Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 1, 219–230

0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.1.219

Reversing the Truth Effect: Learning the Interpretation of Processing Fluency in Judgments of Truth

Christian Unkelbach
Universität Heidelberg

Repeated statements receive higher truth ratings than new statements. Given that repetition leads to greater experienced processing fluency, the author proposes that fluency is used in truth judgments according to its ecological validity. Thus, the truth effect occurs because people learn that fluency and truth tend to be positively correlated. Three experiments tested this notion. Experiment 1 replicated the truth effect by directly manipulating processing fluency; Experiment 2 reversed the effect by manipulating the correlation between fluency and truth in a learning phase. Experiment 3 generalized this reversal by showing a transfer of a negative correlation between perceptual fluency (due to color contrast) and truth to truth judgments when fluency is due to prior exposure (i.e., repetition).

Keywords: processing fluency, truth judgments, familiarity, cue utilization

"The world's highest tree is a spruce." Is this statement true? To the best of my knowledge, it is false; however, researchers have found a simple way to increase people's subjective impression that the statement is true: Repetition. Repeated statements have a higher rated truth or are judged true with a higher probability compared with statements that were not presented before (e.g., Bacon, 1979; Hasher, Goldstein, & Toppino, 1977; Schwartz, 1982). Why does this truth effect occur? Given that the truth effect is partially based on the more fluent processing of previously presented statements (as shown by Reber & Schwarz, 1999), I argue that this experienced fluency is used as a cue in judgments of truth according to the cue's ecological validity. That is, the truth effect occurs because repetition leads to more fluent processing of a statement and people have learned that the experience of processing fluency correlates positively with the truth of a statement.

So far, there are two explanations for the influence of repetition on the rated truth of statements: First, if a statement is recognized as one that has been encountered before, there is an effect of convergent validity (Arkes, Boehm, & Xu, 1991). Although repetition by itself does not increase validity, repetition from a second, independent source does increase the convergent validity of a statement. So, when people remember that they have encountered a statement before but fail to remember the source (i.e., an earlier experimental session), they show a truth effect. However, the truth effect also occurs when people correctly remember that a statement has been repeated within the same experimental session (Bacon, 1979), and thus, there is no logical basis for an effect of convergent validity. Therefore, a second explanation is necessary, which is often referred to as the *familiarity explanation* (e.g., Arkes et al., 1991; Hawkins & Hoch, 1992). Bacon (1979) hypothesized that individuals use familiarity as a basis for recognition, but Hawkins and Hoch (1992) argued that familiarity is actually one's familiarity with the semantic content of a statement, which should increase judged validity independently from the ability to recognize a statement or detect repetition. Yet, Begg, Armour, and Kerr (1985) found no effect of a level-of-processing manipulation at the first presentation of statements, which poses a problem for a semantic familiarity explanation, because deeper encoding should lead to greater familiarity with the semantic content.

Integrating these explanations and most of the available findings, a comprehensive account for the truth effect stems from Begg, Anas, and Farinacci (1992), who showed independent contributions of recollection and familiarity on the rated truth of statements. They proposed that familiarity increases automatically with repetition and that the influence of familiarity on rated truth is unintentional, whereas recollection requires intention. Thus, they classified the familiarity component of the phenomenon as part of a larger group of effects that are caused by stimulus-elicited experiences; these experiences are misattributed to qualities of the stimulus, in this case, the truth of a statement. Accordingly, the truth effect's familiarity part belongs to a class of effects like the false fame effect (Jacoby, Kelley, Brown, & Jasechko, 1989), the revelation effect (Watkins & Peynircioglu, 1990), or the mere exposure effect (Mandler, Nakamura, & Van Zandt, 1987; Zajonc, 1968).

A uniting construct in explanations of these effects is processing fluency, that is, the experienced ease of ongoing perceptual or conceptual cognitive processes (e.g., Jacoby, Kelley, & Dywan, 1989; Whittlesea, 1993). Thus, repeated statements seem to be more valid simply because they are more easily processed than new statements. This was directly tested by Reber and Schwarz (1999), who manipulated processing fluency, not by repeating a statement, but by changing the contrast of the color in which statements were presented. They found that statements with high color contrast were judged to be true more frequently than statements with low color contrast. In addition, by using the continuous

The reported research was supported by a grant of the Deutsche Forschungsgemeinschaft DFG (Fi 294/21). Special thanks go to Myriam Bayer, Daniel Danner, and Martin Stegmüller for their help in all stages of the research and Klaus Fiedler for his support. Helpful comments by Peter Freytag and Rolf Reber are gratefully acknowledged.

Correspondence concerning this article should be addressed to Christian Unkelbach, Psychologisches Institut, Universität Heidelberg, Hauptstrasse 47-51, 69117, Heidelberg, Germany. E-mail: christian.unkelbach@psychologie.uni-heidelberg.de

threshold identification task developed by Feustel, Shiffrin, and Salasoo (1983) in a pilot study, Reber and Schwarz showed that high-contrast strings elicited faster correct identification than low-contrast strings, corroborating the assumption that contrast reliably influences perceptual fluency. Hence, processing fluency is a potent candidate that may underlie the familiarity part of the truth effect, independent from intentional and conscious recollection.

But how does processing fluency influence truth judgments? One possibility is that the experience has an inherent meaning that can immediately be used as input for judging whether a statement is true, much the same way as affective feelings can be used in evaluative judgments (e.g., Schwarz & Clore, 1983). Indeed, there is good evidence that the fluency experience is inherently positive (e.g., Reber, Schwarz, & Winkielman, 2004; Winkielman & Cacioppo, 2001). Alternatively, it has been proposed that the experience is unspecific to begin with—a feeling that there is something "about" the stimulus—and this experience is then interpreted within the given context (e.g., Bornstein & D'Agostino, 1994; Whittlesea, 1993). This approach explains how the same construct can influence a variety of judgments and decisions: Fluently processed stimuli are judged to be frequent (Tversky & Kahneman, 1973), fluently processed names are judged to belong to famous persons (Jacoby, Kelley, Brown, & Jasechko, 1989), and fluently processed items are classified as old in a recognition test (Johnston, Hawley, & Elliott, 1991). Yet, treating processing fluency as an unspecific experience lacks a crucial part, which is pointedly visible in the truth effect. If people experience that there is something "about" a statement, why is this statement then classified as true rather than false? The direction of the impact of processing fluency makes sense intuitively for recognition judgments, when fluency is experienced as familiarity. However, there is no logical or a priori ground for why it should not be the other way round for truth judgments, that is, why fluently processed statements are not classified as false. Just think of any multiple-choice test or any murder mystery; the obvious and easy answers are hardly ever true, or as Sir Arthur Conan Doyle (1891/1990) states through Sherlock Holmes in *The Boscombe Valley Mystery*, "There is nothing more deceptive than an obvious fact" (p. 68).

Thus, it is not difficult to conceive a context in which a fluently processed statement is false. Borrowing from the work of Brunswik (1957), I suggest a learning approach to explain the direction of fluency effects on judgments (Unkelbach, 2006), which has two central assumptions. First, the truth of a statement (or the frequency of an instance or the fame of a name, for that matter) is a distal stimulus quality that cannot be observed or experienced directly; instead, it needs to be inferred from proximal cues. The fluency of one's own cognitive processes is such a proximal cue, and hence, fluency is used as a cue in judgments of truth.[1] Second, people learn that the fluency experience correlates positively with the truth of a statement, and therefore, fluently processed statements are judged to be true.

