Available online at www.sciencedirect.com
ScienceDirect
Organizational Behavior and Human Decision Processes 103 (2007) 68–83
ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
www.elsevier.com/locate/obhdp

Option fixation: A cognitive contributor to overconfidence

Winston R. Sieck [a,*], Edgar C. Merkle [b], Trisha Van Zandt [c]

[a] Klein Associates, Fairborn, OH 45324, USA
[b] Department of Psychology, Wichita State University, Box 34, Wichita, KS 67260, USA
[c] Ohio State University, Columbus, OH 43210, USA

Received 1 July 2004
Available online 12 February 2007

Abstract

The ASC model of choice and confidence in general knowledge proposes that respondents first Assess the familiarity of presented options, and then use the high-familiarity option as a retrieval cue to Search memory for the purposes of Constructing an explanation about why that high-familiarity option is true. The ASC process implies that overconfidence results in part from a tendency to fixate on the high-familiarity option, to the neglect of the other option. If this implication is true, then judgment tasks requiring respondents to evaluate each option independently should result in reduced overconfidence as compared with standard judgment tasks. Two experiments tested this implication, and found that confidence and overconfidence were reduced when respondents evaluated options independently. The findings support the proposal that option fixation contributes to overconfidence, and also clarify the limitations of random error explanations of overconfidence.

Keywords: Confidence; Subjective probability; Likelihood judgment; Calibration; Confirmation bias; Explanation; Dual process; Financial knowledge

Confidence judgment is among the most debated topics in the field of judgment and decision making, and it has played a role in the broader dispute over the extent to which people should be viewed as rational decision makers (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Kahneman & Tversky, 1996). The central issue of the confidence controversy is why people often are found to be overconfident. Is overconfidence due in part to systematic biases in cognitive processing? Or, is it a byproduct of more mundane causes? As an empirical phenomenon, overconfidence is most often studied in the laboratory by the use of general-knowledge test items, such as "Where was Shakespeare born? (a) Stratford-on-Avon, or (b) London." When faced with such an item, respondents first indicate which of the two alternatives is believed to be correct, and then report a probability judgment from 50% to 100% that their responses are correct. An overconfidence effect occurs to the extent that, for a test consisting of many such items, the average of these probability judgments exceeds the actual proportion of correct responses.

Aside from the broader rationality issue, overconfidence is a phenomenon of considerable practical importance. This is in part because of increasing demands to explicitly and accurately communicate probabilistic information in fields involving high uncertainty (e.g., Wilkie & Pollock, 1996). Consider, for example, the field of financial forecasting, which deals with the prediction of such quantities as currency exchange rates, earnings, or stock prices (e.g., Bolger & Önkal-Atay, 2004). Both individuals and corporations stand to make or lose a great deal of money depending on these quantities, and it is thus in their best interest to accurately forecast future financial states (Önkal, Yates, Simga-Mugan, & Oztin, 2003). Confidence judgments become crucial to gauging the certainty of these forecasts: a prediction of

* Corresponding author.
E-mail addresses: wsieck@decisionmaking.com (W.R. Sieck), merkle.14@osu.edu (E.C. Merkle).

0749-5978/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2006.11.001

decreasing stock prices made with 95% certainty is naturally taken more seriously than a prediction made with 60% certainty. Research suggests, however, that such extreme confidence judgments are typically not warranted in the difficult, practical prediction tasks where explicit expressions of uncertainty are most needed (e.g., Thomson, Önkal-Atay, Pollock, & Macaulay, 2003).

Overconfidence is of significant practical importance, yet current explanations for the phenomenon diverge considerably. It is intuitively compelling to consider the overconfidence phenomenon as resulting, at least in part, from cognitive biases in the accumulation or evaluation of evidence. For example, Koriat, Lichtenstein, and Fischhoff (1980) were perhaps the first to propose and test the hypothesis that overconfidence stems at least in part from an inclination to rely more heavily on reasons supporting a chosen answer than on reasons contradicting it. In order to test this proposal, they had subjects in an experimental condition write reasons for and against each of a pair of alternatives given in a general knowledge test, prior to rendering judgments. Consistent with their proposal, subjects in the experimental group were less overconfident than those in a control group. Although Koriat et al.'s initial results were quite promising, there has since been difficulty in replicating them (e.g., Fischhoff & MacGregor, 1982; Yates, Lee, & Shinotsuka, 1992). Other direct evidence in support of the idea that overconfidence stems in part from confirmatory processing has been rather lacking. This is perhaps partly because the overconfidence phenomenon itself has been taken in support of the hypothesis, a notion that has recently come under intense scrutiny. For example, Juslin, Winman, and Olsson (2000) stated that, "With general knowledge items, the idea of an information-processing bias is approaching the status of a dogma, supported by naïve empiricism and selective attention to particular data sets." (p. 385).

The impetus for this statement is that in the last several years, researchers have developed alternative explanations for the overconfidence phenomenon under a general assumption that respondents are unbiased processors of statistical information. One class of explanations has to do with representative sampling of test questions (e.g., Gigerenzer et al., 1991). According to this explanation, overconfidence results from the selection of test questions that are unduly tricky. Some support for the idea that participants have prior conceptions of test trickiness comes from Arkes, Christensen, Lai, and Blumer (1987), who found reduced overconfidence after providing participants with outcome feedback on a few especially tough questions at the beginning of a test. Presumably, the initial feedback prompted respondents to anticipate more tricky questions in the remainder of the test than they would otherwise. Griffin and Tversky (1992) have shown, however, that item selection is insufficient to eliminate overconfidence effects, implying that it does not provide a complete explanation.

Another class of explanations stems from theoretical models that have been developed to explain the overconfidence phenomenon in terms of random error (Erev, Wallsten, & Budescu, 1994). Error models typically assume that a mental representation of degree of uncertainty (an "internal probability") exists and is perfectly calibrated with environmental relative frequencies ("objective probabilities"). However, reported confidence consists of the calibrated internal probability perturbed by a random error term. From within this framework, random error has clearly been shown as sufficient to produce an overconfidence effect (e.g., Juslin, Olsson, & Björkman, 1997). What the random-error models imply is that the presence of an overconfidence effect offers few constraints on the nature of the process. For example, the observation of an overconfidence effect does not necessarily entail the existence of cognitive processing biases.

The findings pertaining to item selection and random error, combined with the dearth of direct evidence for a systematic cognitive bias, have led many to question whether such systematic biases should be considered as contributors to overconfidence at all. For example, based on a fairly comprehensive analysis of existing data, Juslin et al. (2000) concluded that item selection and random error are jointly sufficient to explain observed overconfidence, and that the data do "not support the idea of a cognitive overconfidence bias that is due to, for example, confirmatory search of memory" (p. 393). In sum, overconfidence is an important and complicated phenomenon. There are several potential contributors to overconfidence, and at this point, it is not at all clear whether systematic processing biases ought to be considered among them.

The primary aim of the current study is to test the proposal that there is a systematic tendency to fixate on one option when assessing confidence in general knowledge, as well as the implication that such "option fixation" contributes to observed overconfidence. Option fixation is implicated by a process model of choice and confidence judgment in general knowledge tasks. The second aim of this research is to explore an approach for reducing overconfidence that capitalizes on this hypothesized option fixation. The remainder of this article proceeds as follows. First we will describe the Assess-Search-Construct (ASC) model of choice and confidence in general knowledge. Next, we describe key experimental manipulations, along with accounts by ASC and several alternative models. We then test the accounts in two experiments, and discuss the implications of our findings.

The Assess-Search-Construct (ASC) model

We next turn to describing a model that proposes that, when confronted with a general knowledge question, respondents first Assess the familiarity of the alternatives, then Search memory for the purposes of Constructing an explanation about why the high-familiarity option is true. It is thus referred to as the Assess-Search-Construct (ASC) model.

According to ASC, a person first encodes a test question and makes a rapid familiarity assessment of the presented options within the context of the question stem. Intuitively, familiarity can be thought of as the relative ease with which a presented object (such as a word or picture) is immediately perceived (Benjamin & Bjork, 1996). The feeling of familiarity or perceptual fluency is thus experienced extremely quickly in response to stimulus presentation, and precedes any influence of conscious recollection (e.g., Jacoby, Woloshyn, & Kelley, 2004). This rapid perceptual assessment yields a familiarity signal for each option, given the key words in the question stem. The process tentatively assumes that the high-familiarity option is true, and subjects it to further analysis under that assumption.[1] The overall goal of the analysis is to construct an explanation that describes why the high-familiarity option is true. However, memory search is required in order to obtain the basic facts needed to build the explanation. The process selects the high-familiarity option and key words from the question stem to use as retrieval cues to guide search, but tends not to include the low-familiarity option. In addition to the factual information retrieved, the memory search process produces information about the amount and ease of retrieval (Schwarz & Vaughn, 2002). Such information influences confidence by indicating overall level of knowledge in the domain (Sieck & Yates, 2002, 2003).

