Metacognition Learning (2008) 3:159–174
DOI 10.1007/s11409-008-9024-2

Correcting experience-based judgments: the perseverance of subjective experience in the face of the correction of judgment

Ravit Nussinson · Asher Koriat

Received: 17 September 2007 / Accepted: 19 May 2008 /
Published online: 18 June 2008

Abstract Many of our cognitive and metacognitive judgments are based on sheer subjective experience. Subjective experience, however, may be contaminated by irrelevant factors, resulting in biased judgments. Under certain conditions people exert a metacognitive correction process to remedy such biased judgments. In this study we examine the proposition that even after a judgment has been corrected to avoid the biasing effects on subjective experience, subjective experience itself remains biased. We asked participants to judge the difficulty of anagrams for others. When they were aware of having been exposed to the solutions of some of the anagrams, they corrected their difficulty judgments for these anagrams. Despite this correction, their speeded choices in a subsequent task disclosed their biased subjective experience that these anagrams were easier to solve. Implications for the study of metacognition and for the educational domain are discussed.

Keywords Correction processes · Experience-based judgments · Metacognitive judgments · Subjective experience

Introduction

The study was conducted in the Institute of Information Processing and Decision Making, University of Haifa, Haifa, Israel. We gratefully acknowledge support for this research by the German Federal Ministry of Education and Research (BMBF) within the framework of German-Israeli Project Cooperation (DIP). We thank Limor Sheffer for her help in the analyses of the data.

Ravit Nussinson has previously published under the name Ravit Levy-Sadot.

R. Nussinson · A. Koriat
Department of Psychology, University of Haifa, Haifa, Israel

R. Nussinson (✉)
Institute of Information Processing and Decision Making, University of Haifa, Haifa, 31905, Israel
e-mail: rlevy@research.haifa.ac.il

A great deal of evidence has accumulated suggesting that people rely on their subjective experience in making various kinds of cognitive and metacognitive judgments (Bless and Forgas 2000; Kelley 1999; Koriat et al. 2006). However, when they realize that their subjective experience has been contaminated, people often exert a metacognitive correction process so as to correct for the assumed effects of that contamination on judgment (Strack and Hannover 1996; Wegener and Petty 1995). In this study we examine the proposition that even after the judgment has been corrected to avoid biasing influences, subjective experience itself remains unaffected (Kelley and Jacoby 1996; Strack 1992; Wilson and Brekke 1994). We first review the evidence regarding experience-based judgments and their correction, and then outline the rationale underlying this proposition, pointing out some of its implications in educational settings.

Correction processes in experience-based judgments

In making their judgments people may rely on knowledge and information that they retrieve from memory (information-based judgments). Research suggests, however, that people often base their judgments directly on their subjective feelings (experience-based judgments; see Koriat 1998; Koriat and Bjork 2006; Koriat and Levy-Sadot 1999; Koriat 2008; Rhodes and Jacoby 2007; Schwarz and Clore 1996; Slovic et al. 2002). For example, people rely on the feeling of familiarity associated with a proposition when judging whether that proposition is true or false (Begg et al. 1992). They rely on experienced ease of retrieving an answer when expressing their confidence in the correctness of that answer (Kelley and Lindsay 1993) and on their prevalent affective feeling when judging the likeability of an unfamiliar stimulus (Monahan et al. 2000). Because people often base their judgment blindly on their subjective experience, irrelevant factors that contaminate that experience may lead their judgments astray. For example, in the false-fame effect, participants tend to wrongly judge as famous names to which they were exposed in a previous phase of the experiment, presumably because they rely on the heightened feeling of familiarity of those names (Jacoby et al. 1989a). Similarly, inducing participants to furrow their brows when they attempt to retrieve personal episodes that suggest high self-assurance, leads participants to lower their ratings of self-assurance, presumably because of the increased experience of mental effort associated with recollecting such episodes (Stepper and Strack 1993).

However, under certain conditions people try to correct for the effect of the biased subjective experience on their judgment. Such correction is exerted either by shifting the judgment in the direction opposite to that of the presumed bias, or by re-computing the judgment, relying on alternative, "uncontaminated" bases (Strack 1992).

For example, in a study by Jacoby and Whitehouse (1989) participants performed a yes/no recognition test after studying a list of words. Words that had been preceded by their brief presentation (1 s) immediately prior to their appearance in the test phase were *less* likely to be judged as "old". Jacoby and Whitehouse proposed that participants adjusted their judgments for these words to correct for the heightened feeling of familiarity evoked by the prime.

Similarly, in a study by Schwarz et al. (1991) participants who were asked to recall many past episodes demonstrating self-assertiveness subsequently rated themselves as less assertive than those who were asked to recall a few of such episodes, presumably because of the greater difficulty experienced in recalling many episodes. However, when led to believe that the experienced difficulty had been caused by background music, participants relied more heavily on the retrieved content, reporting higher assertiveness ratings in the many-episodes condition than in the few-episodes condition. These results were seen to suggest a correction by re-computation.

A review of the literature on experience-based judgments suggests that a correction process is exerted only if (a) people are *aware* of a factor that biases subjective experience, rendering it an unrepresentative basis for the judgment, (b) they *believe* that that factor makes subjective experience an unrepresentative basis for judgment, (c) they are *motivated* to reach an unbiased judgment and to invest the needed mental effort to do so, and (d) they have available the *mental resources and time* needed for correction (see Kelley and Jacoby 1996; Strack and Hannover 1996; Wegener and Petty 1995; Wilson and Brekke 1994).

