This work is licensed under a Creative Commons Attribution-ShareAlike 3.0 United States License. Copyright 2010, 2011 Kodi Arfer.

INTUITIVE DECISION-MAKING AS SOCIAL PREDICTION: THE SIMILAR-STRATEGY HYPOTHESIS

A Senior Comprehensive Project
Presented By:
Kodi Arfer

Submitted to the Psychology Department of Allegheny College
in partial fulfillment of the requirements for
BACHELOR OF SCIENCE
7 April 2011

INTUITIVE DECISION-MAKING AS SOCIAL PREDICTION: THE SIMILAR-STRATEGY HYPOTHESIS

Kodi Arfer

Pledge: ____________________

Approved:

____

Robert Hancock, PhD

____

William DeLamarter, PhD

TABLE OF CONTENTS

List of Figures and Tables....................iv

Abstract....................v

Introduction....................2

The similar-strategy hypothesis....................5

Method....................8

Participants....................8

Materials....................8

Procedure....................9

Results....................9

Discussion....................14

Appendices....................16

References....................21

Footnotes....................24

List of Figures and Tables

Figure 1. Mean scores (natural logarithm of response divided by correct answer) for each question and condition (p. 12).

Table 1. Mean responses in Experiment A compared to correct answers (p. 10).

Table 2. Ranges and interquartile ranges of responses in Experiment A (p. 11).

Abstract

Deliberate introspection during decision-making, operationalized as writing down relevant reasons for deciding, has been found to improve decision-making in some cases and impair it in others. The present study tested whether cases of impairment could be explained by how the decision-making strategies of subjects differed from those of criterion judges. Experiment A used a numerical-estimation task for which reasons-writing had previously been found to improve performance. Experiment B used the same task with an added element of social prediction. Previous results with this task were not replicated. There were no significant between-group differences. These negative results suggest that reasons-writing has no more than a weak effect on performance on the task in question.

Intuitive Decision-Making as Social Prediction: The Similar-Strategy Hypothesis

Kodi Arfer

Allegheny College

Intuitive Decision-Making as Social Prediction: The Similar-Strategy Hypothesis

Decision-making is, at its most basic, choosing among alternatives. Often, the exact goal of a decision, the outcome the decider wants to bring about, is ambiguous, as when choosing a college to attend. But even when the goal is clear, the right decision may not be obvious. Worse, we may believe an incorrect choice to be obviously correct. Consider the following problem: “A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?” In a survey of students at Princeton University and the University of Michigan, most replied “ten cents”, even though the correct answer of 5¢ is easy to calculate (Kahneman, 2003). These two choices illustrate two basic decision-making strategies, which may be described as intuitive and deliberative. When choosing intuitively, a decider simply produces a choice without taking much time or apparently expending much effort. Colloquially, we might describe such a strategy as trusting one’s feelings or going by one’s gut instinct. Deliberative decision-making includes any more formal or careful means of choosing, ranging from applying an algorithm to mentally debating the pros and cons of each choice.

There is considerable evidence that intuitive decision-making is flawed and only by exercising great care can we avoid gross errors of judgment (e.g., Kahneman, 2003; Tversky & Kahneman, 1981; Kahneman & Frederick, 2002; Tversky & Kaheman, 1971; Dawes, Faust, & Meehl, 1989; Kogler & Kühberger, 2007). A limitation of these studies is that the decision-making problems they consider can largely be characterized as math problems; they require only one-dimensional quantitative judgments and are amenable to formal reasoning. In other words, they are well-defined problems. In real life, on the other hand, we frequently must choose between alternatives whose salient characteristics are unclear, while our own goals are equally nebulous. Few deliberative strategies are possible in such cases. The most familiar and possibly the most general approach is to "divide and conquer" by considering all the individual reasons we might decide for or against each option. One could implement such a strategy by assigning each reason a numerical weight and then applying a formal heuristic, or by simply dividing reasons into pros and cons and finding "where the balance lies", as Benjamin Franklin suggested to Joseph Priestly (Wilson & Schooler, 1991). In any case, deliberation is distinguished from intuition by decomposing each alternative into reasons and explicitly considering the reasons themselves instead of just the choices as wholes.

