Available online at www.sciencedirect.com

ScienceDirect

Organizational Behavior and Human Decision Processes 106 (2008) 61–76

ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES

www.elsevier.com/locate/obhdp

Alleviating inflation of conditional predictions ☆

Asher Koriat *

Department of Psychology, University of Haifa, Haifa 31905, Israel

Received 28 June 2006
Available online 26 October 2007
Accepted by Robyn Dawes

Abstract

Previous studies indicated that conditional predictions—the assessed probability that a certain outcome will occur given a certain condition—tend to be markedly inflated. Five experiments tested the effects of manipulations that were expected to alleviate this inflation by inducing participants to engage in analytic processing. Rewarding participants for accurate predictions proved ineffective. A training procedure in which participants assessed the likelihood of each of several outcomes before assessing the probability of a target outcome was partly effective in reducing overestimation. Most effective was the requirement to work in dyads and to come to an agreement about the assessed likelihood. Working in dyads helped alleviate prediction inflation even after participants made their individual predictions alone, and its debiasing effect also transferred to the estimates that were made individually on a new set of stimuli. The results were discussed in terms of the factors that make prediction inflation resistant to change.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Conditional predictions; Judgment; Confirmation bias; Debiasing procedures; Metacognition; Overprediction

In many situations in everyday life people need to asses the possible consequences of various events and developments in order to make decisions about their actions. Such assessment may take the form of an open-ended evaluation in which a variety of possible outcomes are considered. Often, however, the assessment may be targeted at a particular outcome and involves predicting the likelihood of that outcome given a specific envisioned condition. Thus, investors often must assess the probability that the value of a certain share will drop following a potential future event. Such targeted predictions have been referred to as conditional predictions (Koriat, Fiedler, & Bjork, 2006). Conditional predictions specify a particular condition as well as a specific potential outcome, and take the form "What is the probability that a specified event will occur given a specified condition?"

Koriat et al. examined the idea that conditional predictions tend to be markedly inflated because the focus on the target outcome highlights aspects of the condition that are consistent with it. Some of these aspects are less likely to come to the fore when that outcome is not mentioned. The backward activation process, in which the stated outcome changes the representation of the condition, leads to overestimation of the probability of the target outcome. It may also result in an overprediction effect: Each of several alternative outcomes may be judged to be very likely, as if there is little competition between them, so that the assessed probabilities of several alternative outcomes may sum up to more than 1.0.

The procedure that was used to examine these ideas is based on the word association task in which people are

☆ The research reported 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 Rinat Gil and Hila Urian for conducting the experiments and Limor Sheffer for her help in the analyses of the data.

* Fax: +972 4 8249431.
E-mail address: akoriat@research.haifa.ac.il

0749-5978/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2007.08.007

presented with a stimulus word and are asked to respond with the first word that comes to mind. In our experiments, we essentially asked participants to make a conditional prediction: to judge the likelihood that a person will respond with a particular target word (outcome) when presented with a certain stimulus word (condition). We presented participants with a list of word-word, cue-target pairs and asked them to estimate the percentage of people who would produce the target word as their first response given the cue word. This task has the advantage that norms are available for a large number of cue-target pairs, listing the actual percentage of responding, and that associative-semantic properties of the pairs can be manipulated that permit investigation of the backward activation account of prediction inflation.

The results of several experiments disclosed a pervasive and strong overestimation bias. That bias was particularly pronounced for pairs with low (but not zero) actual percentage of responding. For such pairs, participants' estimates averaged about 62% when the actual percentage of responding was only 4.4% (Koriat et al., 2006; Experiment 1). A similar pattern of results was obtained independently by Maki (Maki, 2007a,b) across several experiments. He conducted detailed analyses of the function relating estimated probabilities to actual probabilities and found it to have not only an unduly high intercept but also a very shallow slope, indicating insufficient sensitivity to inter-item differences in actual likelihood of responding. What is notable is that estimated probabilities are very low for pairs with a zero association, suggesting a discontinuity in the function relating estimated probabilities to actual probabilities (Koriat, 1981).

To examine the backward activation account of inflated predictions, Koriat et al. (2006) used asymmetrically-associated pairs. For such pairs the association in the forward direction (e.g., *cheddar*-*cheese*, associative strength: .92) is much stronger than that in the backward direction (*cheese*-*cheddar*, associative strength: .05). The results indicated a moderate overestimation when the pairs were presented in the forward direction but a very marked overestimation when the pairs were presented in the backward direction.

The inflation of conditional predictions has much in common with two other biases that have been discussed in the literature—confirmation bias and hindsight bias. Confirmation bias (see Nickerson, 1998) refers to the tendency to justify a conclusion by selectively focusing on evidence that supports it. Such selective focusing has been assumed to underlie the overconfidence in one's answers to general-information questions (see Fischhoff, Slovic, & Lichtenstein, 1977; Koriat, Lichtenstein, & Fischhoff, 1980). Koriat et al. (2006) argued that prediction inflation also derives from a process in which people build a scenario that leads from the condition to the outcome, focusing on supporting evidence. In fact, Fiedler (2000) claimed that the mere specification of a particular future event increases the perceived likelihood of that event. Such overestimation is particularly strong when participants are instructed to imagine or explain the outcome before judging its likelihood (e.g., Carroll, 1978; Hirt & Markman, 1995). Although in Koriat et al's study neither explanation nor imagination of the target outcome was explicitly solicited, participants behaved as if they had to justify the occurrence of that outcome (see Koehler, 1991).

Conditional predictions have also much in common with the hindsight bias (Fischhoff, 1975; for reviews, see Guilbault, Bryant, Brockway, & Posavac, 2004; Hawkins & Hastie, 1990; Blank, Musch, & Pohl, 2007). When people are asked to predict the outcome of a historical event and are then required to recollect their prediction after the outcome is revealed, their recollection of their original predictions tend to shift towards the correct outcome. The prediction inflation is, in a sense, a mirror image of the hindsight bias: Whereas in the hindsight bias the participant's past predictions are distorted in retrospect once the actual outcome is revealed, in prediction inflation the participant's predictions are distorted in the direction of the stated outcome whose future likelihood has to be assessed.

This study focuses on manipulations that may aid in alleviating prediction inflation. How can the overestimation bias be mended? Previous studies examined one hypothesis that derives from the processes assumed to underlie prediction inflation and the related phenomena just discussed. One such hypothesis is that the mere presentation of the target word along with the cue word prevents consideration of alternative responses to the cue word (see Koriat, 1981; Koriat & Bjork, 2005). However, several manipulations that presumably induce consideration of alternative responses failed to improve judgment accuracy. Maki (2007a, Experiment 6) presented the cue word for 10 s prior to seeing the target word, and instructed participants to think about possible response words. Only then was the target word added for rating. In another experiment (Experiment 7) each cue word was accompanied by four of its normative response words, and one of them was selected to be judged. Neither of these manipulations improved discrimination between the items. Focusing on the slope of the function relating judged to actual probabilities of responding, Maki (2007b) observed that manipulations designed to lower the intercept of the function did not increase the slope. For example, an error-correction feedback training did not increase sensitivity to inter-item differences in associative strength, nor did the instruction to rate several alternative responses under the constraint that the ratings should total exactly 100.

Koriat et al. (2006), who focused specifically on the overestimation bias, examined the hypothesis that having participants generate their own association to the cue word should alleviate prediction inflation. Participants first produced an association to the cue word and only then saw the target word and assessed its likelihood. Participants still overestimated the occurrence of the target response even when it differed from the one that they had just generated (Experiment 4). Two additional variations of the generation task also proved ineffective: Neither the production of two associates to the cue word prior to seeing the target (Experiment 5) nor the production of these associates in the presence of that target (Experiment 6) were effective in reducing the prediction inflation markedly. This was true even when neither of the two associates matched the target response.

These results are surprising. The generation manipulation can be assumed to give participants first-hand experience with the task whose outcome they are subsequently asked to predict, making them aware of the likelihood of responses other than the target response. However, for the backward-associated pairs, participants in the generation condition produced a different word from the target word in 95% of the cases, but their predictions of the target (which was revealed immediately after the generation task) averaged around 50%.

