The Role of Explanation in Very Simple Tasks

Eric G. Taylor (etaylor4@illinois.edu)
Department of Psychology, 603 East Daniel St.
Champaign, IL 61820 USA

David H. Landy (dlandy@illinois.edu)
Department of Psychology, 603 East Daniel St.
Champaign, IL 61820 USA

Brian H. Ross (bhross@illinois.edu)
Department of Psychology, 603 East Daniel St.
Champaign, IL 61820 USA

Abstract

Much research on explanation has focused on the ability of explanations to draw upon relevant knowledge to aid in understanding some event or observation. However, explanations may also structure our understanding of events and related tasks more generally, even when they add no relevant information. In three experiments, we show that explanations affect performance in simple, binary decision tasks where they could not possibly add relevant information. Whereas people with no explanation for differences in event probabilities tended to "probability-match," people with an explanation tended to "over-match" (behave more normatively). The results suggest that explanations play a role in structuring our understanding of events, in addition to adding relevant information.

Keywords: explanation, probability matching, decision-making, understanding

Explanations support much intelligent behavior. We explain trends in the stock market in hopes of avoiding future economic woes, explain car failure to diagnose a problem, and we even explain why works of art gives us a chill just to enhance our appreciation (Keil, 2006). In recent years, cognitive scientists have begun to examine the importance of explanation (Lombrozo, 2006; Keil & Wilson, 2000), but despite agreement that explanations serve many goals, the empirical literature has focused on a limited set of tasks and functions. The purpose of this paper is to show a novel (and perhaps unintuitive) case where having an explanation changes performance in order to suggest a broader utility of explanation than currently exists in the literature.

Most work on explanation has examined cases where the explanation provides additional relevant information to help one understand the connection between an observation and other knowledge. For example, category learners often explain the correlations between an exemplar's properties to better understand the category structure (e.g., a bird nests in trees *because* it has wings), and this affects their applications of the category (e.g., Murphy & Wisniewski, 1989). Explanations also improve our understanding of social events, where we often call upon prior social experiences to make sense of others' behavior (Jones & Nisbtt, 1972). Laboratory studies of how explanations draw upon relevant knowledge relate directly to cases in the real world, where, for example, explaining the cause of a social problem (e.g., homelessness, global warming) by incorporating knowledge of social structures affects how we might try to solve that problem.

A major goal in our research program to examine and understand the role of explanation in cognition is to identify and explore the many ways that explanations can influence behavior. Although we are very interested in how explanations invoke relevant knowledge to help us understand events (Hummel, Landy, & Devnich, 2008; Hummel & Ross, 2006; Taylor, Landy, Ross, & Hummel, 2008), in this paper we investigate a different aspect of how explanations may influence performance. We consider whether explanations sometimes affect performance in very simple tasks without adding relevant information.

Our novel theoretical claim is that explanations can affect performance without adding task-relevant information by providing general ways to organize an understanding of a situation or event. We evaluated this idea by examining how explanations impacted behavior on a relatively low-level task, in which additional causal information is of no use. In our view, the explanations served as a task frame, which led participants who received it to structure their understanding of the task differently from those without an explanation.

We chose a binary prediction task, in which participants predict which of two outcomes will occur on the next trial, for many trials. On these tasks, people tend to "probability match," or predict each outcome roughly the percentage of times that the outcome tends to occur (for a review, see Vulkan, 2000). This behavior is non-normative, since predicting the most likely event on each trial maximizes correct predictions.

We added explanations to this paradigm in the following way: Participants in the No Explanation condition were told they would be predicting which of two events would occur on the next trial, from trial to trial, and that one event was more likely than another. Participants in the Explanation

condition were also provided a story explaining why the two events occurred with unequal likelihood, though this explanation did not directly add information about the probabilities of the events. Critically, any differences between conditions in this paradigm could not be due to the explanation adding relevant causal information.

How might explanations of the distribution source affect behavior in the probability matching paradigm? We speculated that explanations would lead the Explanation group to "overmatch"—to predict the more common outcome a greater percentage of times than it actually occurred (to behave more normatively)—more than the No Explanation group. There are many possible reasons for this consistent with our view that explanations provide a way to structure one's understanding of a task: For one, having an explanation might shape one's expectations about the likelihood of the two events, such that on each trial, the more likely event is preferred in the prediction. Alternatively, the explanation might draw attention to the mechanisms causing one event to occur more often, leading to increased confidence in the more likely outcome.

