DOI: 10.1002/ffo2.153

COMMENTARY

WILEY

Exploring scenario planning through controlled experimentation: Commentary on Derbyshire et al. (2022)

Paul J. H. Schoemaker

Wharton, University of Pennsylvania, Pennsylvania, Philadelphia, USA

Correspondence: Paul J. H. Schoemaker, Wharton, University of Pennsylvania, Philadelphia, PA, USA.
Email: schoemak@wharton.upenn.edu

KEYWORDS
confidence ranges, conjunction fallacy, controlled experiments, scenario planning

Funding information
None

The Derbyshire et al. (2022) target article addresses important issues related to testing and validating scenario planning in more scientific ways than through case studies, field reports, and anecdotal inferences. In my recent book on scenario planning (Schoemaker, 2022), I made several pleas to place scenario methodologies on better scholarly foundations. I also offered some further thoughts there about this method's own future, as follows:

> "*As regards future uncertainties that might shape scenario planning itself, my crystal-ball questions would be about: (i) improvements in the methodology itself; (ii) significant shifts in the domains of application; (iii) better integration with quantitative approaches; (iv) quantum enhancements via web-based apps; (v) better connections with big data and AI; and (vi) deeper integration with other academic disciplines*"

Controlled experiments can play valuable roles in advancing all the above items but especially (i), (v), and (vi). Since foresight methods come with many variations, experiments can help assess differences in efficacies. Blending humans and AI in foresight likewise needs more systematic evaluations and experimentation has already proved indispensable in dissecting what scenario planning is about cognitively. Although it will be hard to run tight treatment-control scenario tests with entire organizations, functions, or business units in field settings, cognitive aspects of scenario planning methodologies can clearly be tested experimentally. Care must be taken, however, to understand the limited aims and conclusions of such controlled variations which the target article does only in part. My commentary will discuss this further and a good starting place is to quote from the target article itself.

> "*In* Schoemaker's (1993) *experiments, the participants were MBA students and the questions (i.e., the future events considered) did not pertain to the context in which they may manage and have expertise, but instead to the number of medals won at a forthcoming Olympics for example.* Schoemaker's (1993) *study may therefore also be considered to have limited external validity. A lack of random assignment also limits internal validity in* Schoemaker (1993)".

In actuality, the subjects were active managers in their early to mid-30s completing a part-time MBA at the University of Chicago. This entailed attending up to two evening classes during regular weekdays (not weekends) over a 3-year period in person. Most were sponsored by their organizations and held full-time jobs while enrolled. As such, they would fall—in terms of age and experience—between my full-time campus students and our weekend executive MBA program for senior business leaders. I did not provide this detail in my 1993 article since my experiments focused on general cognitive aspects of selected elements of scenario planning, without any claims of testing full-blown or extended versions as normally conducted inside companies by managers or executives. Also, they were not tested on just trivia questions (like gold medals) but reflected mostly about business issues at work plus various futures facing Apple Inc in personal computers (a topic of interest to most).

1 | WHAT MY EXPERIMENTS TESTED AND WHY

My first experiment asked the managerial MBA students described above to analyze strategic planning issues they encountered at work to assure high familiarity with the scenarios they had to create (i.e., customized to their organization's concerns). The aim was to assess

how scenario thinking would alter their subjective confidence intervals for key parameters that they had themselves previously identified. Each student also asked a colleague at work to read their scenarios and likewise provide confidence ranges. This experimental design compared how a set of scenarios might impact the narratives' creator differently than recipients. This first experiment was intended as an exploratory demonstration pilot rather than designed to be a conclusive large-scale controlled study.

My second experiment investigated a different question, namely whether asking subjects to think in terms of more extreme scenario outcomes might actually shrink their confidence intervals due to lack of credulity. For that task, all subjects assessed the same three unknown uncertainties (of which future Olympic performance in number of US medals was one) and were assigned *at random* to different conditions as explained below. Questions about whether the experience base, demographics, and motivation or effort matter when testing basic cognitive processes have been examined in psychology but do not loom large as explanatory variables. Economists have yet to show, for example, that the kind of cognitive biases documented by Kahneman and Tversky (1984) vanish when subjects are given financial incentives (Hogarth & Reder, 1986). Indeed, their widely documented heuristics and biases experiments have proved remarkably robust across subjects and tasks, even though they have received criticisms as well.[1]

My experiments were conducted from that assumption since my 1993 article primarily aimed to offer a broad overview of scenario planning for the *Strategic Management Journal*. It was based on my 2-year sabbatical with Royal Dutch Shell's scenario planning group in London and my own academic background in decision sciences and strategy. The article's first half provided a conceptual overview of scenario planning, while the second half described four experiments examining different cognitive issues. At that time, the impact of scenario planning and its various design parameters had not received much experimental examination. One curiosity for me was the impact that scenario thinking might have on people's beliefs, degrees of confidence, and problem perceptions. My 1993 article articulated some key theoretical variables to test the impact of scenario planning through controlled experimentation.[2] Second, factors that determine a scenario's overall impact, such as causal strength, coherence, and plausibility, intrigued me. Third, I noted that text comprehension, metaphors (Lakoff, 1987), scripts, schemata, mental models (Schank & Abelson, 1977), and frames (Kahneman & Tversky, 1984) also needed further examination within a scenario context.