As facts are accumulated, the person uses them to construct an explanation for why the high-familiarity option is true. He or she then evaluates this explanation for overall acceptability, which depends on factors such as completeness and coherence (Pennington & Hastie, 1993; Thagard, 2000). Assuming that a minimal threshold for acceptability is achieved, the respondent chooses the high-familiarity option as the correct one.[2] Confidence in that choice is determined by the acceptability of the explanation for why the high-familiarity option is true and the experienced ease of retrieval, as described above.

For example, at one extreme, no information will be successfully retrieved, the person will be unable to construct any explanation and will feel that his or her choice is no more than a guess. At the other extreme, a virtual flood of information on the topic will be accessed, the person will construct an intricate explanation and will feel extremely confident in his or her choice.

In general, ASC predicts poor calibration because confidence and choice are determined by distinct and imperfectly correlated information sources. Choice is driven by familiarity, whereas confidence depends on memory retrieval success and explanation acceptability. Hence, perfect calibration is not expected to be easily or generally attainable. However, specific ASC mechanisms do imply that a potentially sizeable portion of overconfidence results from a systematic processing bias that is ameliorable. Specifically, a key feature of ASC is that a single option is tentatively selected as true based on the familiarity assessment, and then its truth is evaluated via explanation. As long as no reason for rejecting the high-familiarity option surfaces, the low-familiarity option receives scant consideration in the evaluation process.

We call this tendency to evaluate only the high-familiarity option, "option fixation." Option fixation is related to a more general strategy to evaluate options in serial fashion. That is, instead of analyzing options by direct comparison to determine optimality, the person assesses each independently and sequentially in terms of sufficiency. Serial evaluation is perhaps the only practical approach to option evaluation in naturalistic contexts where the space of possible options is quite large and ill-defined, as suggested in Simon's work on satisficing (Simon, 1957). Furthermore, in cases where options are being constructed rather than merely selected, fixing on a particular option allows for adjustment and revision as it is being evaluated (Klein, 1998). Even in general knowledge tasks, singling out the high-familiarity option for evaluation reduces mental effort at little expense to choice accuracy. However, option fixation does contribute to overconfidence. The idea is that the explanation for the high-familiarity option tends to be quite acceptable, so the person has high confidence in its truth.

The high acceptability of intuitive explanations follows from a general tendency for people to believe their explanations to be more detailed and complete than they actually are (Keil, 2003; Rozenblit & Keil, 2002). Were the person to generate one, an explanation for the low-familiarity option would tend to be sufficiently acceptable to reduce the person's overall confidence. This is a consequence of the fact that the more acceptable the explanation is for the high-familiarity option, the weaker the low-familiarity option can be and still reduce overall confidence (cf. McKenzie, 1997). For example, suppose respondents find their explanation for the truth of the high-familiarity option to be completely acceptable, and so they are maximally certain. Then, if they construct even a slightly acceptable explanation for the low-familiarity option, their overall confidence in the initial option would have to be reduced.

[1] Familiarity in ASC is assumed to be a continuous construct. Thus, the probability that two options will be equally familiar is zero.

[2] In standard general knowledge experiments using undergraduates, we assume that the high-familiarity option is virtually always selected as the explicit choice, because nearly any explanation is likely to be deemed minimally acceptable in these settings.

Manipulations and predictions

In this section, we describe the three experimental manipulations we employed in the current study, and discuss ASC's predictions for each. These manipulations were assessment independence, choice prompt inclusion, and foil plausibility.

Assessment independence

Confidence judgments are typically elicited by first requesting that respondents choose the answers they believe are most likely to be correct, and then provide 50–100% probabilities that their choices are in fact correct. Consider the following alternative procedure: The respondent first states which of the two options is correct. Then, the question stem is presented again, but with only one option. The respondent is now asked to imagine that this presented option is the only one they have seen, and report a 0–100% probability that the displayed option is true. Finally, the respondent is again presented with the question stem, along with the other option, and asked to report a 0–100% probability that this option is true. That is, the respondent assesses the probability that each option is true, as if it were presented alone in a true/false test. We thus refer to this latter method as the "Independent Assessment" procedure. The former method is a "Dependent Assessment" procedure, since it introduces a demand to assume that only one of the two options is true when reporting the probability that the chosen option is correct.

According to ASC, overconfidence stems in part from a person's tendency to fixate on the high-familiarity option and only construct an explanation for why it is true. We therefore predict a reduction in overconfidence for independent assessment procedures, which require participants to render probabilities that each of the options is true. The idea is that, when considering each option independently, the respondent will be more likely to select the respective option as a retrieval cue in the memory search, construct an explanation for why that option is true, and then use the acceptability of that explanation as a basis for judgment. That is, it reduces the tendency to fixate on only the high-familiarity option, leading to a reduction in overconfidence.

Choice prompt inclusion

A prompt to explicitly choose between response options is included in both the dependent and independent assessment methods we just described. Another common procedure does not include a choice prompt, but instead asks participants to provide a probability ranging from 0% to 100% that one of the presented options is correct. Several investigators have studied effects of including a choice prompt, and their primary finding is that requiring an explicit choice leads to a reduction in overconfidence relative to when no choice is required (e.g., Ronis & Yates, 1987; Sieck & Yates, 2001; Sniezek, Paese, & Switzer, 1990).

Sieck (2003) reviewed earlier work on the calibration differences found under these choice inclusion/exclusion procedures and provided empirical support for a retrieval account of the phenomenon in a category learning task. He argued that the multiple prompts for judgments would influence a shift in the retrieval cues used, so that information not accounted for at the explicit choice stage would be brought to bear on the confidence judgment, leading to reductions in overconfidence. Essentially the same argument holds for ASC: The second prompt encourages a shift in perspective to meet the demand to render new information, and that leads to a change in the retrieval cues used (Anderson & Pichert, 1978). Respondents are therefore more likely to select the low-familiarity option as a retrieval cue at this point, thereby reducing the tendency to fixate on the high-familiarity option. Note, however, that in the independent assessment procedure, respondents are already influenced to construct an explanation supporting the low-familiarity option. Hence, ASC also makes a novel prediction that independent assessment will moderate the choice prompt inclusion effect, in addition to offering an explanation for known choice prompt inclusion effects. That is, ASC predicts that the choice inclusion effect will be reduced under independent assessment conditions.

Foil plausibility

The incorrect option in a two-alternative task is typically referred to as the foil. Thinking in terms of a true/false test for each option, the foil can appear more or less plausibly "true." And, by experimentally manipulating the plausibility of the foils we can determine the extent to which respondents are sensitive to foil plausibility. Furthermore, ASC makes predictions regarding the interaction between independent assessment procedures and foil plausibility. According to ASC, requiring respondents to independently assess each option will yield heightened sensitivity to the plausibility of the foil. That is, under standard elicitation conditions, ASC predicts that confidence will be relatively insensitive to whether correct options are paired with plausible or implausible foils. The driving force behind this insensitivity to foil plausibility is the proposal that respondents fixate on the high-familiarity option. If each alternative is instead assessed independently and then combined to form the final confidence estimate, the low-familiarity option will have more of an influence on the final judgment. In cases where the foil is highly plausible, independent assessment will reduce the final confidence judgment. On the other hand, final confidence from the independent assessment procedure will stay the same or even increase in cases where the foil is implausible. Hence, the increase in confidence from items with relatively plausible foils to items with relatively implausible foils should be larger with independent assessment methods than with standard, dependent assessment procedures.

Alternative models

The Minerva-Decision Making (MDM) model of confidence calibration presents a competing account (Dougherty, 2001). In the context of general knowledge tasks, MDM assumes that a familiarity assessment is made for each option based on traces in memory. The respondent chooses the alternative with the highest familiarity, and reports a confidence judgment that is the ratio of familiarity for the chosen option to the sum of familiarities for both options (i.e., normalization of the familiarity values). MDM predicts no difference in overconfidence between standard, dependent assessment procedures and independent assessment procedures because the familiarity of each option is already fully incorporated into the confidence judgment. For the same reason, the effects of foil plausibility should be entirely accounted for in the standard procedures. Hence, MDM does not predict an interaction between independent assessment and foil plausibility. Finally, because choice and confidence are based on the same information, MDM does not account for choice prompt inclusion effects. Like MDM, the theory of Probabilistic Mental Models (PMM; Gigerenzer et al., 1991) predicts no differences for the key manipulations described above. If a person does not explicitly know the answer to a question, PMM assumes that the person generates a "reference class" of objects that are similar to the objects in the current question. The reference class yields cues that can be used to answer the current question, and the most valid cue is used to answer the question and give a probability judgment. Because the cue is determined at the choice stage and yields complementary probabilities for each alternative, PMM predicts no reduction in overconfidence or increase in foil sensitivity for the independent assessment procedures. It also does not offer an account of choice inclusion effects.

