant, average effect sizes vary between the categories of the moderator more than predicted by sampling error. This procedure is similar to an analysis of variance that tests for group mean differences or to a multiple regression model that tests for linear effects.

Fixed versus random error models. It is debated whether a fixed-effects (FE) model or a random-effects (RE) model is more adequate for meta-analyses of effect sizes (e.g., Schmidt, Oh, & Hays, 2009). On one hand, FE models are more powerful and therefore more likely to detect effects of moderators. On the other hand, they tend to be more liberal and reject the null hypothesis more often than indicated by the nominal alpha level when the true effect is zero. FE models are thought to be appropriate when an inference is made only about the sample of studies under investigation, whereas RE models are the appropriate method when the goal is to generalize the findings across all possible studies. We fitted both models in the present study for the following reasons. First, there is no consensus yet as to which method should generally be preferred. Second, where the results of both analyses diverge, both can be important: We are interested in the effects of a given moderator on the present sample of studies but also whether the effects of that moderator can be generalized.

Analyses followed the methods introduced by Hedges (1983; Hedges & Olkin, 1985; Hedges & Vevea, 1998; also see Raudenbush, 1994), as implemented for SPSS by Lipsey and Wilson (2001; also see http://mason.gmu.edu/~dwilsonb/ma.html).

Results

Prior to analyses, we checked the distribution of uncorrected effect sizes for outliers. No outliers were indicated by the box plot. We also checked the distribution of uncorrected effect sizes for possible publication bias. As indicated by funnel plots for both the within-items and the between-items criteria, effect sizes are distributed symmetrically, and thus publication bias does not seem to be a problem.

Overall Analyses

In a first step, analyses of the truth effect for all independent effect sizes (32 within items, 70 between items) that were retrieved from 51 studies were separately performed for both components of the effect, collapsing across procedural and participant variables.

Results of overall analyses for both criteria show that both weighted-average *d*s differ from zero (see Table 1). Following the conventions of Cohen (1988), both components of the truth effect are medium-sized effects under both FE and RE assumptions: The effects ranged from $d = -.28$ to .89 for the within-items criterion and from $d = -.18$ to 1.43 for the between-items criterion.

Descriptively, the within-items effect was smaller than the between-items effect, and we tested whether this difference was substantial. Note that, when computed for the same study, the within- and the between-items effects are not independent in two ways: (a) The data are based on the same sample because of the repeated measures designs and (b) the mean subjective truth rating of repeated items in Session 2 is used to compute both effect sizes. To avoid a violation of independence, we chose to compare effect sizes of the within-items criterion with effect sizes of the between-items criterion only for those studies that did not report a within-items criterion. Thus, we excluded between-items effect sizes from analyses that were derived from studies that reported both criteria. Results of the inverse variance weighted one-way ANOVA with a FE model show that, in fact, the within-items criterion is smaller (FE: $d_{\text{within}} = .39$; 95% confidence interval [CI] = 0.32, 0.47; $k = 32$) than the between-items criterion ($d_{\text{between}} = .52$; 95% CI = 0.46, 0.58; $k = 41$), $Q_b(df = 1) = 6.21$, $p = .01$. When a RE model was applied, the difference between both

Table 2. Results of Moderator Analysis Examining the Effect of Change of Wording on Both Components of the Truth Effect

	Type of Repetition	*k*	*d*	95% Confidence Interval		$Q_b(df = 1)$
				Low Estimate	High Estimate	
Within-items criterion						< 1 (< 1)
	Verbatim repetition	30	.39*** (.39)***	0.32 (0.29)	0.47 (0.49)	
	Gist repetition	2	.40† (.40)	–0.44 (–0.09)	0.85 (0.89)	
Between-items criterion						13.35** (5.95)*
	Verbatim repetition	64	.53*** (.53)***	0.47 (0.46)	0.58 (0.60)	
	Gist repetition	6	.21* (.22)†	0.05 (–0.02)	0.37 (0.46)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
†$p < .10$. *$p < .05$. **$p < .001$. ***$p < .0001$.

criteria was only marginally significant (RE: $d_{\text{within}} = .39$; 95% CI = 0.28, 0.51; $k = 32$; $d_{\text{between}} = .53$; 95% CI = 0.44, 0.62; $k = 41$), $Q_b(df = 1) = 3.29$, $p = .07$. This indicates that, in fact, in the given set of studies both criteria differ in magnitude. As the results from the RE model suggest, however, this result should probably not be generalized to all possible studies on the truth effect. Also note that the sample of effect sizes of the between-items criterion used in this comparison differed from the overall set and that the effect size of this subsample was greater than that of the overall sample in both models; perhaps those studies that did not examine the within-items criterion systematically differ from those that did (see Wilson, 2003, for a discussion of confounds in meta-analysis).

We applied homogeneity analyses on both criteria. The results indicate heterogeneity for both the within-items and the between-items criteria (see Table 1). Thus, it is likely that both components of the effect are moderated by other variables. We performed all moderator analyses separately for both criteria. Analyses are grouped into (a) *procedural variables* that represent characteristics of the stimuli, the presentation, and measurement variables and (b) *participant variables* that include the age of the sample, the level of processing, and the influence of further (i.e., more than one) repetitions on the truth effect. For some participant variables, there was too little information given in the studies to include them in the analysis (e.g., gender of participants, individual differences).

Moderator Analyses

Stimulus Variables. Two possible procedural moderators can be characterized as variables of the stimuli: (a) whether statements are repeated verbatim or only the gist of a statement is repeated and (b) whether the proportion of critical items in the overall pool of presented items in the studies influences the effect. Both are of theoretical interest to understand the mediating processes of fluency and discrepancy on the occurrence of the truth effect. For instance, a gist repetition may affect the fluency of a statement (i.e., it is experienced as relatively disfluent as compared to a verbatim repetition). Similarly, a higher proportion of repeated items in a set of statements may reduce the experienced discrepancy against relatively disfluent statements and thereby may reduce the magnitude of the effect.

Verbatim versus gist repetition. We examined whether the statements in experiments were repeated verbatim or whether the gist of the statement was repeated. For example, Arkes and colleagues (1991, Experiment 2) presented text passages about China in a first session and repeated passages containing the same facts as in the text presented first (i.e., they used a gist repetition). As a result, the China-related statements were rated more probably true as compared to a control condition that had not seen the text about China before.

Moderator analyses for changes in wording (gist vs. verbatim repetition) from Session 1 to Session 2 were separately performed for the within-items and the between-items criterion; results are shown in Table 2. Unfortunately, only two effect sizes were available in the gist group for the within-items criterion; results of the Q_b statistic on the within-items criterion can therefore not be reasonably interpreted. We present the results for descriptive reasons only. For the between-items criterion, effects of gist repetition were significantly smaller than those of verbatim repetition under both an FE and an RE model. Interestingly, in the RE analysis, the truth effect from gist repetition did not reach significance. The results suggest that repeating a statement in a different wording than in the first presentation diminishes the truth effect or may even eliminate it.