Wilson and Schooler (1991) tested the efficacy of this kind of reasoning in two studies of preference. College students tasted several brands of strawberry jam and reported how well they liked each, or reviewed information on several psychology courses and reported how likely they were to enroll in each. Some subjects were instructed to list reasons for liking or disliking each jam or class before making their decision. The researchers compared the jam ratings to the ratings of professional taste testers given in *Consumer Reports* magazine and the class ratings to the ratings of faculty and of students who had actually taken the courses. As the researchers had predicted, subjects who analyzed reasons performed worse than control subjects, in the sense that their ratings were less correlated with the relevant expert ratings. Wilson and Schooler explain these findings as a self-persuasion effect: subjects who listed reasons focused on reasons that they considered important, which were not exactly the reasons that would otherwise have determined their choices. The subjects then generated opinions to match the reasons they gave. This explanation is especially plausible given that the existence of self-persuasion by other means (such as cognitive dissonance; see e.g. Brehm & Cohen, 1962) is well established.

Wilson obtained similar results in other studies, several of which are unpublished; most notable is a study described in Wilson (1990). Students rated decorative posters and chose one to take home. Reasons-listing subjects not only chose less popular posters, they also were less likely to be happy with their choice after the experiment. This implies that preferences created by reasons-listing are temporary, making them an unambiguously bad influence.

At heart, however, the idea that the reasons groups in these experiments made *worse* choices than the control groups depends crucially on the metric of choice quality. The metrics used in these cases are questionable because they are essentially subjective. By contrast, MacGregor, Lichtenstein, and Slovic (1988) assessed participants on how accurately they could estimate quantities such as the number of people employed in American hospitals. McMackin and Slovic (2000) note that one of the five experimental conditions of this study is similar to the reasons-listing groups in Wilson and Schooler (1991). Here, the reasons group performed better than the control group.

McMackin and Slovic (2000) hypothesize that, rather than Wilson's results being an artifact of poor dependent measures, the discrepancy between his findings and those of MacGregor et al. (1988) can be explained by differences between the tasks. In particular, McMackin and Slovic cite a "cognitive continuum" postulated by Hammond, Hamm, Grassia, and Pearson (1987), which attempts to describe task characteristics that induce intuitive versus analytic processing. This model predicts, for example, that a task presenting many redundant cues that can be perceptually evaluated will favor intuitive processing. McMackin and Slovic predicted that Wilson's reasons manipulation would worsen performance on an intuitive task and improve performance on an analytic task. The results of their experiment accorded well with this hypothesis. In their *ads task*, which was intended to be intuitive, subjects examined print advertisements and estimated how well fellow students liked them. In the *numbers task*, which was intended to be analytic, subjects estimated quantities of the kind one could find in an almanac, as in MacGregor et al. (1988). The reasons-analyzing group performed worse than the control group on the ads task but better on the numbers task. The ads-task results are especially important as a conceptual replication of Wilson and Schooler (1991) using an objective metric of choice quality.

The similar-strategy hypothesis

The studies so far presented are consistent with the self-persuasion hypothesis of decision-making of Wilson and Schooler (1991) and the cognitive continuum of McMackin and Slovic (2000) and Hammond et al. (1987). However, an alternative, possibly more parsimonious explanation for these experimental results is available. All of the tasks on which the reasons manipulation has been found to worsen performance (rating the taste of jams, the attractiveness of college courses, the aesthetics of posters, and the effectiveness of advertisements) have a feature in common other than the characteristics of the tasks themselves. Namely, the metric of choice quality for each judgment was its similarity to another judgment or to an aggregate of judgments. In the jam and college-course tasks, the subjects' ratings were compared to expert ratings, while in the poster task, the subjects' choices were compared with popular opinion and with the subjects' own opinions at a later date.[1] Only in McMackin and Slovic's (2000) ads task were subjects explicitly asked to approximate other judgments, but in every case success was determined by the accuracy of such an approximation. Notice further that the judgments used to measure choice quality were obtained without asking the judges to write down reasons before deciding.[2] So during the experiments, the control subjects, compared to the reasons-analyzing subjects, were making their decisions in a way more similar to that of the deciders to whom they were compared. For this reason alone, it should come as no surprise that the control groups decided "better".

