



A multinomial modeling approach to dissociate different components of the truth effect

Christian Unkelbach^{a,*,1}, Christoph Stahl^{b,1}

^a *Psychologisches Institut, Universität Heidelberg, 69117 Heidelberg, Germany*

^b *Institut für Psychologie, Universität Freiburg, 79085 Freiburg, Germany*

ARTICLE INFO

Article history:

Received 6 November 2007

Available online 5 November 2008

Keywords:

Processing fluency

Cue learning

Process-dissociation

Truth judgments

ABSTRACT

The subjective impression that statements are true increases when statements are presented repeatedly. There are two sources for this *truth effect*: An increase in validity based on recollection (a controlled process) and increase in processing fluency due to repeated exposure (an automatic process). Using multinomial processing trees (MPT), we present a comprehensive model of the truth effect. Furthermore, we show that whilst the increase in processing fluency is indeed automatic, the interpretation and use of that experience is not. Experiment 1 demonstrates the standard use of the fluency experience and Experiment 2 demonstrates that people can change the interpretation of the experience according to its ecological validity. By implication, the truth effect represents the adaptive usage of feedback received from internal processes.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

If people encounter a statement repeatedly, for example, “Europe’s biggest glacier is the Vatnajökull on Iceland”, their subjective impression increases that the statement is true (e.g., Bacon, 1979; Hasher, Goldstein, & Toppino, 1977). This increase is captured in higher ratings of a statement’s truth when rating scales are used or the higher probability to judge a statement as true when binary true/false decisions are elicited (Begg, Anas, & Farinacci, 1992; Hawkins & Hoch, 1992). As people’s belief in a statement’s truth is central to many areas, ranging from marketing to politics, the mechanisms underlying this “truth” effect are of great practical and theoretical interest. The aim of the present study is to disentangle different components of the truth effect using multinomial models. Furthermore, we aim to show that part of the effect which was heretofore assumed to be automatic is indeed the strategic use of the differential processing experiences elicited by repeated statements and new statements.

1.1. Explaining the truth effect

A standard truth experiment contains a presentation phase t_1 in which statements are presented for a first time. Sometimes, the statements’ truth is rated at t_1 or the items are just presented without eliciting a judgment. Then, after a delay (varying from minutes to days or weeks), a judgment phase t_2 follows in which the truth of old statements (presented at t_1) and new statements (presented at t_2 for the first time) is rated or judged. The truth effect is then evident in higher truth ratings for old statements or the higher probability to judge an old statement to be true, either in comparison to t_1 , or in comparison to new statements at t_2 (cf. Dechêne, 2005, for an overview of different procedures).

* Corresponding author.

E-mail address: christian.unkelbach@psychologie.uni-heidelberg.de (C. Unkelbach).

¹ These two authors contributed equally to this work.

A straightforward reason for the occurrence of the effect is referential or convergent validity (Arkes, Boehm, & Xu, 1991; Hasher et al., 1977). If people remember that they heard or read a statement previously, they have a basis to judge the statement as true (Brown & Nix, 1996). For example, when people remember reading the statement about the glacier in *National Geographic*, they have a reference to judge it as true. Furthermore, when statements from different sources converge, people can also use this convergence as a basis for a “true” judgment.

Yet, just remembering the statement without the particular source also increases the subjective impression that this statement is true. Moreover, even when people cannot recall reading or hearing a statement previously there is an increase in subjective truth (Begg et al., 1992). This non-referential part of the effect will be the component of interest in the following. The initial idea was that repeated exposure increases the subjective familiarity with a statement and this familiarity is used as basis to judge the truth of a statement (Arkes, Hackett, & Boehm, 1989; Bacon, 1979).

To explain this effect, Hawkins and Hoch (1992) hypothesized that it is actually the familiarity with the semantic content of a statement, which in turn allows the retrieval of additional information from memory, thereby creating the subjective impression that a statement is true. However, level-of-processing manipulations increased only the conscious recognition of statements, but not the non-referential part, which renders the idea of semantic familiarity unlikely (Begg, Armour, & Kerr, 1985).

A comprehensive explanation was then offered by Begg and colleagues (1992). Using a process-dissociation procedure (Jacoby, 1991; see also, Kelley & Jacoby, 1998, 2000), they showed that there is indeed an increase in subjective truth due to repeated exposure, independent from an increase based on conscious recollection (i.e., convergent or referential validity). To dissociate controlled from automatic processes in truth judgments, Begg and colleagues' (1992) experiments included a clear indication of the actual truth status when statements were presented for the first time. For example, they informed participants that all statements read by a male voice were true and all statements read by a female voice were false (Begg et al., 1992, Experiment 4). Old statements that were clearly labeled to be false at t_1 were still judged to be true with higher probability than new statements at t_2 . This result, consistent across four experiments, provided clear evidence for a non-referential component of the truth effect. The authors classified this non-referential component as part of a larger group of effects caused by *experiences* elicited by a stimulus, similar to the mere exposure effect (Mandler, Nakamura, & Van Zandt, 1987; Zajonc, 1968), the sleeper effect (Hovland & Weiss, 1951; Pratkanis, Greenwald, Leippe, & Baumgardner, 1988), or the false-fame effect (Jacoby, Kelley, Brown, & Jasechko, 1989).

1.2. Interpreting processing fluency

Following that line of reasoning, Reber and Schwarz (1999) speculated that repeated exposure increases the fluency with which a statement is processed, and that this increase in processing fluency is responsible for the non-referential part of the effect. They manipulated *perceptual* fluency by presenting statements in colors of differing contrast and showed that subjective truth is indeed influenced by perceptual fluency. Unkelbach (2007) then showed the functional equivalence of increased fluency due to repetition and increased fluency due to color contrast (Unkelbach, 2007, Experiment 3; cf. Feustel, Shiffrin, & Salasoo, 1983, for the effect of repetition on processing fluency). The author showed that the non-referential component of the truth effect is based on the interpretation of processing fluency: True/false judgments depended on the differential processing experiences elicited by old and new statements, and the interpretation of this experience could be changed by feedback learning (see also Unkelbach, 2006). In this framework, processing fluency is conceptualized as a cue with ecological validity; statements which are easier to process should indeed be true with higher probability than statements which are difficult to process.

This implies three differences to the account proposed by Begg and colleagues (1992). The first difference is best conveyed by quoting one of their statements: “These effects of familiarity are irrational; there is no logical reason for repetition to affect rated truth or for earlier information to be trusted more than later information.” (Begg et al., 1992; p. 447). The argument presented by Unkelbach (2007) is that processing fluency is a *valid* proximal cue for the distal property of truth, as a positive correlation between a statement’s fluency and its truth is assumed. There are at least three reasons why this should be the case: First, there are countless possible false propositions about properties of the physical world, but only one true proposition (e.g., “The extended right arm of the Statue of Liberty is 46 feet long”; Bacon, 1979). In other words, people should have, in general, a higher chance to be repeatedly exposed to true propositions than to the false propositions. It is important that there is no need for verbatim repetition, as shown by research on the implicit learning of structure (A.S. Reber, 1967; Buchner, 1994). If people read a statement which is similar, but not exactly the same as statements read before, they should nevertheless experience an increase in fluency (Buchner, 1994; Feustel et al., 1983). Second, according to Gilbert (1991), understanding a proposition entails the automatic acceptance of that proposition. The rejection of that proposition is a secondary step that consumes time and resources. Thus, false statements are associated with more cognitive effort, while true, non-rejected statements are associated with fluent processing, leading to a correlation between processing fluency and truth. And third, people should follow a maxim of quality in communication, which creates a prevalence of true information in a representative environment (Grice, 1975).

A critique of this proposed ecological correlation between truth and processing fluency is that there are also many false statements that are frequently repeated (e.g., “The world is flat.”). However, most of such statements are *known* to be false, a reason we include a parameter for actual knowledge in our model (see below). False and frequently repeated statements with ambiguous truth status (e.g., “The Great Wall of China is visible from the moon.”), on the other hand might be judged to be true based on their fluent processing. Yet, these statements benefit from the overall positive correlation between fluency and truth, rather than creating it (see also the discussion in Unkelbach, 2007).

In this respect, the idea of fluency as a valid cue that is interpreted according to its ecological validity also differs from a misattribution account (e.g., Bornstein & D'Agostino, 1994; Fragale & Heath, 2004). As previous research has demonstrated (e.g., Zaragoza & Lane, 1994), misattribution in the sense of source confusion plays a major role in many memory errors. However, in the present framework it is not necessary that the fluency experience is misattributed to a previous encounter from another source, but the experience is used directly as a cue. Thus, the effect is not a memory error, but the application of an interpretation that should be valid in an ecological context, but invalid in the context of a laboratory experiment.

The second difference to the account proposed by Begg and colleagues (1992) is the notion that the non-referential component of the truth effect is automatic. While we agree that fluency due to prior exposure is an automatic process (see Feustel et al., 1983), the *interpretation* of that experience is not. It is possible that the interpretation and use of the experience might become automatic within the routines of daily life (Bargh & Chartrand, 1999), but there is no need to restrict it by the automaticity assumption. It is easy to imagine a context when there is no recollection, but a stimulus (a face, a name, or a statement) elicits a strong and conscious feeling that there is something “about” this stimulus (Higgins, 1998). Depending on the context, one might interpret this feeling as familiarity, but also as liking, or truth (Mandler et al., 1987; Whittlesea, 1993).

