

Prediction of Consumer Behavior by Experts and Novices

J. Scott Armstrong¹

Are those who are familiar with scientific research on consumer behavior better able to make predictions about phenomena in this field? Predictions were made for 105 hypotheses from 20 empirical studies selected from *Journal of Consumer Research*. A total of 1,736 predictions were obtained from 16 academics, 12 practitioners, and 43 high school students: The practitioners were correct on 58.2 percent of the hypotheses, the students on 56.6 percent, and the academics on 51.3 percent. No group performed better than chance.

This article presents a study on the predictive value of scientific knowledge of consumer behavior. It does this by obtaining predictions from people who should be well acquainted with such knowledge, and comparing their predictions with those by people who are unlikely to have this knowledge.²

The first section of the article presents the hypotheses. A description of the prediction study is then presented, followed by results and limitations. Finally, suggestions are provided for improving the predictive value of research on consumer behavior.

Hypotheses

Consumer behavior was expected to be a field in which one could demonstrate gains in predictive validity as a result of scientific research. The *Journal of Consumer Research's* (*JCR's*) style sheet asks for substantive contributions that "lend themselves to generalization." The *Journal's* articles are among the most widely cited of all those published in business and management research. According to the *Social Science Citation Index Journal Citation Report* for 1987, *JCR* ranked first among the 51 business journals as measured by the citation "impact factor." This means that researchers draw upon the research published in *JCR* and that findings from *JCR* are communicated among academics. The *Journal* is also highly regarded by faculty (Luke and Doke 1987). In addition, the field of consumer behavior displays a strong emphasis on empirical testing of hypotheses.

¹ Scott Armstrong is professor of marketing, Wharton School, University of Pennsylvania, Philadelphia, PA. 19104. He thanks the following people for assistance: Kim Rossini and Stuart Neuman aided in writing descriptions of the studies; Martha Lightwood copy edited the survey materials; Wende Gladfelter and Kenneth Weissman administered some of the surveys; Mitzi Vorachek arranged for data collection at Strath Haven High School; Larry Bortner assisted in coding the data; and Kenneth Weissman aided in the analysis of the results and commented on various drafts. Useful comments were received from many people, among them were Dennis A. Ahlburg, David A. Bessler, Russell W. Belk, Stuart Bretschneider, A. S. C. Ehrenberg, George H. Haines, Jr., Steven J. Hoch, Morris Holbrook, Raymond Hubbard, Shelby Hunt, Jacob Jacoby, David L. Kendall, Jerome B. Kernan, Joel Kupfersmid, Donald Lehmann, John D. C. Little, Richard Oliver, Brian Ratchford, William Ross, John R. Rossiter, Terence A. Shimp, and three anonymous reviewers.

² The gain in predictive validity might be viewed as a measure of scientific achievement. I administered a questionnaire to a convenience sample of academics at the Marketing Science Conference at Duke University in March 1989. Most academics agreed with the statement that predictive validity provides one way to assess scientific achievement. On a scale from 1 (disagree strongly) to 5 (agree strongly), the 19 respondents averaged 4. The idea of using the relative predictive ability of experts over naive subjects as an operational measure of scientific achievement also met with substantial agreement (3.8 on the scale).

Three groups thought to have varying knowledge of and ability to predict consumer behavior are academics, marketing practitioners, and consumers in general. A substantial number of academics spend much time studying the scientific work done in consumer behavior. The *Journal* provides a focal point for their efforts. Academics use their scientific knowledge of consumer behavior as a basis for such activities as teaching, consulting for corporations, and testifying in legal and regulatory proceedings. In contrast, marketing practitioners are not likely to be as familiar with this scientific literature. However, practitioners gain expertise through their experience. This expertise might help them make accurate predictions of consumer behavior. Finally, while few scientific studies on consumer behavior reach the general public, consumers' personal experiences should help them predict certain aspects of consumer behavior.

The above discussion implies two hypotheses about consumer behavior predictions. The first deals with the value of expertise, while the second examines the value of scientific knowledge as the source of expertise.

H1: Experts can make more accurate predictions than novices.