The first assumption needs to be treated as a given, whereas the second assumption is directly testable. In my first presentation of this learning approach (Unkelbach, 2006), I showed that the classic finding that fluently processed stimuli are judged to be old in recognition tests (e.g., Johnston, Dark, & Jacoby, 1985) can be reversed if participants learn that the experience of fluency correlates with a stimulus being new rather than old. The same idea is applied here; if participants learn that fluency correlates negatively with the truth of a statement, it should be possible to reverse the truth effect. A similar argument was made on a different level by Skurnik, Schwarz, and Winkielman (2000), who stated that people hold a metacognitive belief that familiarity is diagnostic of truth, "and this belief leads people to infer truth from familiarity" (p. 168). If the metacognitive belief conveys that familiarity/fluency is diagnostic of falsity, the truth effect is also reversed. However, it is not clear where these metacognitive beliefs come from. The learning approach offers a possible explanation: The belief could be the consciously available summary of the learned correlation between processing fluency and truth.

Application of this learning approach to judgments of truth is worthwhile for at least two reasons: First, it would provide further evidence that it is indeed processing fluency that is partly responsible for the truth effect. Second, the direction of the impact of fluency (which is at the heart of the familiarity-based explanation of the truth effect) on truth ratings is explained as the learned interpretation of the experience in a given context. As such, the approach bypasses the notion of familiarity and explains the effect as a direct utilization of the processing experience as a cue. The distinction between familiarity and fluency becomes apparent if one conceptualizes a feeling of familiarity as a feeling that there is something "about" a stimulus (cf. Higgins, 1998), which Begg and Armour (1991) labeled the "ring of truth." Within a given context (or most contexts), this unspecified feeling might be labeled *familiarity*; however, many other interpretations are conceivable. Thus, fluency describes the experience, whereas familiarity is already an interpretation of that experience. Yet, if the clear distinction of familiarity and recollection is valid, as proposed by Jacoby and colleagues (e.g., Jacoby & Kelley, 1992) and demonstrated for the truth effect by Begg et al. (1992), it should not matter whether the explaining construct is termed *familiarity* or *fluency*. However, as said, the term *familiarity* implies an interpretation (namely the one of prior experience); therefore, I use the term *fluency* throughout the remainder of this article, because the term carries less surplus meaning and allows for the possibility that different interpretations of the experience can be learned in different contexts.

In the following, I present three experiments to support the idea that people use processing fluency in judgments of truth according to the learned validity of the experience in a given context. Experiment 1 serves as an introduction to the general paradigm. In this experiment, the truth effect was replicated through a direct manipulation of processing fluency, which was suggested by Reber and Schwarz (1999). Instead of statements being presented repeatedly, the contrast of the color of the font in which the statements were presented and the background was manipulated. This manipulation resolves the confound that is inherently present in an increase in fluency due to prior presentation with effects of referential validity due to recollection. In Experiment 2, the correlation of processing fluency and truth was then manipulated. In a learning phase, participants judged the truth of a statement and were given feedback after each decision. Crucially, there was either a negative or a positive correlation between processing fluency and the truth of a statement. If there is a negative

[1] The idea of fluency as a cue sensu Brunswik (1957) is already present in Jacoby, Kelley, and Dywan's (1989) chapter on memory attributions, and also the correlation of processing experiences with external criteria is mentioned by Begg et al. (1992, p. 456). However, to my knowledge, the full implications of this idea have not been explored.

correlation in the learning phase, the truth effect should be reversed at test. In this second experiment, processing fluency was again manipulated through color contrast. Of course, this creates the problem that people might simply learn contrast as a cue to make judgments of truth. To counter this argument, in Experiment 3 the classic repeated exposure paradigm was used. In a presentation phase, participants saw a large number of statements. The learning phase was then identical to that in Experiment 2; that is, all-new statements were judged, feedback for each decision was given, and fluency (i.e., contrast) correlated positively or negatively with the truth of a statement. At test, no manipulation of processing fluency took place, aside from the fact that half of the statements were already present in the initial phase and half of the statements were completely new. If the positive or negative correlation of processing fluency and truth in the learning phase influenced decisions at test, the only connecting construct would indeed be processing fluency.

Finally, a note on the method of analysis is in order. In most studies, the truth effect is shown as an increase in rated truth of a statement, on average from 4.28 to 4.54 for true statements and from 4.28 to 4.53 for false statements (on a 7-point scale ranging from 1 = *definitely true* to 7 = *definitely false*; data reported by Brown & Nix, 1996). The following experiments deviate from this standard procedure by soliciting, not ratings of validity or truth, but binary decisions whether a statement is true or false. Begg et al. (1992) already used the proportion of "true" ratings in their analysis; however, they converted their scales to binary decisions of "true" and "false" and reported that the results did not differ. A binary decision has the advantage of allowing for immediate feedback on the correctness of a decision, which is important for the learning phase and not possible with rating scales. It is also possible to measure the response latency for a decision, which can be used as a proxy for processing fluency; however, it is important to note that this index is imperfect, because many other factors not related to the experience of fluent processing can contribute to response latencies. Most important, however, binary decisions offer the possibility of estimating signal detection theory (SDT) parameters from the true–false ratings across statements. A signal detection analysis provides two parameter estimates, discrimination ability d' and response bias β. Given that a "true" decision is equivalent to a "yes, a signal is present" decision in a classic SDT experiment, d' is an index of participants' knowledge regarding the presented statements (i.e., their ability to discriminate between true and false statements), whereas β is an index of the truth effect; that is, one would expect that people have a greater tendency to respond "yes, true" to repeated, high-fluency statements compared with new, low-fluency statements. Accordingly, the truth effect is conceptualized as a differential response bias for high- and low-fluency statements. And because d' and β are theoretically independent, an additional advantage is that any knowledge participants might have regarding the actual truth status of a statement is captured by d', leaving β as an estimate of the truth effect unaffected.

Experiment 1

Experiment 1 serves two purposes: first, to show that the manipulation of color contrast allows for replication of the truth effect in a single-exposure paradigm (Reber & Schwarz, 1999) and, second, to introduce the truth effect in terms of SDT. As discussed, the effect is normally shown as an increase in the rated credibility of statements. Aside from these main purposes, easy and difficult statements were used in Experiment 1 to show that the effect emerges only when there is doubt about the distal truth status of a statement, not when there is factual knowledge.

Method

Participants, design, and materials. Twenty psychology students (16 women, 4 men) from the University of Heidelberg, Heidelberg, Germany, participated either for payment or for course credit. The key manipulation pertained to the fluency of a statement (high vs. low). The statements were 60 easy and 60 difficult statements from a wide variety of topics (science, politics, geography, art, and general knowledge). True statements were compiled from a variety of encyclopedia resources, whereas the false statements were made-up facts. All statements had a clear true or false status. Table 1 gives examples for statements from all four categories.