The last difference is that the fluency interpretation account allows symmetrical effects of fluency on judgments of truth; that is, fluently processed statements are judged to be true, but by the same token, non-fluent judgments should be judged to be false. This becomes more apparent if one considers the basis of a fluency experience. Whittlesea and Williams (2001) showed that fluency is experienced only when there is a discrepancy in processing in comparison to a standard or an expectation (see also Hansen, Dechêne, & Wänke, 2008). This discrepancy can be positive, caused by higher fluency than expected, or negative, caused by lower fluency than expected. Thus, in a typical study in which expectancy is likely to be based on the mean fluency across all items, positively discrepant, fluent statements are judged to be true, whereas negatively discrepant, non-fluent statements are judged to be false.

The present research provides further support for this account of the non-referential component of the truth effect. First, we show that true statements are processed faster and at the same time are perceived as more familiar. Second, we provide evidence that the absence of a fluency experience can be used as a cue, by showing that, in addition to rating fluently processed statements as true, people also judge non-fluently processed statements as false. Finally—and most importantly—by explicitly manipulating the validity of the fluency experience as a cue for truth, we demonstrate that people can interpret the fluency experience in a controlled manner.

1.3. A multinomial model of truth judgments

To substantiate these claims, and to formalize the account, we will introduce a model of truth judgments that belongs to the class of multinomial processing tree models (MPT; Riefer & Batchelder, 1988; see Batchelder & Riefer, 1999, for a review). MPT models are based on categorical data such as the true/false classifications of statements, and they allow estimating and separating the contribution of different cognitive processes underlying observed data. Furthermore, MPT models force researchers to specify which processes they assume and how a given process should influence observable data. Thus, otherwise implicit assumptions about processes are made explicit and testable. The process-dissociation procedure developed by Jacoby and colleagues (Jacoby, 1991; Kelley & Jacoby, 1998, 2000) and employed by Begg and colleagues (1992) is but a minimalistic version of a MPT model (see also Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995; Buchner & Wippich, 1996; Draine, Greenwald, & Banaji, 1996). The aim of the present model is to provide a comprehensive formal account of the cognitive processes involved in making a truth judgment. The model is validated by showing that it is capable of accounting for the data from two experiments, and that its parameters respond to experimental manipulations in a psychologically plausible manner.

As described above, in a typical study, participants are presented with a set of true and false statements at t_1 . In some studies, a subset of these statements (e.g., those spoken by a female voice) is said to be true, whereas another subset (e.g., those spoken by a male voice) is said to be false, when in fact, both subsets contain an equal number of true and false items. Participants are then asked to judge the truth or falsity of these old statements at t_2 , mixed with a set of new statements. These truth judgments are then assumed to be determined by the following processes (parameters), which are also illustrated using a processing-tree representation in Fig. 1. The parameters of the model can vary between 0 and 1, and they represent the probability of occurrence of the corresponding cognitive process.

First, when judging the truth of a statement, participants may simply know the factual truth or falsity of a statement and judge it accordingly (the probability of such knowledge is represented by parameter k ; for example, “The world is flat.”). Second, for old statements, participants may recollect whether the statement stems from a true or false source (the probability of source recollection is represented by parameter c). Third, participants may use experienced processing fluency due to a statement's factual truth and judge it accordingly (parameter f_t). Fourth, they may use processing fluency due to prior presentation for old statements (parameter f_p) as well as the lack of fluency for new statements (parameter f_n). Finally, when neither of the above processes leads to a response, participants may simply guess whether a statement is true or false (parameter g).²

² Note that the distinction between fluency due to truth and fluency due to repetition is made only theoretically; empirically, fluency due to truth can also be caused by repetition. Another source, however, is semantic coherence of a statement with prior knowledge (Topolinski, Likowski, Weyers, & Strack, in press).

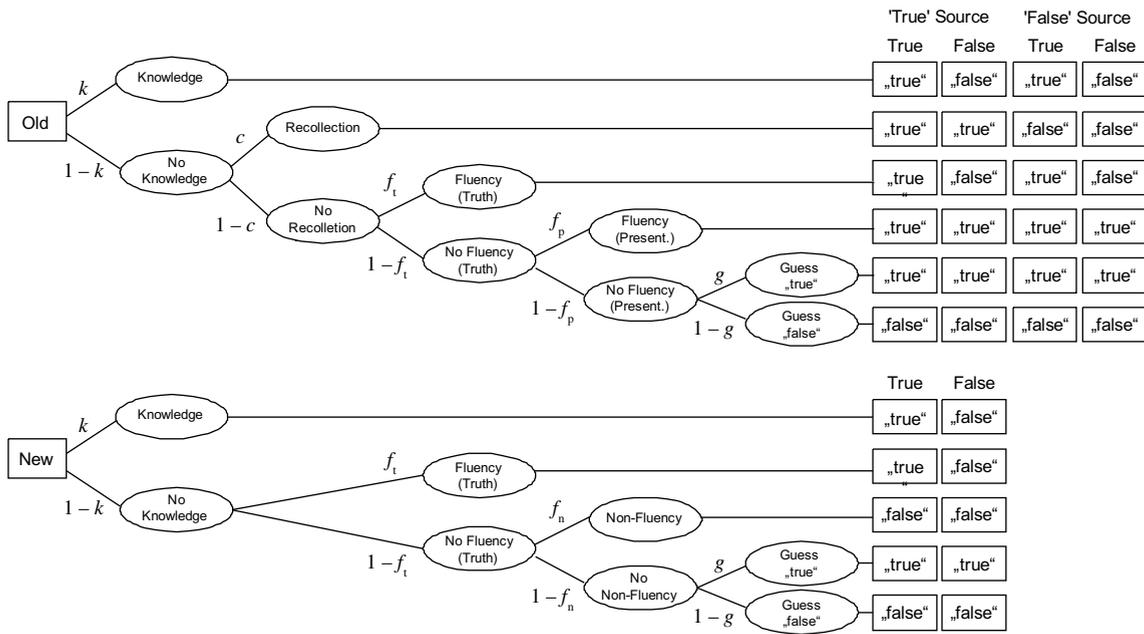


Fig. 1. The multinomial processing tree model of the truth effect. Rectangles represent observable states, ellipses represent latent states of the cognitive system, and the parameters represent transition probabilities between states. The top tree diagram illustrates the processes postulated to occur in response to an old statement at t_2 . Old statements can originate from a ‘true’ source or from a ‘false’ source and can be either true or false (see text). The bottom diagram illustrates the processes for new statements. New statements can be either true or false. The branches of each tree represent the possible combinations of cognitive processes. The responses associated with each branch are given on the right for each statement type. The postulated processes are, from left to right, knowledge of the factual truth value of a statement (parameter k), conscious recollection of the statement’s presentation at t_1 (parameter c), interpretation of the fluency experience based on a statement’s truth (parameter f_t), interpretation of the fluency experience based on a statement’s presentation at t_1 (parameter f_p), or – for new statements – the lack of a fluency experience for statements not presented at t_1 (parameter f_n), and finally, guessing (parameter g).

For illustration, consider responses to an old statement (top diagram in Fig. 1). If knowledge about the factual truth is present, true statements can be correctly judged as true, and false statements can be correctly judged as false with probability k . If, with probability $(1-k)$, knowledge is absent, participants may nevertheless consciously recollect the statement’s presentation at t_1 along with its source with probability c . In this case, judgments will reflect the truth value associated with the source of the statement. When knowledge and conscious recollection are lacking for an old statement, with probability $(1-k)(1-c)$, processing fluency may influence responses. Three different types of fluency experiences are considered. First, as argued above, true statements are on average processed more fluently than false statements. This experience determines responses with probability f_t , and typically, as shown in Fig. 1, fluently processed statements would be judged as true (but see Experiment 2 below). Second, when old and new statements are presented at t_2 , old statements (i.e., those presented at t_1) are processed more fluently and judged to be true with probability f_p , while new statements (i.e., those presented for the first time) are processed less fluently and therefore judged to be false with probability f_n . Finally, if neither of the above processes led to a response, participants simply guess whether the statement is true or false. With probability g , the item is judged “true”, and with probability $(1-g)$, it is judged “false”.

The model in Fig. 1 incorporates theoretical assumptions about the processes underlying truth judgments that are implicitly present in many experiments on the truth effect (e.g., Begg et al., 1992) and makes them explicit. Although some of the assumptions are speculative, for example, that f_p and f_n are separate processes,³ the aim is to present a general model that is applicable to most studies on the truth effect. As a consequence of this comprehensiveness, the full model cannot be computed for most studies of the truth effect without further specifications, because they typically lack experimental conditions that are critical for the identifiability of all the model’s parameters. For example, in a typical study, the response patterns predicted by knowledge (parameter k) and by fluency due to truth (parameter f_t) are identical (i.e., both processes would lead to a “true” response in all conditions), rendering a model-based dissociation impossible. However, identifiable versions of the model for a given data set can be obtained easily by incorporating plausible additional specifications about a subset of the parameters (see below). We derived identifiable and testable versions of the model for the present data sets. The restrictions that were incorporated into the model for each of the present data sets are described in the respective Results sections. The model equations are given in the Appendix A.

³ Another possibility is that it is in fact the difference between f_p and f_n that causes the truth effect, which implies not consecutive, but simultaneous processes. For clarity of presentation, comprehensiveness, and in light of the following data, the present model was chosen.