H2: Academics can make more accurate predictions than practitioners.

In a survey of the editorial board of *JCR*, an overwhelming majority agreed with these hypotheses.³

Design of the Study

I presented descriptions of studies to subjects and asked them to predict the outcomes. This section describes the selection of research studies, the procedure used to describe the studies, the wording of the questions, and the selection of subjects.

Selection of Studies

I examined predictive ability in situations that were of interest to academics by using studies from *JCR*. This would presumably give the academics an opportunity to use their scientific knowledge of consumer behavior for making predictions.

The following criteria were used for the selection of studies from *JCR*:

1. A 10-year period was chosen, starting with the June 1977 issue and using the June issue from each year thereafter. In the event that too few articles met the criteria in a June issue, the September issue was used. Two articles were selected from each year.
2. Only articles with fewer than 10 pages were included.
3. The study had to empirically test at least one hypothesis.
4. The studies were described clearly and would not require much technical knowledge on the part of the subjects.

To obtain 20 articles that met the four criteria, we examined 94. Thirty-two percent exceeded 10 pages, 40 percent did not test any hypotheses, and 6 percent lacked clarity. The selected studies are listed in Exhibit 1.

³ Of the 43 board members who responded, 41 agreed with the hypothesis that experts would predict more accurately than naive subjects, one thought there would be no difference, and one believed that the experts would be the least accurate. Eighty-one percent of the board members agreed that the academics would be more accurate than practitioners; only two members disagreed, one of whom suggested that the academics would be confused by their many theories.

Exhibit 1

The 20 Studies from *Journal of Consumer Research*

Author (Year)
Andreasen (1985)
Belk, Bahn, and Mayer (1982)
Bither and Wright (1977)
Goldberg and Gorn (1978)
Hirschman (1979)
Johnson and Russo (1984)
Kardes (1986)
Kourilsky and Murray (1981)
Krishnamurthi (1983)
Miller and Sturdivant (1977)
Moschis and Moore (1979)
Painton and Gentry (1985)
Petty, Cacioppo, and Schumann (1983)
Scott and Yalch (1980)
Shimp and Bearden (1982)
Smead, Wilcox, and Wilkes (1981)
Swinyard and Coney (1978)
Ursic, Ursic, and Ursic (1986)
Westbrook (1980)
Yalch and Elmore-Yalch (1984)

Descriptions of the Studies

To avoid a task that was too onerous for the subjects, I restricted the description of each study to two pages. This description included only information that was related to the present study. Outcomes were excluded because they were what the subjects were asked to predict.

Six steps were taken, in the following order: (1) a research assistant summarized each of the 20 studies, (2) I edited the summaries, (3) a professional editor copy edited the revised versions, and (4) then I sent each description to the original author, asking if it was correct and fair. Replies were received from authors of all but one article. The authors considered the descriptions to be correct and fair. They also sent suggestions for changes, and these were incorporated in the final descriptions. (5) We pretested each description with at least three subjects. (6) A second research assistant reviewed and proofread the descriptions.

In view of the difficulty of the prediction task, we sent only five of the 20 studies to each subject. The study descriptions were grouped into batches and sent out according to a systematic sampling plan. On average, each subject made about 26 predictions.

Questions

The questions stated the proposed hypotheses and asked the subjects to predict whether each hypothesis tended to be true or false. A "do not understand" response category was included.

There were 105 hypotheses, and two examples are given below.

1. "The more frequently an adolescent interacts with peers about consumption matters, the greater is the tendency to use peer preferences in evaluating products" (Moschis and Moore 1979).

2. "A person will be more satisfied with their recently purchased car if the car met or exceeded his prior expectations" (Westbrook 1980).

The results of the studies supported these two hypotheses.

The original wording was used for all hypotheses, and it typically indicated the original authors' statements on the directional effects of the hypotheses.

A research assistant coded the actual outcomes for each study to enable subjects' predictions to be compared with them. Our survey of the original authors sought verification that this coding of outcomes was correct. After I received the authors' confirmation, a second-research assistant checked the coding. All procedures yielded agreement on all items.