Proportion of critical items. We examined whether the proportion of critical (i.e., repeated) items in the total number of items judged affects the size of the truth effect. Although this proportion varied enormously (i.e., between 18% and 100%), it was 50% in more than half of the cases. As the discrepancy between fluent versus disfluent stimuli is discussed as an important mediator, the proportion of fluent versus disfluent stimuli may be relevant (Dechêne et al., 2009; Hansen et al., 2008). For instance, a higher proportion of disfluent (i.e., nonrepeated) stimuli may lead to greater truth effects by enhancing or strengthening the experienced discrepancy to

Table 3. Results of Moderator Analysis on Statements' Presentation Time for Within- and Between-Items Criterion

	Presentation Time per Statement	*k*	*d*	95% Confidence Interval Low Estimate	95% Confidence Interval High Estimate	$Q_b(df = 2)$
Within-items criterion						3.78 (2.85)
	≤ 8 s	3	.55*** (.55)**	0.29 (0.25)	0.82 (0.85)	
	> 8 s	4	.38*** (.38)**	0.20 (0.16)	0.56 (0.61)	
	Participant paced	17	.28*** (.27)***	0.17 (0.07)	0.39 (0.41)	
Between-items criterion						5.25† (1.30)
	≤ 8 s	5	.49*** (.51)**	0.27 (0.21)	0.71 (0.82)	
	> 8 s	27	.54*** (.53)***	0.46 (0.41)	0.62 (0.65)	
	Participant paced	16	.38*** (.41)***	0.27 (0.24)	0.49 (0.58)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
†$p < .10$. **$p < .001$. ***$p < .0001$.

fluent (i.e., repeated) stimuli. Regression analyses revealed that the proportion of critical items affects neither the within-items criterion, FE: $Q_b(df = 1) < 1$, RE: $Q_b(df = 1) < 1$, nor the between-items criterion, FE: $Q_b(df = 1) < 1$, RE: $Q_b(df = 1) < 1$. However, given the reduced range of variability, this null finding should not be overinterpreted.

Presentation Variables. Presentation variables include the presentation time per statement, the delay between Sessions 1 and 2, the modality of the presentation (i.e., whether statements were presented to the participants visually, auditory, or mixed), and whether the repeated statements in Session 2 were presented to participants homogenously (i.e., in a context without new statements) or heterogeneously (i.e., interspersed with new statements).

Presentation time per statement. In all, 22 studies did not specify the presentation time per statement and therefore were excluded from analysis; for the within-items criterion, $k = 24$ were included, and for the between-items criterion, $k = 48$ were included. We decided to group the available presentation times in three categories because this yielded meaningful and sufficiently large categories: less than or equal to 8 s per statement, more than 8 s per statement, and participant paced.

Presentation time per statements did not affect either criterion in FE and RE models (see Table 3). We conducted follow-up analyses for the within-items criterion and tested whether a presentation time less or equal to 8 s produced significantly greater effects as compared to presentation times more that 8 s. No differences were found using FE or RE models, FE: $Q_b(df = 1) = 1.08$, $p = .29$; RE: $Q_b(df = 1) = 1.08$, $p = .29$. This result corresponds with the finding of Gigerenzer (1984), who experimentally varied whether each statement had to be judged within 5 s versus 10 s and found no difference between these intervals. On a descriptive level, participant-paced presentation seems to produce the smallest effects.

Delay between sessions. Brown and Nix (1996) and Gigerenzer (1984) investigated whether different delays between the sessions affect the truth effect. Gigerenzer varied the delay between sessions (1 vs. 2 weeks). No influence of intersession interval was found. Similarly, Brown and Nix varied whether Session 2 was administered 1 week, 1 month, or 3 months after Session 1; they found no difference in the truth effect for true items. Repeated false items, however, were rated less true than new (also false) items after a 1-week interval; no such difference was found for the 1-month or the 3-month conditions. Thus, experimental evidence so far shows no clear pattern for the influence of intersession interval on the truth effect. Along the same lines, many studies that did not directly examine the influence of the intersession interval obtained the truth effect with delays ranging from no more than minutes (e.g., Arkes et al., 1989; Begg & Armour, 1991; Schwartz, 1982) to several weeks (e.g., Bacon, 1979; Hasher et al., 1977).

For 20 studies, the length of the delay between the sessions was not specified, and therefore these effect sizes were excluded from the analysis. For the between-items criterion, 51 cases were included ($k = 19$ excluded), for the within-items criterion 30 cases were included ($k = 2$ excluded). As shown in Table 4, both criteria were not moderated by the delay between the sessions in both models, between-items criterion FE: $Q_b(df = 2) = 0.66$, $p = .72$; RE: $Q_b(df = 2) = 0.35$, $p = .84$; within-items criterion FE: $Q_b(df = 2) = 3.45$, $p = .18$; RE: $Q_b(df = 2) = 2.75$, $p = .25$. On a descriptive level, the effects size of the within-items criterion was very small when Session 2 was administered on the same day as Session 1. For practical reasons, researchers often tend to realize Session 2 either within the same day or within a week after Session 1. Thus, we conducted additional analyses that compared effect sizes of within-day studies with those of the within-week type. As suggested by a marginally significant effect in the fixed-error model, administering both sessions within the same day ($d_{within} = .25$; 95% CI = 0.07, 0.43; $p < .05$) tends to yield smaller effect, as compared to when the second session is administered at least 1 day after the first ($d_{within} = .44$; 95% CI = 0.31, 0.67; $p < .0001$), FE: $Q_b(df = 2) = 2.91$, $p = .08$. However, this effect does not replicate in a random-error model (within day: $d_{within} = .25$; 95% CI = 0.07, 0.43, $p = .01$;

Table 4. Results of Moderator Analysis of Delay Between Sessions on Within- and Between-Items Criterion

	Session 2 Administered	k	d	95% Confidence Interval Low Estimate	95% Confidence Interval High Estimate	$Q_b(df = 2)$
Within-items criterion						3.44 (2.74)
	Within day	9	.25* (.24)*	0.07 (0.04)	0.43 (0.46)	
	Within week	11	.44*** (.45)***	0.31 (0.29)	0.57 (0.61)	
	Longer delay	10	.44*** (.45)***	0.32 (0.28)	0.56 (0.61)	
Between-items criterion						< 1 (< 1)
	Within day	25	.48*** (.49)***	0.39 (0.37)	0.57 (0.62)	
	Within week	14	.43*** (.44)***	0.32 (0.28)	0.54 (0.59)	
	Longer delay	12	.48*** (.49)***	0.36 (0.32)	0.59 (0.65)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
*$p < .05$. ***$p < .0001$.