As mentioned previously, Koriat et al. (1980) proposed that respondents rely more heavily on reasons that are consistent with a chosen answer than on reasons contradicting it. Yates and colleagues described in more precise terms an "argument recruitment model" that incorporated this kind of a confirmation bias mechanism (Lee et al., 1995; Yates et al., 1992). According to the argument recruitment model, respondents try to recruit pro and con reasons for each option, but with a bias towards the first arguments that happen to be brought to mind. The respondent then evaluates the strengths of the recruited arguments. The respondent chooses the option that is indicated by the balance of the arguments, and reports confidence that depends on how heavily the arguments favor that option. The argument recruitment model and ASC are similar in that both give priority to one option as an "implicit favorite." However, the specific mechanisms by which the favorite is treated are distinct (McKenzie, 1997). The ASC mechanism is fixation on the high-familiarity option, to the neglect of the low-familiarity option. An independent assessment procedure draws attention to each of the options, and thus to the neglected option. The argument recruitment account relies on a confirmation bias mechanism. It maintains that both options are attended to, but that recruited reasons are biased towards the favored option and against the unfavored option. There is no reason to expect that mere attention to the unfavored option will eliminate the bias against it. There is also no reason to expect that independent assessment would increase sensitivity to the plausibility of the foil. With respect to choice prompt inclusion, the argument recruitment model maintains that recruitment occurs prior to rendering choice and confidence judgments. Hence, it has the same difficulty explaining choice prompt inclusion effects as do MDM and PMM. If the recruitment assumption were relaxed to allow for additional reasons to be recruited between choice and confidence, then the extension of the biased process would predict an increase in overconfidence because more arguments for the favored option and against the unfavored option would be generated.

Another important class of models to consider is the error models (Erev et al., 1994; Juslin et al., 1997). The error models typically specify that overconfidence arises as the result of unbiased internal confidence being perturbed by random error. While these models fit some data well, it is difficult to discern their predicted effects on overconfidence under different elicitation methods. Any observed effects might be explained post hoc by assuming additional sources of error, or that particular methods changed the level of random error. However, in and of themselves, they offer little or no *a priori* theoretical guidance on when and when not to expect changes in random error.

In summary, ASC implies that miscalibration occurs because choice and confidence are based on distinct information sources. Also, one mechanism that exacerbates overconfidence is option fixation: only the high-familiarity option is typically selected as a retrieval cue, to the neglect of the low-familiarity option. If this is the case, then confidence elicitation conditions requiring respondents to assess each option independently should result in decreased overconfidence. These are predictions that none of the other models under consideration accommodate. The current experiments are both a test of the ASC proposal of option fixation, and a demonstration of novel elicitation methods that can potentially reduce overconfidence.

Experiment 1

We conducted Experiment 1 to test the effects of assessment independence, choice prompt inclusion, and foil plausibility on overconfidence in a knowledge-intensive task. Specifically, we assessed overconfidence in a test of financial knowledge. The test covers general topics concerning financial management such as bank accounts, insurance, and interest. Because knowledge of financial management is important for one's well-being, lack of understanding can lead to poor financial decisions and real monetary losses (Kapoor, Dlabay, & Hughes, 1991). Furthermore, overconfidence in one's existing financial knowledge removes the inclination to attempt to improve that understanding (Renner & Renner, 2001). Hence, this task possesses a kind of ecological validity that is at least as important as other concerns that have been raised regarding considerations of representative design in overconfidence studies (cf. Keren, 1997).

The manipulations followed a $2 \times 2 \times 2$ design. For the independence factor, participants considered response options either separately or jointly (additional details appear below). For the choice-prompt inclusion factor, participants either made explicit choices between the options and then reported their probabilities, or they reported probabilities without receiving any explicit prompt to choose. Finally, for each respondent, half the questions were selected to have more plausible foils, and the other half had less plausible foils. Based on ASC, it was hypothesized that independently assessing the options should result in greater sensitivity to foil plausibility and yield reduced overconfidence. Furthermore, independent assessment should eliminate the choice prompt inclusion effect described above.

Method

Participants

Study participants were 184 undergraduate students enrolled in an introductory psychology course at the Ohio State University. Experimental participation was part of their course requirement.

Materials

Participants completed a financial knowledge quiz administered in Visual Basic for Applications (VBA). The quiz derived from one that had previously been administered to 4024 high school seniors by the Jump$tart Coalition for Personal Financial Literacy (2002). The current version of the quiz remained largely identical to the original, the only notable difference being that the current version of the quiz used two alternatives per item instead of four. Also, some minor grammatical modifications were made to the questions to correspond to the reduction in choice alternatives.

The alternatives used for each test item in the current version were always (1) the correct alternative and (2) a plausible or an implausible foil. We determined foil plausibility from the proportion of high school seniors who selected each foil when the quiz was originally administered. For each test item, we denoted the foil with the highest proportion of responses a more plausible foil, and the foil with the second-highest proportion of responses a less plausible foil.[3] As an example, consider the following item from the original version of the quiz.

If you have caused an accident, which type of automobile insurance would cover damage to your own car?

3.1%	(a) term
10.5%	(b) comprehensive
*51.3%	(c) collision
35.0%	(d) liability

In this example, the asterisk marks the correct alternative, and the percentages refer to the percentage of seniors who selected the corresponding option in the original study. Alternative (b) was the implausible foil for this question, and alternative (d) was the plausible foil. In our version of the quiz, we presented alternative (c) with either alternative (b) or (d).

Design

The experiment was a 2 (assessment independence) $\times$ 2 (choice prompt inclusion) $\times$ 2 (foil plausibility) mixed design. Assessment independence and choice prompt inclusion were between-subjects factors, and foil plausibility was a within-subjects factor. The choice prompt inclusion and assessment independence factors provided four conditions under which participants could provide confidence judgments. In the "choice, dependent" condition, participants were instructed to choose an option and then give a 50–100% confidence judgment indicating how sure they were that their choice was correct. This is the standard C50 procedure employed in most confidence studies. In the "no choice, dependent" condition, participants were not required to choose an option. These participants instead gave separate confidence judgments that each option was correct such that the two judgments summed to 100%. In the "choice, independent" condition, participants first chose an option and then gave separate, independent probabilities that each option was true. After participants made a choice, only one option at a time was presented on the

[3] The foils with the lowest frequency of responding were too easy to be included, as virtually every question with one of those foils was answered correctly in pretesting.

screen. Participants were instructed to make independent probability judgments by first imagining that the displayed option was the only one that they had seen and then reporting a 0–100% judgment that the displayed option was true. This is essentially similar to a confidence judgment in a true/false test, as opposed to a confidence judgment in a multiple choice test. Because of this, the two judgments were not required to sum to 100%. In the "no choice, independent" condition, participants gave independent probability judgments that each option was true without choosing between the options. In this condition, the test question was first presented along with option (a), and participants gave a probability judgment for option (a) without having seen option (b). Option (b) was then presented after option (a) had been erased from the monitor, and participants gave an independent probability judgment that option (b) was true.

In all four of these conditions, participants were presented with the same 30 question stems in a random order. While the question stems remained the same, the presented options differed between participants depending on foil plausibility. For each participant, foil plausibility on each item was randomly selected, under the constraint that half would be easier and the other half more difficult.

Procedure

The experiment was conducted on a computer. Participants were told that they would answer a number of questions dealing with personal finance, and they were then given instructions specific to their assigned condition. Among other things, these condition-specific instructions asked participants to adhere to different conventions while eliciting probability judgments. In the "choice, dependent" condition, where they made a probability judgment that the chosen option was correct, participants received the following three conventions:

(a) A probability of 50% should mean that your chosen answer is just as likely to be correct as incorrect.
(b) A probability of 100% should mean that you are absolutely certain that your chosen answer is correct.
(c) Increasing probabilities between 50% and 100% should correspond to increasing degrees of certainty that you have picked the correct option.

In the "no choice, dependent" condition, they made probability judgments that each option was correct, participants received the following six conventions:

(a) A probability of 50% should mean that you think that options (A) and (B) are equally likely to be correct.
(b) A probability of 100% should mean that you are absolutely sure that the option being judged is correct.
(c) A probability of 0% should mean that you are absolutely sure that the option being judged is not correct.
(d) Increasing probabilities between 50% and 100% should reflect increasing degrees of certainty that the option being judged is correct.
(e) Decreasing probabilities between 50% and 0% should reflect increasing degrees of certainty that the option being judged is not correct.
(f) Since you know that one of either (A) or (B) is correct, the two probability judgments should add up to 100%.