This observation joins with the finding that when participants were presented with a cue-target pair and asked to estimate the percentage of people who ‘‘would not say the second word in response to the first word, but will say another word instead,’’ their estimates were unduly low, suggesting that the occurrence of the presented target word was again overestimated (Koriat et al., Experiment 2). It would seem that the presentation of the target response along with the cue word largely preempts the experience gained from consideration of other potential responses that might be evoked by the cue word. This is because in making conditional predictions, people assess the strength of support for one outcome almost independently of support for competing outcomes (Robinson & Hastie, 1985; Sanbonmatsu, Posavac, & Stasney, 1997; Van Wallendael & Hastie, 1990). Therefore, the fact that one outcome appears quite plausible does not preclude the possibility that another outcome will also feel very plausible.

Nevertheless, the results of Koriat et al. and Maki stand in sharp contrast with findings indicating that inducing participants to consider alternative outcomes can reduce inflated probabilities (see Hirt, Kardes, & Markman, 2004; Hirt & Markman, 1995). They are also inconsistent with the finding that having participants explain alternative outcomes reduces their confidence in the target outcome (Koehler, 1991). Furthermore, previous results suggested that people tend to rely on their own subjective experience when asked to make predictions for others (Kelley & Jacoby, 1996; Nickerson, 1999). Why then is prediction inflation relatively impervious to the experience gained from generating a different response than the target response?

In this study we attempt to address this question in the framework of the dual-process theoretical framework, which posits a distinction between two modes of processing, labeled System 1 and System 2 by Stanovich and West (2000); see Kahneman, 2003). The former is assumed to be intuitive, heuristic, quick and effortless, whereas the latter is analytic, deliberate, slow, and effortful. Dual process theories have been used to explain a variety of phenomena in cognitive, social and developmental psychology (Epstein & Pacini, 1999; Kelley & Jacoby, 1996; Koriat, Bjork, Sheffer, & Bar, 2004; Koriat & Levy-Sadot, 1999; Sloman, 2002; Strack & Deutsch, 2004). It may be proposed that prediction inflation derives largely from the type of intuitive and automatic mode of processing that characterizes system 1. In this mode, participants respond to the overall overlap between the condition and the outcome without engaging in a critical, systematic evaluation of their assessment. For example, the cue-target pair *cheese-cheddar* gives rise to an immediate feel that *cheddar* is a very plausible response to *cheese* (even though the actual probability is only .05). In order to combat that intuitive feeling, it is necessary to engage in a deliberative, calculated assessment in which the many other potential responses to *cheese* are considered, and then rely on the output of that reasoning to overcome the immediate intuitive feelings (see Koriat et al., 2004). Indeed, previous results suggest that that participants adopt an intuitive, nonanalytic mode of responding as a default but tend to shift to an analytic, deliberate mode when they realize that their immediate ‘‘gut feelings’’ had been contaminated by irrelevant factors or when they experience cognitive disfluency (e.g., Alter, Oppenheimer, Epley, & Eyre, in press; Gilbert, 2002; Jacoby & Whitehouse, 1989; Strack, 1992).

Why then did the generation of one’s own associations fail to mend prediction inflation even when these associations differed from the target (Koriat et al., 2006)? Possibly, the mere feedback that participants gain from generating different responses from the target is not sufficient to overcome the immediate feeling that the target outcome (e.g., *cheese*) is quite likely. What is needed is a manipulation that not only leads participants to engage in an analytic mode of processing but also induces them to rely on that mode in making their predictions. Indeed, the requirement to provide reasons for the occurrence of alternative outcomes has been found to be effective in reducing the judged probability of future events. (Hirt & Markman, 1995). Also, results suggest that predictions that derive from System 1 reasoning are likely to be corrected when participants are accountable for their decisions (Tetlock & Lerner, 1999).

Thus, we explored the effects of three types of manipulation that were intended to induce participants to engage in a deliberative process that might help them overcome the immediate feelings precipitated by the presence of the target. In Experiment 1, participants were given special instructions to try to provide as accurate predictions as they can and were also given monetary incentives for predictions that were very close to the actual percentages. In Experiment 2, participants were given a training session in which they were presented with a series of cue words. For each cue word they were instructed to list six responses that come to mind and to estimate the percentage of people who would give that word as a response in a word-association test. Only then were they asked to assess the likelihood that the target word will be given as a response to that cue word. In Experiment 3, participants worked in dyads and had to come to an agreement about their estimate. The dyadic negotiation was expected to induce a deliberative and critical mode of processing because each member must defend his or her initial judgment and try to convince the other member. Experiments 4 and 5 were designed to obtain additional insights about the effects of working in dyads: Whereas in Experiment 4 participants first provided their independent estimates individually before negotiating on a joint estimate, in Experiment 5 the order of the conditions was reversed in order to examine whether the effects of working in dyads transferred to a situation in which each participant provided his/her own estimate. All of the manipulations used in this study were intended to bring participants, for better or for worse, to reason!

Experiment 1: Inducing motivation for accuracy

The procedure of Experiment 1 was similar to that used by Koriat et al. (2006). However, Participants were instructed to make a special effort to provide accurate estimates, and were told that they would receive a monetary bonus for each prediction that fell within 5 percentage points from the actual norms. Previous research (see Camerer & Hogarth, 1999, for a review) indicated that monetary incentives as such do not necessarily improve judgments. Here, however, the incentives were primarily intended to induce participants to try to provide as precise estimates as they could, under the assumption that the focus on precision should call for analytic reasoning.

Method

Participants

Twenty-four University of Haifa students (17 women, 7 men) participated in the experiment, four for course credit and the rest for payment.

Materials

The materials were the same as those used in Koriat et al. (2006, Experiment 4). A list of 90 Hebrew word pairs was used, consisting of 30 unrelated word pairs (for which associative strength was zero), and 60 asymmetrically associated pairs for which the forward and backward associative strengths averaged .60 and .02, respectively. Half of the asymmetrical pairs were presented in the forward direction and the other half in the backward direction, with the assignment to the two directions counterbalanced across participants.

Procedure and apparatus

The word pairs were displayed on a computer screen. Participants were informed about the procedure of a word association test, and were asked to estimate for each presented pair, the percentage of people who would say the second word (on the left) as the first response to the stimulus word (on the right), in a word association test[1]. Participants were instructed to say their estimate aloud. The experimenter entered their estimate on a keyboard, and 1 s thereafter the next pair was presented.

Participants were instructed to make a special effort to provide an estimate that is as close as possible to the truth. They were told that their estimates will be compared to the norms and that they will receive NIS 1 (about 25 cents) for each prediction that deviates from the norms by no more than 5 percentage points in either direction. When the experiment was completed, the participants were interviewed about the strategy that they had used in making predictions.

Results

We first examine the proportion of items for which the participants' estimates met the criterion of falling within ± 5% from the norm. Although this proportion averaged .25 across all items, it was high only for the unrelated pairs (.68) whereas for the forward and backward pairs it was very low, .05 (range 0–.17) and .03 (range 0–.27), respectively.

We compared these results with those of the control condition in Experiment 4 of Koriat et al. (2006). The procedure for that condition was the same as that of the present experiment except that no special instructions emphasizing accuracy were used and no incentive was included. The proportion of items that fell within ±5% from the norms in that experiment averaged .28 across all items. This proportion amounted to .73, .10, and .02, for the unrelated, forward, and backward pairs, respectively. A Condition

[1] Hebrew is written from right to left.

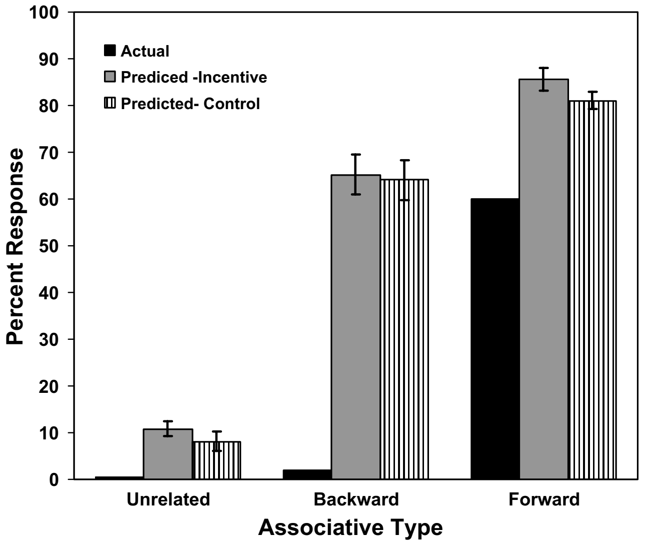

Fig. 1. Mean actual and predicted response percentages for the Incentive and Control conditions for the unrelated, backward and forward pairs. Error bars represent +1 *SEM* (Experiment 1).