Processing fluency was manipulated by color contrast; as mentioned, Reber and Schwarz (1999) and Reber, Zimmermann, and Wurtz (2004) showed that high contrast influences perceptual fluency. Thus, statements were presented in Arial font with blue, red, green, or yellow letters, and high-fluency statements had a high contrast against the white background, whereas low-fluency statements had a low contrast. This was accomplished through manipulation of the RGB (red, green, blue) component of a color. For example, an RGB combination of R = 255, G = 0, and B = 0 results in a strong red color with high contrast against a white background. An RGB combination of R = 255, G = 200 and B = 200 results in a light red color with low contrast against a white background. In the actual experiment, a red high-contrast statement would be assigned a random value between 100 and 120 for the G and the B

Table 1
Examples of True/False and Easy/Difficult Statements Used in All Three Experiments

Validity	Easy	Difficult
True	Sunlight contains ultraviolet radiation.	First Olympic gold medalist (new games) was James Connolly.
	The formula for water is H_2O.	Methuselah was the grandfather of Noah.
	Dolphins belong to the mammals.	Europe's biggest glacier is the Vatnajökull on Iceland.
False	Aristotle was a Japanese philosopher.	The speed of sound is independent from temperature.
	Lead is lighter than aluminum.	The capital of Madagascar is Toamasina.
	Pluto is the biggest planet of the solar system.	Cactuses can procreate via parthogenesis.

component, whereas the R component was fixed at 255. A red low-contrast statement would be assigned a random value between 200 and 220 for the G and the B component. This was done accordingly for each color, with respective changes in the assignment of the RGB components. This resulted in two distributions of high and low contrasts for each color. The presentation of the statements, measures of the responses and response latencies, as well as the manipulations of color contrast were implemented in a computer program written in Microsoft Visual Basic.

Procedure. Experimental sessions included up to 4 participants. Upon arrival, participants were seated in cubicles and read an informed consent form: They were told that they would participate in an experiment to investigate the impact of color on judgments of truth. When they agreed to participate and signed the form, the experimenter started the computer program. In the first part of the experiment, participants responded to the 60 easy statements, of which half were true and half were false. The statements were randomly assigned the status of a high- or low-fluency statement with the restriction that fluency and true–false status were orthogonal. Participants responded to each statement by pressing one of two keys on the computer keyboard; one key in the lower left of the keyboard indicated a "yes, true" response, and another key on the lower right indicated a "no, false" response. The respective question was always on the top of the screen, asking "Is this statement true?" The statements were on the screen until a response was given. The computer recorded the responses and the latencies. After these 60 statements, participants had a short break and were informed that they would continue with the same task, but with more difficult questions. Again, the statements were randomly assigned a high- or low-fluency status such that fluency and truth were orthogonal. Aside from the difficulty of the statements, everything else was identical to the first half. After participants completed this task, they could freely respond to three open questions: whether they suspected any other goals of the experiment besides the ones mentioned in the introduction; whether they noticed anything about the colors; and finally, whether they thought that the colors influenced their decisions, and if so, how. After completing these open questions, they were debriefed, thanked, and paid. Experimental sessions lasted between 15 and 20 min.

Results

Response latencies. Before any analysis, the measured latencies were trimmed at 1,000 ms and 5,000 ms for easy or 7,000 ms for difficult statements; that is, any latency greater than 5,000 ms (or 7,000 ms) was set to 5,000 ms (or 7,000 ms), and any latency less than 1,000 ms was set to 1,000 ms. For easy statements, participants responded on average faster to high-fluency than to low-fluency statements (M = 2,080 ms, SD = 497 vs. M = 2,212 ms, SD = 551, respectively), $t(19) = -4.21$, $p < .001$, $d = 1.93$. The same was true for difficult statements (M = 4,347, SD = 1,266 vs. M = 4,686, SD = 1,233, for high- and low-fluency statements, respectively), $t(19) = -5.16$, $p < .001$, $d = 2.37$. Thus, if one accepts response latencies as a proxy, high-color-contrast statements were more fluently processed than low-fluency statements.

SDT analysis. Overall, participants classified 53% (SD = 4%) of the 60 easy and 54% (SD = 9%) of the 60 difficult statements as true. From the "yes, true" and "no, false" responses, SDT parameters d' and β were estimated (Stanislaw & Todorov, 1999; Swets, Dawes, & Monahan, 2000). Again, d' measures actual discrimination ability, in this case, the ability to discriminate true from false statements, whereas β measures the response bias. A truth effect should be visible in a relatively higher tendency to respond "true" to high- compared with low-fluency statements. For estimation of the parameters, responding "yes, true" a true statement was classified as a "hit," whereas responding "yes, true" to a false statement was classified as a "false alarm." The hit and false-alarm rates as well as the resulting SDT parameter estimates for easy and difficult statements by high and low fluency are given in Table 2.

For easy statements, participants discriminated better between true and false statements when they were presented in high color contrast than in low contrast (i.e., high vs. low fluency), although this difference was not reliable on a standard alpha level, $t(19) = 1.78$, $p < .10$, $d = 0.82$. There was also a general tendency to respond "yes, true" to all statements (β = 0.841); however, the difference between high- and low-fluency statements was not reliable, $t(19) = 0.90$, *ns*.

For difficult statements, the high- or low-fluency status of a statement also influenced the discrimination ability, $t(19) = 2.26$, $p < .05$, $d = 1.04$, such that participants were less able to

Table 2
Hit/False-Alarm Rates and Signal Detection Theory Parameter Estimates From Experiment 1 for Easy and Difficult Statements by High- and Low-Fluency (i.e., Color Contrast)

Statements	β estimates	d' estimates	Hit rate	False-alarm rate
Easy				
High fluency	0.795 (0.436)	2.685 (0.375)	.93 (.04)	.12 (.06)
Low fluency	0.886 (0.485)	2.442 (0.414)	.90 (.05)	.15 (.07)
Difficult				
High fluency	0.900 (0.162)	0.545 (0.379)	.68 (.13)	.48 (.14)
Low fluency	1.005 (0.118)	0.245 (0.386)	.55 (.11)	.45 (.14)

Note. Standard deviations are in parentheses. β values smaller than 1 indicate a tendency to respond "yes, true," whereas values greater than 1 indicate a tendency to respond "no, false." Higher d' values indicate a higher discrimination ability.

discriminate true from false statements when the color contrast was low compared with when it was high. Most important, the fluency manipulation reliably influenced the response bias; participants had a relative tendency to classify high-fluency statements as true and low-fluency statements as false, $t(19) = 2.58$, $p < .05$, $d = 1.18$.[2]

An overview of the open response given at the end of the experiments revealed that the effect is rather unobtrusive. None of the 20 participant guessed that the colors influenced their true/false decisions, although 12 participants mentioned that some statements were difficult to read, which was probably due to extreme cases of the random color distributions. However, no one mentioned a connection between this difficulty and an inclination to respond "no, false," and 3 participants speculated that the colors facilitated or inhibited some memory process, thereby leading to faster or slower responses; for example, 1 participant mentioned that presenting the H_20 statement from Table 1 in blue color speeded up her response.

Discussion

Experiment 1 presents a conceptual replication of the study reported by Reber and Schwarz (1999). By manipulating the color contrast of the presented statements against the white background, it was possible to show a truth effect without repetition: Participants had a tendency to classify high-contrast statements as true and low-contrast statements as false. The replication itself is noteworthy, because in Reber and Schwarz's (1999) initial demonstration, the effect was rather small, and in a recent replication, Parks and Toth (2006) also found only a small effect of perceptual fluency on judgments of truth and familiarity. Here, the effect is reliably and easily found.

On the other hand, no truth effect was reliably observed for easy statements. Beside showing that the impact is greater when there is doubt about the actual truth status, this also shows that the effect is not due to participants' failure to read the statements. For easy statements, they reliably discriminated true from false statements. Thus, they did not just call statements false because they were unable to read them. In addition, for both easy and difficult statements, high color contrast led to faster responses than low color contrast. Again, it is tempting to take response latencies as a proxy for processing fluency and to conclude that faster processing led to "yes, true" responses; however, as already discussed, response latencies need to be treated carefully as an index for fluency (cf. Reber, Wurtz, & Zimmermann, 2004).