Subjective experience following the correction in judgment

In this study we address the question: What happens to the biased subjective experience following a correction process? In agreement with others (see Kelley and Jacoby 1996; Strack 1992; Wilson and Brekke 1994), we propose that the correction process in experience-based judgments is confined to the cognitive or metacognitive judgment but does not extend to the biased subjective experience itself. That is, the biased subjective experience survives the correction operation. In some cases, in fact, the acknowledgment on the part of the participant that his or her subjective experience has been biased need not call for a correction in subjective experience but only in the judgment itself. For example, in the Jacoby and Whitehouse (1989) study described above, although participants corrected their recognition judgments for the primed words, there was no reason for them to correct their valid experience of familiarity invoked by these words. In other cases, even when a correction of subjective experience itself is called for, people may fail to achieve that correction although they succeed in correcting their judgment. For example, when a sentence is heard against a background noise, the noise sounds softer if that sentence had been heard before (Jacoby et al. 1988). Also a word is judged to extend over a longer duration or to be visibly clearer when it was seen before than when it was not (Whittlesea et al. 1990; Witherspoon and Allan 1985). Furthermore, a word feels familiar (and therefore is judged as one that was studied earlier) if its fragment, which serves as cue at test, is designed so as to support its easy retrieval (Lindsay and Kelley 1996). In these cases one may expect that the correction process will also extend to the subjective experience itself rather than being confined only to the judgment, because it is the biased subjective experience which is the very reason for exerting the correction process. In the rest of the introduction we review several ideas and findings that may explain why contaminated subjective experience may persist in the face of the metacognitive correction in judgment.

Why might subjective experience be immune to the correction in experience-based judgments? If people are aware that their subjective experience is biased and attempt to correct their experience-based judgment, why would their subjective experience remain unchanged? The answer seems to lie in the manner in which experience-based judgments are constructed. Ideas and findings from several researchers (Barsalou et al. 2003; Jacoby et al. 1989b; Strack 1992; Whittlesea 1993, 1997, 2002; Whittlesea and Williams 1998, 2000, 2001) suggest that experience-based judgments are the product of a sequence of two inferential processes. In the first inferential process, rudimentary cues concerning the target stimulus are evaluated automatically and unconsciously to yield a global feeling. These cues reside in the immediate, undifferentiated cognitive and motoric response to the stimulus, such as its encoding and retrieval fluency (Benjamin and Bjork 1996; Benjamin et al. 1998; Jacoby and Dallas 1981;

Koriat and Ma'ayan 2005), and the immediate bodily and proprioceptive responses accompanying task performance (Strack and Deutsch 2004). These cues are translated, through an attribution process, into a cognitive feeling (e.g., familiarity), a bodily feeling (e.g., experienced physical intensity) or an affective feeling (e.g., feeling of pride) (Jacoby et al. 1988; Reber et al. 2004a, b; Stepper and Strack 1993; Whittlesea and Williams 1998). Thus, in the false fame effect (Jacoby et al. 1989a, b), for example, the advance exposure to a name increases its fluent processing, and when fluency is attributed to some previous encounter with the name, the result is a feeling of familiarity. In the self-assurance example (Stepper and Strack 1993), mentioned earlier, proprioceptive feedback from the contraction of the corrugator muscle is attributed to the difficulty of the task, and is translated into experienced mental effort. It has been proposed (see Whittlesea et al. 1990) that each of the elementary cues can in principle be attributed to a certain property of the stimulus, of the situation, or of the operation of the cognitive system, and consequently can result in different states of subjective experience. Previous research also suggests that the initial inferential process is automatic, non-analytic, largely unconscious and fast. Therefore its product—sheer subjective experience—has the quality of direct perception (Jacoby et al. 1988; Whittlesea et al. 1990).

In the second inferential process subjective experience is used as a basis for judgment: The familiarity of a name is attributed to the person being famous. Similarly, the mental effort experienced when retrieving self-assurance episodes is attributed to the scarcity of these episodes, lowering one's judgments of self-assurance. Like the first inference, the second inference too is automatic: Subjective experience serves as the default basis for judgment (Bless and Forgas 2000; Strack 1992) and is automatically attributed to the object and to the dimension of the object that are at the focus of attention. Critically, however, whereas subjective experience, which is the basis of the second inference, is inherently conscious, the cues that feed into the first inference are often unconscious. Further, whereas the second inference lends itself to retrospective scrutiny, the first inference does not. Thus, whereas the first inference seems to satisfy all four criteria of automaticity—unintentionality, unawareness, uncontrollability and high efficiency (Bargh 1994; Logan and Cowan 1984)—the second inference is potentially more available to consciousness and more amenable to control. Therefore, when the preconditions for effecting a correction exist, as indicated earlier, the judgment may be corrected, but the biased subjective experience remains encapsulated and immune to such correction (Fodor 1983).

As noted earlier, previous discussions of correction processes in experience-based judgments have alluded to the possibility that subjective experience may be impervious to the effects of correction, (Kelley and Jacoby 1996; Strack 1992; Wilson and Brekke 1994). However no experimental evidence for that possibility has been produced. In this study we examine the proposition that subjective experience is immune to correctional processes to the extent that the biased subjective experience can exert its influence on new judgments when these are made under conditions that minimize the options for a corrective process to occur.

Experiment

In order to experimentally demonstrate the immutability of subjective experience, we asked participants to perform a judgment that is likely to be based on experience. Participants were asked to judge the difficulty of anagrams for others. Their experienced difficulty with some of the anagrams was biased by priming the solution words of these anagrams in a previous phase of the experiment. One group of participants was informed about the bias and was encouraged to correct for it, whereas a second group was not. Assuming that the biased subjective experience persists then we may expect it to influence judgments under conditions that permit little awareness and/or little control. In this study we measured the implicit, automatic effects of biased subjective experience by having participants perform a comparative judgment of the difficulty of two anagrams under time pressure. Our assumption was that under such conditions participants do not have sufficient time to engage in the correction process (see Finucane et al. 2000; Jones and Jacoby 2001). Furthermore, the binary, gross response mode characteristic of comparative judgments was assumed to encourage reliance on the participants' immediate feelings towards the stimuli. We hypothesized that (1) participants in the informed condition would correct their initial judgment for the anagrams for which their subjective experience was biased, and that (2) nevertheless the effects of their biased subjective experience on the speeded comparative judgment task would be similar to those of participants in the uninformed condition who had not corrected their judgment.