(Incentive vs. No-incentive) X Associative Type (forward, backward, unrelated) ANOVA on these means yielded $F(1,42) = 1.09$, ns, for condition, $F(2,84) = 297.99$, $MSE = 0.021$, $p < .0001$, for associative type, and $F < 1$ for the interaction. Thus, there was no sign that the accuracy incentive was successful in increasing the proportion of estimates that fell within the specified range.

We examined next the estimates provided. Fig. 1 presents these estimates (predicted-incentive) as well as those made in Experiment 4 of Koriat et al. (2006) (predicted-control). Included in this Figure are also the actual percentages. As can be seen, the incentive manipulation failed to eliminate the prediction inflation bias. Mean predictions for the forward, backward and unrelated pairs, were all inflated, averaging 85.6%, 65.2% and 10.8%, respectively, compared with 59.9%, 2.0% and 0%, respectively, for the actual percentages. The inflation was significant for each of the associative types, $t(23) = 10.02$, $p < .0001$, $t(23) = 15.05$, $p < .0001$, $t(23) = 6.59$, $p < .0001$,[2] respectively, for the predicted-actual difference.

Consistent with previous results (Koriat et al., 2006), the inflation bias was much stronger for the backward than for the forward pairs. For the backward pairs, the estimates were inflated by a factor of 32. Because no error variance is available for the actual percentages, we performed the analyses on the estimated-actual differences calculated for each participant. A comparison of the difference scores for forward and backward pairs yielded $t(23) = 10.90$, $p < .0001$, indicating a stronger bias for the backward pairs.

The inordinately high estimates observed for the backward pairs could be seen to derive from a simple statistical regression in which small frequencies tend to be overestimated (Fiedler & Armbruster, 1994). However, the results for the unrelated pairs argue against this interpretation: The predictions for these pairs averaged 10.8% (when the actual percentage was zero), much lower than what was found for the backward pairs (65.2%).

Turning next to a comparison between the incentive condition and the control condition, a Condition X Associative Type ANOVA yielded a significant effect for associative type, $F(2,84) = 605.63$, $MSE = 106.31$, $p < .0001$, but $F < 1$ for both condition and the interaction. Mean predictions across all items was 53.9% in the incentive condition and 51.1% in the no-incentive condition (compared with 20.7% for the norms). Thus, the accuracy incentive introduced in Experiment 1 was completely ineffective in alleviating prediction inflation.

Discussion

Experiment 1 was predicated on the assumption that the emphasis on accurate predictions and the use of accuracy incentives might induce participants to engage in an analytic evaluation of potential responses and their relative probabilities. Indeed, in the post-experiment interview, several participants indicated that they performed the task by thinking of responses that they or other people would be likely to make to the cue word, and then compared the perceived likelihood of these responses to that of the target word. Such analytic evaluation would be expected to result in more realistic estimates of the likelihood of occurrence of the target response. Surprisingly, however, there was not even a hint that the incentive manipulation reduced the inflation bias.

It would seem that the deliberate consideration of alternative responses was not successful in overcoming the inflated a-posteriori associations that are activated by the target. Indeed, most participants mentioned in the interview that when the cue and target words were related, they found it difficult to ignore the target word and its association to the cue, and a few participants added that perhaps if they had seen the cue alone they would have probably produced many other responses than the target word. These reports suggest a conflict between heuristic-driven feelings and analytic-based knowledge (see Denes-Raj & Epstein, 1994) with the former winning in influencing the estimation task.

[2] It should be stressed that the analyses of prediction inflation is problematic in the case of the unrelated pairs, because prediction could deviate only in one direction.

Experiment 2: Training in estimating the occurrence of different responses

In Experiment 2 we used a training session that was designed to induce participants to engage in an analytic, systematic process. In that session, participants were required to produce six responses to a cue word and also to assess the likelihood of each of them before estimating the percentage of participants who would produce the target response to that cue word. Thus, unlike the procedure used by Koriat et al. (2006), in which participants produced one or two associates to the cue word, here they generated 6 responses and also estimated the percentage of people who would give each of these. This training procedure was expected to reduce markedly the overprediction effect for the responses generated during training, and to eliminate or reduce the inflation bias for the pairs that were included in the experiment proper.

The design of the experiment involved two conditions. In the intervening training condition, the training was interposed between two administrations of the estimation task. The procedure for each such administration was identical to that of Experiment 1. In the prior training condition, in contrast, participants received the training session first, and then were administered the estimation task.

Method

Participants

Forty Hebrew-speaking University of Haifa undergraduates (24 women and 16 men) participated in the experiment for course credit. They were assigned randomly to the two conditions, with 20 participants in each condition.

Materials, apparatus and procedure

The materials and apparatus were identical to those of Experiment 1. In the intervening training condition, participants first performed the 90-pairs estimation task using a procedure that was identical to that of Experiment 1. They then underwent a training session, and subsequently performed the estimation task again. In the prior training condition, in contrast, participants received the training session before performing the estimation task once.

In both conditions participants first received instructions about the word association task, and were told that they would have to estimate the percentage of people who will give the target word in response to the cue word as the first association that comes to mind. In both conditions participants were given an accuracy incentive as in Experiment 1: They were told that they would receive NIS 0.50 for each prediction that deviates from the norms by no more than 5% points in either direction (this was true for both presentations of the intervening training condition). In the prior training condition, participants were instructed that the training phase is designed to familiarize them with the word-association task and with their own task, which is to assess the likelihood of the target responses. Similar instructions were given to the intervening training group just before the training phase. In both conditions it was emphasized that the training session is intended to improve the accuracy of the participants' estimates and hence to increase their monetary bonus.

The 10 word pairs used for training were chosen so as to represent forward, backward and unrelated associative types. Participants were told that they would first get a chance to practice the word association task themselves. They received a booklet containing 10 training trials. The first five trials included two pages each. At the top of the first page appeared the cue word. Then the same cue word was repeated 6 times, each time with a blank space next to it. Participants were instructed first to write down six different associations that come to mind in response to that word, one in each space. They were then asked to estimate for each such association the percentage of people who are likely to give that association as the first association in response to the cue, and to write their estimate next to the response. When they completed the task, they were asked to turn the page, where the cue word appeared again together with a target response. Participants were asked to give their estimate for that target word. (If the target word was identical to one that they had provided themselves, they were asked simply to copy their previous estimate). This procedure was repeated for four more trials. The forms for the final five trials were similar except that the cue-target pair appeared at the top of a page followed by 6 cue-space pairs. Participants were asked again to give 6 different responses to the cue word, and estimate the percentage of each. Only then were the participants required to go to the top of the page and to estimate the percentage of people who would give the target word as a response.

Following the training session, the participants were instructed to use the insight and experience that they had gained from the training session in making their estimates in the estimation task that followed. In the intervening training condition, participants were shown again the same 90 word pairs, presented in a new random order. Participants were interviewed at the end of the experiment about the strategy that they had used in making predictions, and how they took advantage of the training session.

Results

Training session

We shall examine the results for the training session before analyzing the effects of training on the estimation task. Because of an error, the identity (and condition) of the participants could not be determined for 8 training booklets. However, the results for the remaining 32 participants yielded little differences between the two conditions, and therefore the results for the training session were analyzed across all participants in both conditions.

The average estimates provided by participants decreased monotonically with the ordinal position of their response, averaging 39.2%, 25.5%, 20.4%, 18.7%, 16.4%, 16.1%, respectively, for the first to the sixth responses, $F(5, 195) = 86.20$, $MSE = 35.67$, $p < .0001$.

For each participant, we calculated the sum of the estimates listed for the six responses, and averaged that sum across the 10 training items. These means averaged 137.1% across participants, significantly higher than 100%, $t(39) = 2.57$, $p < .05$. These results indicate that the overprediction effect reported by Koriat et al. (2006) was obtained even during the training session. However, although the overall magnitude of the overprediction effect was similar to that observed by Koriat et al. (2006, Experiment 3b), there was a greater variation in this study: There were only 26.3% of the cases across all participants and items in which the sum of the 6 estimates exceeded 100%. Possibly, the provision of the 6 estimates one after the other led some participants to deliberately ensure that their estimates do not total more than 100%. Indeed, the sum of predictions ranged from 57% to 470% across participants, and was relatively reliable between the first five items and the next five items: Of the 19 participants who evidenced overprediction for the first 5 items, 9 exhibited overprediction in the second set of 5 items as well. In contrast, of the 21 participants who did not evidence overprediction for the first 5 items only 1 did so for the second set, χ^2 $(df = 1) = 9.66$, $p < .005$.