Experiment 1 also demonstrated that signal detection analysis is a valuable tool for showing the truth effect. Because d' and β are theoretically independent, there are fewer constraints on the statements one can use. Any actual knowledge about a domain, and therefore, correct true and false classifications, should only influence estimates of d' and be only tangential for the response bias, as long as true and false statements are equally distributed. The unexpected result is that for both easy and difficult statements, participants discriminated better between true and false statements when the color contrast of a statement was high, even if this difference was only significant for difficult statements. A possible explanation, although not directly testable with the available data, is that low-contrast statements were indeed more difficult to process and thereby hindered the application of relevant knowledge, which resulted in lower d' estimates. Note also that this result corroborates further that a failure to read low-contrast statements and to respond "no, false" accordingly is not responsible for the effect. If low contrast influences decisions uniformly (i.e., responding "false" to all low-contrast statements), there should be no impact on d', because contrast/fluency and truth were orthogonal. That is, such a strategy would result in an equal amount of misses and correct rejections (or hits and false alarms).

Experiment 2

Having introduced the paradigm and the method of analysis in Experiment 1, in Experiment 2 I manipulated the central variable that is assumed to cause the direction of the truth effect, namely the correlation between processing fluency and statement veridicality. Again, the idea is that this correlation can be learned and thereby the interpretation of fluency can be changed in a given context. To do this, in Experiment 2 I used the same procedure that was applied to recognition judgments in my first demonstration of this learning effect (Unkelbach, 2006, Experiment 1). In a learning phase, participants respond to statements and receive feedback about the correctness of their response. The sole between-participants manipulation is then the correlation of processing fluency and truth in the learning phase. In a *classic* condition, all true statements are easy to process (high contrast), and all false statements are difficult to process (low contrast). In a *reversed* condition, however, all false statements are easy to process, and all true statements are difficult to process. Finally, in a *control* condition, truth and processing fluency are again orthogonal. In the test phase, no feedback is given, and truth and fluency are again orthogonal for all conditions. If people learn the correlation of processing fluency and truth, there should be a differential impact of this correlation on the response biases for high- and low-fluency statements in the test phase.

Method

Participants, design, and materials. Forty-six students (10 women, 36 men) from various faculties of the University of Heidelberg participated for payment of €4 (approximately U.S.$5). They were randomly assigned to either the classic, control, or the reversed condition and, thereby, to the respective correlation of fluency with the truth status of the statements in the learning phase. The experiment used the 60 easy statements from Experiment 1 for the learning phase and the 60 difficult statements from Experiment 1 for the test phase. To ensure that it is indeed the contrast (and thereby, the fluency of processing) that is associated with the truth and not a specific color, only blue and red were used in the learning phase, and only yellow and green were used in the test phase. The contrast of each color against the white background was manipulated in the same manner as in Experiment 1. A modified version of the computer program from Experiment 1 was used; the changes pertained to the assignment of high- and low-fluency status to the true and false statements. In the

[2] All reported analyses on the SDT parameter are actually done on the log-transformed values of β, because the distribution of β has a substantial positive skewness. This log transformation was suggested by Tabachnik and Fidell (1996, p. 82). However, all reported effects remain significant at .05 when the analysis is done on the untransformed estimates of β.

classic condition, all true (false) statements had a high (low) color contrast, whereas in the reversed condition, all false (true) statements had a high (low) color contrast. Therefore, fluency and truth were perfectly correlated in both conditions, but with different signs of the correlation. In the control condition, fluency and truth were orthogonal.

Procedure. Experimental sessions included up to 6 participants. Upon arrival, the experimenter seated participants in a cubicle, and they read the same informed consent as in Experiment 1. If they agreed to participate, the experimenter started the computer program. In the first half, participants again responded to the 60 easy questions by pressing a "yes, true" or "no, false" key. Responses and latencies were recorded. Depending on condition, truth status and color contrast were positively, negatively, or not correlated. True and false statements were randomly assigned to high or low fluency for each participant, and color contrasts were randomly created through manipulation of the RGB combination of each font within the parameters specified in Experiment 1. Participants received veridical feedback of whether a response was correct or wrong by a label in the center of the screen that either stated "Correct: That was a true (false) statement" or "Wrong: That was a true (false) statement." The feedback appeared on the screen for 2 s after they pressed one of the response keys, and after a break of 1 s, the next trial started. After the 60 easy questions, participants were informed that the experiment would now continue with more difficult questions and without feedback. In this second half, fluency and truth were orthogonal for all conditions. A statement appeared on the screen until participants responded, and after a break of 1 s, the next trial started. Finally, they answered the same three open questions as in Experiment 1. Upon completion, the experimenter thanked, paid, and debriefed the participants. Experimental session lasted between 14 and 22 min.

Results

Three participants were excluded from the analysis, because they responded uniformly to all either high- or low-fluency statements (2 were in the control, 1 in the reversed condition). Thus, 16, 13, and 14 participants remained in the classic, control, and reversed conditions, respectively.

Response latencies. The measured latencies were again trimmed at 1,000 ms and 5,000 ms for easy questions or 7,000 ms for difficult questions. The resulting means and standard deviations are displayed in Table 3. A 3 (condition: classic, control, reversed) $\times$ 2 (fluency: high vs. low; as repeated measures) mixed analysis of variance (ANOVA) was used for analysis of these data, separately for the learning and the test phase. In the learning phase, there was the expected main effect for fluency; overall, participants responded faster to high-fluency statements (M = 2,110 ms, SD = 660) than to low-fluency statements (M = 2,246 ms, SD = 722), $F(1, 40) = 7.70$, $p < .01$, $d = 0.88$. However, the interaction of condition and fluency was also a highly significant, $F(2, 40) = 24.27$, $p < .001$, which is due to the fact that in the classic and reversed conditions, fluency and truth were perfectly confounded, which was a necessity of the design. That is, participants were faster to respond to true than to false statements. In the test phase, however, when truth and fluency were orthogonal, the only significant effect was the main effect for fluency. Participants responded faster to high-fluency statements (M = 3,865 ms, SD = 1,166) than to low-fluency statements (M = 4,110 ms, SD = 1,151), $F(1, 40) = 10.69$, $p < .01$, $d = 1.03$.

SDT analysis. Overall, participants classified .53 (SD = .05) of the 60 easy and .57 (SD = .10) of the 60 difficult statements as true. Similar to Experiment 1, SDT parameters d' and β were estimated from the "yes, true" and "no, false" responses. Separate SDT analyses of the responses to high- and low-fluency statements in the learning phase were not possible, because truth and fluency were perfectly confounded in the reversed and classic conditions. Overall, participants had a hit rate of .93 (SD = .06) and a false-alarm rate of .12 (SD = .09). The SDT analysis based on these rates of the learning phase replicated Experiment 1: Participants showed a high discrimination ability for the easy statements (d': M = 2.887, SD = 0.687) and an overall tendency to respond "yes, true" (β: M = 0.861, SD = 0.790). Importantly, an ANOVA with condition as factor of these parameter estimates yielded no difference between conditions (Fs < 1.5, *ns*); thus, the correlation between fluency and truth had no influence on the performance during the learning phase.