Table 1
Examples of the 60 true and false statements used in the present experiments

True statements	False statements
First olympic gold medalist (new games) was James Connolly	The fourth man on the moon was Charles Conrad
The sesame plant has thimble-shaped leaves	Cactuses can procreate via pathogenesis
The buzzard is Europe's most frequent bird of prey	The capital of Madagascar is Toamasina

1.4. Pretesting materials

For the following two experiments, we used 120 trivia statements that were created for the experiments reported by Unkelbach (2007). Sixty of these statements are factually true and 60 are factually false. The statements were created with the aim to minimize knowledge about their factual true/false status. Examples of true and false statements are provided in Table 1. As we want to dissociate the contributions of fluency due to repetition and fluency due to a statement's truth, it is necessary to pretest this material. The first test included ten participants who completed a questionnaire in which the 120 statements were presented in a random order. They simply judged whether a given statement was true or not. Participants classified on average 32.30 out of 60 statements ($M = .538$, $SD = .171$) correctly as true and 28.80 out of 60 statements ($M = .480$, $SD = .194$) incorrectly as true. This difference in proportion of "true" judgments between true and false statements indicates an ability to discriminate between true and false statements, albeit not statistically significant, $t(9) = 1.92$, $p < .08$. The mean proportion of "true" judgments was .509 ($SD = .177$), which was clearly not different from .5, $t(9) = 0.16$, *ns.*, showing a lack of response bias.

To demonstrate the validity of the model parameters, we replicated these findings with model-based analyses. As most of the model parameters do not apply to the pretest situation, we used a simple two-parameter version of the model to fit the pretest data. This model was a simple two-high threshold model to account for discrimination between true and false statements. Such discrimination could be based on either knowledge (parameter k) or fluency due to truth (parameter f_i); both processes were not dissociable in the pretest data, and we therefore used a single parameter, k , to capture both effects. If they are not able to discriminate between true and false statements, participants may simply respond by guessing. In sum, discrimination between true and false statements by way of knowledge or fluency due to truth (parameter k), and guessing (parameter g)—was fitted to the pretest data (see Appendix A for details). All model analyses reported in this article were performed using the HMMTree software (Stahl & Klauer, 2007). In applying the model to the pretest data, we corroborated participants' ability to discriminate between true and false statements. The discrimination parameter $k = .06$, was both numerically and statistically greater than zero, $\Delta G^2(1) = 4.09$, $p < .05$. Note again, that this ability to discriminate could be based on knowledge about the factual truth of statements, or due to differential fluency for true and false statements. Second, the guessing parameter $g = .51$ was not biased from the neutral point of .5, $\Delta G^2(1) = 0.40$, *ns.* Thus, the parameters for guessing and discrimination accurately reflect the findings based on an analysis of the raw data (i.e., some discrimination but no guessing bias), supporting the validity of these parameters. The second pretest included 49 participants and was computer-controlled. A computer program presented the 120 statements in a new random order for each participant. Statements were presented in the center of the computer screen and participants judged the familiarity of each statement on a scale from 0 ("not familiar at all") to 100 ("highly familiar") by means of a scrollbar. The question inserted above each presented statement always was: "How familiar is this statement?"⁴ The program recorded the judgments and respective latencies. First, participants were faster to judge true statements ($M = 4850$ ms, $SD = 1064$) than false statements ($M = 4962$ ms, $SD = 1044$), $t(48) = -2.81$, $p < .01$, $d = 0.81$. If one accepts response latencies as a proxy for processing fluency, then, on average, true statements were processed more fluently than false statements. Second, overall, statements were judged to be rather unfamiliar ($M = 44.19$, $SD = 10.24$), $t(48) = -3.97$, $p < .001$, $d = 1.15$, tested against the midpoint of the scale (i.e., 50). And third, an analysis by the factual true/false status of the statements reveals that true statements were on average judged to be more familiar ($M = 46.52$, $SD = 11.09$) than false statements ($M = 41.86$, $SD = 10.45$), $t(48) = 4.89$, $p < .001$, $d = 1.41$, but both means are still significantly on the "unfamiliar" side of the scale, $t(49) = -2.19$, $p < .05$, $d = 0.63$, and $t(49) = -5.45$, $p < .001$, $d = 1.56$, for true and false statements, respectively. Thus, although all statements were rather unfamiliar, true statements were judged to be more familiar than false statements. The size and consistency of this effect across items, as indicated by the small standard deviations as well as visual inspections, suggests that this effect is not due to some outlier statements of which the truth status was known and the familiarity was judged as high.

The result from this second pretest, that true statements were more fluent than false statements, will play an important role in the following experiments. Furthermore, it supports the possibility that participants in the first pretest had no knowledge after all, but were simply using their differential fluency experiences to judge the truth of a statement. As processing fluency is supposedly a valid cue in these 120 statements, this could account for participants' ability to discriminate between true and false statements in the pretest data (i.e., the effect on parameter k). If processing fluency indeed causes the discrimination between true and false statements, then the effect should be malleable: Specifically, manipulations that affect the

⁴ The German instructions "Wie vertraut kommt Ihnen die Aussage vor?" has slightly different connotations compared to the English, "familiar". In particular, it clearly conveys that the question is not about verbatim repetitions.

interpretation of the fluency experience should affect participants' discrimination ability. If, instead, discrimination was due to knowledge, there should be no effects of such manipulations. This issue will be addressed in Experiment 2.

Experiment 1 will demonstrate that the non-referential component of the truth effect depends on the differential fluency experiences elicited by old and new statements. Specifically, we show that the non-referential component consists of a positive fluency effect, that fluently processed statements are judged to be true, but also of a negative fluency effect, that non-fluently processed statements are judged to be false. In other words, participants use the discrepancy between expected and experienced fluency as a cue: Statements that are more fluent than a baseline are judged as true, but statements that are less fluent than a baseline are judged as false.

In a typical study, in which no special expectation is induced or otherwise available about the fluency of processing, we assume that participants compute a baseline as the mean fluency that they experienced across all statements (Whittlesea & Williams, 2001). Thus, it makes sense to say that old statements are more fluent than the baseline, and new statements are less fluent than the baseline. Note that, as long as only two types of statements are used, both processes cannot be dissociated without using a reference point. In our model-based analysis of Experiment 1, we used two different methods, with converging results. First, the formal model allowed us to compute the probabilities for a positive and a negative fluency effect without using an extra baseline when the assumption was made that guessing processes were unbiased (this assumption adequately reflects base rates, and it was supported empirically in the pretest). Second, we used the pretest data as a reference point: In the pretest, all items were new, and therefore, experiences of discrepant fluency resulting from prior presentation could not occur (see Footnote 4).

2. Experiment 1

Experiment 1 followed the standard setup of a truth effect experiment as described above (cf. Begg et al., 1992). At t_1 , participants listened to 30 statements from a male voice and 30 statements from a female voice. They were either told that statements read by the female voice were true and the statements read by the male voice were false, or vice versa. Thus, recollecting the source of an old statement allowed for a "correct" classification as true or false. After a distracter task, participants were presented with 120 statements at t_2 (60 old, 60 new statements) and indicated whether they believed them to be true or false.

2.1. Methods

2.1.1. Participants and design

Seventy-one students (43 women) from various faculties of the University of Heidelberg participated for a monetary compensation of €5 (about \$6.50). They were randomly assigned to one of two between-participants conditions: In one condition, the statements read by the female voice were labeled 'true' and the statements read by the male voice were labeled 'false', and vice versa in the other condition.

Within participants, we manipulated three variables. First, half of the 120 statements were assigned to the *old* and half to the *new* set. Old statements were presented at t_1 and t_2 , while new statements were presented only at t_2 . At t_1 , statements would be presented acoustically, while at t_2 , all statements would be presented visually. Second, 30 of the 60 old statements and 30 of the 60 new statements would be factually true statements, while the remaining statements would be factually false statements. Third, at t_1 , half of the 60 statements were presented by a female voice and the other half by a male voice. Thus, depending on the between-participants condition, they were assigned the status of 'true' or 'false'. Hence, at t_1 , both the 'true' and the 'false' voice presented 15 factually false statements and 15 factually true statements, resulting in 60 statements. At t_2 , all 120 statements were presented visually and participants judged whether a statement was true or false. These judgments should then be influenced by the processes outlined in Fig. 1.

2.1.2. Material

We used the 120 statements which were pretested and discussed above. To implement the voice manipulation, we recorded the readings of all 120 statements by a woman as well as by a man. To fully randomize the assignment to t_1 and t_2 as well as the order of presentation, we digitalized and cut these recordings so that for each statement two sound files existed; one for the male and one for the female voice. These sound files were included in a computer program. The program controlled the random presentation of all statements with the constraints imposed by the design and recorded participants' responses to the statements.

2.1.3. Procedure

Experimental sessions included up to four participants. Upon arrival, they were greeted by an experimenter and seated in a cubicle in front of a PC. They were given a consent form, informing them that they would participate in an experiment to determine how people judge the truth of statements. Then the experimenter started the computer program. The first screen instructed them to put on provided headsets and to listen to a first set of statements. Depending on the condition, the next screen informed them that the statements read by a woman (a man) were true and the statements read by a man (a woman) were false. Next, they heard 60 randomly selected statements from the whole set of 120 statements, with the constraints imposed on the selection by the manipulations described above, constituting the presentation phase at t_1 .

Following this presentation, they performed a distracter task which lasted about 30 minutes. Then they started the t_2 judgment phase. In this phase, the full set of 120 statements was presented in a new randomized order for each participant. The statements were presented visually in the center of the screen and participants had to classify them as “true” or “false” by pressing one of two clearly marked keys. After a key press, the statement disappeared, and, following a 1.5 s interval, the next statement appeared. The computer recorded the decisions. Upon completing the 120 judgments, the computer prompted participants to contact the experimenter, who thanked, paid, and fully debriefed them. Depending on individual speed, a session lasted between 50 and 60 min.