Turning next to the estimates provided for the 10 target responses, these averaged 36.4% ($SD = 15.4$), in comparison with 20.9% for the actual percentage, $t(39) = 6.51$, $p < .0001$. Thus, although each participant listed 5–6 responses that differed from the critical target, the occurrence of the critical target was nevertheless overestimated. Focusing only on the first 5 trials (for which the critical target was revealed only after the listing of 6 responses), in 5.9% of the cases the target was the same as one of the responses that the participant had provided. There were 36 participants who gave both "same" and "different" responses for these pairs. For these participants, the estimated occurrence of the target response was 38.9% ($SD = 22.1$), and 27.0% ($SD = 16.0$), respectively, $t(35) = 2.94$, $p < .01$. However, the actual percentage for those items averaged 41.6% and 10.9%, respectively, $t(35) = 4.25$, $p < .0001$, so that it does not seem that the inflation bias was any weaker when none of the generated responses matched that target than when one of them matched it (see Koriat et al., 2006).

The effects of training on the estimation task

We turn next to the results for the experiment proper. Fig. 2 presents mean predicted percentages for the prior training group and for the first and second presentations of the intervening training group. Also displayed are the corresponding actual percentages. Consider first the results for the intervening training group. The estimates in the first presentation display the typical inflation bias, with the overestimation effect being most pronounced for the backward pairs: Predicted percentages for the forward, backward and unrelated pairs averaged 79.0%, 55.2%, and 12.4%, respectively, compared with 59.9%, 2.0%, and 0%, respectively, for the actual percentages. These results are quite similar to those observed for the control condition in Experiment 4 of Koriat et al. (2006). (The respective mean estimates in that experiment were 81.1%, 64.0%, and 8.1%). The inflation bias was significant for each of the associative classes: $t(19) = 8.05$, $p < .0001$, for the forward pairs, $t(19) = 12.11$, $p < .0001$, for the backward pairs, and $t(19) = 5.35$, $p < .0001$, for the unrelated pairs. Because no error variance is available for the actual percentages, we compared the forward and backward pairs in terms of the estimated-actual differences calculated for each participant. This comparison yielded $t(19) = 11.0$, $p < .0001$, indicating a stronger bias for the backward pairs.

Did the training procedure alleviate the overestimation bias? It seems that it did. A two-way ANOVA comparing the estimates made before and after training yielded $F(1, 19) = 18.50$, $MSE = 264.03$, $p < .001$, for presentation, $F(2, 38) = 135.12$, $MSE = 276.81$, $p < .0001$, for associative type, and $F(2, 38) = 7.57$,

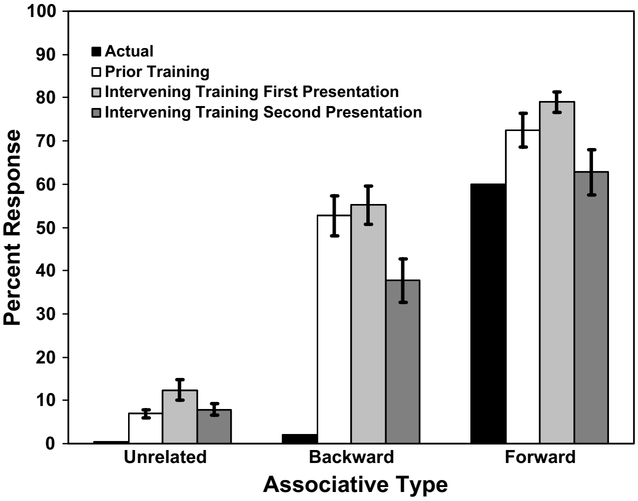

Fig. 2. Mean actual and predicted response percentages for the prior training condition and for both presentations of the intervening training condition. The results are presented separately for the unrelated, backward and forward pairs. Error bars represent +1 *SEM* (Experiment 2).

$MSE = 66.40$, $p < .005$, for the interaction. The overall effect of the training procedure was to reduce estimates by about 13%.

Nevertheless, the training procedure was successful in eliminating the overestimation bias only for the forward pairs, but failed to do so for the backward and unrelated pairs. A comparison of the second presentation estimates with the actual percentages yielded $t(19) = 0.50$, ns, for the forward pairs, but $t(19) = 7.17$, $p < .0001$, for the backward pairs, and $t(19) = 5.62$, $p < .0001$, for the unrelated pairs. A comparison of the forward and backward pairs in terms of the estimated-actual differences calculated for each participant yielded $t(19) = 10.62$, $p < .0001$, indicating a stronger bias for the backward pairs. In the first presentation, estimates were inflated by a factor of 1:1.3 for the forward pairs, and by a factor of 1:27.6 for the backward pairs. The respective values on the second presentation were 1:1.1 and 1:18.9, respectively.

Turning next to the prior training condition, somewhat surprisingly, this training seems to have been largely unsuccessful in reducing the overestimation bias: A two-way ANOVA comparing the estimates made in this condition with those of the intervening training group on the first presentation yielded $F(1, 38) = 1.70$, $MSE = 400.65$, $p < .21$, for condition, $F(2, 76) = 316.78$, $MSE = 143.24$, $p < .0001$, for associative type, and $F < 1$ for the interaction.

Dividing participants in terms of overprediction in the training session

As noted earlier, there were reliable differences between participants in the estimates provided during the training phase. It may be speculated that these differences reflect in part the extent to which participants attempted to engage in an analytic process that helps overcome the experience-based overestimation bias. If so, participants who did not exhibit an overprediction effect during training would be expected to provide lower estimates for the experimental pairs. To examine this hypothesis, the participants in each condition were divided into those whose mean sum of estimates across the 6 responses in the 10 training items exceeded 100% (overprediction) and those for whom that mean was less than 100% (no overprediction). The number of such participants (using only the 32 participants whose condition could be determined) was 10 and 6, respectively, in the prior training group, and 5 and 11, respectively, in the intervening training group. Fig. 3 presents mean estimates provided by each of the two groups in the prior training condition (top panel) and in each of the two presentations in the intervening training condition (bottom panel).

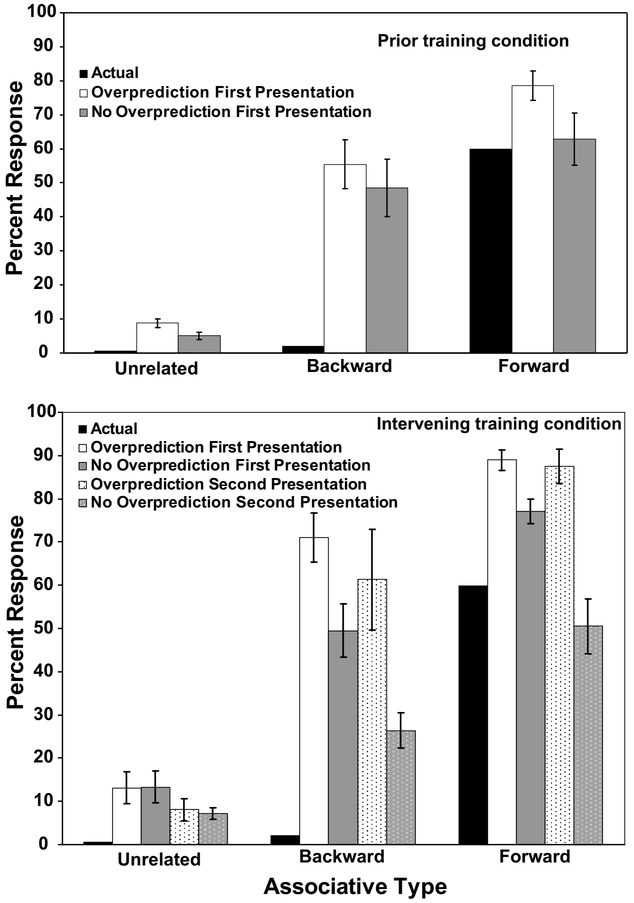

Fig. 3. Mean actual and predicted response percentages for the overprediction and no-overprediction groups, for the unrelated pairs, backward pairs and forward pairs, plotted separately for the prior training condition (top panel) and for the two presentations of the intervening training condition (bottom panel). Error bars represent +1 *SEM* (Experiment 2).