In the "choice, independent" and "no choice, independent" conditions, where they made independent probability judgments that each option was true, participants received the following four conventions:

(a) A probability of 100% should mean that you are absolutely sure that the option is true.
(b) A probability of 50% should mean that you think that the option is just as likely to be true as it is to be false.
(c) A probability of 0% should mean that you are absolutely sure that the option is false.
(d) Increasing probabilities between 0% and 100% should reflect increasing degrees of certainty that the option is true.

Participants in these two conditions were also given an explicit reminder that their confidence judgments for the separate options need not sum to 100%.

After reading the instructions at their own pace, participants immediately started the test. For each test item, the participant: (1) was presented with a financial knowledge question and one or two options, depending on condition; (2) chose an option if their condition required it; and (3) elicited probability judgments via the method required by their condition, as described above. All choices and confidence judgments were entered by the participant using the computer keyboard.

Results and discussion

We present the proportions correct and mean confidence after some explanation of our analytical methods. To compute proportion correct in the "no choice, dependent" and "no choice, independent" conditions, we obtained implicit choices as in past research (e.g., Ronis & Yates, 1987). In particular, we considered the option rated with the higher probability to be the participant's choice. When the probabilities for the two options were equal, we chose one at random. Probability judgments in the "choice, independent" and "no choice, independent" conditions were not required to sum to 100% because the probabilities were that each option was true, independently. That is, the probabilities were not conditioned on exactly one of the two options being true. In these groups, we transformed the probability associated with the chosen option from an independent probability that the option was true to a dependent probability that the option was correct by normalization (McKenzie, 1997). Normalization introduces the condition of exactly one of the two options being correct by setting the sample space to be the total of the independent probabilities. To obtain a participant's conditional confidence that option A is correct, we take C(A)/[C(A) + C(B)], where C(X) is the reported confidence that option X is true. Note that normalization itself does not always improve confidence calibration. Confidence can increase, decrease, or remain unchanged with normalization, depending on whether the confidence judgments sum to less than 100, greater than 100, or 100, respectively, for a given item. This means that, for a given calibration level, bias can increase, decrease, or stay the same as a result of normalization.

Our results are based on 3-factor ANOVA's involving one within- (foil plausibility) and two between- (choice prompt inclusion, assessment independence) subjects factors. All effects are significant at an alpha level of .05 unless noted otherwise.

Mean confidence

Fig. 1 shows mean confidence as a function of assessment independence, choice prompt inclusion and foil plausibility. Mean confidence is collapsed over foil plausibility in panel (a) and over choice prompt inclusion in panel (b). There was a main effect of assessment independence ($F(1,180) = 33.707$, $MSE = 0.013$): mean confidence for dependent judgments was greater than mean confidence for independent judgments. There was also a statistically significant interaction between choice prompt inclusion and assessment independence ($F(1,180) = 5.033$, $MSE = 0.013$): in the dependent condition, average confidence decreased from the no choice to the choice condition, whereas in the independent condition average confidence showed little change.

As predicted by ASC, independent assessment eliminated the choice prompt inclusion effect. There was a main effect of foil plausibility on mean confidence ($F(1,180) = 36.576$, $MSE = 0.001$): mean confidence was lower for questions with plausible foils than for questions with implausible foils. There was a significant interaction between foil plausibility and assessment independence ($F(1,180) = 5.619$, $MSE = 0.001$): average confidence increased more in the independent condition than in the dependent condition as a result of foil implausibility. That is, confidence was more sensitive to foil plausibility in the independent condition. The main

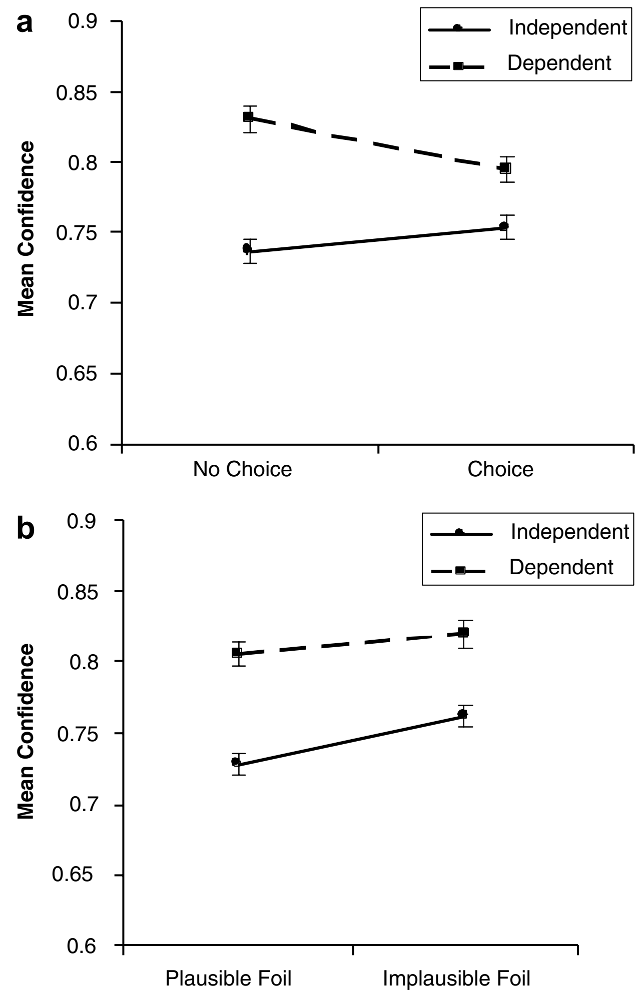

Fig. 1. Mean confidence by choice, dependence, and foil plausibility, Experiment 1.

effect of choice prompt inclusion and effects of other interactions on mean confidence were not significant.

An alternative way to see the effects of foil plausibility on confidence is through effect sizes. Effect sizes have previously been used to examine changes in probability accuracy (Price, 1998). In our experiments, for a given question, conditions that force participants to consider an explanation for each option should increase sensitivity to foil plausibility. Effect sizes measure the magnitudes of change in confidence and in proportion correct from implausible foils to plausible foils. As a result, they are a direct and intuitive measure of sensitivity to foil plausibility. To compute the foil plausibility effect sizes, we subtracted each participant's mean confidence (or proportion correct) for questions with plausible foils from his or her mean confidence for questions with implausible foils. We then divided this difference by the standard deviation of the differences in mean confidence from questions with implausible foils to questions with plausible foils (in the same fashion as a paired-samples *t*-test). Table 1 shows the results for mean confidence,

Table 1
Foil plausibility effect sizes for confidence, proportion correct, and bias by condition, Experiment 1

Confidence dependence	Choice inclusion	
	Choice	No choice
Independent		
Mean confidence	0.673	0.650
Proportion correct	0.659	0.704
Bias	0.448	0.517
Dependent		
Mean confidence	0.154	0.340
Proportion correct	0.677	0.329
Bias	0.577	0.208

proportion correct, and bias by condition (we discuss the proportion correct and bias results below). As shown, mean confidence changes most in independent elicitation conditions and least in the standard "choice, dependent" condition. That is, independent elicitation conditions increase subjects' sensitivity to foil plausibility.

Finally, we conducted an additional analysis in the independent assessments conditions to clarify the mechanisms underlying the reduction in confidence. According to ASC, people should be highly confident in the truth of the high-familiarity option, but they should also be somewhat confident in the truth of the other option (when brought under consideration). In order to test this idea, we examined the distribution of the larger of the two rendered probabilities in the independent assessments condition to determine how confident people are in the truth of the implicitly favored option. Although the full 0–100 range of the scale was used, these probabilities tended to be quite high. In particular, 60.6% of the probabilities in the "choice, independent" condition and 57.7% of the probabilities in the "no choice, independent" condition met or exceeded a judgment of 75%. Also, for all of the judgments that did exceed 75% in these conditions, the unfavored option was found to be reasonably plausible when explicitly considered. Specifically, 22.8% of these probabilities in the "choice, independent" condition and 43.0% of the probabilities in the "no choice, independent" condition exceeded a judgment of 25%, tending to reduce overall confidence.