For the test phase, the hit and false-alarm rates as well as the resulting parameter estimates are given in Table 4, separately for high- and low-fluency statements. The same mixed ANOVAs as for the latencies were performed on the SDT estimates (cf. Footnote 2 for β), with condition and fluency (as repeated measures) as factors. For d', these analyses showed only a main effect for fluency; participants could better discriminate between true and false statements if they had a high color contrast (M = 0.545, SD = 0.501) compared with if they had a low color contrast (M = 0.254, SD = 0.496), $F(1, 40) = 6.86$, $p < .05$, $d = 0.83$. Of greater interest, however, were the response bias results: There were no main effects for fluency or condition, but there was the predicted interaction Fluency $\times$ Condition, $F(2, 40) = 6.00$, $p < .01$. As can

Table 3
Mean Latencies (in ms) in the Learning (Easy) and Test (Difficult) Phases of Experiment 2 for High- and Low-Fluency Statements by Condition

	Learning phase		Test phase	
Condition	High fluency	Low fluency	High fluency	Low fluency
Classic (n = 16)	1,923 (534)	2,363 (695)	3,731 (1,196)	3,980 (1,230)
Control (n = 13)	2,121 (707)	2,340 (843)	3,676 (1,331)	3,860 (1,228)
Reversed (n = 14)	2,314 (727)	2,025 (627)	4,195 (964)	4,492 (951)

Note. Standard deviations are in parentheses.

Table 4
Hit/False-Alarm Rates and Signal Detection Theory Parameter Estimates From the Test Phase in Experiment 2 by High- and Low-Fluency Status and Condition

Fluency	β estimates	d' estimates	Hit rates	False-alarm rates
Classic (n = 16)				
High	0.811 (0.214)	0.548 (0.564)	.73 (.13)	.54 (.18)
Low	0.983 (0.153)	0.235 (0.502)	.61 (.13)	.53 (.17)
Control (n = 13)				
High	0.957 (0.322)	0.538 (0.480)	.67 (.15)	.48 (.17)
Low	1.035 (0.164)	0.163 (0.489)	.56 (.16)	.50 (.14)
Reversed (n = 14)				
High	1.105 (0.563)	0.549 (0.479)	.66 (.07)	.63 (.16)
Low	0.884 (0.165)	0.359 (0.513)	.46 (.17)	.50 (.12)

Note. Standard deviations are in parentheses. β values smaller than 1 indicate a tendency to respond "yes, true," whereas values greater than 1 indicate a tendency to respond "no, false." Higher d' values indicate a higher discrimination ability.

be seen from Table 4, participants in the classic and the control conditions showed a relative tendency to respond "yes, true" to high-fluency statements compared with low-fluency statements. For participants in the reversed condition, this tendency was completely inverted: They showed a bias to respond "yes, true" to low-fluency statements compared to high-fluency statements. In addition to this overall analysis, the truth effect within each condition was tested, that is, the difference in for high- and low-fluency statements. As Table 4 shows, participants in the classic condition evinced the classic truth effect, $t(15) = -3.20$, $p < .01$, $d = 1.65$, whereas the reversed condition showed a less pronounced inverted pattern, $t(13) = 1.97$, $p < .08$, $d = 1.09$. There was no reliable difference in the control condition, $t(12) = 1.18$, *ns*. However, if the classic and reversed conditions are contrasted in an ANOVA, the predicted interaction effect—that is, the differential response bias for high- and low-fluency statements—was highly significant, $F(1, 40) = 11.43$, $p < .01$, $d = 1.07$.

Discussion

Experiment 2 provides further evidence that the interpretation of processing fluency can be learned in a given context (Unkelbach, 2006). If fluency and truth were negatively correlated in the learning phase, processing fluency resulted in a tendency to respond "no, false." Conversely, the classic and control condition showed the pattern of a standard truth effect. This differential impact of the correlation in the learning phase is evidence that processing fluency led to "yes, true" responses in the classic and control conditions but to "no, false" responses in the reversed condition. However, the effect was not significant in the control condition, but in the classic condition it was even more pronounced than in Experiment 1. If one takes into account that the learning phase in the former condition conveyed a zero correlation between truth and fluency, and in latter condition a supposedly existing association was reinforced, this result pattern fits well with the learning idea, although it was not predicted.

Moreover, the corresponding latencies show only a main effect for the color-contrast manipulation in the test phase. As discussed, however, latencies are not an ideal or pure measure for processing fluency, which is immediately visible from the interaction pattern of the response latencies in the learning phase. Although there was a main effect caused by the manipulation of color contrast, the interaction was far more powerful. A possible interpretation of this result is provided by Gilbert's (1991) work on judgments about propositions, from which the prediction is derivable that true statements are processed faster than false statements. According to Gilbert, understanding a proposition entails the automatic acceptance of that proposition. The rejection of that proposition is a secondary step that consumes more time and resources. As this possibility shows, there are many other factors that contribute strategically to response latencies in judgments of truth. What ultimately causes the strong interaction in the learning phase is not at the heart of the discussion, and one might be content with the presence of the fluency main effect. Of greater importance are the results from the test phase and the differential impact of processing fluency on judgments of truth in the classic and reversed conditions.

In addition to the crucial interaction for β, there was the same main effect for d' as in Experiment 1, indicating that the impairment of the discrimination ability is a robust effect. As already mentioned, this effect for d' cannot be due to participants' tendency to call low-fluency statements false, because fluency and truth were orthogonal in the test phase. Such a tendency should be visible in a fluency main effect for β, but as it would create as many misses as correct rejections, it would not influence d'. Aside from some speculations about cognitive resources, it is beyond the scope of the present article to investigate the actual cause of the effect. However, the resource explanation finds some support in participants' answers to the funneled question at the end of the experiment. Five participants mentioned that they had sometimes difficulties reading statements in light colors, and they reported that they could not concentrate as much if a statement was presented in a light color. The answers were also informative in other respects. First, 7 participants correctly noted that they took longer to respond to low-contrast statements. Three mentioned a possible confound, namely that statements in green should lead to "true" responses and statements in red to "false" responses. In hindsight, the choice of the colors was indeed suboptimal in that respect, but in the worst case, this flaw created random noise, because colors were assigned randomly to each statement and the high- or low-fluency status of a statement was also randomly determined. Finally, 1 participant mentioned that he thought that light colors led to "no, false" responses. However, an inspection of this participant's data showed that he did not strategically employ that insight, nor did the results change if he was omitted from the analysis.

Yet, the question of what was actually learned in Experiment 2 remains controversial. From a Brunswikian perspective, one might argue that it is just the color contrast of the statements that is associated with truth and used as a cue, without the necessity of a construct like processing fluency. On the operational level, this distinction makes no difference, because *color contrast* and *processing fluency* are interchangeable terms; that is, it does not matter whether color contrast or processing fluency is learned as a cue. On a theoretical level, it is a major difference. If indeed only the color contrast is learned as an additional cue, it is evidence for an amazing faculty of the cognitive system to use all kinds of cues in a quick and vicarious fashion; however, this explains the effects in Experiment 2 without the notion of processing fluency.

To show that it is indeed processing fluency that is associated with truth, it is necessary to use different manipulations of fluency in the learning and the test phases. Otherwise there is always the alternative explanation that an association between truth and the specific manipulation of fluency is learned. In the present paradigm, the obvious manipulation to enhance processing fluency at test is repetition, as repetition is the standard manipulation for the truth effect. Also, if manipulation of fluency through color contrast in the learning phase influences the response pattern at test, when the crucial difference between items is not high or low contrast, but their old or new status, then processing fluency is the only remaining construct that connects learning and test phases. This would corroborate the claim that processing fluency is responsible for the familiarity-based part of the truth effect and that the interpretation of this fluency is learned.