For the prior training group, indeed no-overprediction participants provided somewhat lower estimates for the experimental pairs (38.8%) than did the overprediction participants (47.6%), $t(14) = 1.41$, $p < .20$. Although the difference was not significant, it might suggest a causal influence of training: Participants who became aware of the competition between alternative responses during training were able to apply that knowledge in making predictions for the experimental pairs. However, this effect could also simply indicate that the overprediction effect is diagnostic of chronic individual differences in the tendency to rely more heavily on subjective experience or on an analytic mode of processing (see Stanovich & West, 2000). Examination of the results for the intervening-training condition suggests that both effects are operative. Thus, individual differences in overprediction in the training session postdicted differences in overestimation: The estimates provided in the first presentation by the no-overprediction and overprediction participants averaged 46.6% and 57.7%, respectively, $t(14) = 2.21$, $p < .05$. The effect was significant for both the backward pairs, $t(14) = 2.17$, $p < .05$, and the forward pairs, $t(14) = 2.56$, $p < .05$. The intervening training, however, appeared to intensify these differences. An Overprediction X Presentation ANOVA on mean estimates yielded $F(1, 14) = 3.65$, $MSE = 82.29$, $p < .08$ for the interaction. The reduction in mean estimates from the first to the second presentation amounted to 18.6% for the no-overprediction group compared with 5.4% for the overprediction group.

Discussion

The results for the intervening training group disclose the most successful means of reducing the inflation bias that has been found so far. The training session succeeded in reducing estimations by about 13% overall, and practically eliminated the inflation bias for the forward pairs.

Nevertheless the results on the whole are quite discouraging given the extensive training that was intended to induce an analytic mode of responding. First, the prior training group yielded no effect of training at all. Practically all participants in this group stated in the postexperiment interview that they had failed to see how the training procedure could help them in providing accurate estimates in the experiment proper. In contrast, some of the participants in the intervening-training condition reported that the training session helped them "put things in proportion" and led them to make lower estimates than they would have made otherwise. Note that indeed, the proportion of participants who exhibited overprediction was smaller for the intervening training group (31%) than for the prior training group (63%), χ^2 $(df = 1) = 3.14$, $p < .08$. Thus, perhaps it was necessary to have participants go through the first presentation in order for them to benefit from the training procedure.

Second, even in the intervening training group the inflation was very marked for the backward pairs in the second presentation, with the percentage of responding overestimated by a factor of almost 20. The impression from the post-experiment interview was that the overestimation bias was particularly strong when participants focused on the target word and tried to evaluate its likelihood. In contrast, a focus on the cue word and on the potential responses that it can induce seemed to help participants overcome in part the a-posteriori associations that stem from the target word.

However, the attempt to generate alternative responses to the cue word, in itself, was not sufficient to override the effect of the a-posteriori associations, as suggested by the results of the training session. These results indicate that some participants exhibited an overprediction effect even during that session. Although the intervening training procedure was effective in reducing overestimation, particularly for the no-overprediction participants, even these participants exhibited inflated predictions for the backward pairs in the second presentation, $t(10) = 6.11$, $p < .0001$. In fact, as noted earlier, Maki (2007b) found an overestimation bias for low-association pairs even when participants were instructed to rate several alternative responses under the constraint that the ratings should total exactly 100. Thus, our attempt to induce an analytic attitude by soliciting estimates for several alternative responses was not sufficient to overcome the inflation bias even among those who succeeded to avoid the overprediction bias during training.

Experiment 3: Working in dyads

Given that participants in the training session did generate alternative responses to the target word and also judged several of these to be quite likely, why did they nevertheless overestimate the occurrence of the target response even on the training items? It would seem that although the training session promoted an analytic process that has the potential of eliminating prediction inflation, participants did not incorporate the consequences of that process strongly enough to override the contaminated subjective experience induced by the presence of the target. How can participants be encouraged to do so?

Experiment 3 examined the possibility that perhaps asking people to work in dyads and to come to an agreement about the estimate not only should make participants consider rational arguments but may motivate them to apply these arguments in producing an accurate estimate. Indeed, the extensive research of Tetlock and his associates on the social contingency model (see Tetlock, 1992; Tetlock & Lerner, 1999) indicates that accountability to others tends to activate effortful and systematic processing and to reduce reliance on easy-to-execute heuristics. Thus, when people feel accountable to other individuals (with unknown views) they tend to engage in preemptive self-criticism, to consider alternative arguments to their own, and to incorporate potential objections into their own position. Also, the attempt to convince each other, by its nature, makes participants appeal to reason, and may be expected to induce a more deliberative mode of processing than when people work alone.

Many previous studies comparing individual and group performance indicated that cooperative groups perform better than independent individuals on a wide range of problem-solving tasks (see e.g., Hill, 1982; Laughlin, Zander, Knievel, & Tan, 2003). Some studies also indicated that decisions benefit from working in groups. For example, Allwood and Granhag (1996) who had participants make their judgments first alone and then in pairs, found less overconfidence bias when working in pairs (see also Sniezek & Henry, 1989). Thus, the hypothesis is that working in dyads should reduce the inflation bias in Experiment 3 because it encourages a deliberative, analytic process and because it endows the outcome of that process with the power of overcoming the illusory convictions generated by the target.

Method

Participants

Forty-eight Hebrew-speaking University of Haifa undergraduates (24 women, 24 men) participated in the experiment, 32 participants were paid for their participation and 8 received course credit. Participants performed the task in pairs and both members of a pair were of the same gender.

Materials, apparatus and procedure

The materials and apparatus were the same as in Experiment 1. The procedure was also the same except that participants worked in pairs and that the estimation task was presented twice, with a different random ordering of the items in each presentation. The two participants sat side by side facing the computer screen. The instructions were the same as those of Experiment 1 but participants were required to come to an agreement about their estimate. They were told that it is likely that their initial estimates will differ, but they should discuss their estimates, try to argue and persuade each other if necessary, and come to an agreement about the final estimate. As in Experiment 2, they were promised NIS 0.5 for each prediction that deviates from the norms by no more than 5% points in either direction (this was true for both presentations). When the two participants agreed on the estimate, the experimenter entered it on the keyboard and initiated the next trial. When the first presentation was over, participants were told that they will repeat the task again using the same list of pairs, but that the pairs will be presented in a new random order.

Results

A preliminary analysis revealed little differences between the men and women dyads. Their estimates averaged 35.0% and 37.6%, respectively. Also, the results for the two presentations were very similar, and the analyses to be reported were therefore pooled across both presentations.

Across the two presentations, mean estimates for the forward, backward and unrelated pairs were 73.5%, 30.4% and 5.0%, respectively (when the actual figures were 59.9%, 2.0% and 0%, respectively). Hence the estimates were still significantly inflated: $t(23) = 5.01$, $p < .0001$, for the forward pairs, $t(23) = 7.37$, $p < .0001$, for the backward pairs, and $t(23) = 10.10$, $p < .0001$, for the unrelated pairs.

Nevertheless, working in dyads was quite successful in reducing the inflation bias in comparison with the predictions obtained in Experiment 1, in which participants performed the task individually. Fig. 4 depicts the mean estimates made in Experiment 1 (Individual) and Experiment 3 (Dyad) as well as the actual percentages. Working in dyads had a pronounced and consistent effect, reducing the estimates by 17.6% on average. A Condition (Individual vs. Dyad) X Associative Type ANOVA yielded significant effects for condition, $F(1,46) = 31.99$, $MSE = 349.41$, $p < .0001$, for associative direction, $F(2,92) = 482.90$, $MSE = 128.10$, $p < .0001$, and for the interaction $F(2,92) = 21.80$, $MSE = 128.10$, $p < .0001$. The effect of working in dyads was significant for the forward pairs, $t(46) = 3.25$, $p < .005$, the backward pairs, $t(46) = 6.01$, $p < .0001$, and the unrelated pairs, $t(46) = 3.42$, $p < .005$.