2.2. Results

The probabilities of “true” judgments are given in Table 2, separately for old and new statements and factually true and false statements; for the old statements, the means are also given separately for the “true” and the “false” voice. As Table 1 clearly shows, the between-subjects manipulation of gender (female/male voice = “true”) had no impact and is omitted from further analysis. First, we analyzed the data using an ANOVA approach: The standard truth effect is evident in the difference between old and new statements, $F(1, 70) = 50.88$, $p < .001$, $d = 1.71$. Participants judged old statements to be true with higher probability ($M = .587$, $SD = .134$) than new statements ($M = .470$, $SD = .159$). Testing these values against the chance level of .5 reveals that old statements were judged to be true significantly above chance, $t(70) = 5.37$, $p < .001$, $d = 1.28$, while the new statements tended to be true below chance, but not significantly so, $t(70) = -1.57$, $p < .12$. Participants had also substantial source memory: Statements made by the “true” voice were judged to be true with higher probability ($M = .756$, $SD = .145$) than statements made by the “false” voice ($M = .415$, $SD = .212$), $F(1, 70) = 138.11$, $p < .001$, $d = 2.81$. Participants also seemed to have some factual knowledge, as shown by a comparison of the factually true ($M = .574$, $SD = .143$) and factually false statements ($M = .521$, $SD = .124$), $F(1, 70) = 29.42$, $p < .001$, $d = 1.30$. However, the effect could be due to differential memory for ‘true’ and ‘false’ sources, as the comparison for old statements from the true voice and the false voice suggests. Thus, the propensity to judge old statements as true represents a mixture of recollection and fluency processes. These will be dissociated in the MPT analysis.

2.3. Model-based analyses

In the above ANOVA analyses of the proportions of “true” responses, we found evidence for a referential component of the truth effect—participants rated items as true when they remembered that it was from a truthful source, and rated them as false when they remembered them to originate from an untruthful source. Second, we found evidence for a non-referential part of the truth effect, in that participants judged old items to be true more often than new items, and there was a tendency, though non-significant, for new items to be judged as true less often than would be predicted by chance. Third, there was an effect of an item’s factual truth: True items were judged as true more often than false items. To disentangle the different components underlying the truth effect, we applied our MPT model to the data.

Departing from the general model depicted in Fig. 1, two simplifying assumptions were made in the first set of analyses. First, as in the pretest, it was not possible to dissociate knowledge (parameter k) and processing fluency due to a statement’s actual truth (f_t). Therefore, a single parameter k was computed that comprises the effects of both of these processes (this choice was arbitrary; similar results were obtained when a single f_t parameter was used instead). Second, reflecting base rates and the pretest findings, guessing was assumed to be neutral, $g = .5$. Hence, the following four processes were considered: (1) Discrimination between factually true and false statements (parameter k), (2) recollection of a statement’s source (parameter c), (3) a fluency experience resulting from a positive discrepancy (parameter f_p), leading to a “true” judgment, and (4) a fluency experience resulting from a negative discrepancy (parameter f_n), leading to a “false” judgment.

To assess goodness-of-fit of the model, the discrepancy was evaluated between empirical category frequencies and those predicted by the model. To quantify and test this discrepancy, the G^2 statistic was computed; G^2 is asymptotically χ^2 -distrib-

Table 2

Average proportions of statements’ classification as “true” as a function of voice (spoken by a “true” voice or “false” voice), factual truth (true vs. false), and repetition status (old vs. new)

	Old statements				New statements	
	Factually true		Factually false		Factually true	Factually false
	‘True’ voice	‘False’ voice	‘True’ voice	‘False’ voice	n/a	
‘True’ voice female ($n = 36$)	.761 (.152)	.467 (.260)	.769 (.156)	.391 (.219)	.500 (.189)	.432 (.187)
‘True’ voice male ($n = 35$)	.758 (.171)	.450 (.223)	.737 (.192)	.352 (.219)	.508 (.149)	.442 (.153)
Total ($n = 71$)	.760 (.161)	.458 (.241)	.753 (.174)	.372 (.218)	.504 (.169)	.437 (.170)

The gender of the “true/false” voice is a between-participants variable. Standard deviations are given in parentheses.

uted (cf. Hu & Batchelder, 1994). As revealed by a sensitivity power analysis (Faul, Erdfelder, Lang, & Buchner, 2007), the goodness-of-fit tests reported below were capable of detecting very small deviations from the model ($w \leq .07$; see Cohen, 1988, Chapter 7) with a power of $1 - \beta = .99$ and $\alpha = .01$.

Initially, to test whether a single positive discrepancy fluency parameter (f_p) might suffice to account for the data, we restricted the parameter for negative discrepancy to zero, $f_n = 0$, and considered a model with only three free parameters—discrimination (k), recollection (c), and (positive) fluency (f_p). However, the three-parameter model was not able to account for the data adequately: The goodness-of-fit test, $G^2(9) = 28.19$, $G^2_{crit}(9) = 21.67$, $G^2 > G^2_{crit}$, revealed a significant discrepancy between the observed frequencies and those predicted by the model. That is, the three parameters were not sufficient to describe the pattern of effects observed in Experiment 1.

In a second step, we therefore dropped the restriction $f_n = 0$, and estimated parameter f_n from the data. That is, we allowed for the possibility that participants judge a statement as “false” when they experience a negative discrepancy between the actual and expected fluency. This model provides a good description of the data: As confirmed by the non-significant result, $G^2(8) = 13.22$, $G^2_{crit}(8) = 20.09$, $G^2 < G^2_{crit}$, the discrepancy between the observed frequencies and those predicted by the model was negligible. This strongly supports that the non-referential part of the truth effect consists indeed of two dissociable sub-components.

Parameter estimates and confidence intervals for the four-parameter model are given in Table 3. The discrimination and parameter was almost identical to that of the pretest data, $k = .05$. This finding supports that discrimination was comparable for the pretest data and the data from Experiment 1 (in addition, it provides empirical justification for using the pretest data as a baseline in a second set of analyses; see Footnote 4). Source recollection probability was estimated at $c = .36$. The probability of a “true” judgment due to a positive fluency discrepancy was $f_p = .28$; the probability of a “false” judgment due to a negative fluency discrepancy was $f_n = .06$.

Next, we tested whether the parameters of the model were significantly different from zero. This is usually done by comparing the goodness-of-fit statistic, G^2 , of the model in which the parameter of interest is free to take on any value between zero and one with the goodness-of-fit of the model in which the parameter is fixed to zero. The difference between the two G^2 values, ΔG^2 , is also asymptotically χ^2 -distributed. Replicating the pretest findings, the discrimination parameter k was significantly greater than zero, $\Delta G^2(1) = 26.03$, $p < .05$.

Turning to the memory parameters, there was substantial recollection of a statement’s source. On the basis of the analyses of the raw frequencies, we would predict an effect on the recollection parameter c . In fact, parameter c was significantly different from zero, $\Delta G^2(1) = 521.95$, $p < .05$, supporting its validity as a measure of conscious source recollection. In sum, both analyses converge to show that—independent from a statement’s factual truth—participants judged a statement to be “true” when it originated from the ‘true’ voice, and they judged a statement to be “false” when it originated from the ‘false’ voice when they were able to recall its source.

Most importantly, when conscious source recollection failed, participants based their judgments of old items on their experiences of fluency. First, they tended to judge old items to be true independent of the voice in which the item was presented. This effect was accurately reflected in the model parameter f_p being significantly greater than zero, $f_p = .28$, $\Delta G^2(1) = 140.34$, $p < .05$. Thus, parameter f_p was validated as a measure of a positive fluency discrepancy resulting from prior presentation.

Moreover, the tendency for a negative fluency effect—that is, an increase in the probability of “false” judgments for new items due to a negative fluency discrepancy—was adequately captured in parameter f_n , which was significantly greater than zero, $f_n = .06$, $\Delta G^2(1) = 14.97$, $p < .05$. Hence, the (non-significant) tendency toward a negative-discrepancy effect for new items that was present in the raw data was paralleled by a significant effect on the model’s parameter f_n . Together with the above finding that a model without parameter f_n is not capable of fitting the data from Experiment 1, this is strong evidence for an effect of negative fluency discrepancy.⁵

The latter two parameters capture the non-referential part of the truth effect and show that the truth judgements are indeed influenced by the differential processing experiences elicited by old and new statements. In sum, these analyses demonstrate the validity of the model and its four parameters for discrimination (k), source recollection (c), positive fluency discrepancy (f_p), and negative fluency discrepancy (f_n). It is thereby shown that the model is a valid measurement model for these processes, and that it can be used to measure these processes in studies of the truth effect.

⁵ One might argue that the restriction of $g = .5$ was not justified, and that this assumption somehow distorted the results. However, given that it accurately reflects the base rate of true and false statements in the experiment, and given that the pretest has revealed unbiased guessing, it seemed a plausible assumption. To demonstrate empirically that guessing was indeed unbiased, and that the restriction did not bias the results, we computed a second set of analyses. In these analyses, we included the pretest data in the model and used it as a baseline (see Appendix for detail). By setting the parameters k and g equal across both data sets, we were able to estimate the guessing parameter from the data. If the above findings were an artifact of restricting $g = .5$, then we would expect that, first, estimates of parameter g should deviate substantially from the neutral point of .5, and, second, the above findings should not replicate when g is no longer restricted. Results showed that the modified model also provided a good fit to the data, $G^2(9) = 13.25$, $G^2_{crit}(9) = 21.67$, $G^2 < G^2_{crit}$, and that the parameter values replicate the four-parameter model. Importantly, both of the fluency parameters were significantly different from zero, $f_p = .26$, $G^2(1) = 46.23$, $p < .05$, and, $f_n = .08$, $G^2(1) = 5.65$, $p < .05$. Similarly, recollection of a statements’ source, $c = .36$, was different from zero, $G^2(1) = 522.85$, $p < .05$, and finally, guessing did not differ from the neutral point, $g = .51$, $\Delta G^2(1) = 0.50$, *ns*. In sum, there was little evidence that the assumption of unbiased guessing was inadequate.