Method

Participants Forty-eight University of Haifa undergraduates whose native tongue was English (mean age = 20.45 years; 18 men) were paid for participating in the experiment. All were American students enrolled in an overseas program in the University of Haifa.

Design The experiment conformed to a condition (uninformed vs. informed + recognition) X type of target anagram (primed vs. unprimed) mixed design, with condition manipulated between participants and type of target anagram manipulated within participants.

Stimulus materials A pool of 238 five-letter English anagrams was created. Each anagram was produced by relocating one of the 5 letters of a word of moderate frequency in the language (from 10 to 45 occurrences per million; as indexed by Thorndike and Lorge 1944). The relocation of the letter was never to a location adjacent to its original one. Based on a norming study in which participants were asked to solve each anagram and to rate its difficulty for other students, 176 anagrams of moderate difficulty were selected. These anagrams were paired to create 88 pairs of anagrams that were of matched difficulty in terms of both proportion of solutions and solution time. When more than two anagrams fulfilled this criterion anagrams were matched so that the judged difficulty of the two anagrams was also similar. The pairs of anagrams were divided into two sets of 44 pairs each (set 1 and set 2). The sets were of very similar difficulty. These sets of pairs of anagrams were used in Phase 2 of the experiment.

In each pair one of the anagrams was identified as anagram A and the other as anagram B. Four lists of 44 five-letter English words were produced. One list consisted of the 44 solution words of anagrams A of set 1, another consisted of the solution words of anagrams B of set 1, another consisted of the solution words of anagrams A of set 2 and yet another consisted of the solution words of anagrams B of set 2. These lists were used in Phase 1 of the experiment.

A post-experiment questionnaire was also composed in which participants were asked to indicate how they had made their ratings of the difficulty of the anagrams for others. The questionnaire included three propositions, each describing a possible basis for the difficulty judgments made (the extent to which the letters of the solution word were scrambled; the frequency of the solution word in the language; the time it took the participant to solve the anagram). For each proposition participants were asked to indicate the proportion of trials for which it applied (on a scale ranging from 1—*seldom* to 4 —*in all cases*; unmarked sentences were coded as 0).

Procedure The experimental procedure used was an extension of the methodology of Kelley and Jacoby's study (1996). In Phase 1, participants read a series of 44 five-letter words, each presented for 2 s (see Fig. 1). In Phase 2, in each of 88 trials participants were presented with two anagrams one after the other: They were asked to solve the first anagram ("target anagram") and say aloud its solution. If correct, the experimenter pressed a key to record solution time. If incorrect, the participant was asked to keep on trying. After solving the anagram or after 20 s have elapsed (in which case the experimenter said the correct answer), participants rated how difficult the anagram was going to be *for other students* to solve. They did so by moving a cursor along a line with the label *very easy* at one end and the label *very hard* at the other end (the markings on the line were later recoded into values from 1 = very easy to 255 = very hard). Next, all participants were presented with a second anagram ("comparison anagram") that matched the target anagram in difficulty according to the results of a previous norming study. Participants had to solve this anagram within 20 s. Then, both the target and the comparison anagrams appeared on the screen and participants had to quickly decide which of them was going to be harder for other students to solve. If the participant did not respond within 3 s, a tone was sounded signaling that he/she should try to respond faster.

The words read by participants in Phase 1 were the solutions of half of the target anagrams (the first in a pair of anagrams presented on Phase 2). Thus for half the pairs of anagrams presented on Phase 2 the solution of the target anagram was presented on Phase 1, and was intended to increase the fluency with which the solutions of these target anagrams came to mind in Phase 2. Whereas participants in the uninformed (U) condition were not told about the connection between the words seen in the two phases, participants in the Informed + recognition (I + R) condition were explicitly instructed to avoid the biasing effect of the prior presentation on their judgments: The instructions for these participants explained that some of the solution words for the first anagram in each pair had

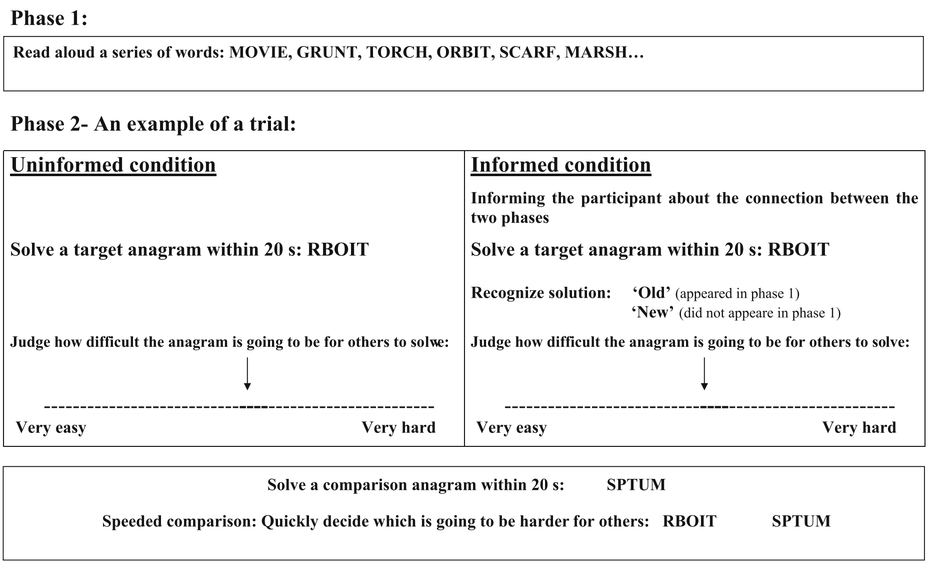


Fig. 1 Outline of the procedure used in the experiment

appeared in the list of words that they had read in Phase 1, and that this would render the solution of these anagrams easier. Because their judgments concerned the difficulty of the anagrams for others who have not seen the solutions before, they should try to avoid the biasing effect of the prior reading on their absolute and relative judgments. After solving the target anagram and before judging its difficulty on the rating line, I + R participants were also asked to decide whether the solution word of that anagram had appeared in the first phase of the experiment.