Subjects

I was interested in obtaining responses from academics likely to be exposed to scientific research on consumer behavior. Systematic sampling was used to select 100 academics from the 1986 membership directory of the Association for Consumer Research (only academic addresses were used). Although the level of expertise varies greatly among the persons in this group, their overall expertise in consumer behavior should be high.

For the practitioner group, I wanted subjects who worked with marketing problems but who were unlikely to be familiar with scientific research on consumer behavior. Systematic sampling was used to select 100 practitioners from the 1984 American Marketing Association membership directory (academic addresses were excluded).

A self-addressed envelope was enclosed in the original mailings, and two postcard reminders were sent. Replies were received from 20 academics and 13 practitioners.

Subjects were asked whether they had previously read each of the studies. Two academics had read most of the studies; because few of their predictions were usable (three or fewer), all responses from these subjects were excluded. Two academic respondents said that they did not understand all of the hypotheses, so they were also excluded. This left 16 academics. One practitioner was excluded because he said that he did not understand the instructions, which reduced the number of practitioners to 12. Six academics and one practitioner reported reading one or more studies, and their responses for these studies were excluded.

Responses from the novices (tenth- and eleventh grade students in honors English classes at Strath Haven High School in Wallingford, PA) were obtained in a group setting. Because these students were expected to have little intrinsic interest, an incentive of \$25 was provided for the individuals in each of the two classes who made the most accurate predictions. All 43 students in the classes responded.

Results

Contrary to the first hypothesis, the experts (academics and practitioners) did not make more accurate predictions. They were correct in 52.6 percent of their predictions, while the novices were correct in 56.6 percent.

The second hypothesis was also false. Academics were not more accurate than practitioners. In fact, the academics' accuracy, 51.3 percent, was the poorest of the three groups.⁴ The results are summarized in Table 1.⁵

⁴ One academic predicted that all hypotheses were false and, as a result, achieved an accuracy of only 13.3 percent for the five studies in his sample. To avoid having results that were unduly influenced by one individual's performance, I excluded his results from the analyses. This improved the academic accuracy from 48.6 percent to 51.3 percent.

Table 1
Predictions by Experts and Naive Subjects

	Percentage Correct*	Percentage expected by chance	Percentage expected by JCR board
Experts:			
Academics	51.3 (360)	58.1	80.0
Practitioners	<u>58.2 (270)</u>	<u>58.8</u>	<u>65.0</u>
Experts' totals	52.6 (630)	58.4	73.7
Naive subjects:			
High school students	56.6 (1,106)	57.9	55.0

* Numbers in parentheses indicate the number of predictions.

Comparisons against Chance

The hypotheses were all worded so that "true" corresponded to the original authors' hypotheses about the direction of the results. This was done because an attempt to randomize the wording produced awkwardly worded hypotheses for many items.

The subjects did not know that the wording reflected the original authors' hypotheses about the direction of the effects, nor did they know the base rate (the proportion of hypotheses for which "true" was the correct answer). Had the subjects known that the base rate exceeded 50 percent, they could have used this to improve their accuracy. For example, by assuming that researchers typically found what they were looking for and, as a result, predicting "true" for all hypotheses, a subject would have been correct for 74.2 percent of the predictions. Subjects who gave a higher percentage of true answers would be expected to achieve a higher level of accuracy. In fact, the three groups differed little in the frequency with which they predicted true. The practitioners predicted true slightly more often (68.1 percent) than the academics (66.7 percent) and the students (66.3 percent).

To establish the accuracy level that one could expect by chance, I assumed that subjects had no predictive ability. This allowed the chance level to be calculated from

$$P_t \times A_t + P_f \times A_f$$

where P_t is the percentage of times true was predicted, A_t is the percentage of times that the actual outcome was true, P_f is the percentage of times false was predicted, and A_f is the percentage of times that the outcome was false.

The expected levels of chance predictions were almost identical among the three groups (see the next to last column in Table 1). No group performed better than chance. It is surprising that predictions by the academics were significantly less accurate than chance ($Z = 2.58$; $p < .01$).