Discussion

Working in dyads was successful in reducing overestimation, and did so across all three types of pairs. In fact, it was no less effective than the manipulation used in Experiment 2 even though participants in Experiment 3 were not induced explicitly to adopt an analytic mode of processing. Possibly, the need to justify one's estimate (Tetlock, 1992) and to convince the other partner inherently results in a greater emphasis on rational, analytic considerations as against emotional considerations that call for reliance on each member's immediate subjective experience. Perhaps, in this respect dyadic interaction is similar to receiving advice from others, which has been assumed to help overcome self-confirmation tendencies (Yaniv, 2004). Furthermore, the active attempt to convince others would seem to increase the likelihood that

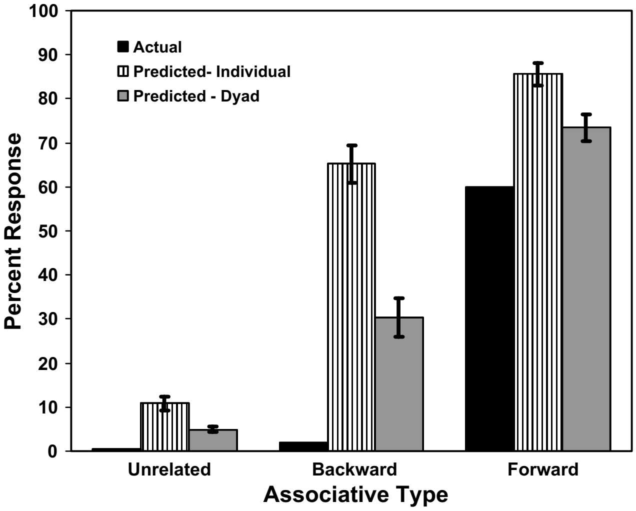

Fig. 4. Mean actual and predicted response percentages for the Dyad (Experiment 3) and Individual (Experiment 1) conditions, for the unrelated, backward and forward pairs. Error bars represent +1 *SEM*.

the voiced arguments will impact judgments and decisions.

Experiment 4: Working in dyads after working alone

Experiment 4 attempted to obtain some insight into the process underlying the effectiveness of the dyadic condition in reducing conditional inflation. In this experiment, participants worked also in dyads, but in the first presentation they provided independent estimates individually and only in the second presentation were they asked to negotiate about a joint estimate. The question of interest was whether in making a joint estimate, participants indeed consistently shifted their individual estimates downward.

Method

Participants

Thirty-two Hebrew-speaking University of Haifa undergraduates (22 women, 10 men) participated in the experiment for payment. They performed the task in pairs and both members of a pair were of the same gender.

Procedure apparatus and materials

The apparatus and materials were the same as in Experiment 3. The procedure was also the same except that in the first presentation each member of the pair sat in front of a different computer and made his or her estimate individually under accuracy incentive instructions. When both participants completed the task, one of them was asked to join the other's computer, and they were presented once again with the list of pairs and asked to come to an agreement about a joint estimate. The procedure was the same as that of Experiment 3 except that the two individual estimates from the first presentation, each in a different color specific to each member, appeared below the cue-target pair. After reaching a joint estimate, each participant was asked to judge whether that estimate will be rewarded, and if the answer was ‘‘no’’, to indicate whether the estimate provided should have been lower or higher in order to be rewarded. These judgments were reported orally, and were entered by the experimenter. In the first presentation, the two participants received the word pairs in the same random order, but in the second presentation the pairs were presented in a new random order.

Results

The absolute discrepancy between the individual estimates provided by the two members of a dyad in the first presentation averaged 19.8% across items and participants. The average of the two individual estimates for each item was compared to the joint estimate. Table 1 presents the means of the averaged individual estimates (first presentation), the means of the joint estimates (second presentation), and the actual percentages. Presented also for comparison are the means from Experiment 1 (individual) and Experiment 3 (joint).

Table 1
Mean actual percentages, averaged individual estimates (first presentation) and joint estimates (second presentation) in Experiment 4, and mean estimates in Experiments 1 and 3 listed for each associative type and across all items

Experiment condition	Associative type			
	Forward	Backward	Unrelated	All
Actual	59.9	2.0	0.0	27.0
Experiment 4—Individual	82.1	58.1	9.8	50.0
Experiment 4—Joint	83.3	44.9	5.6	44.6
Experiment 1	85.6	65.2	10.8	53.9
Experiment 3	73.5	30.4	5.0	36.3

The mean individual estimates in the first presentation were very similar to those obtained in Experiment 1. An Experiment (1 vs. 4) X Associative Type ANOVA yielded significant effects for associative type, $F(2,76) = 565.49$, $MSE = 101.59$, $p < .0001$, but not for experiment or the interaction, $F(1,38) = 1.34$, ns, and $F < 1$, respectively.

Turning next to the second presentation, mean joint estimate (44.6%) was lower than the mean of the individual estimates (50.0%), but was still higher than the mean obtained in Experiment 3 (36.3%). Thus, comparing the individual estimates to the joint estimates, a Condition (Individual vs. Joint) X Associative Type ANOVA (treating condition as a repeated measure) yielded significant effects for condition, $F(1,15) = 23.01$, $MSE = 30.37$, $p < .0005$, for associative type, $F(2,30) = 285.32$, $MSE = 159.28$, $p < .0001$, and for the interaction, $F(2,30) = 22.71$, $MSE = 18.94$, $p < .0001$. The effect of condition was significant for the backward and unrelated pairs $t(15) = 5.34, p < .0001$, $t(15) = 3.28, p < .01$, respectively, but not for the forward pairs, $t(15) = 1.52$.

Thus, once again, working in dyads proved effective in reducing prediction inflation. Although participants first made their estimates individually, and although these estimates were in front of them during the dyadic interaction, the requirement to reach a shared estimate resulted in a reduction in the estimates provided. In fact, across all items, mean joint estimate was lower than the average of the individual estimates for 56.1% items, and was higher for 28.6% items (and equal for the remaining 15.3% items). The respective percentages for the backward pairs were more impressive: 73.3% and 16.0% (and equal for 10.6%), respectively. Thus, possibly working in pairs induced participants to reason about the estimate, and this led them to reduce their estimates in the majority of cases, although it did not do so in all cases. It should be noted that only in 1.3% of the cases was the joint estimate higher than the highest individual estimate, suggesting that although, in principle, the dyadic negotiation could have led both members to increase their estimate, this happened very infrequently. Note also that in 12.1% of the cases, the joint estimate was lower than both of the individual estimates.

Recall that after making their joint estimate, both participants were asked to judge whether they will receive a bonus for making an accurate estimate. In 92.1% of the cases they judged that they will. In fact, however, they won a bonus only in 32.1% of the cases. For the remaining 7.9% cases in which they thought that they will not win a bonus, in 63.0% of them they judged that their estimate should have been lower to win a bonus. This figure was significantly higher than 50%, $t(28) = 2.19$, $p < .05$ (based on 29 participants, because 3 participants indicated that they will always win a bonus), suggesting that participants were partly aware of the conflict between the estimate that was based on their subjective feelings and the one that follows from analytic considerations, which called for lower estimates than what they made.

Turning next to a comparison of the results for the dyadic conditions in Experiments 3 and 4, a two-way ANOVA, Experiment (3 vs. 4) X Associative Type yielded significant effects for experiment, $F(1, 38) = 7.80$, $MSE = 256.10$, $p < .01$, for associative type, $F(2, 76) = 417.77$, $MSE = 123.83$, $p < .0001$, and for the interaction, $F(2, 76) = 3.89$, $MSE = 123.83$, $p < .05$. The effect of experiment was significant for the backward and forward pairs $t(38) = 2.41$, $p < .05$, and $t(38) = 2.51$, $p < .05$, respectively, but not for the unrelated pairs, $t(38) = 0.76$, ns. Clearly, the effectiveness of working in dyads was more pronounced in Experiment 3, when participants did not make their estimates alone before working in dyads. This observation suggests that working in dyads reduces the tendency to engage in a self-confirmation process, and this reduction is stronger when participants do not commit themselves first to their individual estimates.

Discussion

Experiment 4 replicated the observation that working in dyads helps alleviate prediction inflation. What is impressive is that this alleviation occurred even though participants made their individual estimates first and were presented with these estimates when working in dyads. Examination of the results across items confirmed that indeed in most cases the dyadic discussion resulted in reduced estimates in comparison with the estimates made by each person alone. The observation that such was not the case for all items is noteworthy, suggesting that group discussion may not always help to overcome prediction inflation.

This conclusion, however, must be taken with caution because of the differences observed between the dyadic conditions of Experiments 3 and 4. These differences suggest that the conclusions drawn from the results of Experiment 4, in which participants first worked alone, may not generalize to a situation in which participants work in pairs from the beginning. Indeed, examination across items revealed that for 94.4% of the items, mean estimate in Experiment 3 was lower than the average of the individual estimates in Experiment 4, and was higher only for 5.6% of the items. The respective percentages for the backward pairs were more impressive: 100% and 0%! These results suggest that the dyadic situation, in itself, consistently leads people to reduce their estimates. The comparison of the results of Experiment 3 and 4 has important practical implications, suggesting that working in pairs is more effective in alleviating prediction inflation when participants work together from the beginning than when each of them first makes his/her own estimate.

