Running head: DEBIASING CONDITIONAL PREDICTIONS

The Locus of Theory-based Debiasing Effects on Conditional Predictions

Erin Buchanan

University of Mississippi

William S. Maki

University of Arizona

Erin Buchanan

University of Mississippi

Department of Psychology

PO BOX 1848

University, MS, 38677

Phone: 662-915-1075

Fax: 662-615-5398

Email: embuchan@olemiss.edu

Abstract

Associative strength, as assessed in normative studies of free association, is the likelihood of giving one word in response to another. When people are asked to guess that probability, their ratings are inflated relative to the norms. One source of the inflation is the presence of backward associations. Some participants were given an explanation of the influence of backwards associations on predictions of associative strength (theory-based debiasing). All participants then made predictions about associative strengths of many pairs that differed in both forward and backward associative strengths. Predictions increased with increasing forward strength and the slope of that function was increased by high backward strength. The debiasing instructions did not alter that interaction but the debiasing instructions did decrease the average prediction. Debiasing manipulations appear to have their effect not on associative processing but instead on a more controlled, deliberative, re-evaluation of the evident degree of association.

The Locus of Theory-based Debiasing Effects on Conditional Predictions

People often are asked to judge the frequency or probability of events. Such judgments are grossly inflated in many situations. For example, people estimate their competence at levels higher than those supported by their actual levels of skill (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Levin, Momen, Drivdahl, & Simons, 2000). Often, judgments of learning (JOLs) are also inflated leading to inappropriate study strategies (Koriat & Bjork, 2005). Thus, development of methods for remediation of inflated predictions is important for both theoretical and practical reasons (Koriat, 2008; Koriat & Bjork, 2006). The experiment reported here was motivated by the question of whether one such method (theory-based debiasing; Koriat & Bjork, 2006) alters people’s processing of the information being judged or instead decreases overall bias (Maki, 2007b).

Word associations are a convenient domain in which to investigate inflated predictions and their remediation (Koriat, Fiedler, & Bjork, 2006; Maki, 2007a). Word associations are usually measured by the method of free association. In that task, participants are asked to give the first (response) word that comes to mind when presented with a stimulus (cue) word. Averaged over many participants, the forward strength (FSG) of the association between words A and B is indexed by the conditional probability that word B is given in response to word A. The backward strength (BSG) of the pair A-B is given by the (separately normed) probability that word A is given in response to word B. Large databases are available that list the associative strengths of many thousands of cue-response pairs (e.g., Nelson, McEvoy, & Schreiber, 2004). From those norms pairs of words can be selected that differ along both the FSG and BSG dimensions. This disconnection is an important capability because conditional predictions of FSG have been shown to be sensitive to and inflated by the degree of BSG (e.g., Koriat et al., 2006; Maki, 2007a).

Studies of the inflation of conditional predictions within the domain of word associations have shown that that people typically over-estimate FSG, especially for weakly associated pairs. This tendency toward high ratings is most difficult to change (Koriat, 2008; Koriat et al., 2006; Maki, 2007a, 2007b; see also Nelson, Dyrdal, & Goodmon, 2005). For example, the inflation persists even when people are shown a list of common associates for a given cue (Maki, 2007a), when people are given a cue and asked to think about alternatives (Maki, 2007a), and when people overtly generate their own list of associates (Koriat, 2008; Koriat et al., 2006); in all three cases FSG remained grossly over-estimated.

Because of the robustness of the inflation effect, finding a substantial reduction in it is important. Three different methods have proved effective in reducing the inflation effect. Maki (2007b) gave error correction feedback following each judgment for training pairs. During a subsequent transfer phase with novel pairs, participants in the error correction condition gave ratings that were lower and closer to the normed FSGs. Koriat (2008, Experiment 3) had participants work together in dyads to come to a consensus about the predicted FSGs. Relative to working alone, predictions made by the dyads were substantially reduced, especially for the low FSG – high BSG pairs. Koriat and Bjork (2006) studied metacognitive errors (inflated JOLs) in a paired associate learning experiment. One of their two groups (theory-based debiasing) was given special instructions concerning the inflationary effect of BSG on JOLs; the instructions were reinforced with a short conditional prediction exercise in which the participants received feedback on their ratings. The idea is that providing participants with an explanation of the influence of BSG (the theory) should allow them to take BSG into account and thus lower their predictions. The debiasing treatment resulted in lower JOLs given to novel pairs in the transfer phase.