Proportion correct

Fig. 2 presents mean proportion correct by choice prompt inclusion, assessment independence, and foil plausibility. Mean proportion correct is collapsed over foil plausibility in panel (a) and over choice prompt inclusion in panel (b). There was an effect of choice prompt inclusion on proportion correct ($F(1,180) = 4.071$, $MSE = 0.024$), where proportion correct was higher when participants explicitly chose an option. This result arises sometimes, but not consistently, in previous research on choice prompt inclusion effects (Sieck & Yates, 2001; Sniezek et al., 1990). As

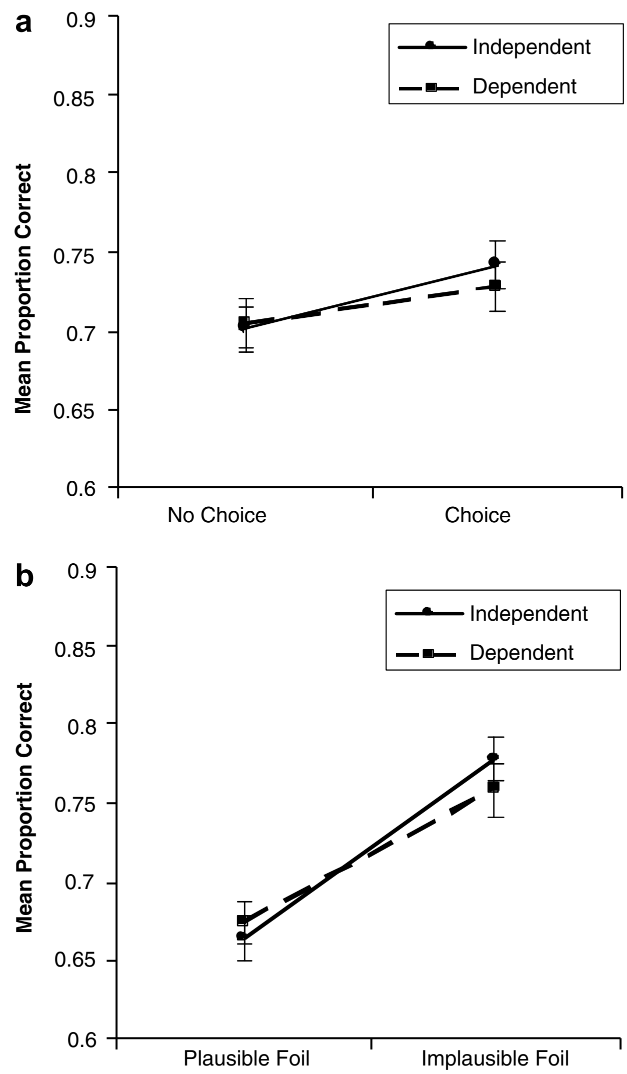

Fig. 2. Mean proportion correct by choice, dependence, and foil plausibility, Experiment 1.

Fig. 2b shows, there was also an effect of foil plausibility on proportion correct ($F(1,180) = 64.152$, $MSE = 0.014$). Proportion correct was higher for the easy questions than for the hard questions, showing that our foil selection methods worked as expected. Effects of assessment independence and of interactions were not significant.

Table 1 presents foil plausibility effect sizes for proportion correct. As for mean confidence, the effect sizes here are standardized differences in proportion correct for questions with implausible foils versus questions with plausible foils. Proportion correct changed by about the same magnitude in all conditions except the "no choice, dependent" condition. The average proportion correct for easy questions in the "no choice, dependent" condition was lower than the average proportion correct for easy questions in the other conditions by about 4 percentage points. An ANOVA shows that foil plausibility had a smaller effect on proportion correct in the "no choice, dependent" condition than in the other 3 conditions ($F(1,182) = 4.004$, $MSE = 0.014$). This difference, which stands out in Table 1, is driving the previously noted main effect of choice prompt inclusion on proportion correct.

Bias (over-/underconfidence)

Fig. 3 presents bias as a function of choice prompt inclusion, assessment independence, and foil plausibility. Bias is collapsed over foil plausibility in panel (a) and over choice prompt inclusion in panel (b). To compute the most commonly used overconfidence measure (bias), we subtract proportion correct from mean confidence. There are main effects of choice prompt inclusion ($F(1,180) = 7.415$, $MSE = 0.022$) and of assessment independence ($F(1,180) = 22.183$, $MSE = 0.022$) on bias, but the interaction between choice prompt inclusion and assessment independence is not significant. As Fig. 3a shows, bias tends to be lower both under independent confidence elicitation and under choice inclusion conditions. The latter effect is established (e.g., Ronis & Yates, 1987), but the former is a novel finding.

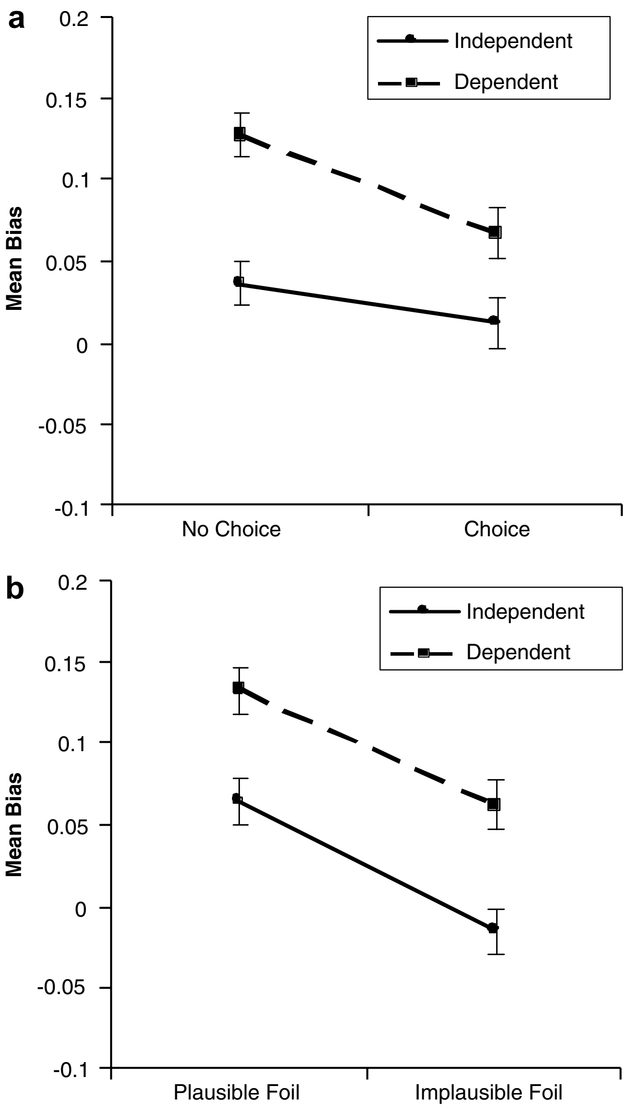

Fig. 3. Mean bias by choice, dependence, and foil plausibility, Experiment 1.

Fig. 3b shows a different view of these data, in which bias is presented by foil plausibility and assessment independence. Bias is greater for questions with plausible foils than for questions with implausible foils, indicating a hard–easy effect for foil plausibility ($F(1,180) = 35.649$, $MSE = 0.015$). The hard–easy effect refers to an empirical phenomenon that overconfidence is often observed when proportion correct is lower than 75%, and that underconfidence is observed when proportion correct is higher than 75–85% (e.g., Suantak, Bolger, & Ferrell, 1996). The effect has been found by comparing more knowledgeable to less knowledgeable people, and by comparing across tasks that differ in difficulty (Lichtenstein & Fischhoff, 1977). To our knowledge, the phenomenon has not been investigated by precisely manipulating foils (or targets) prior to the current study, though such manipulations are important for improving our depth of understanding such hard–easy effects. No interactions were significant; in particular, the hard–easy effect is not reduced with independent elicitation even though confidence was more sensitive to changes in proportion correct with independent elicitation.

Because bias measures the difference between confidence and proportion correct, we examined the hard–easy effect via inspection of mean confidence and proportion correct to determine why the hard–easy effect was not reduced. Changes in confidence due to independent elicitation were offset by changes in proportion correct due to choice inclusion. Independent elicitation conditions and choice conditions reduced bias in two different ways: independent elicitation decreased mean confidence, and choice increased proportion correct. Because mean confidence is usually larger than proportion correct (overconfidence), either or both of these factors can lead to a decrease in bias, which can in turn alter the size of the hard–easy effect. Table 1 presents effect sizes for differences in bias from hard to easy items. These effect sizes are a measure of the magnitude of the hard–easy effect for each condition. Compared to the standard "choice, dependent" condition, independent conditions ("choice, independent" and "no choice, independent") showed smaller hard–easy effects. The effect size for the "no choice, dependent" condition is smaller than all others, suggesting that the hard–easy effect may be reduced under dependent elicitation conditions. Recall, however, that proportion correct in the "no choice, dependent" condition was lower than proportion correct in all other conditions; the hard–easy effect is small for the "no choice, dependent" condition because the change in proportion correct was small compared to the other conditions. The hard–easy effect is smaller for the independent conditions because the change in mean confidence is large compared to other conditions. Combining mean confidence and proportion correct into a single bias measure glosses over these changes and results in non-significance.

Summary

Assessment independence and choice inclusion manipulations affected confidence levels and bias. Independent confidence elicitation conditions reduced participants' confidence judgments such that they were less overconfident. Furthermore, confidence decreased only slightly with foil plausibility, with the greatest decrease occurring in the independent assessment conditions. For the more standard, dependent confidence assessment procedures, there was also a choice prompt inclusion effect on confidence, such that confidence was lower when an explicit choice was made. This is consistent with past work (e.g., Sniezek et al., 1990). The most interesting choice prompt inclusion finding is that the choice inclusion effect was eliminated in the independent assessment elicitation conditions. These results are in accord with the ASC proposal that respondents tend to fixate on one option, consistently using it as a retrieval cue in order to generate an explanation for why that option is true. Because independent elicitation conditions encouraged participants to search memory and construct an explanation for each option independently, those conditions yielded improved calibration.