Comparisons against Experts' Expectations

To provide benchmarks for the evaluation of predictive validity, I conducted a survey of *JCR*'s editorial review board. The questions that were posed are as follows.

⁵ To guard against mistakes, the summaries were done by hand as well as by spreadsheet. Summaries were prepared independently by myself and two research assistants. We also used different approaches to testing statistical significance. Each procedure yielded the same conclusions.

We plan to ask three groups of subjects to predict the outcomes of hypotheses for 100 studies in a sample drawn from *JCR*. Specifically, they would be asked to state whether a hypothesis would tend to be true or false. Please estimate the ability of these groups to make correct predictions of the direction of the results. (Someone with no information at all would expect to be correct about 50 percent of the time by chance alone.) (a) How often would you expect correct predictions by academic experts with a strong interest in consumer behavior, as judged by membership in the Association for Consumer Research? (b) How often would you expect correct predictions by practitioners with a serious interest in marketing, as judged by membership in the American Marketing Association? (c) How often would you expect correct predictions by naive subjects, who would be selected from among intelligent high school students?

Responses were received from 43 board members, an 86 percent return rate. Their responses appear in the last column of Table 1 and serve as a basis for comparison.⁶ The levels of accuracy are surprising in light of these expectations. In particular, the academics' accuracy level (51.3 percent) was much lower than the board members had expected (80 percent).

Additional Statistical Tests

It was possible that differences in the number of hypotheses predicted by each subject may have affected the results. To examine this, I calculated an accuracy score for each individual. I then made a comparison between experts and novices using the extended median test (Siegel and Castellan 1988). This test reduces the impact of extreme responses. The students tended to be more accurate than the academics, but the superiority was not statistically significant.

The predictions were then summarized for each hypothesis (rather than by using individual predictions). Students were most accurate for 31 of the hypotheses, practitioners for 31, and the academics for only 20 (the rest were ties). These results were consistent with those on the poor predictive validity of the academics, but the difference was not statistically significant.

When the study was used as the unit of analysis (i.e., examining the average predictive validity by study), the results again failed to support the hypotheses. Here, the high school students were more accurate than the academics in 11 of the 20 studies. The students did better than the practitioners in 10 of 20 studies.

Limitations

Researchers may be able to make predictions only about their own specialties within the field. If their specialties were narrow, their knowledge might not have been useful in analyzing studies from other specialties.⁷ On the other hand, academics speak of consumer behavior as a body of knowledge, some teach courses on consumer behavior, and textbooks treat consumer behavior as a field.

The emphasis that *JCR* places on "new knowledge" suggests that these studies might be new areas and, as a result, the experts might not be aware of them. This does not seem highly plausible because, in 19 of the 20 studies, the authors made extensive use of prior research in developing their hypotheses and interpreting their results.⁸

⁶ The board members' responses differed greatly. The range was from 40 percent to 90 percent for academics' accuracy (question [a]), from 15 percent to 90 percent for practitioners' accuracy (question [b]), and from 20 to 80 percent for high school students' accuracy (question [c]).

⁷ William Ross examined the materials and classified the studies into six categories: cognitive psychology (seven studies), developmental psychology (five), social psychology (three), anthropological/ sociological (three), ethics (one), and marketing (one).

⁸ The exception (Ursic et al. 1986) was based on an implied, commonsense hypothesis.

The criteria used in selecting papers may have led to biases. Shorter and less complex papers were selected from JCR. Academics might have had an advantage over the other two groups if the papers had been more complex.

No information was obtained on the ways in which subjects made predictions. The academics might not have shown sufficient care in applying their knowledge effectively. Of course, this same argument could be used for the practitioners and novices.