Table 5. Results of Moderator Analysis on Presentation Modality for Within- and Between-Items Criterion

	Presentation Mode	k	d	95% Confidence Interval Low Estimate	95% Confidence Interval High Estimate	$Q_b(df = 2)$
Within-items criterion						< 1 (< 1)
	Visual	21	.43*** (.43)***	0.32 (0.32)	0.53 (0.53)	
	Auditory	5	.43** (.43**)	0.22 (0.22)	0.63 (0.63)	
	Mixed	2	.40** (.40**)	0.17 (0.17)	0.63 (0.63)	
Between-items criterion						2.66 (1.01)
	Visual	38	.51*** (.52)***	0.44 (0.42)	0.58 (0.61)	
	Auditory	24	.51*** (.52)***	0.43 (0.39)	0.61 (0.64)	
	Mixed	4	.68*** (.67)***	0.48 (0.38)	0.87 (0.95)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses. The fixed-effect and the random-effect model revealed the same results in the analysis of the within-items criterion because the random error variance component was zero.
**$p < .001$. ***$p < .0001$.

within week: $d_{within} = .45$; CI 95% = 0.30, 0.59, $p = .0001$), $Q_b(df = 2) = 2.57$, $p = .11$). A tendency toward a smaller within-items effect for statements that were shown on the same day could be explained by a higher proportion of participants who remembered their first truth judgment for the respective statements; they might not have shown a strong increase in subjective truth for the second judgment because of a desire for consistency.

Modality of presentation. The truth effect has been reported with visual (e.g., Arkes et al., 1989; Begg et al., 1992; Hawkins & Hoch, 1992), with auditory (Gigerenzer, 1984; Hasher et al., 1977), and with mixed (i.e., visually and auditory; Bacon, 1979; Begg & Armour, 1991) presentations. It is yet unclear whether modality of presentation systematically influences the truth effect. Four studies did not specify the modality of presentation and therefore were excluded from analyses, leaving $k = 28$ effect sizes for the analysis of the within-items criterion and $k = 66$ for the analysis of the between-items criterion. Results (see Table 5) show that the truth effect is not affected by modality (the within-items criterion results for mixed presentation have to be interpreted with caution because of the small number of effect sizes).

Homogenous versus heterogeneous context. In most studies the truth effect has been examined in heterogeneous lists of mixed and repeated statements; it has rarely been studied in homogeneous lists (Dechêne et al., 2009; Schwartz, 1982). Results of the moderator analysis on the available effect sizes of these different list types are presented in Table 6. Unfortunately, Q_b statistics for the between-items criterion could not be computed because only one case of a homogeneous list was reported. More studies are certainly needed to investigate the truth effect under homogeneous presentation conditions. The results for the within-items criterion suggest that the truth effect may be reduced in a homogeneous context. This finding provides initial support for the important role of context in determining comparison standards (Dechêne et al., 2009; Hansen et al., 2008).

Measurement Variables

Scale. We examined whether the scale used to measure the truth ratings influences the truth effect. Most studies ($k = 26$) used a 7-point Likert-type scale with values ranging from *false* (1) to *true* (7). Ten studies used a 7-point scale with values from *true* (1) to *false* (7). Three studies used a 6-point scale (higher values indicated higher truth ratings). Five

Table 6. Results of Moderator Analysis of Homogenous Versus Heterogeneous Presentation for Within- and Between-Items Criterion

	List Type	k	d	95% Confidence Interval Low Estimate	95% Confidence Interval High Estimate	$Q_b(df = 1)$
Within-items criterion						5.65* (4.44)*
	Heterogeneous	29	.42*** (.43)***	0.35 (0.33)	0.50 (0.52)	
	Homogeneous	3	.08 (.07)	−0.20 (−0.24)	0.35 (0.39)	
	List Type	k	d	Low Estimate	High Estimate	Q_b
Between-items criterion						—
	Heterogeneous	65	.53*** (.51)***	0.48 (0.44)	0.58 (0.58)	
	Homogeneous	1	−.09 (−.09)	−0.62 (−0.76)	0.43 (0.55)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
*$p < .05$. ***$p < .0001$.

Table 7. Results of Moderator Analysis of Scale Type on Within- and Between-Items Criterion

	Scale	k	d	95% Confidence Interval Low Estimate	95% Confidence Interval High Estimate	$Q_b(df = 2)$
Within-items criterion						7.94* (6.00)*
	1–7 (false–true)	23	.33*** (.33)***	0.24 (0.22)	0.43 (0.44)	
	1–7 (true–false)	2	.40** (.41)**	0.18 (0.12)	0.63 (0.69)	
	1–6 (false–true)	6	.62*** (.62)***	0.44 (0.42)	0.79 (0.82)	
	16 cm (false–true)	—				
	Dichotomous	—				
	Scale	k	d	Low Estimate	High Estimate	$Q_b(df = 4)$
Between-items criterion						24.50** (11.27)*
	1–7 (false–true)	40	.39*** (.41)***	0.32 (0.32)	0.45 (0.49)	
	1–7 (true–false)	12	.64*** (.65)***	0.53 (0.49)	0.74 (0.79)	
	1–6 (false–true)	6	.66*** (.65)***	0.48 (0.43)	0.83 (0.88)	
	16 cm (false–true)	5	.66*** (.65)***	0.49 (0.42)	0.83 (0.89)	
	Dichotomous	7	.54*** (.53)**	0.32 (0.27)	0.75 (0.79)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
*$p < .05$. **$p < .001$. ***$p < .0001$.

studies used a continuous 16 cm scale with higher values indicating higher truth rating. Finally, 6 studies used a dichotomous measure. Only a single study used a 9-point scale (*false* to *true*); it was not included in this analysis. Results (see Table 7) suggest that the scale used to measure truth ratings influences both components of the truth effect. This finding was replicated in FE and RE models. For the within-item criterion, results suggest that a 6-point scale evokes a larger effect as compared to a 7-point scale. This is indeed the case, as indicated by a follow-up test, FE: $Q_b(df = 1) = 7.65$, $p < .05$; RE: $Q_b(df = 1) = 5.75$, $p < .05$.

Results for the between-items criterion suggest that the widely used 7-point scale from *false* to *true* evokes smaller effects than do the other scales. A follow-up analysis revealed that this difference was significant, FE: $Q_b(df = 1) = 23.57$, $p < .001$; RE: $Q_b(df = 1) = 10.52$, $p = .001$. Effect sizes obtained using dichotomous responses do not differ from those collected on the other scales, FE: $Q_b(df = 1) < 1$; RE: $Q_b(df = 1) < 1$.

It has been discussed whether scales with a midpoint (i.e., odd scales) are superior to scales without one (i.e., even scales), or vice versa (e.g., Krosnick, 1991; Krosnick & Fabrigar, 1997). Even scales force the participant to choose a clear direction of judgment, whereas odd scales allow participants to choose an indifference point. Thus, we were interested to determine whether the truth effect can be captured better by an even scale that may foster the use of fluency

Table 8. Results of Moderator Analysis of Scale Type (Odd vs. Even) for Within- and Between-Items Criterion

				95% Confidence Interval		
	Scale	*k*	*d*	Low Estimate	High Estimate	$Q_b(df = 1)$
Within-items criterion						7.61* (5.91)*
	Odd	26	.35*** (.34)***	0.26 (0.25)	0.43 (0.44)	
	Even	6	.62*** (.62)***	0.45 (0.42)	0.79 (0.82)	
Between-items criterion						4.03* (1.65)
	Odd	52	.46*** (.47)***	0.40 (0.38)	0.52 (0.55)	
	Even	13	.61*** (.60)***	0.47 (0.42)	0.75 (0.78)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
*$p < .05$. ***$p < .0001$.