As just noted, for half the pairs of anagrams the target anagram was primed whereas for the other half it was not. Comparison anagrams were never primed. Assignment of a pair to the primed vs. unprimed conditions was counterbalanced across participants. Furthermore, for half the participants a certain anagram in the pair served as the target anagram whereas for the other participants its matched anagram served as the target. These balances were crossed, yielding the four possible lists of words for Phase 1 (as described in *Stimulus materials*), and they were fully crossed with the two experimental conditions. At the end of the experiment, participants completed a questionnaire about how they had made their ratings of the difficulty of the anagrams for others.

Participants in both conditions were expected to experience the primed target anagrams as easier to solve. Following Kelley and Jacoby (1996), our first hypothesis was that U participants would rate primed targets as less difficult for others than unprimed targets, whereas I + R participants would tend to correct their judgments for primed targets, giving primed and unprimed targets similar ratings.

Second, assuming that I + R participants do yield evidence for a correction process, we expected that under time-pressure they should nevertheless choose its matched comparison anagram as the one that will be harder for others to solve. Thus, although the difficulty judgments of the I + R participants were expected to differ from those of the U participants, we expected the two groups to exhibit a similar pattern of choices in the speeded comparison task. These results would accord with the hypothesis that the biased subjective experience of the I + R group remains unchanged in the face of the correction.

Results

A significance level of .05 was adopted. Observations with response time shorter than 250 ms (0.6%), assumed to be anticipations, were omitted from the analysis. In order to ascertain that choices in the speeded comparison task were indeed fast, only observations with comparison responses within 2000 ms (84.9%) were included in the analysis.[1]

The priming manipulation was effective: Participants were faster solving primed (MED= 2,476 ms) than unprimed target anagrams (MED=4,288 ms), $F(1, 46)=72.70$, MSE=11, $p<0.001$, $\eta^2=0.61$, and the proportion of correct solutions was higher for the primed (0.92) than for the unprimed targets (0.85), $F(1, 46)=42.52$, MSE=.003, $p<0.001$, $\eta^2=0.48$.

Ratings of difficulty for others: evidence for a correction process We should note first that I + R participants were moderately accurate in the recognition memory test in which they judged whether each anagram had been presented in Phase 1 (71% Hits, 22% False Alarms). This performance was comparable to that reported by Kelley and Jacoby (1996).

[1] Practically the same pattern of results was found when all responses (i.e., those in the range 250–3000 ms) were included in the analysis.

An initial analysis on the ratings of difficulty for others was conducted on all responses, including items that I + R participants failed to identify as “old” in the memory test. A Condition (U vs. I + R) X Type of Target Anagram (primed vs. unprimed) analysis of variance (ANOVA) yielded a non-significant Condition X Anagram Type[2] interaction, $F(1, 46)=2.03$, MSE=83.54, $p=0.16$, $\eta^2=0.04$. I + R participants assigned higher difficulty ratings to target anagrams ($M=116.80$) than U participants ($M=103.31$), $F(1, 46)=6.31$, MSE=639.22, $p<0.05$, $\eta^2=0.13$. More important, both U participants and I + R participants assigned primed anagrams lower difficulty ratings ($M=102.89$) than unprimed anagrams ($M=117.22$), $F(1, 46)=58.97$, MSE=83.54, $p<0.001$, $\eta^2=0.56$. This pattern suggests that participants in both conditions relied on their subjective experience, judging the primed anagrams as easier for others to solve.

The results, however, were different when the memory responses were taken into account. As noted in the Introduction, one of the pre-conditions for the exertion of correction is awareness of the contaminating effects on subjective experience. Therefore, I + R participants may be expected to correct their judgments only for those anagrams that they had classified as “old”. Thus, the ANOVA just reported was repeated, using as the within-participant factor for I + R participants the subjective status of the item as “new” or “old” rather than its objective status.

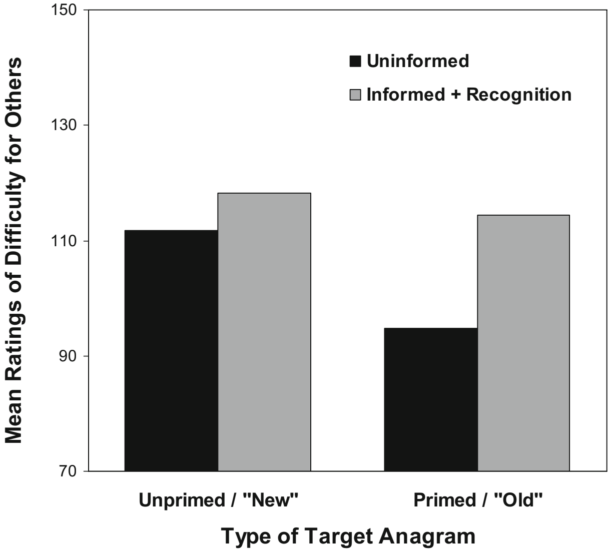
The results of the Condition (U vs. I + R) X Anagram Type (primed vs. unprimed for the U condition, classified as “old” vs. classified as “new” for the I + R condition) on difficulty ratings for others yielded significant effects for both condition and anagram type as before, $F(1, 46)=5.83$, $MSE=705.60$, $p<0.05$, $\eta^2=0.11$, $F(1, 46)=30.44$, MSE=86.05, $p<0.001$, $\eta^2=0.40$, respectively. This time, however, the Condition X Anagram Type interaction was significant, $F(1, 46)=11.93$, MSE=86.05, $p<0.005$, $\eta^2=0.21$ (see Fig. 2). The effect of anagram type was significant for the U condition, $F(1, 23)=73.54$, $p<0.001$, MSE=47.09, $\eta^2=0.76$, with ratings of difficulty averaging 94.81 for primed anagrams, and 111.80 for unprimed anagrams. The I + R condition, in contrast, yielded a nonsignificant difference, $F(1, 23)=1.47$, MSE=125.01, $p=0.24$, $\eta^2=0.06$, with ratings of difficulty averaging 114.44 and 118.35, respectively, for anagrams classified as “old” and as “new”. These results suggest that I + R participants corrected their ratings for those anagrams that they classified as “old”.