Experiment 5: Working alone after working in dyads

The primary aim of Experiment 5 was to test the hypothesis that the beneficial effects of working in dyads transfer to a situation in which participants work individually.[3] Evidence consistent with this hypothesis will support the idea that dyadic interaction indeed affects mode of reasoning. Therefore, after working in dyads, each participant provided his/her own estimate on a new set of items. A secondary aim of Experiment 5 was to obtain some preliminary information about the nature of the arguments that are raised during the dyadic interaction. Therefore the conversation between the two participants during the dyadic session was tape recorded.

Method

Participants

Twenty Hebrew-speaking undergraduates (all women) participated in the experiment for payment.

Apparatus and materials

The apparatus was the same as in previous experiments. The 60 related pairs used in the previous experiments were divided into 4 sets of 15 pairs each that were closely matched in terms of forward and backward associative strengths. These were combined to form two lists of 30 pairs each, such that in each list one set was presented in the forward direction and one in the backward

[3] I am grateful to an anonymous reviewer for proposing this experiment.

direction, with direction counterbalanced across participants. Each list included also 15 different unrelated pairs from the original list, so that there were 15 pairs of each associative type in each list. One list was used for the dyadic session and the other was used for the individual session, with the assignment of the lists to each session counterbalanced across participants. The pairs were randomly ordered for each participant for each session.

Procedure

The procedure for the dyadic session was the same as in Experiment 3, with the exception that the conversation between the two participants was recorded. The procedure for the individual session that followed was similar to the individual condition in Experiment 4. As in Experiment 4, participants in both sessions also indicated at the end of each trial whether they expected to receive a bonus, and if not whether their estimate should be lower/higher to be rewarded. (The results from these questions were not informative and will not be reported.)

Results

We first examine the joint predictions made in the first session. The estimates for the forward, backward and unrelated pairs averaged 69.5%, 30.6% and 4.5%, respectively (when the actual percentages were 59.9%, 2.0% and 0%). Estimates for the backward and unrelated pairs were still inflated, $t(9) = 6.04$, $p < .001$, and $t(9) = 5.12$, $p < .001$, respectively. However, they were numerically lower even than the joint estimates of Experiment 3 (73.5%, 30.4%, and 5.0% for the forward, backward and unrelated, respectively). In fact, the estimates for the forward pairs did not differ significantly from the actual percentages, $t(9) = 1.36$, ns. Perhaps the fact that participants knew that they were recorded increased their need to be accountable, resulting in lower estimates than in Experiment 3.

Did the beneficial effect of working in dyads transfer to the task of working alone? To compare the joint and individual estimates, we first averaged the individual estimates made by the two members of each dyad in the second session. The means of the averaged estimates were 69.9%, 32.7% and 5.3%, for the forward, backward and unrelated pairs, respectively. These means were very similar to those obtained for the dyadic condition. Indeed, a 2-way ANOVA, Condition (Dyad vs. Individual) X Associative Type yielded, $F(2,18) = 90.76$, $MSE = 233.79$, $p < .0001$, for associative type, but $F(1,9) = 1.36$, $MSE = 13.91$, $p < .28$, for condition, and $F < 1$ for the interaction.

We also compared the individual estimates made in the second session (without averaging them) to the individual estimates observed in Experiment 4 (which were made prior to the dyadic condition). (The latter, it may be recalled, averaged 82.1%, 58.1%, and 9.8%, for the forward, backward and unrelated pairs, respectively). A 2-way ANOVA, Experiment (5 vs. 4) X Associative Type, yielded, $F(1,50) = 17.08$, $MSE = 425.61$, $p < .0001$, for experiment, $F(2,100) = 433.08$, $MSE = 134.31$, $p < .0001$, for associative type, and $F(2,100) = 10.31$, $MSE = 134.3$, $p < .0001$, for the interaction. The estimates of Experiment 5 were lower than those of Experiment 4 for the forward, backward and unrelated pairs, $t(50) = 2.48$, $p < .05$, $t(50) = 4.70$, $p < .0001$, and $t(50) = 2.50$, $p < .05$, respectively. The interaction suggests that the strongest difference was observed for the backward pairs.

We will comment only briefly on the impressions gained from examination of the recorded dyadic conversations. These impressions should be treated with caution in the absence of similar data from a condition in which participants work alone. As might be expected, the majority of the arguments involved rational considerations. Many of these were objections by one member to the high estimate proposed by the other members (“10% is a lot; that means that 10 people out of 100 will respond with that word”). In many cases participants listed their own personal associations, and then tried to infer the percentage of participant who would respond with the target word (“how many other associations did we mention? Five? and some people may give other associations that we did not even think of. So we must reduce the estimate”). Some participants commented that the presence of the target makes the estimation task difficult (“It is a problem that we see the target word, because it is difficult to get it out of your mind”). For a few of the (backward associated) pairs, some participants explicitly stressed the asymmetric association (“If it were *cradle-baby* then perhaps many people would respond with *baby* to *cradle* but not many will say *cradle* in response to *baby*”). It is interesting to note that some participants stated that perhaps they should not give estimates lower than 5% for some (unrelated) pairs in order to maximize their chances to win a bonus (but nevertheless, as noted earlier, mean estimates for the dyadic condition were lower than that in the individual condition of Experiments 1 and 4 even for the unrelated word pairs, see Table 1). In most cases, then, participants appealed to reason and attempted to argue for lowering the initial estimates.

Discussion

The results of Experiment 5 are analogous to those of Hirt et al. (2004). In their study they found that considering an alternative in one domain had the subsequent effect of reducing likelihood estimates in a completely unrelated domain. They interpreted this result as indicating that the generation of alternatives induces a mental simulation mind-set that generalizes to another domain. Similarly, the transfer of debiasing from the dyadic condition in Experiment 5 to the individual condition also indicates that working in dyads does not merely affect overt responding but results in a change in mode of reasoning. A similar argument has been made with regard to the effects of accountability: It was proposed that accountability can produce a qualitative change in mode of processing (Tetlock & Lerner, 1999).

General discussion

There has been some interest in recent years in evaluating the efficacy of judgments and decisions that are made automatically and immediately as against those that are based on a deliberate and careful analysis (see McMackin & Slovic, 2000). Several researchers proposed that nonconscious information processing is often more efficient than conscious analysis (Lewicki, 1986) and that thinking about reasons for decisions can sometimes lead to decisions of poorer quality (Wilson & Schooler, 1991). For example, Dijksterhuis and his colleagues (Dijksterhuis, Bos, Nordgren, & van Baaren, 2006; Dijksterhuis & Nordgren, 2006) reported several experiments suggesting that it is not always advantageous to engage in thorough deliberation before choosing, and that in fact a deliberation without attention yields better decisions. These suggestions run counter to the conventional wisdom that careful analysis leads to decisions of better quality (Janis & Mann, 1997; Koriat et al., 1980).

Clearly, whether an intuitive, non-deliberative mode of processing yields better decisions and judgments than an analytic mode of processing depends on the nature of the task (Dijksterhuis & Nordgren, 2006; Dijksterhuis et al., 2006; McMackin & Slovic, 2000). In fact, studies in metacognition have documented many situations in which illusions of competence and illusions of knowing derive precisely from reliance on one's own immediate subjective experience. These studies, suggest that in some cases intuitive feelings can lead metacognitive judgments astray (Benjamin, Bjork, & Schwartz, 1998; Chandler, 1994; Fischhoff et al., 1977; Jacoby & Whitehouse, 1989; Koriat, 1998; see Koriat, 2007 for a review). In such cases, faulty metacognitive judgments can be alleviated either by educating subjective experience itself or by inducing a shift to an analytic mode of processing (Koriat & Bjork, 2006; Koriat et al., 2004).