Table 3
Parameter estimates (and 95% confidence intervals) for the data from Experiment 1

Process	Parameter	Estimated value
Knowledge	k	.05 (.03 .07)
Source recollection	c	.36 (.33 .39)
Positive fluency	f_p	.28 (.23 .33)
Negative fluency	f_n	.06 (.03 .09)

Parameter values represent the probabilities of the respective cognitive processes.

2.4. Discussion

We presented a model of the truth effect and used an MPT method to test the model in Experiment 1. This model represents an extension of the process-dissociation model used by Begg and colleagues (1992). In addition to this earlier model, we included parameters for knowledge, guessing, and positively as well as negatively discrepant fluency experiences. This extended model presented a good description of the data and was in line with analyzing the probabilities of “true” judgments in a conventional way. Using this method, we dissociated different components of the truth effect. First, participants classified statements as true or false based on their knowledge or the fluency-experience due to a statement’s truth. Second, participants classified statements as true or false based on their memory for the source of that item. In this specific case, we manipulated the sources to be credible or not. The estimated memory parameter c is equivalent to the effect of convergent validity (Arkes et al., 1991), or the *controlled* component as demonstrated in traditional process-dissociation analyses of the effect (Begg et al., 1992). Third, participants classified statements as true or false based on the experience elicited by that statement. While previous approaches focused almost exclusively on the increase in rated truth observed for *old* statements (e.g., Hawkins & Hoch, 1992), we showed that this non-referential component consists of two sub-processes, as evinced by the f_p and f_n parameters. Given that repetition increases processing fluency (Feustel et al., 1983), the present method showed that participants not only used the fluency experience elicited by old statements to judge them as true, but they also used the relatively non-fluent processing of new statements to judge them as false (within Experiment 1 as well as compared to a pretest baseline; see Footnote 4). The parameter estimates also show that the influence of the positive discrepancy was stronger than the negative discrepancy, or non-fluent processing.

This result also explains a recent finding by Hansen et al. (2008), who failed to find an effect of negative discrepancy on truth judgments. As these authors employed rating scales instead of dichotomous judgments, and their effects were rather small, this failure represents most likely a power issue. Rating scales are usually less suited to detect a truth effect (see Begg et al., 1992; Unkelbach, 2007); thus, they probably were able to detect the stronger effect of positive discrepancy, but not the smaller effect of negative discrepancy. On the other hand, one might argue that the assumption of a negative discrepancy effect is not necessary, and that the two parameters are redundant. However, a model with only one fluency parameter was not able to account for the data from Experiment 1. In light of this finding, and given previous work on differential processing experiences (e.g., Whittlesea and Williams, 2001), the present model, including the possibility for a negative discrepancy, appears to be both appropriate and comprehensive.

We have conceptualized the non-referential components as the interpretation and use of processing fluency according to its ecological validity. Now the question remains whether the influence of these experiences is indeed automatic—as implied by Begg and colleagues (1992)—or, rather, under participants’ control.

3. Experiment 2

Experiment 2 will demonstrate that the truth effect’s non-referential part reflects the controlled interpretation of processing fluency. To do this, we provided an explicit interpretation of the fluency experience; if the non-referential part of the truth effect is susceptible to manipulations by instructions, we have evidence that the effect is not automatic (for an overview how such naïve beliefs influence the interpretation of fluency experiences, cf. Schwarz, 2004). This manipulation will also answer whether participants’ ability to discriminate between true and false statements (i.e., the effects on parameter k) in the pretest and Experiment 1, are indeed due to knowledge or based on statements’ processing fluency due to truth. Factual knowledge should not be influenced by interpretations of processing fluency.

At t_1 , we informed participants that either all statements presented in that phase were true, or that all statements were false. The fluency experience resulting from the presentation at t_1 and the fluency experience resulting from a statement being true should be functionally equivalent, as this is the basic tenet of the present approach. Therefore, participants should apply the information about the statements presented at t_1 to all statements at t_2 . That is, they should judge all fluently processed statements as true when told that “all old statements are true” (i.e., in the old = true condition), and they should judge all fluently processed statements as false when told that “all old statements are false” (i.e., in the old = false condition). As true statements are processed more fluently, they should be judged as true more often in the old = true condition, and they should be judged as false more often in the old = false condition, as compared to false items. The old = false condition is thus equivalent to communicating a false ecological validity (i.e., interpretation) of the fluency cue.

However, the use of the information given in the instruction (e.g., “all old statements are false”) is also possible when a statement’s earlier presentation is recollected. If substantial recollection is present, we would thus not be able to dissociate its effects from those of fluency due to prior presentation, and, consequently, from those of fluency due to a statement’s truth. Thus, we attempted to minimize the impact of conscious recollection and fluency due to presentation at t_1 by switching to a visual presentation with a very short presentation time.

In sum, we were only using the setup of a memory experiment as such that participants could misinterpret the fluency due to a statement’s truth as being due to prior presentation at t_1 . Given that, they should apply the explicit instruction about the truth status of old statements (all true/all false) to factually true statements, thereby demonstrating that the use and interpretation of the fluency experience is a controlled process.

3.1. Methods

3.1.1. Participants, design, and materials

Sixty students (47 women) from various faculties of the University of Heidelberg participated for a payment of €4 (approximately \$5). They were randomly assigned to one of the two between-participants instruction conditions; participants in the old = true condition received instructions that all statements from t_1 would be true, whereas participants in the old = false condition received instructions that all statements from t_1 would be false. The same 120 statements as in Experiment 1 were used. The manipulation of statements being old (presented at t_1 and t_2) or new (only presented at t_2) and statements being factually true or false were again realized as within-participants manipulations. Thus, there were 30 factually true statements and 30 factually false statements at t_1 , while at t_2 , all 120 statements were presented. We also implemented two major changes in the computer program from Experiment 1: At t_1 , the program now presented the statements visually, and before the presentation the program presented an instruction stating that all the following statements were either true or false, depending on the respective between-participants condition.

3.1.2. Procedure

Experimental sessions included up to four participants. Upon arrival, they were seated in a cubicle in front of a computer. The experimenter handed them a consent form which informed them that they would take part in a study investigating how people form judgments of truth. If they agreed to participate, the experimenter started the computer program. Depending on the condition, the program provided information that all following presented statements were either *true* or *false*. Then, participants could start the presentation by pressing the space bar. The computer program randomly selected 30 factually true and 30 factually false statements. The statements were presented at the center of the screen for 1.5 s; after a delay of 1 s, the next statement was presented. This rapid presentation aimed to reduce effects of recollection to a minimum. After the t_1 presentation phase, participants engaged in a distracter task for about 20 min, after which the t_2 judgment phase started. This phase was identical to Experiment 1. Upon completing their judgments for the 120 statements, the experimenter thanked, paid, and fully debriefed participants about the hypotheses underlying the present research. Sessions lasted between 35 and 45 min, depending on individual speed.

3.2. Results

The average proportions of “true” judgments are displayed in Table 4 for factually true and false statements, separately for old and new statements. Two effects are immediately visible from this data. First, participants clearly understood the instruction and judged more statements as true when they were told that all old statements were true ($M = .653$, $SD = .118$) than when they were told that all old statements were false ($M = .375$, $SD = .106$), $F(1,58) = 92.61$, $p < .001$, $d = 2.53$. However, Table 4 also clearly shows that participants applied the instruction differentially to factually true and false statements, resulting in a highly significant interaction, $F(1,58) = 163.55$, $p < .001$, $d = 3.36$. If we analyze these data by factually true and false statements, this pattern becomes even more apparent: Factually true statements were judged to be true when the instruction was that *old* statements were true ($M = .771$, $SD = .143$); if the instruction was that old statements were false, factually true statements were much less likely to be judged as true ($M = .265$, $SD = .127$), $t(58) = 14.49$. Factually false statements were not affected by the instructions about old statements ($M = .536$, $SD = .118$ vs. $M = .484$, $SD = .137$), $t(58) = 1.55$, *ns*. Thus, participants applied the instruction about “old” statements to factually true statements, and judged them as true or false accordingly. The only bridging construct between “old” and “true” in this framework is then indeed

Table 4

Average proportions of statements’ classification as “true” for factually true and false statements as a function of the instruction at t_1 and statements’ old/new status (standard deviations in parentheses)

	Old statements		New statements	
	True	False	True	False
“Old statements are true”	.778 (.147)	.519 (.140)	.763 (.155)	.552 (.149)
“Old statements are false”	.262 (.156)	.481 (.143)	.268 (.126)	.488 (.160)

processing fluency. This will become even more apparent when we apply the MPT model to the data to decompose the factors underlying the truth effect.

3.3. Model-based analyses

Analyses of the proportions of “true” judgments revealed that there were more “true” judgments in the old = true condition as compared to the old = false condition; importantly, this was the case only for factually true statements but not for factually false statements. This pattern implies that participants were somehow able to discriminate between true and false statements. Response tendencies alone are not sufficient to account for this pattern, although response tendencies (“guessing”) are likely to be also affected by the instructions. More importantly, the observed difference between true and false statements is unlikely to be based on knowledge about factual truth or falsity of statements because such knowledge would have led participants to respond correctly in both conditions, regardless of the instruction. Instead, we believe that the differential fluency experienced for true and false statements (as observed in the pretest) was used as a cue on which participants’ judgments were based, and that this cue was used in a controlled manner, leading to “true” responses in the old = true condition, and to “false” judgments in the old = false condition. This process is represented in the model by parameter f_t , and we will demonstrate that a model accommodating this process can account for the data from Experiment 2.