Table 9. Results of Moderator Analysis of Data Collection Modality for Within- and Between-Items Criterion

				95% Confidence Interval		
	Data collection modality	*k*	*d*	Low Estimate	High Estimate	$Q_b(df = 1)$
Within-items criterion						< 1 (< 1)
	Paper and pencil	21	.42*** (.43)***	0.33 (0.31)	0.54 (0.55)	
	Computer	2	.27 (.21)	−0.32 (−0.34)	0.73 (0.53)	
Between-items criterion						17.74*** (7.21)*
	Paper and pencil	42	.59*** (.59)***	0.53 (0.51)	0.66 (0.69)	
	Computer	14	.30*** (.33)**	0.18 (0.16)	0.43 (0.50)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
*$p < .05$. **$p < .001$. ***$p < .0001$.

more than an odd scale. We grouped effect sizes into the two categories (i.e., both variants of the 7-point scale and the 9-point scale were classified as odd; the 6-point scale and the dichotomous responses were classified as even; the 16 cm scale was excluded because it cannot be categorized as odd vs. even). Results (see Table 8) show that measuring truth judgments with an even scale yields greater effects on both criteria. Note that, in the RE model, the scales do not significantly differ for the between-items effect.

Modality of experiment and data collection. We examined whether the data collection mode, that is, paper and pencil ($k = 30$ studies) versus computer ($k = 11$ studies), influences the truth effect. Ten studies did not report mode of data collection, and it was not possible to infer the mode of data collection that was used; these studies were excluded from analysis. Unfortunately, there were again too few within-items effect sizes in one condition (computer collection); the results are given for descriptive reasons. Results of both models (see Table 9) show that the modality of data collection affects the between-items effect: Conducting a truth effect experiment on the computer produces smaller effects than with paper and pencil. The reasons for this effect are unclear, and we can only speculate that there may be confounds with restricted presentation times in this analysis or differences in the presentation of statements (i.e., sequentially vs. all statements on the same sheet) that may foster or diminish the experience of discrepant fluency. However, other reasons are also possible, and further research is needed on this point.

Other Variables. Beyond procedural aspects, other variables may also affect the size of the truth effect and may function differentially on the processes that are involved in the occurrence of the within- and the between-items effect. Research has concentrated on cognitive processing styles and memory processes. Level of cognitive processing was mainly examined by manipulation of high versus low involvement or cognitive load. Participants' age was investigated because it may indirectly reflect the influence of memory on the truth effect because older adults tend to have impaired memory capacities. Furthermore, we conducted analyses on the influence of more than one repetition of statements on the truth effect. Other variables, such as individual judgment or processing styles (e.g., need for cognition, faith in intuition), were examined so infrequently that we were not able to investigate them here.

Level of Processing. An important question in research on the truth effect is whether the effect is moderated by level of processing. Hawkins and Hoch (1992) manipulated in Session 1 whether participants had to judge comprehensibility (i.e., a low involvement manipulation) or subjective truth (i.e., high involvement) of consumer-related claims and found that the truth effect (indicated by the between-items

Table 10. Results of Moderator Analysis of Level of Processing at Session 1 for Between-Items Criterion

	Level of Processing	k	d	95% Confidence Interval Low Estimate	95% Confidence Interval High Estimate	$Q_b(df = 1)$
Between-items criterion						12.88** (4.35)*
	High	51	.44*** (.45)***	0.37 (0.37)	0.49 (0.54)	
	Low	19	.63*** (.62)***	0.54 (0.49)	0.72 (0.75)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
*$p < .05$. **$p < .001$. ***$p < .0001$.

Table 11. Results of Moderator Analysis of Age of Participants for Between-Items Criterion

	Age	k	d	95% Confidence Interval Low Estimate	95% Confidence Interval High Estimate	$Q_b(df = 1)$
Between-items criterion						1.51 (0.79)
	Young	66	.49*** (.49)***	0.44 (0.42)	0.54 (0.57)	
	Old	4	.64*** (.64)**	0.41 (0.33)	0.88 (0.96)	

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
**$p < .001$. ***$p < .0001$.

criterion) was larger under the low involvement condition (also see Hawkins, Hoch, & Meyers-Levy, 2001). In following this classification, the task of performing truth judgments at Session 1 was coded as a high level of processing, whereas merely reading the statements, or performing other superficial tasks, was coded as a low level of processing. A levels of processing effect is supported by the results of the moderator analyses (see Table 10), showing that the truth effect on the between-items criterion is increased with low as compared to high levels of processing at Session 1. Note that this moderator analysis cannot be performed for the within-items criterion because (with one exception) all studies implemented the level of processing manipulation in Session 1 and did not obtain the necessary Session 1 truth ratings for the low level of processing condition.

Age of Participants. The mean age of young participants was $M = 22.54$ ($SD = 4.11$) and $M = 72.86$ ($SD = 5.09$) for older participants. Note that only studies that directly investigated the influence of age on the truth effect reported the age of the sample. However, most studies use student samples, suggesting a mean age between 18 and 30 years; thus, we included these studies in the group of younger participants. Law and colleagues (1998) found that older adults were especially susceptible to show the truth effect, especially after a longer delay (Skurnik et al., 2005). However, there are also findings that show no differences in the truth effect between old and young participants in a fluency manipulation (Parks & Toth, 2006). Data were not available for the within-items criterion. Descriptively, the truth effect was larger for older adults (see Table 11); however, results of the moderator analysis suggest that the truth effect is not dependent on participants' age.

Further Repetitions. Hasher and colleagues (1977) tested whether, after an initial repetition, truth ratings continue to be affected by further repetitions. Unfortunately, very few effect sizes were available for both criteria, and therefore the Q_w statistic cannot be reasonably interpreted (within-items criterion) or even computed (between-items criterion). Nevertheless, there appears to be a tendency for an effect of further repetitions on the between-items criterion: Mean effect sizes of a second repetition were not different from zero for the within-items criterion but were significantly different from zero for the between-items criterion (see Table 12). Also, descriptively, the effects of a second and third repetition were very small for the within-items criterion; in contrast, the magnitude of the between-items criterion remained constant beyond the first repetition.