Koriat (2008) explained his debiasing result in terms of the dual process framework (e.g., Stanovich & West, 2000). Among the dimensions along which the processes differ (Stanovich & West, 2000) are “fast” vs. “slow”, acquisition by “personal experience” vs. “formal tuition”, and “associative” vs. “rule-based.” To this list Koriat added “intuitive” vs. “analytic.” Koriat attributed the inflated predictions of FSG to the associative/intuitive process operating by default in making these judgments. But when exposed to a debiasing treatment, the rule-based/analytic process is engaged; effective debiasing treatments lead participants to rely on this more deliberative mode of processing in making their predictions. Thus reasoning in dyads (Koriat, 2008), being given an explanation about backward associations (Koriat & Bjork, 2006), and being informed about erroneous judgments (Maki, 2007b) all can be seen as engaging the slower, more deliberative, analytic mode of processing.

In our view, however, what is missing from this story is evidence that debiasing treatments do *not* influence the fast/intuitive/associative process that gives rise to the subjective experience responsible for the inflated judgments. We believe that a signature of associative processing in the domain of predictions of FSG is the FSG x BSG interaction discovered by Maki (2007a, Experiment 1). The word pairs in that experiment varied along three levels of FSG crossed in a factorial design with two levels of BSG. The predictions of FSG were a linear function of actual FSG with a high intercept and shallow slope. The slope was steeper for the high BSG pairs indicating that BSG amplifies the effect of FSG. If some debiasing treatment were to influence associative processing, we predict that the manifestation of that effect would be the reduction in the amplifying effect of BSG and hence a diminished FSG x BSG interaction.

The problem in all three instances considered here is that the materials selected for use in the studies did not allow the observations needed to measure the interaction. Maki (2007a) could not have observed an altered interaction because the BSGs were controlled (all limited to small values < .10). Koriat (2008) could not have observed the interaction because he compared pairs with high FSG and low BSG to pairs with low FSG and high BSG; this is an incomplete factorial with two degrees of freedom that allow either an interaction or its absence. Koriat and Bjork (2006) studied similar pairs in the same incomplete factorial design so their treatment also could have influenced associative processing in addition to any effect on more deliberative processing. The open question then is whether debiasing treatments influence the intuitive/associative mode of processing in addition to any influence on the more analytic mode of processing.

In the experiment reported here, we crossed FSG (low vs. high) and BSG (low vs. high) in a complete factorial design. Our experiment employed theory-based debiasing (Koriat & Bjork, 2006). We anticipated two possible outcomes. Debiasing instructions could alter the manner of processing associations and thus diminish the FSG x BSG interaction (Maki, 2007a). Alternatively, such instructions could cause the participants to be more analytic and cautious in their responding and thus reduce the overall level of predictions (Maki, 2007b).

Method

*Participants*

Participants were recruited from the University of Mississippi human subjects’ pool and received class credit for their participation. Participants (122 female, 36 male) were all college aged (18-24 yr). There were 84 participants in the Standard instructions group and 74 participants in the Debiasing instructions group.

*Materials*

Four sets of cue-response pairs, *N*=24 each, were selected from the Nelson et al. (2004) association norms; selection was random but subject to the following constraints. Each set was classified as being low or high in FSG and low or high in BSG. For both FSG and BSG, *x*, low associative strength was defined as 0.00 < *x* ≤ 0.20 and high associative strength was defined 0.50 < *x* ≤ 0.80. The resulting mean low and high FSGs were 0.06 and 0.63, and the resulting mean low and high BSGs were 0.06 and 0.64.

These 96 *test pairs* were arranged in a list in sets of 24. Each set contained an equal number (*N*=6) of each type of pair: low-FSG/low-BSG, high-FSG/low-BSG, low-FSG/high-BSG, and high-FSG/high-BSG. Pairs within the four sets were then randomly ordered. The order of the 96 pairs was reversed to form a second list to control for order effects.

Eight more *debiasing pairs* were selected from the association norms for use in the debiasing instructions. Seven of the pairs varied considerably in FSG (range 0.01 – 0.54) but were all very high in BSG (range 0.75-0.88). The eighth pair was high in FSG (0.92) and low in BSG (0.06). An additional three *practice words* (LOST, OLD, ARTICLE) were drawn from the norms for use in illustrating the free association task.

Two booklets were prepared. Each began with an introduction to the concept of association and a short free association list containing the three practice words. The second part of each booklet described the associative rating procedure and presented a sample rating form containing a list of the eight debiasing pairs. The rating instructions were the same as those used in previous studies (Maki, 2007a, 2007b): “Assume 100 college students from around the nation gave responses to each CUE word. How many of these 100 students do you think would have given the RESPONSE word?” Ratings were marked on a 10-point scale (0-9) with each response defined as 0-9, 10-19, …, 80-89, 90-100 (as in Maki, 2007a). The third part of each booklet consisted of a listing of the 96 test pairs with adjacent rating choices (0-9), 24 per page; the rating scale and rating instructions appeared as reminders at the top of each page. The booklets differed only with respect to the ordering of the 96 test pairs (one the reverse of the other).