Experiment 2

The results of Experiment 1 support the idea that independent assessment encourages participants to develop an explanation for why each option is true, thereby reducing overconfidence and increasing sensitivity to foil plausibility. The independent assessment conditions used in Experiment 1 might be strengthened, however. Although participants were instructed to give confidence judgments separately for each option, they were not directly required to construct an explanation for why each was true. Experiment 2 replicates and extends the independent assessment findings of Experiment 1. Specifically, we sought to increase the likelihood that participants would use each option as a retrieval cue and search memory for facts to construct an explanation for why the option is true. As we describe below, we accomplished this by requiring participants to write out an explanation for why each option was true. Based on ASC, we expected that explicitly explaining why each option is true during an independent elicitation process would result in less overconfidence and more sensitivity to foil plausibility.

Method

Participants

Study participants were 141 undergraduate students enrolled in an introductory psychology course at the Ohio State University. Experimental participation was part of their course requirement.

Materials

The materials were exactly the same as in Experiment 1.

Design

The experiment was a 3 (elicitation method) $\times$ 2 (foil plausibility) mixed design. Elicitation method was a between-subjects factor, and foil plausibility was a within-subjects factor. Two of the three elicitation methods in Experiment 2 were exactly the same as the choice conditions in Experiment 1: these were the "choice, dependent" and "choice, independent" conditions. Recall that in the "choice, dependent" condition, participants first chose an option and then gave a 50–100% confidence judgment indicating how sure they were that their choice was correct. In the "choice, independent" condition, participants first chose an option and then gave separate, independent confidence judgments that each option was true. The new condition in Experiment 2 was the "explain, independent" condition. For each test item, participants in this condition first chose an option and then considered the two options separately and independently. Following a choice between the two options, option (a) was presented by itself on the monitor. Participants were instructed to assume that the displayed option was true and to type an explanation in response to the question, "Why is this option true?" Following the explanation, participants reported a 0–100% probability judgment indicating their perceived likelihood that the displayed option was in fact true. Option (a) was then cleared from the monitor, and the procedure was repeated for option (b).

Procedure

Experiment 2 was conducted on the same computers as Experiment 1, and the instructions for the "choice, dependent" and "choice, independent" conditions were the same as described in Experiment 1. Participants in the "explain, independent" condition were told that they would answer questions dealing with personal finance, and they were also told that they would have to write explanations for each presented option in those questions. Following these instructions, participants received the same conventions regarding probability judgment elicitation as described for the other independent conditions in Experiment 1.

After reading the instructions, participants in the "choice, dependent" and "choice, independent" conditions proceeded as in Experiment 1. For each test item, participants in the "explain, independent" condition: (1) were presented with a financial knowledge question and two options; (2) chose an option; (3) typed explanations for why option A was true, under the assumption that they knew A to be true; (4) reported a probability judgment that A was in fact true; (5) repeated steps (3) and (4) for option B. Participants answered the same 30 questions as in Experiment 1, half of which contained a plausible foil and half of which contained an implausible foil. Test items and options were randomized as they were in Experiment 1.

Results and discussion

The results we present below are based on 2-factor ANOVA's involving one within- (foil plausibility) and one between- (elicitation method) subjects factor. Probability judgments in the independent elicitation conditions ("choice, independent" and "explain, independent") were not required to sum to 100%, and they were normalized as we described in Experiment 1.

Mean confidence

Fig. 4 presents mean confidence by elicitation method and foil plausibility. There was a main effect of elicitation method ($F(2,138) = 21.137$, $MSE = 0.015$), where mean confidence in the "choice, dependent" elicitation method is larger than in the other two conditions. There was also a main effect of foil plausibility ($F(1,138) = 5.103$, $MSE = 0.002$), where mean confidence for items with implausible foils was larger than mean confidence for items with plausible foils. Fig. 4 shows that this effect is small, and it is most pronounced in the "explain, independent" condition. The interaction

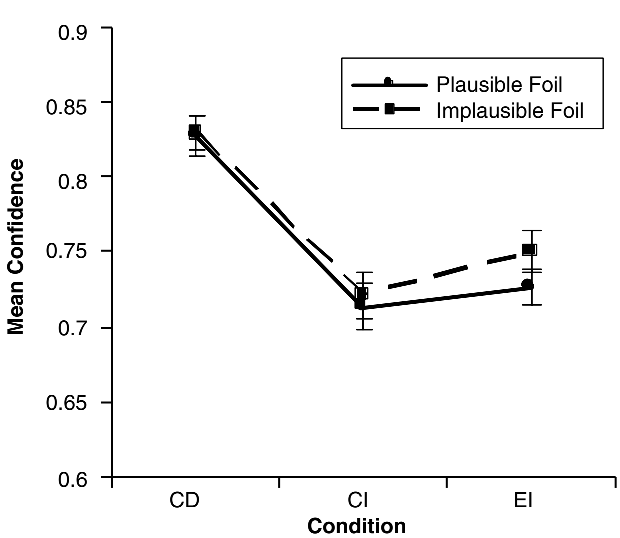

Fig. 4. Mean confidence by condition and foil plausibility, Experiment 2. Condition abbreviations: CD, choice, dependent; CI, choice, independent; EI, explain, independent.

between elicitation method and foil plausibility was not significant.

Foil plausibility effect sizes for confidence in the "choice, dependent", "choice, independent", and "explain, independent" elicitation methods are $d = 0.031$, $d = 0.126$, and $d = 0.404$, respectively. We computed these effect sizes as in Experiment 1: the difference between mean confidence for implausible and plausible foils divided by the standard deviation of the differences. These effect sizes are smaller than the ones we observed in Experiment 1. The effect size for the "choice, dependent" condition remains very small, and that for the "choice, independent" condition is noticeably smaller than it was in Experiment 1 (though the effect for "choice, independent" is still larger than for "choice, dependent"). The effect size for the "explain, independent" condition is largest, signifying the greatest change in confidence based on foil plausibility. It is not obvious why the effect sizes have decreased in comparison to Experiment 1, but the ordering of effects in the "choice, dependent" and "choice, independent" conditions remains the same. Based on Fig. 4 and the effect sizes, confidence is most sensitive to foil plausibility under the "explain, independent" condition.

As in Experiment 1, we conducted an additional analysis in the independent assessment conditions to test the proposals that: (1) people are highly confident in the truth of the high-familiarity option, and (2) people are also somewhat confident in the truth of the other option once it is explicitly brought under consideration. As in Experiment 1, the full 0–100 range of the scale was used, and yet these probabilities tended to be quite high. In particular, 50.1% of the probabilities in the "choice, independent" condition and 54.2% of the probabilities in the "explain, independent" condition met or exceeded a judgment of 75%. Also, for all of the judgments that did exceed 75% in these conditions, the low-probability option was found to be reasonably plausible when explicitly considered. Specifically, 27.5% of these probabilities in the "choice, independent" condition and 28.3% of the probabilities in the "no choice, independent" condition exceeded a judgment of 25%, and so tended to reduce overall confidence.

Proportion correct

Fig. 5 presents mean proportion correct as a function of elicitation method and foil plausibility. There was a main effect of elicitation method ($F(2,138) = 3.711$, $MSE = 0.020$), where proportion correct is lower in the "choice, independent" condition than in the other two conditions. Although this effect is apparent in both the ANOVA and the effect sizes (presented below), it is not immediately obvious why proportion correct in the "choice, independent" condition changed from Experiment 1. Had the Experiment 2 "choice, independent" proportion correct remained the same as the Experiment 1

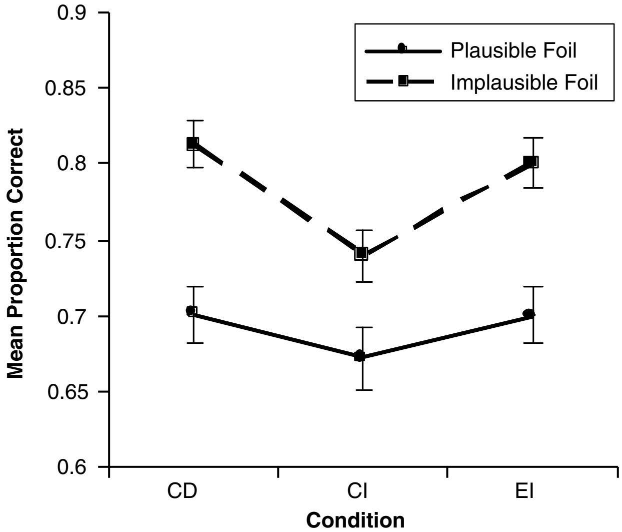

Fig. 5. Mean proportion correct by condition and foil plausibility, Experiment 2. Condition abbreviations: CD, choice, dependent; CI, choice, independent; EI, explain, independent.