The MPT model applied here deviates from the model used in Experiment 1. First, the data from Experiment 2 allows dissociating knowledge and fluency due to truth, as they should lead to different patterns of responses in the old=false condition. In this condition, knowledge would still lead to a correct judgment, whereas fluency would now be interpreted opposite to its ecological validity, such that fluently processed statements are interpreted as old and therefore inferred to be *false*; analogously, statements that are experienced as non-fluent would be interpreted as new and therefore inferred to be *true*. Thus, separate parameters k and f_t could be estimated.⁶

Second, because the instruction manipulation is expected to influence guessing, two separate guessing parameters were estimated, a $g_{\text{old} = \text{true}}$ and a $g_{\text{old} = \text{false}}$ parameter. Because of this additional guessing parameter, it was necessary to impose an additional restriction. Therefore, we were not able to separate the differential processing experiences f_p and f_n in Experiment 2. This distinction is not central to Experiment 2, and it was already shown in Experiment 1 that both processes contribute to truth judgments. We still allowed for the occurrence of both processes but we could not compare the magnitude of the effect of the fluency experience caused by old and by new statements. Instead, the difference in “true” judgments between old and new statements that could be ascribed to prior presentation was modeled by a single parameter f_p in Experiment 2. The model equations are given in the [Appendix A](#).

The assumptions incorporated into the model were tested with a goodness-of-fit test. The power of this test to detect very small deviations from the model ($w = .06$) was again very good, $1 - \beta = .99$, with $\alpha = .01$ and $N = 7200$, $G^2_{\text{crit}}(2) = 9.21$. The resulting $G^2(2) = 2.70$, $G^2 < G^2_{\text{crit}}$, revealed that the model, including our additional assumptions, described the data well.

Next, parameter estimates and hypotheses tests were computed. Results are given in [Table 5](#). Interestingly, when fluency due to truth is separated from knowledge in a measurement model, no substantial evidence remained for knowledge, $k = .01$, $\Delta G^2(1) = 1.20$, *ns*.⁷

However, there was strong evidence for an effect of fluency due to truth, $f_t = .23$, $\Delta G^2(1) = 408.59$, $p < .05$. This result implies that a simpler and more parsimonious model without the process represented by parameter f_t cannot account for the data. The effect on parameter f_t supports our interpretation that the above interaction was caused by a fluency experience that was associated with true statements more often than with false statements: Factually true statements were more likely to be judged as “true” in the old = true condition and as “false” in the old = false condition, as compared to factually false statements. That is, participants misinterpreted the fluency experience due to a statement’s truth as evidence for the statement being old, and judged it according to the respective explicit rule (either old = true or old = false).

The remaining parameter estimates were as expected. The shortened presentation was successful in reducing conscious recollection of statements to zero, $c < .01$, $\Delta G^2(1) = 0$, *ns*. This was also true for fluency due to prior presentation, $f_p < .01$, $\Delta G^2(1) = 0$, *ns*. Third, guessing parameters reflected an expected bias towards the “true” response in the old = true condition, $g_{\text{old} = \text{true}} > .5$, $\Delta G^2(1) = 244.72$, $p < .05$, and an expected bias towards the “false” response in the old = false condition, $g_{\text{old} = \text{false}} < .5$, $\Delta G^2(1) = 156.40$, $p < .05$.

3.4. Discussion

The results from Experiment 2 clearly demonstrate that the interpretation of the fluency experience can be changed by providing simple instructions about the validity of the experience. True statements were judged “true” more often in the

⁶ We first considered a model with two different f_t parameters, one for each experimental condition. This model achieved a good fit to the data, $G^2(1) = 2.69$, $G^2 < G^2_{\text{crit}}$. However, both parameters could be set equal across conditions without significant reduction in model fit, $\Delta G^2(1) = 0.01$, $p = .92$, and we therefore proceeded with the simpler model in the interest of parsimony.

⁷ A possible problem in estimating the knowledge parameter arises from theoretically possibly conflicts of factual knowledge (e.g., that an old statement is indeed true) and instructions (e.g., all old statements are false). However, as we have shown previously, factual knowledge regarding the statements was minimal and rather based on the statements’ processing fluency. Thus, we are confident that knowledge did not contribute to performance in the present experiments.

Table 5
Parameter estimates (and 95% confidence intervals) for the data from Experiment 2

Source	Parameter	Estimated value
Knowledge	k	.01 (.00 .03)
Source recollection	c	<.01 (.00 .19)
Fluency due to presentation at t_1	f_p	<.01 (.00 .09)
Fluency due to truth	f_t	.23 (.20 .26)
Guessing (instruction old = true)	g_{ot}	.70 (.64 .76)
Guessing (instruction old = false)	g_{of}	.34 (.28 .39)

Parameter values represent the probabilities of the respective cognitive processes.

old = true instruction condition than in the old = false instruction condition. Importantly, note that the instructions were about “old” statements. However, as our pretests have shown, the factually true statements are indeed easier to process and judged to be more familiar. Thus, the feasible explanation for the data is that participants did not distinguish between fluency due to factual truth and fluency due to prior presentation. Instead, they judged all fluently processed statement as true or false according to the instruction provided at t_1 . This is equally true for old and new statements; thus, participants must have applied the explicit interpretation provided for “old” statements to factually true statements. As the only connecting construct between old and true statements is processing fluency, this strongly supports the fluency interpretation account of the non-referential part of the truth effect. Furthermore, it shows that the interpretation and use of that processing experience is not automatic after all.

This account is supported by the proposed MPT model, which fitted the present data well. First, it showed that participants' guessing strategies were influenced by the information given at t_1 , leading to more “true” responses in the old = true and more “false” responses in the old = false conditions. Second, it showed that the very short presentation time was successful in reducing conscious recollection to basically zero. Similarly, the parameter for factual knowledge was also zero. And most importantly, it showed that participants used the fluency of a statement that originates from its truth to make differential judgments at t_2 , based on the instruction about old statements given at t_1 . As a corollary, this implies that the knowledge in the pretest and in Experiment 1 was probably due to the correct usage of the fluency experience due to truth, and not due to knowledge in the common denotation of the word.

Ideally, we would have seen a similar interpretation effect for old and new statements as well. A straightforward explanation for the lack of a fluency effect due to prior presentation in this experiment is indeed the difficult presentation at t_1 ; as the confidence intervals of the respective parameters show, memory performance in this experiment was indeed poor. Moreover, the explicit instruction that *all* statements are either true or false might have reduced the level of processing invested in the statements presented at t_1 .

Another question is why fluency due to truth has not been observed more frequently. The answer lies in the fact that both factual knowledge and fluency due to truth lead to the same “true” responses in almost every truth experiment, similar to our pretest and Experiment 1. Only if fluency due to truth and knowledge lead to different outcomes—as in Experiment 2—can the two processes be dissociated.

In sum, the experiment provides further support that the non-referential part of the truth effect is due to the interpretation of the fluency experience. Furthermore, it shows that fluency due to repetition and fluency due to truth are functionally equivalent. In hindsight, this is not even surprising, as fluency due to truth is possibly based on prior exposure as well (see discussion above). And finally, Experiment 2 strongly suggests that while the increase in fluency might be automatic due to basic cognitive processes (Feustel et al., 1983; Reber, Wurtz, & Zimmermann, 2004), the interpretation of the fluency experience is open to willful and conscious control, as the simple provision of a verbal instruction was sufficient to completely reverse the impact of the processing experience associated with true statements.

4. General discussion

The present approach conceptualized the truth effect as the interpretation of the differential fluency experiences elicited by old and new statements. The interpretation is based on the correlation between the experience and an external criterion, here, the factual truth or falseness of a statement. Thus, processing fluency is used as a proximal cue to judge the distal criterion of truth (Brunswick, 1957). We thereby follow the idea that the truth effect is based on an inferential process (Skurnik, Yoon, Park, & Schwarz, 2005; Winkielman & Schwarz, 2001), which draws on an experience that is neutral and malleable to begin with (Leboe & Whittlesea, 2002; Mandler et al., 1987). We formalized these assumptions into a comprehensive MPT model; this model, with slight adaptations, provided a good description of the data obtained the pretests and in two experiments.

The model represents our theoretical assumptions about the processes underlying the truth effect in a formal way that allows for the derivation of quantitative hypotheses. Further, it provides a framework for the empirical testing of such hypotheses. Across two experiments, in which we manipulated fluency by repeated presentation (Experiment 1), as well as the interpretation and use of the fluency experience regardless of its source (Experiment 2), the model has fared well. In Experiment 1, repeated presentation affected the f_p and f_n parameters as predicted. In Experiment 2, a manipulation of the interpretation of the fluency experience affected true statements more than false statements, and this was reflected

in the f_t parameter as predicted. Thus, the critical parameters of the model have been successfully validated. The other parameters—recollection (parameter c), guessing (parameter g), and knowledge (parameter k)—represent psychological processes whose role in the truth effect are not under debate; in part, they were successfully validated in previous models of which the present model is but an extension (e.g., Begg et al., 1992). In our view, this model represents an important starting point for a more formal investigation of the truth effect.