Setting aside, for the moment, how exactly explanations might structure the task, note that any difference across conditions would suggest that explanations add something more to cognition than task-relevant causal information. Furthermore, if explanations cause differential behavior in what is considered to be a relatively low-level task (fish and pigeons show the same behavior as humans; Behrand & Bitterman, 1961; Bullock & Bitterman, 1962), then the effects of explanation could be impressively far reaching.

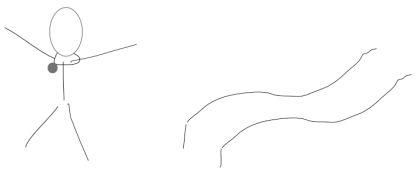
Experiment 1–Basic Probability Matching

The goal of Experiment 1 was to investigate whether explanations serve partly to structure people's understanding of basic tasks (like the probability matching task). If so, then people with explanations should perform differently than those without. In the case of probability matching, we predicted that the Explanation condition would show more over-matching than the No Explanation condition. Further experiments would narrow in on the particular reasons for the explanation advantage.

Method

Participants and Design Forty-six University of Illinois undergraduates participated for course credit, twenty-five randomly assigned to the Explanation condition and twenty-one to the No Explanation condition.

Materials Two line drawings were shown to participants during the experiment, one representing a medal winner from the Olympics and another representing the Great Wall of China (see Figures 1a and 1b.) All other instructions were displayed in text on a computer screen.



Figures 1a and 1b: Drawings used in Experiments 1 and 2 depicting an Olympic medal winner and the Great Wall.

Procedure The experiment was conducted on Macintosh computers. Participants signed a consent form prior to the experiment and then read the instructions. For the Explanation condition, instructions stated that a commemorative coin was produced for the 2008 Olympics and that a mistake was made in manufacturing so the side with the medal winner tended to come up more often than the side with the Great Wall. Participants were asked to make predictions for a sequence of coin flips as to whether, on the next trial, the coin would come up with the medal winner or the Great Wall. After their prediction, they would be shown the outcome of the flip. Finally, they were told that a counter at the top of the screen would indicate their overall performance after each trial. A black line would indicate their performance level on the previous trial.

The No Explanation condition was identical, except that participants were not told about the coin; instead, they were asked to predict which of two line drawings (either a person or two wavy lines) would appear on the next trial, for a sequence of trials. Furthermore, they were told that the drawing of a person tended to appear more often than the drawing of the two wavy lines.

There were 100 prediction trials. For 70 trials, the outcome was the person and for 30 the outcome was the two wavy lines. Subjects were not told how many trials the person would appear. On each trial, participants were asked to press the "P" key to predict the medal winner (person) or the "W" key to predict the Great Wall (two wavy lines.) After they entered their choice, participants viewed the outcome and their performance level, and then they pressed the "N" key to continue. Every 20 trials, they were reminded the purpose of the experiment. The Explanation condition was told, "Remember to choose the side of the coin that you think will come up next," and the No Explanation condition was told, "Remember to choose the drawing that you think will come up next."

After the prediction phase, participants were told that the experiment contained 100 trials and were asked to estimate how many of these trials resulted in the person side up. Next, participants answered questions regarding their strategies during the predictions task. They were given a list of options in a text file and told to delete the strategies they did not use. They were also given the same set of strategies again and asked which they thought was the best strategy for going about the task. The strategy reports did not lead to

consistent results across experiments, so they are omitted from our analyses and discussion. At the end of the session, participants were debriefed.

Due to experimenter error, 4 participants in the Explanation condition and 1 participant in the No Explanation condition did not complete the frequency judgments and strategy report tasks. In addition, 2 participants in the Explanation condition and 3 participants in the No Explanation did not give a frequency judgment (answered "I don't know.") Finally, 2 participants in the Explanation condition and 5 participants in the No Explanation condition entered a range of values for their frequency judgment (e.g., "between 60 and 80 trials.") For these participants, we used the mean of the endpoints in our analyses.