*Procedure*

Participants were tested in groups of 15-20. Each group was randomly assigned to receive either the standard or debiasing instructions. Each participant was randomly assigned one of the two booklets. All groups began by reading a short paragraph describing the free association task and then performed the free association demonstration. After writing their responses to the three practice words, the experimenter asked for a show of hands for those responding with FOUND to LOST. In the next show of hands, responses to OLD were split mainly between NEW and YOUNG. Participants reported a variety of responses to ARTICLE, the most frequent of which was NEWSPAPER. Then, all the groups were introduced to the associative rating task as an alternative way of measuring associations. All the groups used the 10-point scale to rate the eight debiasing pairs. The experimenter then used a show of hands to estimate the groups’ average rating for each pair. Finally, all the groups rated the 96 test pairs.

The critical manipulation that distinguished the debiasing-instruction groups from the standard-instruction groups occurred during the discussion of the ratings of the eight debiasing pairs. For the standard instruction groups, the experimenter gave some examples of what the correct ratings would be for individual word pairs (such as a 5 or 6 for the pair SISTER-BROTHER which has a forward strength of 0.54). For the debiasing instruction groups, the theory-based debiasing instructions were modeled after those reported by Koriat and Bjork (2006). After the rating of the debiasing pairs, the experimenter explained how word pairs actually have two associative relationships, both forwards and backwards. Examples were given illustrating how a backwards relationship can unduly influence a person’s rating of a forward relationship.  For example, STEAK-SIRLOIN has a very weak forward relationship, *p*(STEAK🡪SIRLOIN) = 0.01, and should only be rated 0 or 1 on the rating scale.  However, STEAK-SIRLOIN has a very strong backwards relationship, *p*(SIRLOIN🡪STEAK) = 0.81, that would likely lead to an inflated rating.  After STEAK-SIRLOIN was explained, both SISTER-BROTHER (high FSG, high BSG) and CHEDDAR-CHEESE (high FSG, low BSG) were discussed to show different combinations of forward and backward strength.  Then participants were allowed to complete the 96 ratings on the experimental word-pairs.

Results

In the free association demonstration, the participants’ responses to the three cue words were very close to those in the Nelson et al. (2004) norms. Response probabilities were calculated from the responses given on the booklets. The probability of FOUND in response to LOST was 0.626 (normed FSG=0.747). In response to OLD, NEW was given with *p* = 0.417 (normed FSG = 0.473) and YOUNG was given with *p* = 0.215 (normed FSG = 0.236). The most frequent response to ARTICLE was NEWSPAPER, *p* = 0.411 (normed FSG = 0.419).

Our participants over-estimated the FSGs for the eight debiasing pairs used in the rating demonstration. Ratings were averaged and converted to FSGs. The average predicted FSG was 0.624 (average normed FSG = 0.244). These two results indicated that participants understood the concept of free association, the judgment task, and that our groups were not different before instruction from previous findings of overrating judgment pairs (Maki, 2007a).

For each participant in each group, ratings were averaged across the 24 test pairs in each combination of FSG and BSG and converted to a 100-point scale. The mean ratings are shown in Figure 1 as a function of FSG (also represented on a 100-point scale). In each panel, the filled symbols represent the results from the low BSG pairs and the open symbols represent the results from the high BSG pairs. The left panel shows the ratings for the Standard instruction group and the right panel shows the ratings for the Debiasing instruction group. Figure 1 shows three principal findings. First, there was a pronounced interaction between forward and backward strength; the slope of the function relating ratings to forward free association probabilities was increased by higher backward associations. Second, the magnitude of that FSG x BSG interaction was unchanged by the debiasing instructions. Third, the debiasing instructions did have an effect, namely to reduce ratings overall by about nine points (on the 100-point scale).

The reliability of these observations was tested with a 2 x 2 x 2 mixed analysis of variance using interactions with subjects as error terms.1 The within-subjects factors were FSG (low vs. high) and BSG (low vs. high). The between-group factor was Instructions (standard vs. debiasing). The overall effect of FSG was significant, *F*(1, 156) = 955.74, η2partial = 0.86, *p* < .001; each of the four functions in Figure 1 show a reliable increase from low to high FSG, smallest *t* (73) = 14.70*, p* < .001. The overall effect of BSG was significant, *F*(1, 156) = 374.10, η2partial = 0.71, *p* < .001; high BSG resulted in higher ratings than did low BSG. The FSG x BSG interaction was also significant, *F*(1, 156) = 69.63, η2partial = 0.31, *p* < .001; the effect of FSG on ratings was amplified by BSG.