Additional evidence suggesting that I + R participants exerted a correction process comes from the correlation between solution times and difficulty ratings across the target anagrams (see also Kelley and Jacoby 1996). If I + R participants corrected their difficulty ratings, relying less heavily than U participants on their subjective experience with the anagrams, then they should yield a lower correlation between solution time and judged anagram difficulty. This was indeed the case (Table 1). The within participant Pearson correlations were Fisher z-transformed to approximate them to the normal distribution and were then subjected to a Condition (U vs. I + R) X Anagram Type (primed vs. unprimed for U condition, classified as “old” vs. classified as “new” for I + R condition) mixed ANOVA. This analysis yielded $F(1, 46)=5.35$, MSE=0.15, $p<0.05$, $\eta^2=0.10$, for condition, $F(1, 46)=14.60$, MSE=0.04, $p<0.001$, $\eta^2=0.24$, for anagram type, and $F(1, 46)=0.47$, MSE=.04, $p=0.50$, $\eta^2=0.01$, for the interaction.

In sum, our first hypothesis was supported: Uninformed participants rated primed targets as less difficult for others than unprimed targets, reflecting the biasing effects of priming on subjective experience. Informed participants, however, corrected their judgments for those anagrams that they classified as “old”: Their ratings of the difficulty of these anagrams were similar to those they gave to the anagrams that they classified as “new”.

[2] Hereafter we would use the term ‘Anagram Type’ as shorthand for the term ‘Type of Target Anagram’.

Fig. 2 Mean difficulty for others ratings. For the uninformed (U) condition results are presented separately for primed and unprimed target anagrams. For the informed + recognition (I + R) condition results are presented separately for target anagrams classified as “old” and for target anagrams classified as “new”. Note: The titles *Unprimed* and *Primed* refer to type of target anagram for the Uninformed condition. The titles *New* and *Old* refer to target anagrams that were classified as “old” and “new” by participants in the Informed + Recognition condition



The speeded comparison task We now turn to the results for the speeded classification task. In that task, the participants were presented with the target anagram and its comparison anagram and were asked to judge which of them would be harder for other students to solve. If I + R participants continue to experience target anagrams classified as “old” as relatively easy to solve despite their corrected judgments, they should tend *not* to choose them as harder for others, and their speeded judgments should not differ from those of the U participants who are also expected to yield the same effect.

Figure 3 presents the proportion of choosing the target anagram as harder as a function of condition and type of target anagram. It can be seen that indeed I + R and U participants exhibited very similar response patterns on the speeded comparison task, both exhibiting an effect of type of anagram. A two-way ANOVA, Condition X Anagram Type (primed vs. unprimed for the U condition, classified as “old” vs. classified as “new” for the I + R condition) on the proportion of choosing the target anagram as harder yielded $F(1, 46)=14.14$, MSE=0.008, $p<0.001$, $\eta^2=0.24$, for anagram type. Neither the effect of condition nor the interaction were significant, $F(1, 46)=0.16$, $MSE=0.004$, $p=0.69$, $\eta^2=0.004$, and $F(1, 46)=0.89$, MSE=0.008, $p=0.35$, $\eta^2=0.02$, for both, respectively. The target anagram was selected as the harder anagram for others to solve less when it was primed or classified as “old” (0.40) than when it was unprimed or classified as “new” (0.47).

Table 1 Mean Pearson correlation between time to solve target anagrams and their difficulty

Condition	Type of target anagram				
	Unprimed	Primed	Total	Classified as “New”	Classified as “Old”
U	0.70	0.64	0.67		
I + R			0.57	0.62	0.52
Total			0.62	0.66	0.58

The values are based on raw Pearson correlations and not on the *z*-transformed ones

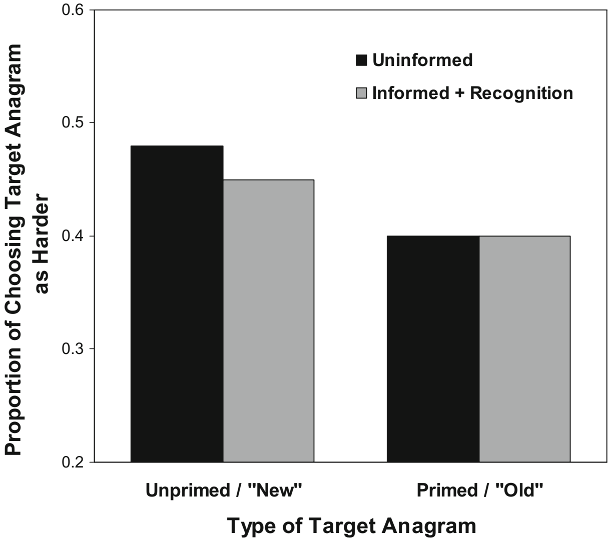


Fig. 3 Mean proportion of choosing target anagram as harder for others. For the uninformed (U) condition results are presented separately for primed and unprimed target anagrams. For the informed + recognition (I + R) condition results are presented separately for target anagrams classified as "old" and for target anagrams classified as "new". Note: The titles *Unprimed* and *Primed* refer to type of target anagram for the Uninformed condition. The titles *New* and *Old* refer to target anagrams that were classified as "old" and "new" by participants in the Informed + Recognition condition

These results suggest that the correction process operates only on the judgment, leaving subjective experience unchanged. In fact, when the target anagram was unprimed (for U participants) or classified as "new" (for I + R participants) the tendency to choose it as the harder one for others was at chance level, reflecting the matched difficulty level of target anagrams and their corresponding comparison anagrams. When the target anagram was primed (for U participants) or classified as "old" (for I + R participants), participants in both conditions were equally likely to choose its matched comparison as being harder for others to solve ($F<1$), disclosing their similar perception of the target anagram as easier.