Results and Discussion

Predictions As predicted, the Explanation condition predicted the more frequent outcome greater than 70% of the time (78.5 trials, SD = 9.7) and on more trials than the No Explanation condition (71.9 trials, SD = 11.3). The Explanation condition average was significantly different from 70, $t(24) = 4.40$, $p < .01$, but the No Explanation condition average was not, $t(20) = .77$, $p = .45$. The difference between conditions was significant, $t(40) = 2.10$[1], $p < .05$.

Frequency Judgments The conditions did not differ in their average frequency judgments, suggesting that the increase in predictions for the more likely event was not due to inflation in perceived frequency of that event. On average, participants in the Explanation judged the frequency of the person drawing to be 73.8 (SD = 7.2), compared to 73.3 (SD = 9.6) in the No Explanation condition. The groups did not differ in their average frequency estimations, $t(27) = .31$, $p = .76$.

The results from Experiment 1 show that simply having an explanation can affect performance in a basic cognitive task without adding relevant information. This perhaps unintuitive outcome is consistent with the idea that explanations serve partly to structure one's understanding of a task, and thus, lead to differences in behavior.

Experiment 2–Probability Matching With Bets

We had two goals for Experiment 2: first, to replicate the findings from Experiment 1, and second, to increase the power of the difference between conditions by allowing participants to place a bet on each prediction, which could then be used to weight the individual predictions.

[1] Throughout the paper, degrees of freedom for between subjects tests with unequal sample sizes were the Welch-Satterthwaite values. This may cause degrees of freedom to differ within the same experiment across tests, since they are dependent on the variances of the samples.

Method

Participants and Design Twenty-two University of Illinois undergraduates participated for course credit, equal numbers assigned to the Explanation and No Explanation conditions.

Materials and Procedure The materials and procedure were identical to Experiment 1, except that participants in both conditions made bets on their predictions. After each prediction, they were told to bet 1, 2, or 3 chips (not corresponding to monetary value) by pressing the key corresponding to their bet. If they were correct (incorrect), they would win (lose) the amount of chips bet. The performance bar at the top of the screen was adjusted corresponding to the magnitude won or loss on each trial.

Results and Discussion

Predictions Experiment 2 replicated the results from Experiment 1. Participants in the Explanation condition predicted the person side up, on average, on 88.5 trials (SD = 11.4), whereas participants in the No Explanation condition predicted the person side up on 77.1 trials (SD = 17.3) of trials. The No Explanation condition average did not differ from 70, $t(10) = 1.36$, $p = .20$, but the Explanation condition average did, $t(10) = 5.38$, $p < .01$. The difference between conditions only approached significance, $t(17) = 1.83$, $p = .08$, although, when summing the wins and losses across bets, the average for the Explanation condition (2.21, SD = .63) was greater than that of the No Explanation condition (1.49, SD = .94), $t(17) = 2.11$, $p < .05$.

Frequency Judgments As in Experiment 1, there were no differences in the frequency estimates, suggesting that the Explanation advantage is not due to belief in a greater likelihood of the more common event. On average, participants in the Explanation condition judged the frequency of the person drawing to be 74.9 (SD = 5.8), compared to 73.4 (SD = 7.3) in the No Explanation condition. The groups did not differ significantly, $t(18) = .65$, $p = .62$.

Experiment 3–Diagnosis with Multiple Cues

In Experiment 3, we generalized our results with the basic probability matching task to a slightly richer scenario where people made predictions based on the presence of a diagnostic cue. Before each prediction, participants viewed one of two possible cues, which were associated with unique (and opposite) outcomes for 70% of the trials. The outcomes were reversed for the remaining trials. Participants were told to make their prediction of the outcome based on the cue, but as in Experiments 1 and 2, only the Explanation condition was told why the cues tended to lead to particular outcomes. Generally, the task was isomorphic to two intermixed basic probability matching tasks—one task for trials with cue A, and another task with cue B.

Method

Participants and Design Thirty-five University of Illinois undergraduates participated for course credit, twenty randomly assigned to the Explanation condition and fifteen to the No Explanation condition.