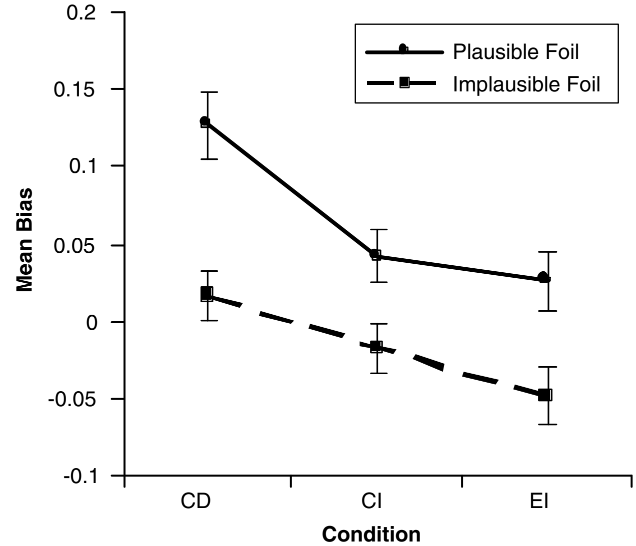

Fig. 6. Mean bias by condition and foil plausibility, Experiment 2. Condition abbreviations: CD, choice, dependent; CI, choice, independent; EI, explain, independent.

"choice, independent" proportion correct, the effect of elicitation method on proportion correct would have disappeared. Thus, we suspect that the effect can be attributed to random error. There was also an effect of foil plausibility on proportion correct ($F(1,138) = 50.948$, $MSE = 0.012$). As expected, proportion correct for items with plausible foils was lower than proportion correct for items with implausible foils. The interaction between elicitation method and foil plausibility was not significant.

Foil plausibility effect sizes for proportion correct in the "choice, dependent", "choice, independent", and "explain, independent" elicitation methods are $d = 0.676$, $d = 0.447$, and $d = 0.672$, respectively. On the whole, these effect sizes are of the same magnitude as those in Experiment 1. Going from items with implausible foils to items with plausible foils, proportion correct decreased by about the same magnitude in the "choice, dependent" and "explain, independent" conditions. In the "choice, independent" condition, proportion correct decreased by a somewhat smaller magnitude. As we mentioned above, proportion correct for items with implausible foils in the "choice, independent" condition dropped by about 5 percentage points as compared to proportion correct in the "choice, independent" condition for Experiment 1. It is this change that causes the effect size to be different from the other two conditions.

Bias

Fig. 6 presents bias (over-/underconfidence) as a function of elicitation method and foil plausibility. There is an effect of elicitation method on bias ($F(2,138) = 9.447$, $MSE = 0.018$), where bias is highest in the "choice, dependent" condition and decreases to a low in the "explain, independent" condition. There is also an effect of foil plausibility on bias ($F(1,138) = 37.128$, $MSE = 0.013$), where bias is higher for items with plausible foils than for items with implausible foils. The interaction between elicitation method and foil plausibility was not significant. Fig. 6 shows that participants in the "explain, independent" condition were underconfident for hard foils. This underconfidence is statistically reliable ($t(47) = -2.65$, $p < 0.01$). These underconfidence results are not inconsistent with ASC, however, in that overconfidence was further reduced under the "explain, independent" condition. ASC does not predict perfect calibration resulting from this or other elicitation procedures, because choice and confidence are presumed to rely on distinct and imperfectly correlated information sources. ASC only predicts a directional effect of reduced overconfidence, which we found.

Foil plausibility effect sizes for bias in the "choice, dependent", "choice, independent", and "explain, independent" elicitation methods are $d = 0.652$, $d = 0.413$, and $d = 0.461$, respectively. As noted in the Experiment 1 results, these effect sizes can serve as a measure of hard–easy effects due to foil plausibility. The effect size for the "choice, dependent" condition is appreciably higher than the effect sizes for the other two conditions, which are about equal. From this we may conclude that the hard–easy effect can be reduced under independent elicitation conditions, as compared to standard dependent elicitation conditions.

Summary

The goal of Experiment 2 was to examine whether overconfidence reductions could be made stronger by increasing the chances that the low-familiarity option would be used as a retrieval cue. We introduced a new condition where, for each option, participants explained why the option was true before rendering a confidence judgment that the option was, in fact, true. Average confidence was more sensitive to foil plausibility, and bias was lower (tending towards underconfidence) for this explanation-based assessment procedure than in other conditions.

General discussion

The findings from Experiments 1 and 2 provide initial support for the ASC proposal that overconfidence is due in part to option fixation. That is, ASC proposes that participants assess the familiarity of each option in the context of the question stem. They then repeatedly select the high-familiarity option as a retrieval cue in cycles of memory search, i.e., they fixate on the high-familiarity option. The memory search produces facts for building an explanation about why the high-familiarity option is true, and the resulting explanation tends to generate high confidence. In Experiment 1, we tested the hypotheses that choice prompt inclusion and assessment independence reduce overconfidence, and increase sensitivity to foil plausibility. In accord with the predictions, we found that independent assessment methods reduced both mean confidence and overconfidence. Furthermore, independent assessment increased confidence sensitivity to foil plausibility and eliminated standard choice inclusion effects. Finally, we found that confidence in the truth of the preferred option often exceeded 75%, suggesting that people were often highly confident in their intuitive explanations for the favored option.

Experiment 2 introduced a new independent assessment condition. In this "independent explanations" condition we asked participants to assume, for each option, that they had found the option to be true, and write an explanation stating why the option was true. The new condition brought about a further reduction in overconfidence and also increased sensitivity to foil plausibility. Finally, people again appeared to be highly confident in their favored explanations. The complete set of findings yields support for the ASC proposal of option fixation, as well as for the new method of independent assessment for probability elicitation. ASC provides a plausible account of the cognitive processes underlying choice and confidence, and independent assessment methods show initial promise to reliably reduce overconfidence. We describe theoretical and practical implications in more detail below.

Memory search and retrieval in ASC

ASC is related to dual process theories of reasoning and decision making, with familiarity as its core System 1 component, and explanation as its core System 2 component (e.g., Evans, 2003; Sloman, 1996). However, it is technically a triarchic theory, since it emphasizes the importance of memory search as a primary component that integrates the familiarity assessment and explanation construction systems. Memory search and retrieval is a core feature of ASC. It is the locus of option fixation. Furthermore, the experienced ease and amount of retrieval also have direct effects on confidence, in addition to the acceptability of the constructed explanation. For example, Sieck and Yates (2002, 2003) have shown that retrieval can increase or decrease confidence, depending on the amount retrieved and experienced retrieval difficulty. These retrieval effects thus provide a possible explanation for the ineffectiveness of listing reasons for and against each option as a method for reducing overconfidence. In particular, the reasons task demands the adoption of retrieval cues that disrupt the memory process, because knowledge is not stored in terms of reasons for and against each option. It appears to be especially difficult to retrieve reasons against the implicit favorite, which can lead to an increase in confidence that option is correct (Sanna, Schwarz, & Small, 2002; Yates et al., 1992). In contrast, confirmation bias mechanisms do not address retrieval, and so expect judgment improvement from reasons listing tasks.

The retrieval component of ASC also suggests explanations for some other interesting findings in the literature, such as dud alternative effects (Windschitl & Chambers, 2004). The dud alternative effect refers to the finding of greater confidence in a condition with four options, two of which are highly implausible, than in a two-option condition without the highly implausible options.[4] According to ASC, a respondent tends to fixate on the high-familiarity option, but the low-familiarity option is also expected to be occasionally used as a retrieval cue. Further, the more options that are included, the more likely it is that some of them will be adopted as retrieval cues, thereby leading to an increase in the overall amount of information retrieved. And, the additional successful retrieval of information about the domain of the question leads to an increase in confidence, over and above the confidence generated from the acceptability of the explanation for why the high-familiarity option is true. This retrieval-based explanation for the dud alternative effect needs to be directly tested.

---

[4] Windschitl & Chambers (2004) established the dud alternative effect for an arbitrary confidence scale, but not for probability scales such as are used here. They proposed that respondents' explicit consideration of normative probability would eliminate or even reverse the effect. For example, with the two extra alternatives, the equiprobable value drops from 50% to 25%, and the effect could easily wash if people tend to anchor on those values. This, of course, does not diminish the theoretical importance of the phenomenon.