Note that the previous conclusion is based on the assumption that the comparison anagrams that were matched with the primed target anagrams and those that were matched with the target anagrams that were classified as "old" were of similar difficulty. Indeed, an analysis conducted on the median solution time for the comparison anagrams confirmed that such was the case ($F(1, 46)=0.80$, MSE=1,761,817.1, $p=0.38$, $\eta^2=0.02$, MED= 3,873 ms for primed targets, MED=4,115 ms for targets classified as "old").

Further analyses: subjective reports regarding the bases for difficulty judgments How did I + R participants correct their difficulty ratings? Did they re-compute their judgments on the basis of alternative bases other than their experienced difficulty or did they adjust their ratings to oppose the bias in subjective experience (Strack 1992)? Some insight into this question is provided by the results of the post-experiment questionnaire in which participants marked the propositions that describe the basis of their judgments, and indicated the proportion of trials for which that proposition applied (on a scale ranging from 1—*seldom* to 4—*in all cases*; unmarked sentences were coded as 0). Table 2 presents the mean reported frequency of using each basis.

If I + R participants switched to an alternative basis in making their corrected judgments, then they should report greater reliance on such factors as (a) the frequency of the solution word in the language and/or (b) the extent to which the letters of the solution word were scrambled (see Kelley and Jacoby 1996) compared with U participants. The results, however, do not support this possibility. A condition (U vs. I + R) X Basis (solution time vs. degree of

Table 2 Means and SD (in parentheses) of rated extent of reliance on 3 bases for judgment as a function of condition

Condition	Basis for Judgment		
	Solution time[a]	Degree of scrambling[b]	Word frequency[c]
U	2.96 (0.36)	2.33 (1.30)	2.04 (1.40)
I + R	2.46 (0.88)	2.50 (1.41)	1.79 (1.41)
Total	2.71 (0.71)	2.42 (1.32)	1.92 (1.40)

Scales ranged from 1—*seldom* to 4—*in all cases*

[a] "I based my difficulty ratings on the time it took me to solve the anagram: The longer it took me to solve an anagram - the higher the difficulty rating"

[b] "I based my difficulty ratings on the extent to which the letters of the solution word were scrambled: the more scrambled the letters – the higher the difficulty rating"

[c] "I based my difficulty ratings on the frequency of the word in the language: The less frequent the word is – the higher the difficulty rating"

scrambling vs. word frequency) mixed ANOVA yielded $F(1, 46)=1.12$, MSE=1.22, *ns*, $\eta^2=0.02$, for condition, $F(1, 46)=5.16$, MSE=1.49, $p<0.01$, $\eta^2=0.10$, for basis, and $F(2, 92)=0.91$, MSE=1.49, $p=0.41$, $\eta^2=0.01$, for the interaction. In general, participants reported relying primarily on solution time. However, compared with U participants, I + R participants reported relying less on solution times, $F(1, 46)=6.60$, MSE=0.45, $p<0.05$, $\eta^2=0.13$, but did not differ from U participants in the extent of reliance on degree of scrambling, $F(1, 46)=0.19$, MSE=1.77, $p=0.67$, $\eta^2=0.004$, or on word frequency, $F(1, 46)=0.38$, MSE=1.98, $p=0.54$, $\eta^2=0.008$. This pattern is consistent with the assumption that I + R participants corrected their ratings by adjusting them rather than by re-computing them.

Discussion

This study provides the first empirical support for the proposition that even after a correction in judgment has been exerted to avoid the influence of a contaminated subjective experience on judgment, subjective experience itself remains unaffected (Kelley and Jacoby 1996; Strack 1992; Wilson and Brekke 1994). The experiment was designed to show first that participants who are informed about the contaminating effect of priming on judged difficulty of anagrams, will correct their judgments, assigning similar ratings to primed and unprimed anagrams. Second, however, under conditions that minimize the opportunities for corrections, their judgments would still exhibit the biasing effects of priming on judged anagram difficulty, even following correction.

Let us first examine the results that pertain to the correction process. When participants realized that the prior reading of the solution words may have rendered their subjective experience unrepresentative for judgment (i.e., when they were informed) and when they were also aware of the prior appearance of the particular solution (i.e., when they classified the anagram as "old") they exerted a correction of the experience-based judgments. Importantly, this pattern of results was only observed when the results were analyzed using participants' subjective classification of the target anagrams as "old" or "new"; not when they were analyzed using the objective status of the target anagrams (primed vs. unprimed). This result underscores the idea that awareness of the biasing factor (Jacoby and Whitehouse 1989; Kelley and Jacoby 1996; Schwarz and Clore 1983; Strack and Hannover

1996) and an intuitive theory that that factor makes subjective experience undiagnostic (Kelley and Jacoby 1996; Schwarz et al. 1991; Schwarz and Clore 1983; Wänke et al. 1995) are both necessary for the correction process to occur. Note, however, that in the Kelley and Jacoby (1996) study the corrective pattern was obtained even when the analysis was performed on the basis of the objective status of the target anagrams. Because recognition performance (i.e., awareness of the biasing factor) was similar across the two studies, the reason for this discrepancy may seem unclear. Still, it might be that the more demanding nature of the I + R task in our study (i.e., the addition of a speeded classification task on every trial) limited the resources available to participants for exerting a correction process. This may have resulted in a smaller adjustment than in the Kelley and Jacoby study of the difficulty ratings of primed anagrams that were correctly recognized as "old". When averaged with the ratings of primed anagrams that were mistakenly recognized as "new" this would have resulted in a low difficulty ratings for primed anagrams, and with seemingly no signature of correction.