None of these effects were modulated by instructions. Neither the interaction of FSG and Instructions nor the interaction of BSG and Instructions were significant, both *F*s < 1. Of particular note, the three-way interaction of FSG, BSG, and Instructions was not significant either, *F*(1, 156) = 1.34, η2partial = 0.01.

The effect of the debiasing instructions was confined to an overall decrease in ratings that was independent of the other variables, *F*(1, 156) = 30.92, η2partial = 0.16, *p* < .001. As shown in Figure 1, the intercepts of the functions relating predicted FSG to normed FSG were reduced by about the same amount by the debiasing instructions.

Discussion

The present results join with earlier findings (Koriat, 2008; Koriat et al., 2006; Maki, 2007a, 2007b) in showing that predictions of word associations are greatly inflated relative to word association norms (Nelson et al., 2004). The present results also join with other findings (Koriat, 2008; Koriat & Bjork, 2006; Maki, 2007b) in showing that the amount of inflation can be reduced by appropriate debiasing techniques. The present results, however, go a step further in showing that theory-based debiasing instructions reduce the overall level of prediction without affecting associative processing. The interaction between forward and backward associative strength shown in Figure 1 is unchanged by the debiasing instructions even though the average prediction is significantly reduced.

We emphasize that our results are not a null outcome; a debiasing effect *was* found but it was confined to the intercept of the functions relating predictions to normed associative strength. Moreover, other aspects of our results are consistent with previously reported results from studies of free association (Nelson et al., 2004) and associative judgments (Maki, 2007a, 2007b). Our participants’ free association responses during the free association demonstration were close to normative values. Our participants performed as in other studies (e.g., Koriat, 2008) by showing substantially inflated predictions. In addition, the slopes and intercepts of the functions relating predictions to normed associative strengths (Figure 1) were similar to those previously observed (Maki, 2007a, 2007b), and the present results replicate in each group the FSG x BSG interaction reported by Maki (2007a, Experiment 1).

As noted in the Introduction, Koriat (2008) followed Stanovich and West (2000) in distinguishing between two modes of processing (called System 1 and System 2 by Stanovich and West). System 1 was seen to be automatic, fast, and associative. System 2 was seen to be controlled, slow, and rule-based. System 1 processing of associated word pairs gives rise to a subjective feeling of association responsible for inflated judgments. Debiasing treatments engage System 2 processing thus triggering a more deliberate consideration of the association that reduces the inflation. It appears from our results shown in Figure 1 that the debiasing instructions do indeed, as has been previously argued (Koriat, 2008; Koriat & Bjork, 2006), influence System 2 processing (the intercepts) while leaving System 1 processing (the slopes and interaction) unchanged. But why do the debiasing instructions fail to influence the associative, System 1 processing? Fodor (1983), in his essay on modularity, characterized his hypothesized input modules as being “mandatory” and “fast”; in those regards, the notion of an input module is analogous to System 1. Fodor also claimed that there was “limited central access” to the information processed by such modules. One possible answer to our question, then, is to complete the analogy and propose that an associative processor operates in accord with the properties of System 1 and is opaque to more central cognitive operations that correspond to System 2. Our main results then follow: theory-based debiasing instructions act to reduce the overall bias toward over-estimates of associative strength but the debiasing leaves unchanged the associative processing that causes the strong subjective impression of association.

References

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*, 83-87.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge : The MIT Press.

Koriat, A. (2008). Alleviating the inflation of conditional predictions. *Organizational Behavior and Human Decision Processes*, *106*, 61-76.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology: Learning, Memory, & 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, & Cognition*, *32*, 1133-1145.

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

Levin, D.T., Momen, N., Drivdahl, S.B., & Simons, D.J. (2000). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition*, *7*, 397-412.

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.

Nelson, D. L., Dyrdal, G. M., & Goodmon, L. B. (2005). What is preexisting strength? Predicting free association, similarity ratings, and cued recall probabilities. *Psychonomic Bulletin & Review*, *12*, 711-719.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402-407.

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

Author Note

Erin Buchanan, Department of Psychology, University of Mississippi. William S. Maki, Department of Psychology, University of Arizona.

We thank Ruth Maki for helpful comments on the manuscript. Order of authorship is alphabetical.