In short, I predict that whenever a person estimates a judgment, whether they think of the task as one of social prediction or not, and whether they are estimating the previously collected judgments of multiple people or their own future judgment, their estimate will be more accurate the more similar their decision-making strategy is to that of the criterion judges. I call this idea the *similar-strategy hypothesis* (SSH). It can be seen as an implication of a broader idea, which is that similar decision-making circumstances in general increase estimation accuracy. The situation would then be analogous to memory. It has been shown, as in the famous "diving experiment" of Godden and Baddeley (1975), that recall is best in the same circumstances as encoding.

As for my claim of the superior parsimony of the SSH, consider again the alternative. Without the SSH, it appears that *both* Wilson's self-persuasion hypothesis and Hammond et al.'s cognitive continuum are necessary to explain the previously discussed findings. The former postulates a very counterintuitive effect of deliberation, and the latter makes many large, highly general claims about information processing. By contrast, the SSH makes a weak, almost obvious claim, and is a natural implication of observed properties of social prediction.

Particularly salient is the observation that people use their own judgments to predict the judgments of others. In a series of five experiments, Epley, Keysar, Van Boven, and Gilovich (2004) demonstrated that perspective-taking is accomplished with a strategy of "egocentric anchoring and adjustment". Predictors first make their own judgment, then adjust it to compensate for differences between themselves and the person or people whose judgments they are trying to predict. For example, Epley et al. (2004) had college students try to distinguish Coca-Cola from Pepsi by taste alone. They did no better than chance. Other students were told which was which, tasted both, and then were asked to estimate the accuracy of uninformed tasters. They significantly overestimated the accuracy of their peers, unless they had a monetary incentive to estimate accurately. Similarly, many studies have demonstrated the false-consensus effect: the tendency for people to think that their own opinions and behaviors are common (Mullen, 1985).

One broad idea that can be taken to imply the SSH is simulation theory. Among psychologists who study social-inference processes, particularly mind-reading (the ability to perceive the mental states of others), there are two prevailing views. According to *theory theory*, we determine others' mental states by (unconsciously) applying a set of general rules that attribute certain inner states to certain observable phenomena, akin to a psychological theory.[3] According to *simulation theory*, we reason about other minds by vicariously experiencing their circumstances. That is, we produce judgments about others' feelings and actions by pretending we were in their shoes and observing our own reaction (Gallese & Goldman, 1998). Epley et al.'s (2004) egocentric anchoring and adjustment can be taken as evidence for simulation theory. Some neuroscientific support comes from studies of the anterior insula and frontal operculum. There are regions of these areas that are active while feeling disgust, happiness, and pain and while perceiving faces expressing these emotions. Lesions in these regions impair both the ability to feel these emotions and to recognize them in others (Keysers & Gazzola, 2009). It thus appears that we vicariously feel the emotions of others.

To directly test the SSH, I employed the numbers task of McMackin and Slovic (2000). Half the participants provided a literal replication of McMackin and Slovic's experiment: they estimated unknown quantities, and some were asked to write down reasons before deciding. The remaining participants were asked to estimate these estimates; none were told of the reasons manipulation, but some were asked to write down reasons before making their own estimates. Of these two groups of meta-estimators, I predicted that each would predict the judgments of the group using the same strategy better than the judgments of the group using the other strategy.

Method

Participants

Participants were undergraduate students at Allegheny College enrolled in lower-level psychology courses. Students were compensated with course credit at the professor's discretion.