Experiment 3

The general outset of Experiment 3 is more akin to classic experiments on the truth effect than are the previous experiments. In an initial presentation phase, participants see 60 difficult statements and are told that some of these statements are true and some are false. To minimize encoding and processing of the statements, which could lead to effects of referential validity (Arkes et al., 1991; Hasher et al., 1977), the statements are on the screen for a very short time, and participants make no judgments concerning the statements. In one condition, statements are on the screen for 1 s, and in another condition, for 3 s. Then, after an unrelated task to create a delay after the initial presentation, participants do the learning phase from Experiment 2 with all-new statements. This phase again implements the crucial manipulation of the correlation of fluency and truth. In a *classic* condition, fluency and truth correlate positively, whereas in a *reversed* condition, the correlation is negative. Immediately after the learning phase, the test phase starts with 120 difficult statements. Sixty statements were present in the initial phase and are therefore old or repeated. Sixty statements are completely new, and across these 120 statements, truth and old versus new status are orthogonal. Participants' task is simply to judge whether they believe that a statement is true, and all statements are present in black on a white background. Thus, there is no perceptual manipulation of fluency; however, the idea is that old statements are more fluently processed and that people experience that there is something "about" these statements, without explicitly remembering that they have seen these statements. The interpretation of that experience should be dependent on the correlation of fluency and truth in the learning phase. That is, a differential response bias for people in the classic and reversed conditions to old and new statements is expected.

In addition to these truth judgments, participants judge whether a statement was present during the initial phase of the experiment, immediately after they judged whether a statement is true. It might seem problematic to elicit both kinds of judgments in close succession, because there is the possibility of mutual dependencies. For example, Fragale and Heath (2004) reported data that show that people attribute information they believe to credible sources; therefore, people might tend to "remember" statements they have classified as true. However, in the present case, people know that remembering a statement is no basis for a true or false decision. Moreover, Bacon (1979) as well as Brown and Nix (1996) elicited both judgments for the same statement and reported no problems with this methodology. Besides this additional measure, the main point is still to show a transfer from fluency on a perceptual basis (created by color contrast) to fluency on a memory basis (created by prior presentation).

Method

Participants, design, and materials. Fifty-three students (26 women, 27 men) of the University of Heidelberg participated for payment of €4 (approximately U.S.$5). They were randomly assigned to one of the four conditions resulting from the combination of the correlation of fluency with truth (classic vs. reversed) and the display time in the initial presentation phase (3 s vs. 1 s). The experiment consisted of three phases: a presentation phase, a learning phase, and a test phase. For the learning phase, the same set of easy statements as in the previous two experiments was used. For the presentation and test phases, another 60 statements (half true, half false) were added to the already existing set of 60 difficult statements; true statements were again created by selecting facts from encyclopedia resources and false statements were made-up facts. By and large, the new set was very similar to the already existing set, examples of which are given in Table 1. The presentation of the statements and the assessment of the dependent variables were again done by a program written in Microsoft Visual Basic.

Procedure. Experimental sessions included up to 6 participants. Upon arrival, participants were seated in a cubicle and given a consent form that informed them that they would participate in a study to investigate how people judge the truth of a statement. If they agreed to participate, the experimenter started the computer program. The first screen informed them that they would see a rapid presentation of statements. They were also informed that half of the statements were true and half were false. For each participant, the 120 difficult statements were randomly assigned the status of old or new; an "old" statement would appear in the presentation phase and in the test phase, whereas "new" statements would appear only in the test phase. Note that the classification of old/new is functionally equivalent to the high/low-contrast manipulation in Experiment 2. The random assignment was done under the constraint that old versus new status and truth were orthogonal. Participants started the presentation by clicking a button; depending on the condition, each of the 60 "old" statements was onscreen for 1 s or 3 s, with a break of 500 ms between the statements. The statements were presented in the center of the screen in black color against the white background, and no response was required. After this first phase, participants did an unrelated evaluative conditioning experiment, which caused a break of 25–30 min between the presentation and the learning/test phases. When they were done with the conditioning experiment, participants continued with the learning phase. This phase was identical to the learning phase in Experiment 2; fluency was manipulated by color contrast and depending on condition, fluency was either perfectly positively (classic) or negatively

(reversed) correlated with truth. That is, either all high-contrast statements were true (classic condition) or all high-contrast statements were false (reversed condition), and vice versa for low-contrast statements. After responding to the 60 statements in the learning phase, participants continued with the test phase. A new random order of all 120 difficult statements (30 true/old, 30 true/new, 30 false/old, 30 false/new) was created for each participant. All statements were presented in black color in the center of the screen. A statement remained on the screen until participants responded, using the same "yes, true" or "no, false" keys as in the previous experiments. Following that initial response, the statement stayed on the screen, but a question label appeared asking whether this statement was present in the initial presentation. Participants could now respond with the same keys, but the labels on the screen changed to "old" and "new." Responses and latencies were measured. When they had responded to all 120 statements, the experimenter thanked, paid, and debriefed participants about the hypothesized effects.

Results

One participant was excluded from the analysis because she responded uniformly to all statements. There remained 28 participants in the classic and 24 in the reversed condition.

Response latencies. Before analysis, the latencies from the learning phase were trimmed at 1,000 ms and 5,000 ms and from the test at 1,000 ms and 7,000 ms. The resulting means are displayed in Table 5. A 2 (condition: classic vs. reversed) × 2 (fluency: high vs. low; as repeated measures) mixed ANOVA was used for analysis of the data. For the learning phase, the exact same pattern as in Experiment 2 was found: Participants were faster to respond to high-fluency statements (M = 2,376 ms, SD = 661) than to low-fluency statements (M = 2,508 ms, SD = 716), $F(1, 50) = 9.27$, $p < .01$, $d = 0.86$. The interaction of fluency and condition was again highly significant, $F(1, 50) = 135.25$, $p < .001$, $d = 3.29$, which is again due to participants' faster responses to true statements compared with false statements.

For the test phase, the resulting means are displayed in the right columns of Table 5. The same analysis as for the learning phase was conducted; however, the experimental factors were now the condition (classic vs. reversed) and the status of the statement (old vs. new): Unexpectedly, participants were slower to respond to old (M = 4,694 ms, SD = 818) than to new statements (M = 4,276 ms, SD = 832), $F(1, 50) = 57.74$ $p < .001$, $d = 2.15$. No other effect was significant (Fs < 1, *ns*).

Table 5
Mean Latencies (in ms) in the Learning (Easy) and Test (Difficult) Phase of Experiment 3 for High- and Low-Fluency Statements by Condition

	Learning phase		Test phase	
Condition	High fluency	Low fluency	Old statements	New statements
Classic (n = 28)	2,230 (663)	2,719 (807)	4,750 (951)	4,285 (972)
Reversed (n = 25)	2,547 (630)	2,261 (505)	4,629 (644)	4,266 (651)

Note. Standard deviations are in parentheses.

SDT analysis of true versus false judgments. Overall, participants classified .54 (SD = .03) of the 60 easy statements in the learning phase and .52 (SD = .14) of the 120 difficult statements in the test phase as true. From the "yes, true" and "no, false" responses, SDT parameters d' and β were estimated in the learning phase and separately for old and new statements in the test phase. In the learning phase, participants had a hit rate of .94 (SD = .05) and a false-alarm rate of .13 (SD = .08). The corresponding SDT parameter estimates of the learning phase showed again participants' high discrimination ability for the easy statements (d': M = 2.982, SD = 0.782) and an overall tendency to respond "yes, true" (β: M = 0. 588, SD = 0.427). Yet, there was no difference in d' and β between the classic and the reversed condition (Fs < 1, *ns*).