Materials Two drawings of "red blood cells" were shown to participants during the experiment, one very round cell and the other very large. All other instructions were displayed in text on a computer screen.

Procedure The experiment was conducted on Macintosh computers. Participants signed a consent form prior to participating. The procedure was similar in structure to that of the previous two Experiments, but the cover story was new. For the Explanation condition, instructions stated that participants would learn to measure genetic markers for particular traits. They would be shown drawings of bloods cells coming from patients who have either a gene that generally causes them to be taller than average, or a gene that promotes having a strong immune system. The gene that causes tallness usually (but not always) also causes blood cells to be larger than average. The gene that improves the immune system usually also causes red blood cells to be particularly round. Participants' task was to observe the shape and size of an individual's blood cell and then predict whether that individual was either taller than average or has a strong immune system. To predict taller than average, they should push the "T" key, and to predict strong immune system, they should push the "I" key. After their prediction, they would be shown the correct answer—"The outcome was T (or I.)" Finally, they were told that a counter at the top of the screen would indicate their overall performance after each prediction. A black line would indicate their performance level on the previous trial.

The No Explanation condition was identical, except that participants were not told that the shapes referred to red blood cells, nor that they were using the shapes to predict the traits "taller than average" and "good immune system"; instead, they were told simply that if the shape is particularly large, the outcome is likely to be "T." If the shape is particularly round, the outcome is likely to be "I."

There were 120 prediction trials. For round cue trials, 42 (70%) of the outcomes were "I," and for large cue trials, 42 (70%) of the outcomes were "T"; the other trials had the opposite outcome. Participants were not told the actual number of trials the more likely outcome would appear. On each trial, participants were asked to press the "T" key or the "I" key to predict the outcomes "T" or "I." After they entered their choice, participants viewed the outcome and their performance level and then they pressed the "N" key to continue. Every 20 trials, they were reminded the purpose of the experiment. The Explanation condition was told, "Remember to choose the trait you think the next patient will have. Press T if you think they will have a gene that tends to make them taller than average. Press I if you think they will have one that encourages them to have a strong immune system," and the No Explanation condition was told, "Remember to choose the result that you think will come up next. Results will be either T or I."

After the prediction phase, participants were told that the experiment contained 120 trials, on 60 of which the cue was "round" and on the other 60 the cue was "large." Separately for each cue, they were asked to guess how many of the 60 trials resulted in the "T" outcome. Then, they answered questions regarding their strategies during the predictions task. They were asked to give two strategy reports, one for trials when the cue was "round" and another for when the cue was "large." After the strategy questionnaire, participants were given a debriefing form and dismissed.

Three participants in the Explanation condition and one participant in the No Explanation gave frequency estimates of 60 for both of the cues. Since these estimates were likely due to confusion regarding the estimation task, they were removed from the analyses.

Results and Discussion

Predictions The predictions data were analyzed for each cue, separately, and then collapsed across the two cues. For the "round" cue, the Explanation condition predicted a strong immune system greater than 70% of the time (88.5% of trials, SD = 11.1) and on more trials than the No Explanation condition (76.2% of trials, SD = 15.1). The Explanation condition average was significantly different from 70, $t(19) = 7.44$, $p < .01$, but the No Explanation condition average was not, $t(14) = 1.60$, $p = .13$. The difference between conditions was significant, $t(25) = 2.66$, $p < .05$.

For the "large" cue, the Explanation condition predicted tall greater than 70% of the time (84.5% of trials, SD = 16.5) and on more trials than the No Explanation condition (75.1% of trials, SD = 15.6). The Explanation condition average was significantly different from 70, $t(19) = 3.92$, $p < .01$, but the No Explanation condition average was not, $t(14) = 1.27$, $p = .22$. The difference between conditions was only marginally significant, $t(31) = 1.72$, $p < .10$.

Collapsed across cue, the Explanation condition predicted the more likely outcome greater than 70% of the time (86.5% of trials, SD = 13.6) and on more trials than the No Explanation condition (75.7% of trials, SD = 15.1). The Explanation condition average was significantly different from 70, $t(19) = 5.41$, $p < .01$, but the No Explanation condition average was not, $t(14) = 1.45$, $p = .17$. Finally, the Explanation condition predicted significantly more trials with the person side up than did the No Explanation condition, $t(28) = 2.19$, $p < .05$.