Materials

All subjects answered the same five questions. The questions were essentially the same as in McMackin and Slovic (2000) (correct answers given in parentheses; figures not cited are McMackin and Slovic's):[4]

A. What is the area of the United States in square miles? (3,540,940)

B. How many cigarettes were consumed in the US in 1998? (451,000,000,000; Mackay & Eriksen, 2002, p. 30)

C. What is the circulation of *Reader's Digest* in the US? (7,114,955; Audit Bureau of Circulations, 2009)

D. How long is the Amazon River, in miles? (3,900)

E. What was the population of the US in the year 1900? (76,200,000)

Procedure

The study consisted of two simultaneous experiments. Experiment A was a literal replication of McMackin and Slovic (2000), in which participants were told to estimate the answer to each question. Some subjects (*intuitive direct estimators*) were given no special instructions as to how to decide, whereas others (*deliberative direct estimators*) were told to write down reasons before deciding. In Experiment B, subjects were given the same questions but were asked to estimate the average estimates of their peers, rather than the actual answers to the questions. As in Experiment A, some subjects (*deliberative meta-estimators*) were told to write down reasons before deciding, whereas others (*intuitive meta-estimators*) were not, although none of these participants were told anything about the decision-making strategies used by the direct estimators.

Upon arrival at the premises of the experiment, each participant was handed a packet containing complete instructions and a response sheet. There were four sets of packets, two for each experiment and two for each decision-making strategy, and all four sets were shuffled into a single stack; thus, subjects were randomly assigned to each experiment and strategy. Participants worked on the questions at separate desks in a classroom. After turning in their answers, subjects were fully debriefed.

Results

Responses were elicited from 79 students. One subject who provided no estimates and another whose estimates were all less than 10 were excluded from analysis, as were 23 estimates, supplied by 9 subjects in reasons-writing conditions, that lacked accompanying reasons.

To test if Experiment A literally replicated McMackin and Slovic's (2000) results, the same analyses were performed. In general, the results differed from those of McMackin and Slovic (2000). Table 1 shows the geometric mean of all subjects' responses in Experiment A. McMackin and Slovic found that for questions A through D, the mean for the deliberative group was closer to the correct answer than that of the intuitive group, and on question E, the means were very similar, 35,678,172 for the intuitive group versus 35,455,911 for the deliberative group. Here the deliberative group outperformed the intuitive group on questions A, B, and E, but the intuitive group performed better on questions C and D. Table 2 presents the range and interquartile range (IQR) for each condition. These ranges are calculated not as differences of extreme scores but as natural logarithms of quotients of extreme scores. McMackin and Slovic found that the intuitive ranges were greater for every question but B and that the intuitive IQRs were greater for every question but B and C. Here, in stark contrast, the deliberative ranges were greater for every question but D and E, and the deliberative IQRs were greater for all questions.

Table 1
Mean Responses in Experiment A Compared to Correct Answers

	Geometric mean		
Question	Intuitive	Deliberative	Correct answer
A	340,801	623,690	3,540,940
B	199,334,958	498,163,504	451,000,000,000
C	1,432,224	1,042,132	7,114,955
D	2,144	1,705	3,900
E	26,462,124	53,632,283	76,200,000

Table 2
Ranges and Interquartile Ranges of Responses in Experiment A

	Range		Interquartile range	
Question	Intuitive	Deliberative	Intuitive	Deliberative
A	10.41	12.61	4.01	5.70
B	15.24	16.73	5.28	5.90
C	11.92	13.82	3.74	4.01
D	8.22	7.86	2.04	3.22
E	10.82	6.91	2.13	3.41

Each question was analyzed in a 2 (experiment) × 2 (decision-making strategy) between-subjects ANOVA, the dependent variable for each subject being the natural logarithm of the subject's answer divided by the correct answer. The logarithmic transformation was used to compensate for the strong positive skew of the data (Howell, 2009). None of the models had significant fit (all $F(3)$s < 1.5, all ps $> .2$). Indeed, there were no significant main effects or interactions (all $F(1)$s < 3.5, all ps $> .06$; see Figure 1 and appendices). In sum, there was no evidence that responses varied between conditions.