Correspondence should be addressed to William S. Maki, Department of Psychology, University of Arizona, Tucson, AZ 85721. Email addresses: Erin Buchanan (embuchan@olemiss.edu) and William S. Maki (wmaki@email.arizona.edu).

Footnotes

1 Analyses of variance using interactions with items (pairs) as error terms yielded the same pattern of significance for all main effects and interactions.

==================== EXTRA TEXT ==========================

Materials for the comparison of interest here consisted of 30 word pairs that were high in FSG (0.60) and low in BSG (0.02) and 30 word pairs that were reversed and thus low in FSG (0.02) and high in BSG (0.60).

But what can be inferred from that result is limited by the use of a design in which two combinations, high FSG – high BSG and low FSG – low BSG, are missing. Maki (2007a, Experiment 1) reported an interactive effect of FSG and BSG on associative ratings; the slope of the function relating ratings to FSG was increased by high levels of BSG. The missing points in Koriat’s design preclude detecting such an interaction or lack thereof. The missing cells in the design provide two degrees of freedom so the interaction could have been preserved or destroyed by the dyad manipulation. The possibilities are illustrated in Figure 1. Panel A shows a FSG x BSG interaction. Panel B shows how the interaction could be preserved given Koriat’s dyad effect. Panel C shows how the interaction could also be eliminated by the dyad manipulation. Thus it is difficult to ascertain whether working in dyads alters processing of associative information (by changing the interaction) or reduces some overall bias toward attaching large numbers to associated pairs.

The same interpretive problem afflicts other studies in which the incomplete factorial design is used.

task and use materials similar to those described above: high FSG – low BSG contrasted with low FSG – high BSG.

The resulting ratings were compared to those obtained in a previous experiment in which participants worked individually to predict FSGs for the same pairs of words (Koriat, 2008, Experiment 1).

thereby demonstrating inflated predictions and engaged the group in a discussion of the ratings

Koriat (2008, Experiment 3) reported a large debiasing effect when participants worked in dyads to come to consensus on predicted FSGs. The debiasing effect was larger for pairs with low FSG and high BSG than for pairs with high FSG and low BSG. On 100-point scales, the average predictions for participants working alone were 65.2 and 85.6 (compared to normed values of 2 and 60, respectively). The average predictions for participants working in dyads were 30.4 and 73.5 (compared to the same normed values of 2 and 60). On the surface that seems like a large slope effect: 65.2 - 30.4 vs. 85.6 - 73.5. However, the difference is much smaller when referenced to the calibration line based on normed FSGs. Working in dyads caused a drop of 44.9% of the possible distance to the calibration line for the low FSG pairs: 100 x (65.2 - 2)/(30.4 - 2); the corresponding drop for the high FSG pairs was 52.7%: 100 x (85.6 - 60)/(73.5 - 60). We think it reasonable to adjust for the calibration line especially because the participants in Koriat’s experiments were paid for predictions that fell within 5% of the normed values; adjusting predictions for the high BSG pairs the same amount as for the low BSG pairs would have resulted in a monetary loss to the participants in the dyads. The compensatory strategy might have involved being more cautious about being cautious for the pairs that aroused strong feelings of association. The net effect of the dyad manipulation was to move all judgments proportionately closer to the calibration line by about the same relative amount. Thus, we conclude that Koriat’s results are consistent with our claim that debiasing manipulations do not affect associative processing and instead exclusively influence how the results of that processing are judged.

Galton, F. (1879). Psychometric experiments. *Brain*, *2*, 149-162.

Garskof, B. E., & Forrester, W. (1966). The relationships between judged similarity, judged association, and normative association. *Psychonomic Science*, *6*, 504-504.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, *103*, 592-596.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley & Sons.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*, 215-233.

Maki, W. S. (2008). A database of associative strengths from the strength sampling model: A theory-based supplement to the Nelson, McEvoy, and Schreiber word association norms. *Behavior Research Methods*, *40*, 232-235.

Maki, W. S., & Buchanan, E. (2008). Latent structure in measures of associative, semantic, and thematic knowledge. *Psychonomic Bulletin & Review*, *15*, 598-603.

McFadden, D. L. (2000/2003). Economic choices. In T. Persson (Ed.), *Nobel Lectures, Economic Sciences 1996-2000.* Singapore: World Scientific Publishing.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure? *Memory & Cognition*, *28*, 887-899.

Sanbonmatsu, D. M., Posavac, S. S., & Stasney, R. (1997). The subjective beliefs underlying probability overestimation. *Journal of Experimental Social Psychology*, *33*, 276-395.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273-286.

Train, K. E. (2003). Discrete Choice Methods with Simulation. New York: Cambridge University Press.

Yellott, J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*, 109-144.