For the test phase, the hit and false-alarm rates as well as the resulting SDT estimates are given by condition in Table 6, separately for old and new statements. Mixed 2 (condition: classic vs. reversed) × 2 (statement: old vs. new; as repeated measures) ANOVAs were performed on these estimates. For d', this analysis showed no reliable effect, only a slightly higher discrimination ability for participants in the reversed condition compared with the classic condition, $F(1, 50) = 2.12$, $p < .15$, $d = 0.41$. Of greater interest, however, are the results for the response bias. First, there was a slight overall tendency to respond "yes, true" in the reversed condition compared with the classic condition, $F(1, 50) = 2.16$, $p < .15$, $d = 0.42$. Analyzing each condition separately, neither the classic truth effect nor the reversal reached a standard level of significance, $t(27) = -0.96$, *ns*; $t(23) = 2.03$, $p < .06$, $d = 0.85$. However, analyzing both conditions in a mixed 2 (condition: classic vs. reversed) × 2 (statement: old vs. new; as repeated measures) ANOVA shows the predicted interaction for β to be significant; participants in the classic condition had a slight tendency to respond "yes, true" to old statements compared with new statements, whereas participants in the reversed condition had a tendency to respond "yes, true" to new statements compared with old statements, $F(1, 50) = 7.24$, $p < .01$, $d = 0.76$. Including presentation time (whether statements were on the screen for 1 s or 3 s in the initial presentation phase) in this analysis left the interaction unchanged, $F(1, 50) = 7.35$, although the main effect for condition increased a little, $F(1, 50) = 2.64$. All other effects remained nonsignificant (Fs < 2, *ns*).

SDT analysis of old versus new judgments. In addition to judging whether a statement was true or false, participants also judged whether a statement was presented earlier or not, that is, whether it was an old or new statement. The overall SDT estimates from this recognition task show that participants could not discriminate between old and new statements (d': $M = -0.016$, SD = 0.375) and had a tendency to respond "new" (β: M = 1.392, SD = 1.541). Performing 2 (condition: classic vs. reversed) × 2 (presentation time: 1 s vs. 3 s) ANOVAs on these estimates showed no effect for d' at all, although an increase in d' was expected with longer presentation time (all Fs < 1, *ns*). However, there was a main effect for presentation time on β, such that participants in the 1-s condition had a greater tendency to respond "new" than did participants in the 3-s condition (M = 1.743, SD = 2.19 vs. M =

Table 6
Hit/False-Alarm Rates and Signal Detection Theory Parameter Estimates From the Test Phase in Experiment 3 by Old and New Statements and Condition

Statements	β estimates	d' estimates	Hit rates	False-alarm rates
Classic (n = 28)				
Old	0.992 (0.117)	0.135 (0.394)	.53 (.15)	.48 (.12)
New	1.032 (0.204)	0.058 (0.359)	.50 (.15)	.48 (.15)
Reversed (n = 24)				
Old	1.009 (0.253)	0.240 (0.343)	.60 (.17)	.51 (.15)
New	0.902 (0.181)	0.192 (0.455)	.58 (.19)	.51 (.15)

Note. Standard deviations are in parentheses. β values smaller than 1 indicate a tendency to respond "yes, true," whereas values greater than 1 indicate a tendency to respond "no, false." Higher d' values indicate a higher discrimination ability.

1.067, $SD = 0.149$), $F(1, 50) = 3.38$, $p < .08$, $d = 0.52$. All other effects were not significant (Fs < 2).[3]

Discussion

It was possible to systematically influence the response tendency to old and new statements by manipulating the correlation of truth with a visual property (i.e., color contrast) in a learning phase. Experiment 2 offered the alternative explanation that the learning phase simply conveyed the validity of an additional cue, namely color contrast. Experiment 3 leaves no room for such an explanation; the only remaining link between prior presentation and color contrast is processing fluency. If high color contrast correlated with truth in the learning phase, participants showed the classic differential bias between old and new statements at test (although not as strongly as in Experiments 1 and 2); if, however, low color contrast correlated with truth, this differential response tendency was reversed. Differing from Experiment 1 and 2, the within-condition comparisons yielded no significant results. However, one needs to take into account the very short presentation time and the rather low power in comparison with other experiments demonstrating the truth effect (Parks & Toth, 2006; Reber & Schwarz, 1999). The overall effect was weaker than in Experiment 2, but then again, such weaker effects are to be expected when learning and application contexts differ and the manipulation of fluency (i.e., presentation time) is comparatively weak. The point, however, was not to show a practical application or the strength of such a reversal, but to demonstrate that the reversal is possible at all.

An unexpected finding is that participants took longer to respond to old statements than to new statements. Although latencies are only an imperfect proxy for fluency, this pattern presents a problem, because one might argue that new statements were actually more fluently processed. Yet, there are at least two arguments against such an objection. The first is a pragmatic argument: Although the overall recognition ability was close to zero, some participants probably remembered some of the statements. However, this did not speed up their decisions, because they knew that half of the old statements were false as well. Therefore, the slow overall responses to old statements might be due to participants who pondered over statements they remembered. Unfortunately, this notion is not testable, because it is not possible to discriminate between true hits in the recognition task and random hits, which precludes a statement-by-statement analysis. This problem calls for a process-dissociation procedure (Jacoby & Kelley, 1992) in future research, as used by Begg et al. (1992), in which participants are told that all statements from the initial presentation are either false or true.

The second argument is more theoretically based on the difference between subjective fluency and objective speed (cf. Reber, Wurtz, & Zimmermann, 2004). Although the initial processing of an old statement might be experienced as fluent due to prior presentation, the source might not be clear, which calls for processes of attribution (what is the cause of the experience?) and interpretation (what does the experience mean?). Thus, responses to old statements can be slower, although the initial experience is more fluent. This reading of the effect also fits well with the notion that the fluency experience is a feeling that there is something about the statement (Higgins, 1998) or that a statement has a ring of truth (Begg & Armour, 1991).

The short presentation time in the initial phase was successful insofar as participants could not reliably discriminate between old and new statements, thereby reducing influences of active recognition (i.e., referential validity) in the present paradigm. Some of the recognized statements might nevertheless contribute to the slower response latencies. Moreover, the manipulation of presentation time had no impact on d', but only on β, which is an indication that the presentation was just too fast in both conditions for an actual recognition effect. However, participants' tendency to respond "new" in the 1-s condition can be interpreted as that they simply did not feel "entitled" to recognize that many statements.

Altogether, Experiment 3 showed that processing fluency is partly responsible for the truth effect, and the impact of fluency on judgments of truth depends on the learned interpretation of that fluency experience.

General Discussion

The outset of the present research was that the truth effect is partly based on the (mis)attribution of some stimulus quality (i.e.,

[3] According to Bacon (1979), the crucial variable is the subjective belief in whether a statement is old or new; an ANOVA on the estimates with participants' subjective old versus new judgments as a classification variable instead of the actual old or new status resulted in no differential response biases for judged old and new statements (β: $M = 0.986$, $SD = 0.430$ vs. $M = 0.999$, $SD = 0.224$, respectively), nor was there an interaction with condition.

the experienced processing fluency of the stimulus) to the truth of the statement. This notion places the truth effect in a larger family of effects that are based on the differential interpretation of experiences caused by a stimulus. A broad theoretical model to explain such effects is provided by Whittlesea and colleagues' SCAPE (selective construction and preservation of experience) model (e.g., Whittlesea & Leboe, 2000; Whittlesea & Williams, 2001). According to this model, cognitive processes such as recognition and recall are constantly monitored and continuously evaluated. This evaluation process results in an experience or a feeling when there is a discrepancy (i.e., variations in processing fluency) between the evaluation and expectancies concerning the process. However, the feeling resulting from this discrepancy is nonspecific, and the discrepancy triggers a search for an explanation. Although in many cases the context provides a natural explanation, for example, a feeling of familiarity, the model leaves room for different interpretations of variations in processing fluency. The truth effect fits well into this model; the experienced variations are not attributed to prior exposure, resulting in a feeling of familiarity, but to some other quality of the statement, namely, that a statement is true.