Frequency Judgments The frequency judgments did not differ between conditions, both when the cue was "large" and when it was "round." When the cue was "large,"

participants in the Explanation condition judged the frequency of the "T" outcome, on average, to be 42.4 (70.7%; *SD* = 10.4), compared to 38.9 (64.8%; *SD* = 9.5) in the No Explanation condition. The groups did not differ in their average frequency estimations, $t(29) = .98$, $p = .34$. When the cue was "round," participants in the Explanation judged the frequency of "T" outcome, on average, to be 26.6 (44.3%; *SD* = 14.6), compared to 23.5 (39.2%; *SD* = 11.2) in the No Explanation condition. The groups did not differ significantly, $t(29) = .66$, $p = .51$.

Collapsed across cue, for each participant we computed the estimated frequency of the more likely outcome by averaging the estimate of "T" when the cue was "large" and 60 minus the frequency estimate of "T" when the cue was "round" (since all trials that were not "T" were "I"). The average estimate for the Explanation condition was 37.9 (63.2%; *SD* = 8.0), compared to 37.7 (62.8%; *SD* = 10.0) in the No Explanation condition. The group averages were not significantly different, $t(25) = .06$, $p = .95$.

The overall pattern of results was similar to that of Experiments 1 and 2. Participants in the Explanation condition over-matched (predicted the more likely outcome more than 70% of trials), but participants in the No Explanation condition did not. Also, collapsed across cues, the Explanation condition predicted the more likely outcome more often than the No Explanation condition. Finally, the frequency judgments of the two groups were not significantly different.

General Discussion

Many previous studies have shown that explanations are crucial for thinking and reasoning tasks, in which the explanation helps to understand some observation by drawing upon relevant prior knowledge. However, our findings suggest that explanations are not merely information couriers, since they also affect performance (indeed, improve normative responding) on even very simple tasks where additional information is not at all useful. Based on these results, we suggest that one role explanations play in cognition is to help to organize a person's understanding of a situation or event, so that having an explanation leads to differences in behavior relative to not having an explanation.

How Explanations Might Structure Understanding

As we mentioned earlier, explanations may help to shape our understanding of an event in many possible ways. The goal for our discussion is to consider a few possibilities in more detail, and to suggest how future research could explore their implications.

Increased Rational Responding One possibility is that giving an explanation for the differences in the event likelihoods tended to engage more analytic processes in the Explanation condition than in the No Explanation condition. In effect, this might have raised the number of participants in the Explanation condition who thought deeply about the task and decided consciously to endorse the normative strategy—to choose the more likely event on every trial. Previous research shows that people using the normative response pattern do tend to be higher in cognitive ability, suggesting a relation between high-level reasoning and normative responding (West & Stanovich, 2003). Conveniently, this pattern could be observed in the data by comparing the number of strictly normative participants in the two conditions.

In fact, we found very small and highly similar levels of normative participants across conditions. In Experiment 1, both conditions had 1 such participant. In Experiment 2, we found 3 and 2 normative participants in the Explanation and No Explanation conditions, respectively. In Experiment 3, we found zero normative participants. These data suggest that the explanation condition was not more likely to endorse the normative strategy, suggesting that a shift in rational reasoning was does not account for the effect.

Mental Simulation Another possibility is that having an explanation for the difference in outcome likelihoods allows one to mentally simulate the event (e.g., the coin flip) before each prediction, and this leads to a bias in predicting the more likely outcome, perhaps because it is more natural to simulate. This account applies to Experiments 1 and 2, where the coin flip is a discrete, simulable event, but less to Experiment 3 where simulating the relation between the shape of a blood cell and a medical trait seems less natural.

Whether or not all cases of explanation affecting performance are due to mental simulation, there are ways to test the role of simulation in explanation-based predictions. For example, one could directly manipulate the ease of simulating the events and look for an influence of simulation ease on levels of normative responding. Another method is to have participants perform a task that would either facilitate or work against the particular simulation (see Barsalou, 2008, for a review of simulation effects), where the prediction is that simulation-consistent behaviors lead to more predictions of the likely outcome. Current studies in our lab are beginning to address these issues.