Summary

Overall, the truth effect was of medium size. The effect was moderated by response format and presentation duration but not by modality of presentation or the proportion of critical (i.e., repeated) statements. The truth effect was smallest on the widely used 7-point Likert-type scale (1 = *false*, 7 = *true*), and it was larger when an even scale was used ($d = .61$) as compared to an odd scale ($d = .46$). The effect tended to be moderated by statements' presentation time, such that participant-paced presentation tended to yield smaller effects as compared to experimenter-paced presentation. The truth

Table 12. Overall Analyses of Effect Sizes for Second and Third Repetition for Within- and Between-Items Criterion

		k	*d*	95% Confidence Interval Low Estimate	95% Confidence Interval High Estimate	*z*	Q_w
Within-items criterion							
	Second repetition	7	.16† (.17)†	–0.01 (–0.03)	0.33 (0.37)	1.93 (1.65)	8.52
	Third repetition	2	.16 (.16)	–0.08 (–0.08)	0.41 (0.41)	1.28 (1.28)	0.02
Between-items criterion							
	Second repetition	6	.44*** (.48)**	0.26 (0.22)	0.63 (0.73)	4.76 (3.68)	9.01
	Third repetition	1	.68** (—)	0.35 (—)	1.02 (—)	4.04 (—)	—

Note: Fixed-effects values are presented outside parentheses, and random-effects values are within parentheses.
†$p < .10$. **$p < .001$. ***$p < .0001$.

effect did not depend on the modality (auditory, visual, or mixed) of the statements' presentation. The proportion of repeated items on the overall set of judged statements also did not influence the effect.

A few additional moderators could be investigated only for between-items effect sizes (there were too few relevant within-items effect sizes). The truth effect was greater under lower (*d* = .63) as compared to higher (*d* = .44) levels of processing at Session 1. The effect was further moderated by the data collection modality: A paper and pencil procedure (*d* = .59) yielded a larger effect as compared to a computer-based procedure (*d* = .30).

Last but not least, the truth effect was larger for the between-items (*d* = .49) than for the within-items criterion (*d* = .39), a finding that is consistent with the notion that they may reflect different underlying processes. This notion is further supported by moderator analyses: As summarized below, the within-items and the between-items criteria were differently affected by a subset of the investigated moderators.

Within-Items Criterion. The within-items truth effect was moderated by the delay between sessions: When Session 2 was administered on the same day as Session 1, a significantly smaller effect (*d* = .25) was observed, as compared to when a longer delay was used (*d* = .44). The within-items truth effect was not affected by a verbatim (*d* = .39) versus gist (*d* = .40) repetition. A within-items truth effect was not found with a homogeneous test list (*d* = .08), but only with a heterogeneous list of old and new statements in Session 2 (*d* = .42). Descriptively, additional repetitions (i.e., two or three repetitions) did not reveal a considerable truth effect on the within-items criterion (*d*s = .16).

Between-Items Criterion. The results obtained for the between-items criterion differed from those obtained for the within-item criterion with regard to the two moderators of delay and verbatim versus gist repetition. In contrast to the within-items effect, the delay between sessions did not influence the between-items effect. Also, in contrast to the within-items effect, the between-items effect was moderated by verbatim versus gist repetition: The effect was larger with a verbatim (*d* = .53) as compared to a gist (*d* = .21) repetition. Finally, and again in contrast to the within-items criterion, the effect of the second and subsequent repetitions on the between-items criterion was still of considerable size (*d* = .44), descriptively, and comparable to the effect of the first repetition.

Discussion

In summary, some general conclusions can be drawn. First, the truth effect is of medium size (with CIs ranging between *d* = .32 and *d* = .55, depending on the choice of criterion and type of analysis). Second, the effect is robust against variations in presentation duration and modality, the proportion of repeated statements, and participants' age. Third, the effect is smaller for scales that offer an indifference point (i.e., scales with an odd number of points). These findings are consistent with the widely held view that repetition, largely via the subtle cues provided by increased processing fluency, affects truth judgments in a wide range of situations. Before we turn to comparing the within- and between-items effects, some additional findings are discussed: the levels of processing effect and the role of interindividual differences.

Levels of Processing

The between-items truth effect was larger under low as compared to high levels of processing, that is, shallower cognitive processing during the statement's first encounter led to a more pronounced truth effect. Hawkins and Hoch (1992) argued that high levels of processing foster more elaborative thoughts about the statements; such elaborations are likely to have a stronger influence on subsequent truth judgments than low-level heuristic cues such as fluency. In contrast, people who engaged in shallower cognitive processing are less likely to elaborate; their judgments tend to be based more on peripheral heuristic cues such as a statement's familiarity (e.g., Hawkins & Hoch, 1992; Hawkins et al., 2001) or, we might add, its fluency. Note that the present results are restricted to effects of level of processing at the first encounter of a statement; there is a paucity of research on the influences of level of processing at the time of judgment.

The levels of processing effect is reminiscent of research on the "Spinozan" account introduced by Gilbert and colleagues (Gilbert, Krull, & Malone, 1990; Gilbert, Tafarodi, & Malone, 1993). This account holds that "seeing is believing": When information is encoded, it is automatically and effortlessly "tagged" as true; elaborative processes are required to modify this tag. If elaboration is not possible—for example, under time pressure or cognitive load—people tend to encode false information as true, even if it was explicitly identified as false (Gilbert et al., 1993). Thus, similar to the present result that low levels of processing facilitate the truth effect, Gilbert and colleagues have shown that conditions of effortless or superficial cognitive processing facilitate the acceptance of new information as true.

Note that the Spinozan account does not distinguish between effects on encoding versus judgment because it is not concerned with the effects of repetition. To investigate levels of processing effects on the experience (and use) of fluency during truth judgments, future research should concentrate on the influence of level of processing at the time of the judgment.

Individual Differences

As revealed by the present meta-analysis, there is a lack of research on the influence of individual differences on the truth effect. It is plausible that individuals' predispositions such as general skepticism (e.g., Hurtt, 1999; Obermiller & Spangenberg, 1998) or the tendency toward an intuitional–experiential thinking style (e.g., Epstein, Pacini, Denes-Raj, & Heier, 1996) may influence their susceptibility to the truth effect: Individuals with a highly intuitive style of thinking may be more sensitive to their metacognitive experiences, such as processing fluency, and may therefore exhibit a stronger truth effect. In contrast, people with a greater tendency to display general skepticism may be less susceptible to the truth effect. Given that the present meta-analytic review has firmly established the existence of a substantial, robust, medium-sized effect, future research focusing on individual differences is undoubtedly worthwhile.

Two Components of the Truth Effect

We now turn to discussing the differences between the between- and within-items effects in terms of contextual effects on the comparison standard. But first, let us again point out an important limitation of the meta-analytic findings: Almost all studies implemented a heterogeneous context with both repeated and nonrepeated statements, from which both the between-items and most of the within-items effect sizes were computed for the present analyses; only a small number of studies used the homogeneous context (for these studies, only the within-items criterion was computed). As a consequence, there is considerable overlap between the processes that determine the between- and within-items effects; they cannot be interpreted as reflecting distinct cognitive processes. As we pointed out above, it is not possible to purely disentangle underlying cognitive processes using the meta-analytic method. Such an effort to disentangle the underlying processes requires the use of controlled experiments, preferably in combination with comprehensive formal models of the truth effect (e.g., Unkelbach & Stahl, 2008).