A possible alternative explanation for the truth effect would be that the fluency experience might have an inherent meaning, similar to other subjective experiences like anger, joy, hunger, thirst, or pain. Indeed, previous research indicates that processing fluency is an inherently positive experience (Reber, Schwarz, & Winkielman, 2004; Reber, Winkielman, & Schwarz, 1998; Winkielman & Cacioppo, 2001). Given that most people evaluate true information as more positive than false information, the non-referential part of the truth effect could be based on the experience's positivity. The present data, however, as well as other studies (Unkelbach, 2007, Experiments 2 and 3), show that its denotative use is open to different interpretations. As we clearly demonstrated in Experiment 2, participants could use the processing fluency experience elicited by a statement to judge it as false. As said, this result does not preclude a positive connotation of the experience, but it speaks strongly against the idea that the truth effect is based on an inference from the experienced positivity to the truth of a statement. The same is true for models in which *familiarity* invariably leads to a “true” judgment.

From this inferential cue-learning conception followed the question of how automatic the fluency-based component of the truth effect really is. In Experiment 2, the inherent fluency of true statements was interpreted according to the explicit instruction provided within the experiment about “old” statements. Thus, we might distinguish between an automatic increase in processing fluency and the subsequent non-automatic use in judgments; but before we explore this distinction in more depth, there are two more open questions that warrant discussion.

First, why do the results differ so markedly from other experiments employing process-dissociation procedures, the most relevant being Begg and colleagues' (1992) work? The answer lies naturally in the different procedures. While in Experiment 2 participants were told that *all* old statements were either true or false, the classic process-dissociation procedure experiment informs participants that half of the statements are true or false, depending on the source (as in our Experiment 1). This procedure forces participants to exactly recollect the source of an item, because at t_1 , both supposedly true and false statements are provided. In these cases, the fluency experience does not provide a valid cue about the truth of a statement, as it might be due to presentation from a “true” source, a “false” source, or the statements' factual truth. In contrast, in the setup of Experiment 2, all old statements were said to be either true or false, and the experience of fluency could be directly interpreted, depending on the given instruction.

Second, how does the present approach fit within existing models of memory and consciousness? As said, we conceptualize the fluency-based component of the truth effect as the interpretation of an experience elicited by the stimulus. A broad theoretical model to explain such effects is the SCAPE model (selective construction and preservation of experiences) by Whittlesea and colleagues (e.g., Leboe & Whittlesea, 2002; Whittlesea & Williams, 2001). According to this framework, mental events are constantly monitored and evaluated. If a mental event is surprising or discrepant, for example, if processing is surprisingly fluent, there is a need to explain that discrepancy. According to that model, for example, a feeling of familiarity arises when the discrepancy is attributed to a source in the past, while recognition based on recollection requires the reconstruction of the context in which the item was encountered. Although the present approach differs conceptually from a misattribution explanation, the difference in relation to the SCAPE model lies largely in the terminology used. We employed the term interpretation; however, there is no problem for the argument when the term attribution is used. In such a framework, the effect works as follows: First, a statement is fluently processed. If the context of the statement is accessible, it is recognized, the experience is explained, and an informed decision concerning the truth is possible. If the contextual details are not accessible, the experience needs explanation. In earlier presentations of the model, Whittlesea (1993) proposed an automatic attribution to a source in the past, resulting in a feeling of familiarity. Here, we simply argue that the experience is explained according to its ecological validity in the given context. In most cases, processing fluency is indeed caused by a source in the past, and thus, the experience can be used as a cue for previous encounters with a stimulus. However, many other interpretations are conceivable (Mandler et al., 1987; Reber, Zimmermann, & Wurtz, 2004). Although we did not present a strict test for the ecological validity assumption in the present experiments (but see Unkelbach, 2006, 2007), the data are fully consistent with that notion.

Now again, is the interpretation of the experience a conscious and controlled, or an unconscious and automatic process? As we have shown here, it is possible to systematically influence the impact of the experience on judgments by explicit instructions, implying a controlled process. Furthermore, the conscious application of an explicit rule to subjective mental experiences was already demonstrated in other paradigms (Skurnik, Schwarz, & Winkielman, 2000; Skurnik et al., 2005; Winkielman & Schwarz, 2001). Yet, the argument in Unkelbach (2006, 2007) for the interpretation of the experience according to its ecological validity rested on a feedback learning process, rather than on an explicit rule. From a cue learning perspective, this represents no contradiction, as even basic conditioning processes seem to depend on consciousness and inferences (Lovibond & Shanks, 2002). For example, the interpretation of a cue could be controlled in the beginning and become automatic with frequent usage (cf. Bargh & Chartrand, 1999; Bargh & Williams, 2006). The difference between the explicit instruction and the learning approach is simply the way in which the validity of the cue is conveyed. The learning approach represents a more realistic setting, in which the validity must be learned, implicitly or explicitly, from an environment in which only noisy cue-criterion correlations exist. The provision of an explicit and valid rule is an instance that might never happen, but serves only as a pointed example that the inferential stage is far from automatic. However, there is also

the possibility of a mixture of both. For example, people learn and relearn the usage of depth cues in perception (e.g., Jacobs, 2002); while in the beginning, there is a need to actively select the appropriate cues, adaptation processes allow to let go of that control and the selection and usage of cues becomes increasingly automatic (for a similar view on the usage of heuristics, cf. Gigerenzer & Goldstein, 1996).

Furthermore, if one follows the reasoning of Mandler (1980), as well as the ideas of Leboe and Whittlesea (2002), the process underlying the seemingly dissociated parts of the truth effect, recollection and experience, might be the same, but based on different kinds of information. These authors argue against a functional dichotomy of recognition by recollection and recognition by feelings of familiarity (i.e., based on fluency). Instead, both ways are supposedly based on an inferential process. This fits well with the present result that apparent knowledge about the truth of statements (cf. the pretests) is actually based on the fluency of these true statements (cf. Experiment 2). The process is then an inferential one for both components, only that one is based on an experience, whereas the other is based on the recollection of specific details, for example, that you have indeed read the introductory statement in *National Geographic*. In fact, the Vatnajökull is Europe's volumetric largest glacier. For the present purposes, however, the functional difference between the components is assumed theoretically and found empirically. The *actual* underlying structure of memory is beyond the scope of the present work.

5. Conclusions

We introduced a methodology that allows specifying the underlying components of the truth effect that is based on, but also going beyond, the process-dissociation procedure introduced by Jacoby (1991). Using this methodology, we introduced a comprehensive model of the truth effect, building on previous explanations of the truth effect (Begg et al., 1992; Unkelbach, 2007). Based on this model, we showed that the fluency-based part of the truth effect is malleable according to the conveyed interpretation of the fluency experience. As Begg and colleagues (1992) have noted, repetition *per se* does not provide a logical basis for judgments of truth. Otherwise, Wittgenstein's (1955) ironically suggested strategy of buying several of the same newspapers to ensure that the content is true would indeed be effective. We want to add now that repetition might not represent a logical basis *per se*, but in an uncertain world, the processing fluency associated with repetition nevertheless represents a valid proximal cue to judge the distal property of truth. Thus, the truth effect is not illusory after all, but a byproduct of an adaptive learning mechanism; and in most cases, people are well off to consider a statement as true when it is fluently processed.

Acknowledgments

Both authors contributed equally to this paper. This research was supported by a Feodor-Lynen Research Fellowship awarded by the Alexander von Humboldt–Foundation to Christian Unkelbach. Electronic mail should be sent to christian.unkelbach@psychologie.uni-heidelberg.de.

Appendix A.

A.1. Pretest

In the pretest, true and false statements were judged as either “true” or “false”. For true statements, the probability of a “true” judgment is $p(\text{“true”}|\text{true})$. Given the binary response format, the probability of a “false” judgment for true items is given by $p(\text{“false”}|\text{true}) = 1 - p(\text{“true”}|\text{true})$. Analogously, the probability of a “true” judgment to a false statement is given by $p(\text{“true”}|\text{false})$, and the probability of a “false” judgment for false items is given by $p(\text{“false”}|\text{false}) = 1 - p(\text{“true”}|\text{false})$. The structure of the data is illustrated in Table A1.

Because $p_{12} = 1 - p_{11}$ and $p_{22} = 1 - p_{21}$, the data yields only two independent empirical probabilities. An identifiable model for these data can therefore contain a maximum of two free parameters. A simple model with two processes—discrimination between true and false statements by way of knowledge (parameter k), and guessing (parameter g)—was fitted to the pretest data. All other parameters (i.e., c, f_t, f_n, f_p) were restricted to zero. The model is given by the following equations:

$$p_{11} = k + (1 - k)g, \quad (1a)$$

$$p_{12} = (1 - k)(1 - g), \quad (1b)$$

$$p_{21} = (1 - k)g, \text{ and} \quad (1c)$$

$$p_{22} = k + (1 - k)(1 - g). \quad (1d)$$

Table A1
Structure of the pretest data

Truth value	Response	
	“True”	“False”
True	p_{11}	p_{12}
False	p_{21}	p_{22}

Table A2
Structure of the data from Experiment 1

Item type	Truth value	Response	
		“True”	“False”
‘True’ voice	True	p_{11}	p_{12}
	False	p_{21}	p_{22}
‘False’ voice	True	p_{31}	p_{32}
	False	p_{41}	p_{42}
New	True	p_{51}	p_{52}
	False	p_{61}	p_{62}

A.2. Experiment 1

In Experiment 1, old and new statements were presented. Of the old statements, half were presented by a ‘true’ voice, while the other half were presented by a ‘false’ voice. Half of the statements from each of these sets were factually true and the other half were false. Thus, there were six different statement types: True and false old statements from the ‘true’ voice, true and false old statements from the ‘false’ voice, and true and false new statements. Table A2 illustrates the structure of the data. As the p_{i2} are given by $p_{i2} = 1 - p_{i1}$, the number of independent empirical probabilities, including those from the pretest data, is eight. Furthermore, using the model equations given below, three of these probabilities can be expressed as a function of the other p_{i1} . This leaves five independent equations, implying that an identifiable model can have a maximum of five free parameters. The five-parameter model used in the second set of analyses (see Footnote 3) is given below. The four-parameter model that was used in the first set of analyses is easily derived from the five-parameter model by setting $g = .5$.