An additional five ANOVAs had been planned to directly test the SSH, but their theoretical justification depended on obtaining significant results from the first five ANOVAs, so they were not run.

Discussion

Experiment A failed to replicate McMackin and Slovic (2000). The blame may lie in part with differences in sample size: only 38 subjects provided usable data in Experiment A. This number would suffice for many experimental studies of pure behavior, but McMackin and Slovic had 143 participants in their numbers task.

Within-group variance of responses to these questions, as measured by interquartile

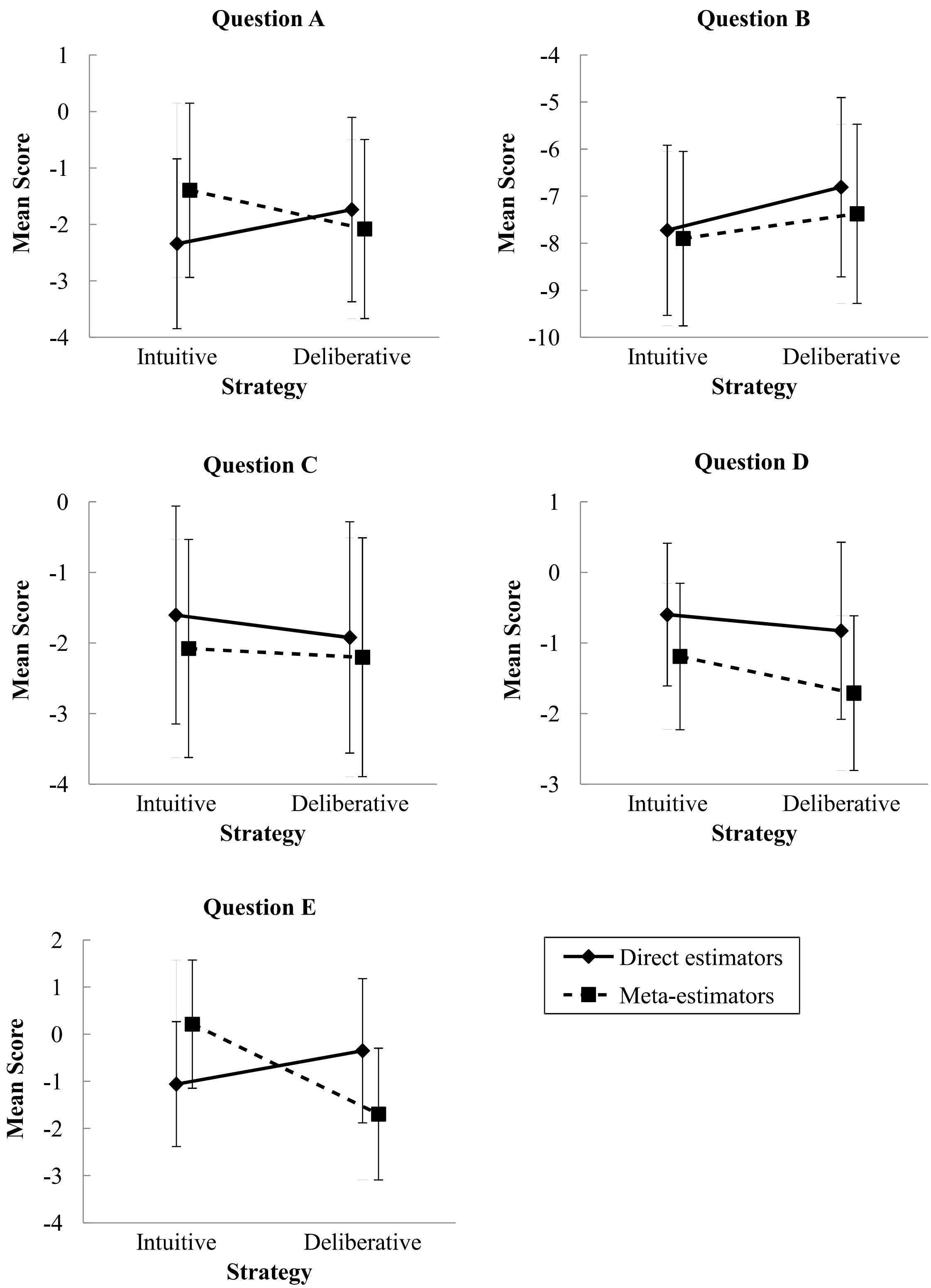

Figure 1. Mean scores (natural logarithm of response divided by correct answer) for each question and condition. Error bars depict 95% confidence intervals.

ranges and the width of confidence intervals for group means, was considerable. This comes as no surprise considering the many variables that influence estimates but were not controlled by the task design, particularly participants' relevant knowledge. For example, examining the reasons supplied for Question E suggests that, as one would expect, many subjects estimated the population of the United States in 1900 by adjusting downwards from an estimate of the current figure. But several subjects misremembered 6.5 billion, an approximation of the current *world* population, as an approximation of the current *national* population, and thus supplied an estimate of 1 billion or more. Similarly, given that the target audience of *Reader's Digest* is disjoint from the population from which subjects were drawn, many subjects were unaware of the immense popularity of that periodical. Within-group variance in this task might be reduced considerably by supplying all subjects with reference figures useful for making each estimate.

Another probable source of variance is subjects' differing familiarity with large numbers in general. The legend of the wheat and chessboard, and the passage in Plato's *Meno* in which Socrates teaches a slave that doubling the side lengths of a square quadruples its area, attest to our naive expectation that figures will remain small. It is