Despite these limitations, the present distinction between the two criteria makes an important first contribution to research on context effects. This is because, first, the overlap between the cognitive processes involved in the between- and within-items effects is not perfect, and important theoretical distinctions remain such as the prediction of a negative truth effect for disfluent items. Second, despite considerable overlap, important empirical differences were observed. As predicted, the between- and within-items effects were of different magnitude, they were affected by different moderators, and a tendency toward a negative truth effect was found.

Dissociations. Support for the notion of different underlying processes comes from important empirical dissociations between the within-items criterion and the between-items criterion. Three main dissociations emerged: First, the between-items effect was of greater magnitude than the within-items effect. This observation is consistent with the above analysis that judgments of nonrepeated (and therefore disfluent) items are driven toward the "false" end of the scale in a heterogeneous context. Obviously, this line of reasoning implies that when judgments for nonrepeated statements are compared between the first and second sessions, we should observe a negative effect, and this negative effect should account for the difference in magnitude between the two criteria. We conducted such an analysis on the eligible set of effect sizes ($k = 29$) and obtained the predicted negative effect of $d = –.07$ (with CIs ranging from –.14 to .01). Although this effect is not significantly different from zero because of the smaller number of eligible effect sizes, it is of comparable magnitude to the difference between the between- and within-items effects (see Table 1). We conclude that the negative truth effect accounts for much, but not all, of the difference between the between- and within-items effect sizes.

Second, increasing the delay between sessions increases the within-items effect but not the between-items effect. This suggests that, with increasing delay, the attribution of fluency to the statement's previous encounter (i.e., as familiarity) is less likely to be successful, perhaps because of decreasing source memory accuracy. Instead, the experienced fluency is then attributed to a statement's truth, in the sense of convergent validity. Alternatively, a shorter delay might help participants remember their first-session responses to the statements, and a desire for behavioral consistency (i.e., a tendency to repeat the first-session response) might then diminish the truth effect. Other explanations are possible, but all explanations have in common that delay moderates the within-items truth effect via its influence on memory processes other than those that support processing fluency. If delay affected fluency directly, such an effect should also be reflected in the between-items criterion. However, a moderation by delay was not observed for the between-items effect. This suggests that delay does not affect fluency directly and that the between-items effect is less susceptible to explicit memory processes such as those discussed above. Instead, the fact that the between-items effect is independent of delay suggests that it may come about mainly because of the differential fluency experiences provided in a heterogeneous context.

This is consistent with a third difference between the two criteria that emerged with regard to the moderating role of verbatim versus gist repetition. The moderator did not affect the within-items criterion; if, as suggested above, the within-items effect relies more on explicit memory processes, this could explain why in this case a repetition of a statement's gist—its conceptual meaning—is sufficient and why an additional repetition of the exact wording is not beneficial. The between-items criterion, in contrast, exhibited larger effects for verbatim repetition than gist repetition, suggesting that the underlying processes are low-level perceptual processes typical of implicit memory effects (for an overview, see Roediger, 1990).

In sum, moderation analyses yielded different patterns for the between- and within-items effects. The findings are consistent with the notion that there are different processes underlying the two criteria: Although both an increase in fluency because of repetition and an experience of a discrepancy with a comparison standard play a role in both criteria, the discrepancy experience in the heterogeneous context is also boosted by the difference between relatively fluent (i.e., repeated) and relatively disfluent (i.e., nonrepeated) statements at the time of judgment; this difference is fully reflected only by the between-items criterion. As most of the studies employed heterogeneous contexts, the truth effect in these studies was influenced by a discrepancy between repeated and nonrepeated statements, increasing the magnitude also of within-items effect sizes computed for these studies. Future research on the truth effect should carefully distinguish between the effects of repetition itself and the effects of the presence of nonrepeated, disfluent items at test. Moreover, research should concentrate on both of the underlying cognitive processes—the increase in fluency and the experience of discrepancy—separately.

The Discrepancy-Attribution Hypothesis. The present results suggest that the roles of discrepancy and of the comparison standard in fluency judgments require further attention. According to the discrepancy-attribution account by Whittlesea and Williams (1998, 2000, 2001a, 2001b), feelings of familiarity result when the processing of a stimulus is experienced as unexpectedly fluent (i.e., more fluent than expected on the basis of a comparison standard), and this feeling is attributed to a source in the past. Thus, if a surprisingly fluently processing experience is interpreted as resulting from an event in the past (i.e., a prior encounter with that stimulus), a feeling of familiarity is elicited. Also well demonstrated is the influence of processing fluency on other memory-based judgments such as feelings of remembering (e.g., Jacoby & Whitehouse, 1989) and recall (e.g., Roediger & McDermott, 1995). Furthermore, there is evidence that the discrepancy-attribution account generalizes to recognition (Westerman, 2008). Beyond memory-based judgments, the discrepancy-attribution account can explain the present findings on judgments of truth as well as those of other fluency-based judgments such as liking judgments (Alter & Oppenheimer, 2009; Willems & Van der Linden, 2006). But despite the success of the discrepancy-attribution account in explaining a wide variety of judgment effects, little is known about the experience of discrepancy (and, for that matter, about the attribution process). Investigations into these processes are necessary to fill this void. The same is true for the comparison standard that plays an important role in the discrepancy-attribution hypothesis (cf. Whittlesea & Leboe, 2003). We distinguished here between a comparison standard based on internal expectancies (as in the homogeneous context) and one generated on the fly from external stimuli (as in the heterogeneous context). The results suggests that there may be substantial differences between the two types of comparison standards; for instance, comparison standards based on internal expectancies are likely to depend on individual characteristics to a greater degree than standards based on external stimuli. More research is needed that further investigates the effects of this and other factors on the comparison standard and, thereby, on the role of fluency on judgments.

Relation to Other Fluency-Based Effects: Mere Exposure

The truth effect is one out of a class of fluency-based effects. Another prominent example of this class is the mere exposure effect (e.g., Bornstein, 1989; Bornstein & D'Agostino, 1992, 1994; Zajonc, 1968). Here, instead of judgments of truth, the experience of fluency informs judgments of liking. According to the processing fluency/attribution model (Bornstein & D'Agostino, 1992, 1994), prior exposure facilitates subsequent retrieval from memory, and this experience of ease is interpreted as positive. As a consequence, repeated stimuli are rated more favorably than new stimuli. Just as in the case of the truth effect, corresponding results were obtained using perceptual fluency manipulations: Highly visible stimuli are preferred over less visible stimuli (Reber et al., 1998; Winkielman & Cacioppo, 2001).

According to the hedonic fluency hypothesis (Reber, Schwarz, & Winkielman, 2004), easily processed stimuli are generally preferred because they automatically elicit positive affect. However, it remains to be shown whether positive affect is also involved in the emergence of the truth effect.

On one hand, there is evidence for a link between positive feelings and truth judgments (Garcia-Marques, Mackie, Claypool, & Garcia-Marques, 2004). On the other hand, recent evidence suggests that, when positivity is dissociated from fluency, it is fluency, not positivity, that affects truth (Unkelbach, Bayer, Alves, Koch, & Stahl, 2009); furthermore, it has been shown that the interpretation of fluency as "true" is learned (Unkelbach, 2007) and involves controlled cognitive processing, at least to some extent (Unkelbach & Stahl, 2008). Therefore, it is not yet clear whether and how a direct positive–true link exists and the extent to which this possible association is established automatically or by controlled processing. Thus, the role of affect in the truth effect is not fully understood, and future research on the truth effect and fluency effects in general is needed to resolve this question.