Finally, our I + R participants reported less reliance on solution time compared with U participants but no stronger tendency to rely on the two alternative bases. This suggests that they corrected their judgments by adjustment and not by re-computation.

Let us now turn to the results pertaining to the proposition that biased subjective experience survives the correction of judgment. One might have expected that after correcting the difficulty judgments for the effect of priming subjective experience itself would change: The relative ease with which the anagram's solution came to mind (which was misattributed to the difficulty of the anagram) would now be properly attributed to a prior encounter with the solution in Phase 1 and would instead give rise to an experience of familiarity. The results, however, suggest otherwise. They suggest that the correction process affects only the judgment without modifying subjective experience itself: When I + R participants classified target anagrams as "old" they tended to choose their comparison anagrams as the ones that would be harder for others to solve. This implies that they continued to experience "old" anagrams as easy despite having corrected their judged difficulty.

In this study we provided support for a proposition already alluded to by previous discussions of correction processes in experience-based judgments—namely—that subjective experience may be impervious to the effects of correction (Kelley and Jacoby 1996; Strack 1992; Wilson and Brekke 1994). Although some studies did show that awareness of the source of bias is sometimes insufficient to correct illusory subjective experiences (much like what happens in perceptual illusions, see Jacoby et al. 1992; Whittlesea et al. 1990), evidence that subjective experience is immune to the correctional processes was missing. Here we support this proposition and show that the biased subjective experience can exert its influence on new judgments when these are made under conditions that minimize the options for a corrective process to occur.

Implications for the study of metacognition

The results reported in this study have certain meta-theoretical implications for metacognitive processes in general. According to the crossover model (Koriat 2000, 2007; Koriat and Levy-Sadot 1999), subjective experience plays a unique function in bridging between two distinct modes of operation of the cognitive system: an implicit-automatic mode and an explicit-controlled mode. Sheer subjective feelings, such as the feeling of knowing or subjective confidence, are shaped by implicit, unconscious processes (see Jacoby et al. 1989a, b; Whittlesea et al. 1990). Once formed, however, these feelings become part and parcel of explicit-controlled processes, and can serve to guide action

(Schwarz and Clore 1996). The effects of priming on judgments of difficulty (in the U group) illustrate how implicit influences (increased processing fluency) translate into conscious-controlled responses (judgments of difficulty for others) through the mediation of subjective experience (experienced ease of solution). The correction processes (used by the I + R group) reflect largely the explicit-controlled mode of operation, as suggested by our discussion of the conditions under which such processes occur. Seen from this perspective, while subjective experience may affect explicit judgment and behavior, it is itself relatively opaque to the influence of higher-order conscious, analytic processes.

Are there nevertheless conditions under which subjective experience may be influenced and modified by higher-order processes? According to some commonly held assumptions, the cognitive system tends to attribute the basic information (e.g., fluency) to the factor to which attention is directed (Whittlesea et al. 1990). To the extent that higher-order, controlled processes can affect the focus of attention, then, they might be expected to also affect subjective experience (see Lazarus and Lazarus 1994). This implies that conscious processes can influence the nature of the subjective experience that is formed in response to the basic information, but cannot modify subjective experience after it has been shaped. This speculation deserves investigation.

Our findings have important implications in many communicative and educational settings in which a knowledgeable person attempts to communicate some information to another, less knowledgeable person. In that case what one knows may contaminate one's own subjective experience, leading to the judgment that the information is comprehensible to the recipient of that information. A classic example is that of an instructor whose familiarity with the material that he has to teach makes him believe that the material is easy for the students, and that he has succeeded in communicating that material to the students (see Koriat and Bjork 2005). Our results suggest that although the instructor may be able to correct for his biased subjective experience when he has sufficient time and motivation to do so, his contaminated subjective experience will remain unchanged, and is liable to impair the effectiveness of his teaching when he is not on guard.

Our findings also imply that the correction process results in dissociation in consciousness between "what one knows" (i.e., the corrected judgment) and "what one feels" (i.e., the biased subjective experience) (see Marcel 1993). They suggest that this dissociation is most likely to be reflected in judgment and in behavior when the conditions for the exertion of a correction process are not met (see also Greenwald et al. 1998; Haidt 2001).

Acknowledgements This study was conducted at the Institute of Information Processing and Decision Making, University of Haifa, Haifa, Israel. We gratefully acknowledge support for this research by the German Federal Ministry of Education and Research (BMBF) within the framework of German-Israeli Project Cooperation (DIP). We thank Limor Sheffer for her help in the analyses of the data.

References

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Jr Wyer, & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Barsalou, L. W., Niedenthal, P. M., Barbey, A., & Ruppert, J. (2003). Social embodiment. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43 (pp. 43–92). San Diego, CA: Academic Press.

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. doi:10.1037/0096-3445.121.4.446.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology. General*, *127*, 55–68. doi:10.1037/0096-3445.127.1.55.

Bless, H., & Forgas, J. P. (2000). The message within: Toward a social psychology of subjective experiences. In H. Bless, & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 372–392). Lillington, North Carolina: Edwards Brothers.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*, 1–17. doi:10.1002/(SICI)1099-0771(200001/03)13:1>1::AID-BDM333<3.0.CO;2-S.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. doi:10.1037/0022-3514.74.6.1464.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Bulletin*, *108*, 814–834.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*, 240–247. doi:10.1037/0278-7393.14.2.240.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology. General*, *110*, 306–340. doi:10.1037/0096-3445.110.3.306.