surprising that millions of smokers can over the course of a year smoke hundreds of billions of cigarettes, and that a country whose main landmass spans no more than 3,500 miles in any direction has an area in the millions of square miles. Subjects who happen to be familiar with this tendency to underestimate may be able to compensate for it. If so, calling subjects' attention to the dangers of underestimation immediately before they perform the task, perhaps by telling them the wheat-and-chessboard story, might further reduce within-group variance.

Of course, one might also think of the failure to replicate and the insignificant fit of the linear models in terms not of large error variance but of weak effects. Notice that for three of the five questions, the estimated effect size of experiment was larger than that of strategy. This is surprising in light of the many times an effect of reasons-writing has been demonstrated in the literature versus the novelty of the meta-estimation manipulation. Originally, my greatest concern with the design of the experiment was that the instruction to meta-estimate would be mostly ignored. In fact, it may have made more a difference in subjects' estimates than reasons-writing. Perhaps the moral is that reasons-writing, rather than meaningfully helping or hurting judgment, is of little consequence, since it takes either enormous samples or judgments with no objective criterion of correctness for reasons-writing to have any measurable effect.

Appendix A

Analysis of Variance of Scores (Question A)

Source	df	MS	*F*	η_p^2	*p*
Experiment	1	2.570	.147	.002	.702
Strategy	1	9.756	.003	< .000	.959
Experiment × Strategy	1	0.704	.673	.010	.415
Error	70	16.418			
Total	74				

Appendix B

Analysis of Variance of Scores (Question B)

Source	df	MS	*F*	η_p^2	*p*
Experiment	1	2.570	.157	.002	.694
Strategy	1	9.756	.594	.008	.443
Experiment × Strategy	1	0.704	.043	.001	.837
Error	71	16.418			
Total	75				

Appendix C

Analysis of Variance of Scores (Question C)

Source	df	MS	*F*	η_p^2	*p*
Experiment	1	2.381	.222	.004	.639
Strategy	1	0.813	.076	.001	.784
Experiment × Strategy	1	0.157	.015	< .001	.904
Error	63	10.749			
Total	67				