The present data supplement this idea by offering an explanation of why such experienced variations, the fluent or nonfluent processing of a stimulus, result in judgments that a statement is true rather than false, or in other words, why the bell that rings when a statement is fluently processed has a ring of truth (Begg & Armour, 1991). The central argument is that there is indeed a positive correlation between the experience and the truth of a statement, and this correlation can be learned as a cue validity, similar to cue validities that can be learned in depth perception (e.g., Jacobs, 2002) or in decision making (e.g., Gigerenzer & Goldstein, 1999). Thus, the truth effect occurs because processing fluency is a cue with ecological validity in judgments of truth. The reversal is then produced by creating a context in which the interpretation of the variations in processing fluency have reversed implications and the feeling that there is something about the statement results in a judgment that the statement is false. Although the data from Experiment 2 can be explained as the utilization of an additional cue (i.e., color contrast), because learning and test used the same manipulation of processing fluency, there is no room for such an explanation in Experiment 3. In the learning phase of Experiment 3, processing fluency was manipulated through color contrast, but at test, the only difference between statements was their old versus new status. The influence of the correlation in the learning phase on the judged truth of old and new statements convincingly shows that processing fluency is used as a cue, according to its ecological validity.

An open question is how the proposed positive correlation between processing fluency and truth (i.e., the validity of the cue) comes about outside the laboratory. Again, theoretically it is possible to think of an environment in which fluency or familiarity is a cue for the falseness of a statement, for example, in a hostile environment, in which most statements are lies or deceptions. A possible reason for a positive correlation has already been mentioned: Gilbert (1991) contrasted Descartes's idea that understanding a proposition and accepting that proposition are two independent processes with Spinoza's argument that the acceptance of a proposition is an automatic part of the comprehension of the proposition. Consequently, people automatically believe in propositions they understand, as they believe in the objects they see in the physical world. According to Gilbert (1991) as well as Gilbert, Krull, and Malone (1990), the second account receives more support from psychological research. So, if understanding that a statement is false entails a secondary process, whereas true statements are automatically accepted, there is a built-in advantage in processing for true over false statements, which possibly causes the correlation between experienced fluency and truth. Beside this more philosophically grounded explanation, there is also a pragmatic reason why there should be a positive correlation. Prior exposure to a stimulus (or an idea, a concept, etc.) results in facilitated processing of that stimulus. This is shown not only by experiments on fluency effects (e.g., Jacoby & Whitehouse, 1989), but also by many priming studies (e.g., Klauer & Musch, 2003; Klauer, Musch, & Eder, 2005) and research on repetition effects (Feustel et al., 1983). Given that, the only necessity is that people are more frequently exposed to the same true than to the same false statements. This necessity is normatively fulfilled by Grice's maxim of quality in interpersonal communications, according to which speakers should present only truthful information (Grice, 1975). However, there is also a practical reason why people are more frequently exposed to the truth. There are infinite false propositions regarding physical reality, but only one true proposition. To take the example from the beginning, there are many possible false statements that involve the world's highest tree (a birch, an oak, a cedar, etc.), but there is only one true statement: The highest tree is currently a 112-m sequoia in the Redwood National Park (however, the all-time record is an Australian 132-m eucalyptus tree; *TreeRecords*, n.d.). Consequently, there is a good chance that people encounter the same true statements, concepts, or ideas more frequently or repeatedly than they do false statements, concepts, or ideas, and therefore, these true items are more fluently processed, creating the proposed positive correlation between processing fluency and truth. To the same extent, as false statements are repeated (such as with rumors, urban legends, or false propaganda), they may gain credibility as well; yet, these cases should be the exception rather than the rule, and they exploit the positive correlation between fluency and truth rather than create them.

If one accepts this proposed correlation, the idea that people learn differential interpretations of processing fluency in varying contexts offers a parsimonious explanation of the truth effect. People show the truth effect because they notice a variation in processing and interpret this processing fluency according to its correlation with truth, which, in the real world, is a positive one.

References

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

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

Begg, I. M., 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. M., & 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.

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.

Brunswik, E. (1957). Scope and aspects of the cognitive problem. In H. Gruber, K. R. Hammond, & R. Jessor (Eds.), *Contemporary approaches to cognition* (pp. 5–31). Cambridge, MA: Harvard University Press.

Doyle, A. C. (1990). The Boscombe Valley mystery. In *The illustrated Sherlock Holmes: The adventures of Sherlock Holmes* (pp. 65–84). London: Peerage Books. (Original work published 1891)

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112,* 309–346.

Fragale, A. R., & Heath, C. (2004). Evolving informational credentials: The (mis)attribution of believable facts to credible sources. *Personality and Social Psychology Bulletin, 30,* 225–236.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (pp. 75–96). New York: Oxford University Press.

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

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59,* 601–613.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.). *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.

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.

Higgins, E. T. (1998). The aboutness principle: A pervasive influence on human inference. *Social Cognition, 16,* 173–198.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences, 6,* 345–350.

Jacoby, L. L., & Kelley, C. M. (1992). A process-dissociation framework for investigating unconscious influences: Freudian slips, projective tests, subliminal perception, and signal detection theory. *Current Directions in Psychological Science, 1,* 174–179.

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

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

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

Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 3–11.

Johnston, W. A., Hawley, K. J., & Elliott, J. M. (1991). Contribution of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 210–223.

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). Mahwah, NJ: Erlbaum.

Klauer, K. C., Musch, J., & Eder, A. B. (2005). Priming of semantic classifications: Late and response related, or earlier and more central? *Psychonomic Bulletin and Review, 12,* 897–903.

Mandler, G., Nakamura, Y., & Van Zandt, B. J. (1987). Nonspecific effects of exposure on stimuli that cannot be recognized. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 646–648.

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

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition: An International Journal, 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., Wurtz, P., & Zimmermann, T. D. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition: An International Journal, 13,* 47–60.

Reber, R., Zimmermann, T. D., & Wurtz, P. (2004). Judgments of duration, figure-ground contrast, and size for words and nonwords. *Perception & Psychophysics, 66,* 1105–1114.

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

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523.

Skurnik, I., Schwarz, N., & Winkielman, P. (2000). Drawing inferences from feelings: The role of naive beliefs. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 162–175). Philadelphia: Psychology Press.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers, 31,* 137–149.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1–26.

Tabachnik, B. G., & Fidell, L. S. (1996). *Using multivariate statistics.* New York: HarperCollins.

TreeRecords. (n.d.). Retrieved April 8, 2006, from http://www.planet-wissen.de/pw/Artikel,,,,,,,A9FD48ED9EF344AEE0340003BA04DA2C,,,,,,,,,,,,,,,.html

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207–232.

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science, 17,* 339–345.

Watkins, M. J., & Peynircioglu, Z. F. (1990). The revelation effect: When disguising test items induces recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1012.

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. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129,* 84–106.

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

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, 9,* 1–27.

Received May 8, 2006
Revision received July 21, 2006
Accepted July 31, 2006 ■