Strength and Believability Previous research shows that the strength, or believability of an explanation impacts judgments related to the explananda. For example, Fugelsang et al. (2004) gave people either a strong or weak explanation for the relation between some causal variable and an outcome and then had people observe contingencies between the variable and the outcome. After viewing the same contingency data, people with a stronger explanation gave higher ratings of causal power than those with a weak explanation. If explanations affect judgments of causal power, they might also affect sequential predictions. Specifically, people with a stronger explanation may predict the more likely outcome on a greater number of trials than

those with a weak explanation. Along the same lines, one could view our No Explanation condition as the Extremely Weak Explanation condition, in which case our current results are attributable to explanation strength.

A simple way to test this idea is to generate explanations with more and less strength and look for differences in performance as a function of strength. Also, to test the role of strength in our current in experiments, one could ask participants before the prediction task for an estimate of the number of trials the more likely event will occur. If estimates of frequency parallel causal power judgments, then one would predict higher frequency estimates in the Explanation Condition than the No Explanation Condition. We are currently running a version of Experiment 2, where in place of the predictions task, people estimated how many trials (out of 100) the picture of the person would appear. Considering only people who gave an estimate greater than 50 (those who understood from the instructions that the person would appear more often), the average estimate from the Explanation condition is 68.4 (SD = 9.4) compared to 69.5 (SD = 5.8) in the No Explanation condition, which is not a significant difference, $t(20) = .37$, $p > .1$.

In each of the accounts we considered, the explanation was purported to structure the task by adding cognitive resources other than particular, task-relevant information. Whether these resources include general reasoning procedures, mental simulations, or top-down biases for interpreting data, the simple presence of an explanation appears to be a catalyst for higher cognitive processing. That is, explanations affect the structure, as well as the content, of thought.

Conclusion

Explanation is a powerful cognitive function. Previous research on explanation has concentrated on the ability of explanations to call upon relevant knowledge to improve our understanding of some event, and this knowledge often affects people's judgments in related tasks. Although we agree that explanations are crucial for connecting everyday observations to knowledge, we suggest that explanations have other functions beyond adding relevant information. Explanations may shape our understanding of events and scenarios, such that behavior in related tasks is often different (and sometimes normatively improved) compared to behavior without an explanation. Future research will need to explore this view with developments in theory and new empirical findings.

Acknowledgments

This research was supported by AFOSR Grant # FA9550-07-1-0147. We thank John Hummel, Erin Jones, and the Ross-Hummel lab meeting for their helpful comments regarding this research.

References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Behrend, E. R. & Bitterman, M. E. (1961). Probability-matching in the fish. *American Journal of Psychology, 74*, 542-551.

Bullock, D. H. Bitterman, M. E. (1962). Probability-matching in the pigeon. *American Journal of Psychology. 75*, 634-639.

Fugelsang, J., Stein, C., Green, A., & Dunbar, K. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory. *Canadian Journal of Experimental Psychology, 58*, 132-141.

Hummel, J. E. & Ross, B. H. (2006). Relating category coherence and analogy: Simulating category use with a model of relational reasoning. In *Proceedings of the Twenty Fourth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hummel, J. E., Landy, D. H., & Devnich, D. (2008). Toward a process model of explanation with implications for the type-token problem. *AAAI Fall Symposium on Naturally-Inspired Artificial Intelligence.* Arlington, VA.

Jones, E. E. & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79 –94). Morristown, NJ: General Learning Press.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology. 57*, 227-254.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*, 464-470.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. *Tiberghien (Ed.), Advances in cognitive science, Volume 2: Theory and applications* (pp. 23–45). Chichester: Ellis Horwood.

Taylor, E. G., Landy, D. H., Ross, B. H., & Hummel, J. E. (2008). Generating explanations. *Forty-ninth Annual Meeting of the Psychonomic Society Chicago.*

Vulkan N. (2002). An economist's perspective on probability matching. *Joumal of Economic Surveys, 14*, l01-118.

West, R. F. & Stanovich, K. E. (2003). Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Memory & Cognition, 31*, 243- 251.