Mechanisms underlying confidence

The critical issue guiding the present research is whether or not there is a cognitive basis for overconfidence. Early research suggested that there is, and showed initial promising evidence to that effect. More recently, however, the balance of evidence and argument has shifted away from the possibility of a cognitive basis. For example, several of the models considered in the introduction suggest that overconfidence is determined primarily or completely by non-cognitive sources. For example, the theory of Probabilistic Mental Models proposes that confidence will be quite well calibrated, so long as test items are randomly selected. Minerva-DM predicts little overconfidence as long as there is sufficient experience that is well encoded. The strongest conclusion coming from the work on random error models is that there is no justifiable basis for assuming a systematic cognitive tendency underlying overconfidence. Note that this conclusion stems from the ability of random error models to yield good quantitative fits to patterns of calibration data, with an emphasis on explaining simple effects like overconfidence and the hard–easy effect.

An alternative position is that overconfidence is a complicated phenomenon that is determined by multiple contributors, some of which are likely to be cognitively based. The accumulation of evidence for a number of overconfidence effects, such as the choice prompt inclusion effects and independent assessments effects examined here attest to the validity of this view. Random error models appear to provide little insight into such effects, suggesting that there is a need for models that attempt to describe psychological mechanisms. Hence, our research also highlights the need to reconsider cognitive process models for understanding the full range of extant overconfidence phenomena, as well as for establishing new effects. In particular, despite the predictive success of the random error models, little progress has been made in specifying leverage points for influencing confidence and its calibration (Sieck & Arkes, 2005). Yet there is enormous practical value to be had in promoting quality judgment in extremely difficult domains such as financial forecasting. Hence, more emphasis should be given to the exploration of models that specify psychological mechanisms for confidence.

Practical implications

For a given test question, the method of independent confidence elicitation was intended to encourage participants to generate an explanation for why each option was true and then judge each separately. Following the experiment, we combined these separate confidence judgments to create a fused judgment that is based on evidence pertaining to both options. Thus, our finding of reduced overconfidence under independent elicitation conditions was external to the participant; we combined participants' confidence judgments after the experiment. This differs from other methods of reducing overconfidence that are more internal to the participant. For example, Lichtenstein and Fischhoff (1980) were moderately successful at reducing overconfidence by giving participants elaborate instructions about confidence or large amounts of feedback about their calibration. Arkes et al. (1987) reduced participants' overconfidence by manipulating perceived difficulty of test items and by telling participants that they would have to justify their answers following the test.

Creation of independent elicitation conditions that lead to internal improvements in calibration is an important topic of future study. One possible approach would be to use "scaffolding" as described in the education literature (e.g., Palinscar, 1986). Scaffolding refers to the idea of providing strong support for performance that is then gradually withdrawn. For example, respondents could start out going through the full procedure as specified in the independent explanations method. That is, considering each option separately, writing why each is true, and reporting the independent judgments. They might additionally be asked to report the confidence in their choices. The fused values from the independent assessments could then be used as a form of direct confidence feedback to participants that does not require any external outcomes. Once they are able to directly report confidence in choice that matches the fused independent assessments, those independent judgments could be removed. Furthermore, after individuals have engaged in practicing the explanation exercise sufficiently to promote the mental habit of attempting to construct an explanation for why each option is true, then the explanation writing task could be removed as well. At this point, the "scaffold" has been completely removed, leaving only a trained judge performing a standard judgment task.

Scaffolding depends on a number of specific cognitive decision skills to be mastered to produce well calibrated judgments, and attempting to acquire them in a single step is unlikely to be particularly effective. Extensive studies are needed to determine the effectiveness of such calibration training interventions, as well as how they might be combined with other existing training ideas for judgment improvement. In the meantime, the independent assessment procedures described here, particularly the independent explanations method, show encouraging initial promise and should be tested in more applied research contexts.

References

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17*, 1–12.

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes, 39*, 133–144.

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

Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *International Journal of Forecasting, 20*, 29–39.

Dougherty, M. R. P. (2001). Integration of the ecological and error models of overconfidence using a multiple-trace memory model. *Journal of Experimental Psychology: General, 130*, 579–599.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101*, 519–527.

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences, 7*, 454–459.

Fischhoff, B., & MacGregor, D. (1982). Subjective confidence in forecasts. *Journal of Forecasting, 1*, 155–172.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Griffin, T., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (2004). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. In D. Balota & E. J. Marsh (Eds.), *Cognitive psychology: Key readings* (pp. 322–337). New York, NY: Psychology Press.

Jump$tart Coalition for Personal Financial Literacy. (2002). From bad to worse: Financial literacy drops further among 12th graders. Posted on the Jump$tart Website: http://www.jumpstart.org/upload/news.cfm?recordid = 99.

Juslin, P., Olsson, H., & Björkman, M. (1997). Brunswikian and Thurstonian origins of bias in probability assessment: On the interpretation of stochastic components of judgment. *Journal of Behavioral Decision Making, 10*, 189–209.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological Review, 107*(2), 384–396.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582–591.

Kapoor, J. R., Dlabay, L. R., & Hughes, R. J. (1991). *Personal finance*. Boston: Irwin.

Keil, F. C. (2003). Folkscience: coarse interpretations of a complex reality. *Trends in Cognitive Sciences, 7*, 368–373.

Keren, G. (1997). On the calibration of probability judgments: Some critical comments and alternative perspectives. *Journal of Behavioral Decision Making, 10*, 269–278.

Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge: MIT Press.

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

Lee, J.-W., Yates, J. F., Shinotsuka, H., Singh, R., Onglatco, M. L. U., Yen, N.-S., et al. (1995). Cross-national differences in overconfidence. *Asian Journal of Psychology, 1*(2), 63–69.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*, 159–183.

Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance, 26*, 149–171.

McKenzie, C. R. M. (1997). Underweighting alternatives and overconfidence. *Organizational Behavior and Human Decision Processes, 71*, 141–160.

Önkal, D., Yates, J. F., Simga-Mugan, C., & Oztin, S. (2003). Professional vs. amateur judgment accuracy: The case of foreign exchange rates. *Organizational Behavior and Human Decision Processes, 91*, 169–185.

Palinscar, A. S. (1986). The role of dialogue in providing scaffolded instruction. *Cognitive Science, 10*, 151–177.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition, 49*, 123–163.

Price, P. C. (1998). Effects of a relative-frequency elicitation on likelihood judgment accuracy: The case of external correspondence. *Organizational Behavior and Human Decision Processes, 76*, 277–297.

Renner, C. H., & Renner, M. J. (2001). But I thought I knew that: Using confidence estimation as a debiasing technique to improve classroom performance. *Applied Cognitive Psychology, 15*, 23–32.

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes, 40*, 193–218.

Rozenblit, L. R., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*, 521–562.

Sanna, L. J., Schwarz, N., & Small, E. M. (2002). Accessibility experiences and the hindsight bias: I knew it all along versus it could never have happened. *Memory & Cognition, 30*, 1288–1296.

Schwarz, N., & Vaughn, L. A. (2002). The availability heuristic revisited: Ease of recall and content of recall as distinct sources of information. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 103–119). Cambridge: Cambridge University Press.

Sieck, W. R. (2003). Effects of choice and relative frequency elicitation on overconfidence: Further tests of an exemplar-retrieval model. *Journal of Behavioral Decision Making, 16*, 127–145.

Sieck, W. R., & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making, 18*, 29–53.

Sieck, W. R., & Yates, J. F. (2001). Overconfidence effects in category learning: A comparison of connectionist and exemplar memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1003–1021.

Sieck, W. R., & Yates, J. F. (2002). Retrieval effects on confidence in general knowledge. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th annual conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Sieck, W. R., & Yates, J. F. (2003, November). *Cues and processes underlying confidence in general knowledge*. Paper presented at the annual meeting of the Society for Judgment and Decision Making, Vancouver.

Simon, H. A. (1957). *Models of man: Social and rational*. New York: Wiley.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.

Sniezek, J. A., Paese, P. W., & Switzer, F. S., III (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46*, 264–282.

Suantak, L., Bolger, F., & Ferrell, W. R. (1996). The hard–easy effect in subjective probability calibration. *Organizational Behavior and Human Decision Processes, 63*, 201–221.

Thagard, P. (2000). *Coherence in thought and action*. MIT Press.

Thomson, M. E., Önkal-Atay, D., Pollock, A. C., & Macaulay, A. (2003). The influence of trend strength on directional probabilistic currency predictions. *International Journal of Forecasting, 19*, 241–256.

Wilkie, M. E., & Pollock, A. C. (1996). An application of probability judgment accuracy measures to currency forecasting. *International Journal of Forecasting, 12*, 25–40.

Windschitl, P. D., & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 198–215.

Yates, J. F., Lee, J. -W., & Shinotsuka, H. (1992, November). *Cross-national variation in probability judgment*. Paper presented at the Annual Meeting of the Psychonomic Society, St. Louis.