Although the sample sizes were large with respect to predictions (1,736), hypotheses (105), and total respondents (71), they were limited with respect to subgroups of respondents (e.g., 16 academics) and the number of studies (20). On the other hand, statistically significant results were obtained. Given that these results were in the wrong direction, it would require a very large sample to reverse the findings.⁹

The above limitations raise the possibility that the poor performance of the academics was due to the design of the study. Other limitations, however, favored the academics. One is that the experts were more selective in their forecasts. Forecasters can typically improve their accuracy by refusing to make predictions in difficult cases. Thus, while the high school students made predictions for 98 percent of the hypotheses, the practitioners and academics made predictions for 91.8 and 89.9 percent, respectively. Other conditions favoring academics were as follows: (1) the situations were selected by academic researchers, (2) it is likely that those academics who responded had more expertise than the nonresponding academics, and (3) academics may have read the papers and forgotten that they had done so.

Discussion

This study examined the value of scientific research in making predictions about consumer behavior. The reasons for success or failure were not studied. For example, the subjects' failure to obtain good predictions could have occurred because the research described in the 20 papers had little replicability.¹⁰ Hubbard and Armstrong (1994) found that only 15 percent of published replications in the major marketing journals provided full support for the original findings, while 60 percent provided conflicting results. Perhaps the academics had difficulty in predicting outcomes because they were able to think of competing hypotheses for each situation. Indeed, it often happens that more than one hypothesis seems applicable. Huck and Sandler (1979) found plausible hypotheses to explain the results for all but one of 95 published studies.

Other possible explanations are that the scientific research on consumer behavior does not yield findings that can be generalized, that the generalizations do not yield unambiguous predictions, that the findings are not effectively written, that the scientists do not understand the research in this area, that researchers cannot effectively use the knowledge, or that they refuse to believe the findings. It is interesting that, for the 14 papers read by six of the academics, their predictive accuracy was only 69.4 percent.

Jacoby (1978, p. 87) in his 1975 presidential address to the Association for Consumer Research, concluded that too large a proportion of the consumer research literature is of little value. He proposed that researchers in consumer behavior should demonstrate the predictive value of their findings. Given the results of this study, Jacoby's advice still seems relevant.

Journals might take steps to favor publication of research that demonstrates predictive value. One way to encourage this would be through a structured reviewers' rating sheet that asks whether tests of predictive validity

⁹ For example, if the sample size were doubled, and the direction of the results turned out to be opposite to those obtained in this study, the conclusions would not change.

¹⁰ None of the 20 studies in our sample was replicated. This was confirmed by a letter that was sent to the authors of each study.

have been made, whether the study has been replicated, whether the findings are clear enough so that others can use it, and whether the conditions under which the findings apply have been well specified.¹¹

The scientific research on consumer behavior has not been shown to produce findings that can be generalized so that they are useful for prediction.¹² This study does not, however, argue against research on consumer behavior. A lack of general findings implies a need for research in each situation. Hoch (1988) found that, although marketing managers were, in general, less accurate than novices in making predictions about consumer interests and opinions, those who were familiar with research specific to these situations were more accurate.

Conclusions

Contrary to the hypotheses, experts were not more accurate than novices, and academics were not more accurate than practitioners in this study based on 1,736 predictions about consumer behavior. The expectations of all *JCR* board members were incorrect for at least one of the two hypotheses in this study.

None of the subject groups performed better than chance. The academics did significantly worse than chance.

The results differed significantly from the expectations of *JCR* board members. This was especially so for the academics; the board members expected 80 percent accuracy, but the academics achieved only 51.3 percent. This was significantly worse than chance.

This study should be extended by using a larger sample of experts, additional studies on consumer behavior, and different procedures for selecting the studies and the subjects. Future research might also try to understand why academics are unable to provide more accurate predictions. Knowledge of consumer behavior should be useful. Currently, however, we have little understanding of the conditions under which scientific knowledge of consumer behavior has predictive value.