Appendix D

Analysis of Variance of Scores (Question D)

Source	df	MS	*F*	η_p^2	*p*
Experiment	1	9.129	1.782	.027	.187
Strategy	1	2.348	.458	.007	.501
Experiment × Strategy	1	0.353	.069	.001	.794
Error	65	5.124			
Total	69				

Appendix E

Analysis of Variance of Scores (Question E)

Source	df	MS	F	η_p^2	p
Experiment	1	0.022	.003	$< .001$	.960
Strategy	1	6.405	.726	.011	.397
Experiment × Strategy	1	30.365	3.443	.048	.068
Error	68	8.820			
Total	72				

References

Audit Bureau of Circulations (2009, December 31). ACCESS ABC: eCirc for Consumer Magazines [Web front-end to a database]. Retrieved from `http://abcas3.accessabc.com/ecirc/magtitlesearch.asp`

Brehm, J. W., & Cohen, A. R. (1962). *Explorations in cognitive dissonance*. Oxford, England: Wiley.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1674.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327–339.

Fantino, E., & Esfandiari, A. (2002). Probability matching: Encouraging optimal responding in humans. *Canadian Journal of Experimental Psychology, 56*, 58–63.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*, 493–501.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*, 325–331.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytic cognition in expert judgment. *IEEE Transactions on Systems, Man and Cybernetics, 17*, 753–770.

Howell, D. C. (2009). *Statistical methods for psychology* (7th ed.). Belmont, CA: Cengage Wadsworth.

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

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 49–81). New York: Cambridge University Press.

Keysers, C., & Gazzola, V. (2009). Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology, 19*, 666–671.

Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition, 113*, 123–127.

Kogler, C., & Kühberger, A. (2007). Dual process theories: A key for understanding the diversification bias? *Journal of Risk and Uncertainty, 34*, 145–154.

MacGregor, D., Lichtenstein, S., & Slovic, P. (1988). Structuring knowledge retrieval: An analysis of decomposed quantitative judgments. *Organizational Behavior and Human Decision Processes, 42*, 303–323.

Mackay, J., & Eriksen, M. (2002). *The tobacco atlas*. Brighton: Myriad Editions Limited. Retrieved from `http://www.who.int/tobacco/statistics/tobacco_atlas/en`

McMackin, J., & Slovic, P. (2000). When does explicit justification impair decision making? *Applied Cognitive Psychology, 14*, 527–541.

Mullen, B. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology, 21*, 262–283.

Simon, H. A. (1979). Rational decision-making in business organizations. *The American Economic Review, 69*, 493–513.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61*, 394–403.

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

Tversky, A., & Kaheman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.

Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.

Wilson, T. D. (1990). Self-persuasion via self-reflection. In J. M. Olson & M. P. Zanna (Eds.), *Self-inference processes: The Ontario Symposium* (Vol. 6, pp. 43–67). Hillsdale, NJ: Erlbaum.

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

Footnotes

1. Actually, each subject's *opinion* of the poster she chose was not collected after the experiment. Instead, the experimenters asked each subject whether she had kept the poster, whether she had hung it up, and whether she planned to take it home at the end of the current semester. But surely the subjects made these decisions without writing down reasons.

2. See Wilson and Schooler (1991) for a comparison of the decision-making processes of the subjects and those of the criterion judges.

3. In this context, the term "theory of mind" is common, but ambiguous: it can refer either to our capacity for mind-reading, however implemented, or to the implicit folk theory whose existence is posited by theory theory.

4. My answer for C is substantially lower than McMackin and Slovic's figure of 15,126,664, possibly because the actual circulation has dramatically decreased over the intervening decade. Question B in McMackin and Slovic (2000) is "How many cigarettes are consumed in the US each year?" and the given answer is 604,100,000,000; no source is cited, and I consider this figure suspect. In practice, it should not have mattered very much what numbers I took as correct answers, since in McMackin and Slovic (2000), subjects' mean estimates were typically off by an order of magnitude.