Jacoby, L. L., Kelley, C. M., Brown, J., & Jasechko, J. (1989a). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*, 326–338. doi:10.1037/0022-3514.56.3.326.

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

Jacoby, L. L., Lindsay, D. S., & Toth, J. P. (1992). Unconscious influences revealed: Attention, awareness, and control. *American Psychologist*, *47*, 802–809.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology. General*, *118*, 126–135. doi:10.1037/0096-3445.118.2.126.

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language*, *45*, 82–102. doi:10.1006/jmla.2000.2761.

Kelley, C. M. (1999). Subjective experience as basis of "objective" judgments: Effects of past experience on judgments of difficulty. In D. Gopher, & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 515–536). Cambridge, MA: The MIT Press.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, *35*, 157–175. doi:10.1006/jmla.1996.0009.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*, 1–24. doi:10.1006/jmla.1993.1001.

Koriat, A. (1998). Metamemory: The feeling of knowing and its vagaries. In M. Sabourin, F. Craik, & M. Robert (Eds.), *Advances in psychological science* (Vol. 2 (pp. 461–469). Hove, England: Psychology Press.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, *9*, 149–171. doi:10.1006/ccog.2000.0433.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge, UK: Cambridge University Press.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*, 187–194. doi:10.1037/0278-7393.31.2.187.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*, 1133–1145. doi:10.1037/0278-7393.32.5.1133.

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 483–502). New York, NY: Guilford Press.
Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*, 478–492. doi:10.1016/j.jml.2005.01.001.
Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology. General*, *135*, 36–69. doi:10.1037/0096-3445.135.1.36.
Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky, & R. A. Bjork (Eds.), *A handbook of memory and metamemory* (pp. 117–136). Mahwah, NJ: Lawrence Erlbaum Associates.
Lazarus, R. S., & Lazarus, B. N. (1994). *Passion and reason: Making sense of our emotions*. New York, NY: Oxford University Press.
Lindsay, D. S., & Kelley, C. M. (1996). Creating illusions of familiarity in a cued recall remember/know paradigm. *Journal of Memory and Language*, *35*, 197–211.
Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327. doi:10.1037/0033-295X.91.3.295.
Marcel, A. J. (1993). Slippage in the unity of consciousness. In J. Wiley (Ed.), *Experimental and theoretical studies of consciousness* (pp. 168–186). Chichester, England: John Wiley & Sons.
Monahan, J. L., Murphy, S. T., & Zajonc, R. B. (2000). Subliminal mere exposure: Specific, general and diffuse effects. *Psychological Science*, *11*, 462–466. doi:10.1111/1467-9280.00289.
Reber, R., Schwarz, N., & Winkielman, P. (2004a). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*, 364–382. doi:10.1207/s15327957pspr0804_3.
Reber, R., Zimmermann, T. D., & Wurtz, P. (2004b). Judgments of duration, figure-ground contrast, and size for words and nonwords. *Perception & Psychophysics*, *66*, 1105–1114.
Rhodes, M. G., & Jacoby, L. L. (2007). Toward analyzing cognitive illusions: Past, present, and future. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 379–393). New York, NY: Psychology Press.
Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*, 195–202. doi:10.1037/0022-3514.61.2.195.
Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513–523. doi:10.1037/0022-3514.45.3.513.
Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 433–465). New York, NY: Guilford Press.
Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahnemann (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York, NY: Cambridge University Press.
Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. *Journal of Personality and Social Psychology*, *64*, 211–220. doi:10.1037/0022-3514.64.2.211.
Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. L. Martin, & A. Tesser (Eds.), *The construction of social judgments* (pp. 249–275). Hillsdale, NJ: Erlbaum.
Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247. doi:10.1207/s15327957pspr0803_1.
Strack, F., & Hannover, B. (1996). Awareness of influence as a precondition for implementing correctional goals. In P. M. Gollwitzer, & J. A. Bargh (Eds.), *The Psychology of action: Linking cognition and motivation to behavior* (pp. 579–596). New York, NY: Guilford Press.
Thorndike, E. L., & Lorge, I. (1944). *The teachers' word book of 30,000 words*. New York, NY: Columbia University Press.
Wänke, M., Schwarz, N., & Bless, H. (1995). The availability heuristic revisited: Experienced ease of retrieval in mundane frequency estimates. *Acta Psychologica*, *89*, 83–90. doi:10.1016/0001-6918(93)E0072-A.
Wegener, D., & Petty, R. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, *68*, 36–51. doi:10.1037/0022-3514.68.1.36.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*, 1235–1253. doi:10.1037/0278-7393.19.6.1235.

Whittlesea, B. W. A. (1997). Production, evaluation and preservation of experiences: Constructive processing in remembering and performance tasks. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 37 (pp. 211–264). San Diego, CA: Academic Press.

Whittlesea, B. W. A. (2002). False memory and the discrepancy-attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology. General*, *131*, 96–115. doi:10.1037/0096-3445.131.1.96.

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*, 716–732. doi:10.1016/0749-596X(90)90045-2.

Whittlesea, B. W. A., & Williams, L. S. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica*, *98*, 141–165. doi:10.1016/S0001-6918(97)00040-1.

Whittlesea, B. W. A., & Williams, L. S. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*, 547–565. doi:10.1037/0278-7393.26.3.547.

Whittlesea, B. W. A., & Williams, L. S. (2001). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*, 3–13. doi:10.1037/0278–7393.27.1.3.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*, 117–142. doi:10.1037/0033-2909.116.1.117.

Witherspoon, D., & Allan, L. G. (1985). The effect of a prior presentation on temporal judgments in a perceptual identification task. *Memory & Cognition*, *13*, 101–111.