References

- Andreasen, Alan R. (1985), "Consumer responses to dissatisfaction in loose monopolies," *Journal of Consumer Research*, 12 (September), 135-141.
- Belk, Russell W., Kenneth D. Bahn and Robert N. Mayer (1982), "Developmental recognition of consumption symbolism," *Journal of Consumer Research*, 9 (June), 4-14.
- Bither, Stewart W. and Peter Wright (1977), "Preferences between product consultants: Choices versus preference functions," *Journal of Consumer Research*, 4 (June), 3947.
- Goldberg, Marvin E. and Gerald J. Gorn (1978), "Some unintended consequences of TV advertising to children," *Journal of Consumer Research*, 5 (June), 22-29.
- Hirschman, Elizabeth C. (1979), "Differences in consumer purchase behavior by credit card payment system," *Journal of Consumer Research*, 6 (June), 58-65.
- Hoch, Stephen J. (1988), "Who do we know: Predicting the interests and opinions of the American consumer," *Journal of Consumer Research*, 15 (December), 315-324.
- Hubbard, Raymond and J. Scott Armstrong (1994), "Replications and extensions in marketing: Rarely published but quite contrary," *International Journal of Research in Marketing*, 11, 233-248.

¹¹ The *Journal* did not use a structured rating sheet for referees until 1988, and the current rating sheet does not directly address these issues.

¹² One way to interpret this result is to conclude that the knowledge has not been translated into theories that have predictive validity.

- Huck, Schuyler and H. M. Sandler (1979), *Rival Hypotheses*. New York: Harper & Row.
- Jacoby, Jacob (1978), "Consumer research: A state of the art review," *Journal of Marketing*, 42, 87-96.
- Johnson, Eric J. and J. Edward Russo (1984), "Product familiarity and learning new information," *Journal of Consumer Research*, 11 (June), 542-550.
- Kardes, Frank R. (1986), "Effects of initial product judgments on subsequent memory-based judgments," *Journal of Consumer Research*, 13 (June), 1-10.
- Kourilsky, Marilyn and Trudy Murray (1981), "The use of economic reasoning to increase satisfaction with family decision making," *Journal of Consumer Research*, 8 (September), 183-188.
- Krishnamurthi, Lakshman (1983), "The salience of relevant others and its effect on individual and joint preferences: An experimental investigation," *Journal of Consumer Research*, 10 (June), 62-72.
- Luke, Robert H. and E. R. Doke (1987), "Marketing journal hierarchies: Faculty perceptions" *Journal of the Academy of Marketing Science*, 15, 74-78.
- Miller, Kenneth E. and Frederick D. Sturdivant (1977), "Consumer responses to socially questionable corporate behavior: An empirical test," *Journal of Consumer Research*, 4 (June), 1-7.
- Moschis, George P. and Roy L. Moore (1979), "Decision making among the young: A socialization perspective," *Journal of Consumer Research*, 6 (September), 101-110.
- Painton, Scott and James W. Gentry (1985), "Another look at the impact of information presentation format," *Journal of Consumer Research*, 12 (September), 240-244.
- Petty, Richard E., John T. Cacioppo and David Schumann (1983), "Central and peripheral routes to advertising effectiveness: The moderating role of involvement," *Journal of Consumer Research*, 10 (September), 135-146.
- Scott, Carol A. and Richard F. Yalch (1980), "Consumer response to initial product trial: A Bayesian analysis," *Journal of Consumer Research*, 7 (June), 32-41.
- Shimp, Terence A. and William O. Bearden (1982), "Warranty and other extrinsic cue effects on consumers' risk perceptions," *Journal of Consumer Research*, 9 (June), 38-45.
- Siegel, Sidney and N. John Castellan, Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
- Smead, Raymond J., James B. Wilcox and Robert E. Wilkes (1981), "How valid are product descriptions and protocols in choice experiments?" *Journal of Consumer Research*, 8 (June), 37-42.
- Swinyard, William R. and Kenneth A. Coney (1978), "Promotional effects on a high- versus low-involvement electorate," *Journal of Consumer Research*, 5 (June); 4148.
- Ursic, Anthony C., Michael L. Ursic and Virginia L. Ursic (1986), "A longitudinal study of the use of the elderly in magazine advertising," *Journal of Consumer Research*, 13 (June), 131-133.
- Westbrook, Robert A. (1980), "Intrapersonal affective influences on consumer satisfaction with products," *Journal of Consumer Research*, 7 (June), 49-54.
- Yalch, Richard F. and Rebecca Elmore-Yalch (1984), "The effect of numbers on the route to persuasion," *Journal of Consumer Research*, 11 (June), 522-527.