The present results show that the truth effect and the mere exposure effect share some moderating variables. Some of the results obtained by Bornstein (1989) in a seminal meta-analysis on the mere exposure effect correspond with our findings on the truth effect: Both effects occur independently from exposure duration (except for a tendency toward a smaller truth effect for participant-paced presentation that is not easily interpreted in terms of duration). Furthermore, both effects are moderated by list type: They are found mainly under heterogeneous presentation conditions. The latter finding may stimulate research on the similarity of contextual influences on the fluency experience in the truth and mere exposure effects (e.g., Dechêne et al., 2009; Willems & Van der Linden, 2006).

Somewhat in contrast to the mere exposure effect, we found no clear influence of further repetitions on the truth effect. (Note, however, that the effect of a second repetition on the between-items criterion is consistent with the mere exposure literature that also used this criterion.) The lack of a clear effect may be a result of the small number of studies available for that specific analysis. It may also be argued that a judgment about truth is qualitatively different from a liking judgment: Whereas a liking judgment is by definition based on subjective experiences, a truth judgment can always be seen before the background of an objective value. This may lead to a ceiling effect in judgments because a statement cannot be "truer" than "definitely true" by definition. In either case, beyond noting important similarities, future research should focus on qualitative differences between different types of fluency-based judgments (for an overview of similarities, see Alter & Oppenheimer, 2009).

Methodological Note

The present observations suggest some obvious methodological recommendations. First, researchers interested in effects of processing fluency should select the critical comparison—between items or within items—in a principled manner, based on theoretical considerations. Second, a principled choice of judgment context and study design is probably even more important: Researchers should be sensitive to the possible contextual effects on the comparison standard that drives fluency effects on judgments. More generally speaking, preferring a within-participants over a between-participants design may not only affect statistical power but also affect the context in which observations are made, which in turn may substantially alter psychological processes involved. Systematic experimental investigations of this possibility are highly desirable not only in the truth effect domain.

Practical Relevance

Effects of repetition have received high prominence in the fields of advertising and persuasion (e.g., Batra & Ray, 1986; Holden & Vanhuele, 1999; Nordhielm, 2002; Roggeveen & Johar, 2002, 2007; Schumann, Petty, & Clemons, 1990). Here, communication effectiveness is the focus of research on applications of repetition, and repetition effects are seen as based on two factors or phases that influence message response (Berlyne, 1970): The first is a "wear-in" phase of positive habituation and increasingly positive responses. It may be followed by a "wear-out" phase, characterized by too much repetition, onset of tedium, and therefore decreasing message effectiveness. The findings of the present analysis mainly contribute to our understanding of the "wear-in" phase because data are scarce for the effects of more than one repetition (although there is a tendency that the truth effect wears off after one or two repetitions).

An important factor for advertising effectiveness is the consumers' level of processing (e.g., Anand & Sternthal, 1990; Greenwald & Leavitt, 1984; Hawkins & Hoch, 1992). This factor is thought to influence whether people encode surface features of stimuli (i.e., low level of processing) or whether they evaluate the semantic content of a stimulus (i.e., higher level of processing). The present analysis showed that lower levels of processing reveal greater effects of repetition on truth judgments; that is, a low level of processing at the time of initial encoding leads to a more pronounced truth effect. Thus, in case of an advertising strategy based on repetition, it is beneficial for advertisers when consumers engage in shallower processing. Similar results have been obtained for liking judgments on advertisements (Nordhielm, 2002). Although the present meta-analysis did not obtain a sufficient number of effect sizes to investigate this factor, source credibility may also be highly influential (e.g., Petty & Cacioppo, 1986). More research is needed to investigate the relative roles and the interplay of levels of processing at initial encoding versus at judgment.

A related point is the strategy of advertising variation (e.g., Schumann et al., 1990): This approach holds it to be beneficial when an advertisement's content varies over repetitions. However, the present results suggest that an exact, verbatim repetition is more effective in changing consumers' belief in a marketing claim. The present findings suggest that perhaps a combination of repetition of core elements (i.e., to benefit from implicit memory effects) and variation of peripheral elements (i.e., to counter wear-out effects of tedium) may be most effective. Finally, the finding that the truth effect is largely independent of duration of stimulus exposure, delay, and other variables leads to the conclusion that it is a reliable phenomenon with relevant practical impact on advertising and persuasion processes.

Conclusions

Two main conclusions can be drawn. First, the present review has even more firmly established the existence of the truth effect: Repeated presentation increases participants' subjective judgments of a statement's truth. The effect is of medium size and occurs under a variety of conditions. Second, two components of this effect can be distinguished, both theoretically and empirically: On a theoretical level, the psychological processes underlying the truth effect differ between homogeneous and heterogeneous measurement contexts; empirically, the within-items effect and the between-items effect were of different magnitude, and they were differently affected by moderating variables. These conclusions support an important role of fluency (cf. Alter & Oppenheimer, 2009), and they suggest that fluency effects are moderated by the context in which judgments are performed. The next generation of research on fluency effects on judgments should focus on such contextual moderation.

Authors' Note

Alice Dechêne and Christoph Stahl equally contributed to this work.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The authors received no financial support for the research and/or authorship of this article.

Notes

1. To the best of our knowledge, the zipper was invented in Switzerland, and thus the statement, "The zipper was invented in Norway," is false.
2. This distinction between referential validity and convergent validity is mirrored in remember versus know judgments regarding specific memories (Gardiner & Richardson-Klavehn, 2000).
3. There is theoretical debate as to whether the process by which fluency affects truth judgments is one of misattribution (e.g., Whittlesea & Williams, 1998) or whether it represents a valid use of environmental cues (e.g., Hertwig, Herzog, Schooler, & Reimer, 2008; Unkelbach, 2007). This distinction is not in the focus of the present review, and we therefore remain neutral in this debate.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Alter, A. L., & Oppenheimer, D. M. (2009). Using the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*, 219-235.

Anand, P., & Sternthal, B. (1990). Ease of message processing as a moderator of repetition effects in advertising. *Journal of Marketing Research, 26,* 345-353.

*Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology, 27*, 576-605.

*Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making, 2*, 81-94.

*Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 241-252.

Batra, R., & Ray, M. L. (1986). Situational effects of advertising repetition: The moderating influence of motivation, ability, and opportunity to respond. *Journal of Consumer Research, 12,* 432-445.

*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 Behavioural Science, 17*, 199-214.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception and Psychophysics, 8,* 279-286.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. *Psychological Bulletin, 106*, 265-289.

Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal Personality and Social Psychology, 63,* 545-552.

Bornstein, R. F., & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition, 12,* 103-128.

*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.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage.