In the task that we used in this study—predicting the occurrence of targets as responses to cues in a word-association task—the inflated estimates seem to derive from uncritical reliance on the illusory feelings that are produced by the presence of the target. Hence, one potential way for alleviating inflated predictions is to induce participants to engage in a deliberative mode of processing that may override or reduce the effects of contaminated subjective experience. In fact, the task used in this study is similar to the numerical estimation task for which McMackin and Slovic (2000) expected deliberate reasoning to improve the quality of judgments, in contrast to "intuitive" tasks for which reasoning was expected to disrupt judgments. In the former task, participants were required to estimate various quantities (e.g., "How many cigarettes are consumed in the US each year?"). It was observed that participants who were asked to think about the reasons for their estimates provided estimates that were generally closer to the truth than control participants who were not explicitly required to do so. Thus, we expected that in a similar manner any manipulation that induces an analytic mode of processing should help improve the accuracy of conditional predictions.

Three such manipulations were explored: Motivating participants to provide as accurate and precise estimate as they could and rewarding them for accurate predictions, providing training designed to induce an analytic, systematic analysis, and having participants work in pairs. Two conditions proved effective in reducing prediction inflation, first, intervening training when this was introduced between two administrations of the prediction task (Experiment 2), and the requirement to work in dyads and to come to an agreement regarding the prediction (Experiment 3–5). The benefits from both manipulations were clearly more marked than those achieved by the manipulations that were included in the previous investigation (Koriat et al., 2006). However, the magnitude of the prediction inflation remained high even in these conditions, and furthermore, little benefit was observed in Experiment 1and in the prior-training condition of Experiment 2.

Perhaps the most important conclusion from this study and from previous attempts to alleviate the inflation of conditional predictions (Koriat et al., 2006; Maki, 2007a,b) is how difficult it is to mend convictions that stem from the presence of the outcome whose likelihood is to be assessed. This difficulty seems to derive not only from the failure to engage in the kind of analytic reasoning that has the potential of producing more accurate estimates but also from the failure to benefit from the outcome of that reasoning in overcoming immediate subjective feelings. This conclusion is based on several observations. First, merely motivating participants to be accurate and rewarding them for estimates that are close to the norms had no effect whatsoever (Experiment 1). Presumably, the accuracy instructions did influence participants' mode of processing, as suggested by their post-experimental verbal reports, but this was not effective enough to affect their predictions.

Second, the results of the training procedure used in Experiment 2 clearly indicate that participants were able to engage in a detailed and careful analysis of the task, and that their analyses could have easily led them to make more realistic predictions. Surprisingly, however, even during the training session itself participants tended to commit overpredictions and to overestimate the likelihood of the target response. Furthermore, despite the beneficial effects of intervening training, the magnitude of overestimation remained marked for the backward pairs, suggesting that analytic reasoning was not sufficient to eliminate the effects of a-posteriori activations.

Finally, the dyadic condition of Experiments 3–5 proved the most effective. Presumably, dyadic interaction induces a change in mode of reasoning, as suggested by the transfer effect observed in Experiment 5. In addition, however, the considerations that emerge in the context of that interaction are more likely to impact the ultimate prediction than when such considerations are raised when working individually on the task. Presumably, the need to justify one's estimate and the attempt to convince each other not only activate a mode of reasoning in which an appeal is made to rational, verbalizable considerations, but also endows these considerations with the power to overcome biased subjective convictions.

The results of Experiments 4 are instructive, suggesting that group discussion has a directional effect, inducing participants to relinquish their earlier (individual) predictions in favor of lower predictions. Nevertheless, the effects on prediction were more modest than those observed in Experiment 3 in which participants worked in pairs without first committing themselves to their individual predictions. In fact, the results of Experiment 3 suggested that practically in every case the requirement to come to a joint prediction resulted in lower predictions for the backward pairs than the average of the predictions that participants would have likely provided when working alone.

Most impressive are the results of Experiment 5 indicating transfer of debiasing from a dyadic condition to an individual condition. These results are perhaps the most direct support for the claim that dyadic interaction results in a shift in mode of reasoning. Although several observations suggest that this shift is indeed from system 1 towards system 2 processing, more direct evidence for this claim is needed.

It is of interest to inquire whether dyadic interaction can prove beneficial in other situations in which people tend to fall prey to illusions and biases that stem from contaminated subjective feelings (e.g., Koriat & Bjork, 2006; Koriat et al., 2004). It may be speculated that working in dyads should be particularly effective in improving the quality of decisions in tasks that were defined by McMackin and Slovic (2000) as "analytic" in contrast with those classified by them as "intuitive." They found that thinking about reasons before deciding improved decision quality for the former tasks but disrupted that quality for the latter tasks. Perhaps a similar interactive pattern will be observed in comparing a condition in which participants work in dyads with one in which they work individually.

In conclusion, the present study provided some clues regarding the procedures that are likely to help in alleviating prediction inflation. A comparison of these procedures with those that proved ineffective in previous studies (Koriat et al., 2006; Maki, 2007a) suggests that the key to mending inflated predictions lies in endowing analytic-based reasoning with the power to overcome heuristic-driven feelings. The present study suggests some of the ways in which this can be achieved.

References

Allwood, C. M., & Granhag, P. A. (1996). Realism in confidence judgments as a function of working in dyads or alone. *Organizational Behavior and Human Decision Processes, 66*, 277–289.

Alter, A. L., Oppenheimer, D. M., Epley, N., Eyre, R. N. (in press). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General.*

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.

Blank, H., Musch, J., Pohl, R. F. (Eds.) (2007). The hindsight bias, (Special Issue). *Social Cognition, 25*.

Camerer, C. F., & Hogarth, R. (1999). The effects of financial incentives in economics experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*, 7–42.

Carroll, J. S. (1978). The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology, 14*, 88–96.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition, 22*, 273–280.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology, 66*, 819–829.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311*, 1005–1007.

Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1*, 95–109.

Epstein, S., & Pacini, R. (1999). Some basic issues regarding dual-process theories from the perspective of cognitive-experiential self-theory. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 462–482). New York: Guilford Press.

Fiedler, K. (2000). On mere considering: The subjective experience of truth. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition* (pp. 13–36). Philadelphia, PA: Psychology Press.

Fiedler, K., & Armbruster, T. (1994). Two halfs may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology, 66*, 633–645.

Fischhoff, B. (1975). Hindsight # foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 552–564.

Gilbert, D. T. (2002). Inferential correction. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 167–184). New York: Cambridge University Press.

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology, 26*, 103–117.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*, 311–327.

Hill, G. W. (1982). Group versus individual performance: Are N+1 heads better than one? *Psychological Bulletin, 91*, 517–539.

Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mindset through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology, 40*, 374–383.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of personality and Social Psychology, 69*, 1069–1086.

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

Janis, I. L., & Mann, L. (1997). *Decision making: A psychological analysis of conflict, Choice and commitment*. New York: Free Press.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.

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

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*, 499–519.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory & Cognition, 9*, 587–598.

Koriat, A. (1998). Illusions of knowing: The link between knowledge and metaknowledge. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 16–34). London, England: Sage.

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.

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.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*, 643–656.

Koriat, A., Fiedler, K., & Bjork, R. A. (2006). Inflation of conditional predictions. *Journal of Experimental Psychology: General, 135*, 429–447.

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: Guilford Press.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107–118.

Laughlin, P. R., Zander, M. L., Knievel, E. M., & Tan, T. S. (2003). Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective reasoning. *Journal of Personality and Social Psychology, 85*, 684–694.

Lewicki, P. (1986). *Nonconscious social information processing*. London: Academic Press.

McMackin, J., & Slovic, P. (2000). When does explicit justification impair decision making? *Applied Cognitive Psychology, 14*, 527–2000.

Maki, W. S. (2007a). Judgments of associative memory. *Cognitive Psychology, 54*, 319–353.

Maki, W. S. (2007b). Separating bias and sensitivity in judgments of associative memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*, 231–237.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737–759.

Robinson, L. B., & Hastie, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 443–456.

Sanbonmatsu, D. M., Posavac, S. S., & Stasney, R. (1997). The subjective beliefs underlying probability overestimation. *Journal of Experimental Social Psychology, 33*, 276–295.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–396). New York: Cambridge University Press.

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes, 4*, 1–28.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–665.

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.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. Zanna (Ed.). *Advances in experimental social psychology* (Vol. 25, pp. 331–376). New York: Academic Press.

Tetlock, P. E., & Lerner, J. S. (1999). The Social contingency model: Identifying empirical and normative boundary conditions on the error-and-bias portrait of human nature. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 571–585). New York: Guilford Press.

Van Wallendael, L. R., & Hastie, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & Cognition, 18*, 240–250.

Wilson, T. D., & Schooler, J. S. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60*, 181–192.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*, 1–13.