Parameters were equated across the male and female voice conditions. The influences of knowledge and fluency due to truth (parameters k and f_t) could not be separated on the basis of the data. Parameter f_t was therefore restricted to zero; as a consequence, parameter k may reflect not only knowledge but also fluency due to truth. In addition to knowledge and guessing, source recollection (parameter c) and fluency due to prior presentation (parameters f_p and f_n) were assumed to contribute to truth judgments, as detailed in the following equations:

$$\begin{aligned}
 p_{11} &= k + (1 - k)c + (1 - k)(1 - c)f_p + (1 - k)(1 - c)(1 - f_p)g, \\
 p_{12} &= (1 - k)(1 - c)(1 - f_p)(1 - g), \\
 p_{21} &= (1 - k)c + (1 - k)(1 - c)f_p + (1 - k)(1 - c)(1 - f_p)g, \\
 p_{22} &= k + (1 - k)(1 - c)(1 - f_p)(1 - g), \\
 p_{31} &= k + (1 - k)(1 - c)f_p + (1 - k)(1 - c)(1 - f_p)g, \\
 p_{32} &= (1 - k)c + (1 - k)(1 - c)(1 - f_p)(1 - g), \\
 p_{41} &= (1 - k)(1 - c)f_p + (1 - k)(1 - c)(1 - f_p)g, \\
 p_{42} &= k + (1 - k)c + (1 - k)(1 - c)(1 - f_p)(1 - g), \\
 p_{51} &= k + (1 - k)(1 - f_n)g, \\
 p_{52} &= (1 - k)f_n + (1 - k)(1 - f_n)(1 - g), \\
 p_{61} &= (1 - k)f_n + (1 - k)(1 - f_n)g, \text{ and} \\
 p_{62} &= k + (1 - k)(1 - f_n)(1 - g).
 \end{aligned}$$

The pretest data was modeled using Equations 1a–1d as introduced above; parameters k and g were equated across both data sets.

Table A3
Structure of the data from Experiment 2

Condition	Item type	Truth value	Response	
			“True”	“False”
Old = True	Old	True	p_{11}	p_{12}
		False	p_{21}	p_{22}
	New	True	p_{31}	p_{32}
		False	p_{41}	p_{42}
Old = False	Old	True	p_{51}	p_{52}
		False	p_{61}	p_{62}
	New	True	p_{71}	p_{72}
		False	p_{81}	p_{82}

A.3. Experiment 2

In Experiment 2, old and new statements were presented that could be either true or false. Participants were told that all old items had the same truth value; old items were said to be true in the old = true condition and false in the old = false condition. The data structure is illustrated in Table A3. As the p_{i2} are again given by $p_{i2} = 1 - p_{i1}$, the number of independent empirical probabilities is eight. This number is further reduced because two empirical probabilities can be expressed as a function of the remaining probabilities, yielding six independent equations.

Assuming that the instruction manipulation does not affect knowledge but does affect the interpretation of fluency, the effects of knowledge can be dissociated from those of fluency due to truth. Thus, we incorporated a parameter k that captured participants' knowledge, and a parameter f_t that captured effects of the interpretation of fluency that is based on a statement being true. Given that the same material was used in both conditions, these parameters were restricted to be invariant across conditions. Given that conditions did not differ with regard to the presentation at t_1 , recollection (parameter c) and fluency due to presentation at t_1 (parameter f_p) were equated across conditions. Guessing was allowed to vary across conditions (old = true: parameter g_{ot} ; old = false: parameter g_{of}). The model is given by the following equations:

$$\begin{aligned}
 p_{11} &= k + (1 - k)c + (1 - k)(1 - c)f_t + (1 - k)(1 - c)(1 - f_t)f_p + (1 - k)(1 - c)(1 - f_t)(1 - f_p)g_{ot}, \\
 p_{12} &= (1 - k)(1 - c)(1 - f_t)(1 - f_p)(1 - g_{ot}), \\
 p_{21} &= (1 - k)c + (1 - k)(1 - c)(1 - f_t)f_p + (1 - k)(1 - c)(1 - f_t)(1 - f_p)g_{ot}, \\
 p_{22} &= k + (1 - k)(1 - c)f_t + (1 - k)(1 - c)(1 - f_t)(1 - f_p)(1 - g_{ot}), \\
 p_{31} &= k + (1 - k)f_t + (1 - k)(1 - f_t)(1 - f_p)g_{ot}, \\
 p_{32} &= (1 - k)(1 - f_t)f_p + (1 - k)(1 - f_t)(1 - f_p)(1 - g_{ot}), \\
 p_{41} &= (1 - k)(1 - f_t)(1 - f_p)g_{ot}, \\
 p_{42} &= k + (1 - k)f_t + (1 - k)(1 - f_t)f_p + (1 - k)(1 - f_t)(1 - f_p)(1 - g_{ot}), \\
 p_{51} &= k + (1 - k)(1 - c)(1 - f_t)(1 - f_t)g_{of}, \\
 p_{52} &= (1 - k)c + (1 - k)(1 - c)f_t + (1 - k)(1 - c)(1 - f_t)f_p + (1 - k)(1 - c)(1 - f_t)(1 - f_p)(1 - g_{of}), \\
 p_{61} &= (1 - k)(1 - c)f_t + (1 - k)(1 - c)(1 - f_t)(1 - f_p)g_{of}, \\
 p_{62} &= k + (1 - k)c + (1 - k)(1 - c)(1 - f_t)f_p + (1 - k)(1 - c)(1 - f_t)(1 - f_p)(1 - g_{of}), \\
 p_{71} &= k + (1 - k)(1 - f_t)f_p + (1 - k)(1 - f_t)(1 - f_p)g_{of}, \\
 p_{72} &= (1 - k)f_t + (1 - k)(1 - f_t)(1 - f_p)(1 - g_{of}), \\
 p_{81} &= (1 - k)f_t + (1 - k)(1 - f_t)f_p + (1 - k)(1 - f_t)(1 - f_p)g_{of}, \text{ and} \\
 p_{82} &= k + (1 - k)(1 - f_t)(1 - f_p)(1 - g_{of}).
 \end{aligned}$$

References

- 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 & Memory*, 5, 241–252.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, 54, 462–479.
- Bargh, J. A., & Williams, E. L. (2006). The automaticity of social life. *Current Directions in Psychological Science*, 15, 1–4.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6, 57–86.
- 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., 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 (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: Harvard University Press.
- Buchner, A. (1994). Indirect effects of synthetic grammar learning in an identification task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 550–566.
- Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General*, 124, 137–160.
- Buchner, A., & Wippich, W. (1996). Unconscious gender bias in fame judgments? *Consciousness and Cognition: An International Journal*, 5, 197–220.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NY: Erlbaum.
- Dechêne, A. (2005). *Hängt die Wahrheit von der Formulierung ab? Einflüsse von syntaktischer Veränderung und Verarbeitungstiefe auf den Truth Effect*. [Does Truth depend on wording? The impact of syntactic variations and depth of processing on the truth effect] Unpublished Masters Thesis, Universität Basel, Switzerland.
- Draine, S. C., Greenwald, A. G., & Banaji, M. R. (1996). Modeling unconscious gender bias in fame judgments: Finding the proper branch of the correct (multinomial) tree. *Consciousness and Cognition: An International Journal*, 5, 221–225.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- 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. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 46, 107–119.
- 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.
- Hansen, J., Dechêne, A., & Wänke, M. (2008). Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology*, 44, 687–691.
- 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.
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15, 635–650.
- Hu, X., & Batchelder, W. H. (1994). The statistical analysis of engineering processing tree models with the EM algorithm. *Psychometrika*, 59, 21–47.
- Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, 6, 345–350.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.
- Jacoby, L. L., 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.
- Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, 98(2), 127–140.
- Kelley, C. M., & Jacoby, L. L. (2000). Recollection and familiarity: Process-dissociation. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory*. Oxford University Press. pp. 215–228.
- Leboe, J. P., & Whittlesea, B. W. (2002). The inferential basis of familiarity and recall: Evidence for a common underlying process. *Journal of Memory and Language*, 46, 804–829.
- Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 3–26.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252–271.
- 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.
- Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effect: III. The sleeper effect is dead. Long live the sleeper effect. *Journal of Personality and Social Psychology*, 54, 203–218.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855–863.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness & Cognition*, 8, 338–342.
- Reber, R., 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 of affective judgments. *Psychological Science*, 9, 45–48.
- Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring 'fringe' consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness & Cognition*, 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.
- Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95, 318–339.
- Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14, 332–348.
- 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, US: Psychology Press.
- Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, 31, 713–724.
- Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, 39, 267–273.
- Topolinski, S., Likowski, K. U., Weyers, P., & Strack, F. (in press). The face of fluency: Semantic coherence automatically elicits a specific pattern of facial muscle reactions. *Cognition & Emotion*.
- Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, 17, 339–345.
- Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 219–230.
- Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1235–1253.
- 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., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings of 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 & Social Psychology*, 81, 989–1000.
- Winkielman, P., & Schwarz, N. (2001). How pleasant was your childhood? Beliefs about memory shape inferences from experienced difficulty of recall. *Psychological Science*, 11, 176–179.
- Wittgenstein, L. (1955/1977). *Philosophische Untersuchungen* [Philosophical Investigations]. Frankfurt, Germany: Suhrkamp.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9, 1–27.
- Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 934–945.