*Dechêne, A. (2005). *Hängt die Wahrheit von der Formulierung ab? Einflüsse von syntaktischer Veränderung und Verarbeitungstiefe* *auf den truth effect* [Influence of syntactical changes and level of processing on the truth effect]. Unpublished master's thesis, University of Basel, Switzerland.

*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., & Wänke, M. (2006a). [Truth effect and faith in intuition]. Unpublished raw data.

*Dechêne, A., & Wänke, M. (2006b). [Truth judgments under processing constraints]. Unpublished raw data.

*Doland, C. A. (1999). Repeating is believing: An investigation of the illusory truth effect. *Dissertation Abstracts International*, *60*(4-B), pp. 1879. (UMI No. 9927617).

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measure designs. *Psychological Methods, 1*, 170-177.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology, 71,* 390-405.

Garcia-Marques, T., Mackie, D. M., Claypool, H. M., & Garcia-Marques, L. (2004). Positivity can cue familiarity. *Personality and Social Psychology Bulletin, 30*, 585-593.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229-244). New York: Oxford University Press.

*Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency-validity relationship. *American Journal of Psychology, 97*, 185-195.

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,* 601-613.

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,* 221-233.

Greenwald, A., & Leavitt, C. (1984). Audience involvement in advertising: Four levels. *Journal of Consumer Research, 11,* 581-591.

Hansen, J., Dechêne, A., & Wänke, M. (2008). Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology, 44,* 687-691.

Hansen, J., & Wänke, M. (2008). It's the difference that counts: Expectancy/experience discrepancy moderates the use of ease of retrieval in attitude judgments. *Social Cognition, 26*, 447-468.

*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. (1992). Low-involvement learning: Memory without evaluation. *Journal of Consumer Research, 19*, 212-225.

*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.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin, 93*, 388-395.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. D. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486-504.

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1191-1206.

Holden, S. J. S., & Vanhuele, M. (1999). Know the name, forget the exposure: Brand familiarity versus memory of exposure context. *Psychology and Marketing, 16*, 479-496.

Hurtt, R. K. (1999). *Sceptical about scepticism: Instrument development and experimental validation.* Salt Lake City: University of Utah.

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous over night: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology, 56,* 326-338.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118*, 126–135.

*Kim, C. (2002). The role of individual differences in general scepticism in the illusory truth effect. *Dissertation Abstracts International, 63*(05), 1913A. (UMI No. 3053399).

Krosnick, J. A. (1991). The stability of political preferences: Comparisons of symbolic and non-symbolic attitudes. *American Journal of Political Science, 35*, 547-576.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 141-164). New York: John Wiley.

Law, S. (1999). Investigating the truth effect in young and elderly consumers: The role of recognition and source memory. *Dissertation Abstracts International, 60*(1-A), Jul 1999, pp. 0191. UMI No. AAMNQ35219).

*Law, S., & Hawkins, S. A. (1997). Advertising repetition and consumer beliefs: The role of source memory. In W. D. Wells (Ed.), *Measuring advertising effectiveness: Advertising and consumer psychology* (pp. 67-75). Mahwah, NJ: Lawrence Erlbaum.

*Law, S., Hawkins, S. A., & Craik, F. I. M. (1998). Repetition-induced belief in the elderly: Rehabilitating age-related memory deficits. *Journal of Consumer Research, 25*, 91-107.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Mitchell, J. P., Dodson, C. S., & Schacter, D. L. (2005). fMRI evidence for the role of recollection in suppressing misattribution errors: The illusory truth effect. *Journal of Cognitive Neuroscience, 17*, 800-810.

*Mitchell, J. P., Sullivan, A. L., Schacter, D. L., & Budson, A. E. (2006). Misattribution errors in Alzheimer's disease: The illusory truth effect. *Neuropsychology, 20,* 185-192.

*Mutter, S. A., Lindsey, S. E., & Pliske, R. M. (1995). Aging and credibility judgment. *Aging and Cognition, 2,* 89-107.

Nordhielm, C. L. (2002). The influence of level of processing on advertising repetition effects. *Journal of Consumer Research, 29,* 371-382.

Obermiller, C., & Spangenberg, E. R. (1998). Development of a scale to measure consumer skepticism toward advertising. *Journal of Consumer Psychology, 7,* 159-186.

Parks, C. M., & Toth, J. P. (2006). Fluency, familiarity, aging, and the illusion of truth. *Aging, Neuropsychology, and Cognition, 13*, 225-253.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). New York: Academic Press.

Raudenbush, S. W. (1994). Random effect models. In H. Cooper & L. V. Hedges (Eds.), The *Handbook of research synthesis* (pp. 301-321). New York: Russell Sage.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*, 338-342.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8,* 364-382.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science, 9*, 45-48.

Roediger, H. L., III. (1990). Implicit memory: Retention without thinking. *American Psychologist, 45*, 1043-1056.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803-814.

*Roggeveen, A. L., & Johar, G. V. (2002). Perceived source variability versus familiarity: Testing competing explanations for the truth effect. *Journal of Consumer Psychology, 12*, 81-91.

Roggeveen, A. L., & Johar, G. V. (2007). Changing false beliefs from repeated advertising: The role of claim-refutation alignment. *Journal of Consumer Psychology, 17,* 118-127.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99,* 400-406.

Schmidt, F. L., Oh, I., & Hayes, T. L. (2009). Fixed versus random effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 13,* 97-128.

*Schwartz, M. (1982). Repetition and rated truth value of statements. *American Journal of Psychology, 95*, 393-407.

Schumann, D. W., Petty, R. E., & Clemons, D. S. (1990). Predicting the effectiveness of different strategies of advertising variation: A test of the repetition-advertising hypothesis. *Journal of Consumer Research, 17,* 192-202.

Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General, 135*, 409-428.

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31*, 713-724.

*Toppino, T. C., & Brochin, H. A. (1989). Learning from tests: The case of true-false examinations. *Journal of Educational Research, 83*, 119-124.

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science, 12*, 339-345.

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgements of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 219-230.

Unkelbach, C., Bayer, M., Alves, H., Koch, A., & Stahl, C. (2009). *Fluency and Positivity as possible causes of the truth effect*. Manuscript submitted for publication.

Unkelbach, C., & Stahl, C. (2008). A multinomial modelling approach to dissociate different components of the truth effect. *Consciousness and Cognition*, *18*, 22-38.

Westerman, D. (2008). Relative fluency and illusions of recognition memory. *Psychonomic Bulletin & Review, 15,* 1196-1200.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1235-1253.

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language, 49*, 62-79.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica, 98*, 141-165.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 547-565.

Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis: I. The heuristic basis of feeling of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 3-13.

Whittlesea, B. W. A., & Williams, L. D. (2001b). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 14-33.

Willems, S., & Van der Linden, M. (2006). Mere exposure effect: A consequence of direct and indirect fluency-preference links. *Consciousness and Cognition, 15*, 323-341.

Wilson, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science, 587*, 69-81.

Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology, 81*, 989-1000.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement, 9*, 1-27.