Beyond these cognitive factors, I also highlighted the need to examine how scenarios can be made to have *emotional* impact (e.g., through drama or interactive videos, or simulation) such that people could actually *feel and experience* possible futures in their gut as well (as per Deighton et al., 1989). Understanding and unraveling all these factors was a tall order then and the experiments reported in Schoemaker (1993) only addressed some of the cognitive issues mentioned above. The benefit of the experimental approach is that we can isolate a few factors deemed to be important and use controlled variation to assess their dynamics and impact on dependent variables of interest. Statistical methods and careful experimental controls also allow us to assess if and how various factors interact with each other (using factorial designs) so that scenario planning will not just remain a subjective intervention or black art. Experimentation entails judicious tradeoff between high internal validity afforded by controlling conditions and reduced external validity by not reflecting all real-world factors that could matter.

2 | A RECAP OF MY FIRST TWO EXPERIMENTS

To address some of the authors' misstatements about my own 1993 experiments as well as the presumed failed replications by Phadnis and others, let's recap what I was specifically testing to highlight that each experiment explored a different issue. My first experiment tested the following hypothesis (copied verbatim from the 1993 article):

H1: Asking subjects to construct multiple scenarios concerning the possible values of key uncertainties will widen their subjective confidence ranges about them.

As reported in more detail in table 3 of my 1993 article, systematic effects were found on statistical confidence ranges for both the MBA subjects and their chosen colleagues. The 50% ranges widened 56% on average after scenario construction for subjects and 67% for their chosen colleagues. These stretch factors were 44% and 55%, respectively for the 90% confidence ranges. Each of these four stretch results was significant at the .05 level or better under a one-tailed test. Interestingly, it did not matter much whether the scenarios were self-constructed or not. (My 1993 Appendix details the exact instructions).

I summarize these findings here to take issue with the characterizations of my experiments by the authors as quoted at the start. This first experiment (out of four in total) relied entirely on target variables selected by my evening student subjects who were full-time managers in local companies. Each developed scenarios about business issues they were intimately familiar with at work. In addition, each of these 65 students then handed their self-constructed scenarios to one knowledgeable colleague and invited that person to provide 50% and 90% confidence ranges for the same target variables as originally selected by the subject. As such, the criticism that the subjects lacked knowledge about the questions tested or were inexperienced in business do not apply to this experiment.

2.1 | Experiment II

Although the first experiment established statistically that subjective confidence ranges on average stretched when managerial subjects construct scenarios themselves about strategically relevant uncertainties, some of the ranges also narrowed at times for subjects. But such shrinkages were the exception rather than the norm and thus not strong enough to push the mean stretch ratio below one when comparing after and before ranges for either the 50% or 90%

judgments. But the possibility that ranges might shrink (see Q1 column in my original table 3) is intriguing and relevant. Indeed, such seasoned scenario practitioners as Pierre Wack (1985) from Shell and Clem Sunter (1987) from Anglo-American in South Africa, emphasized that scenario analysis may reveal that particular combinations of outcomes are unlikely or implausible. Recognizing this can be valuable in strategic planning since it concerns the important subject of scenario implausibility. In the Anglo America's case, it resulted in one global scenario being dropped from the original four suggested by their 2by2 scenario matrix. In that spirit, my second experiment was designed to test the following hypothesis.

H2: Forcing subjects to consider extreme values for key uncertainties will at some point reduce their subjective confidence ranges due to incredulity

Table 4 of my 1993 article summarized the relevant statistics in considerable detail, contrasting moderate versus more extreme departures from a subject's initial guesses. The Olympic medals question referred to in the authors' target article was one of three questions tested using a between subject design. Subjects who wrote scenarios about the Olympics showed significantly ($p < .05$) wider ranges for the scenario group than the control group. Although this finding provides additional support for *H1*, it also revealed that a more extreme scenario condition tested (which asked for ±50% deviation from the initial best guess) did not induce wider 50% confidence ranges (in support of *H2*). Also, subjects' 90% confidence ranges were not significantly wider in the ±50% condition than in the ±30% one.

Scenarios and ranges for another question tested, about the year-end level of the Dow Jones stock index, revealed a similar mixed pattern. The 90% ranges were noticeably wider ($p < .07$) for the scenario group but not so for the 50% ranges. Within the Dow Jones scenario group, however, the more extreme scenario version (50% vs. 30% departure from the initial best guess) did not produce significantly wider 50% or 90% confidence ranges. The third question tested, concerning George Bush's final election result, most strongly supported *H2*. The scenario group perceived *less* uncertainty in the election outcome than the control group. Within the Election scenario group, those contemplating ±10% deviations from expectation perceived *less* uncertainty than those considering ±5% deviations from the initial mean guess. The historical pattern that US presidential elections have usually been quite close may have made the ±10% scenarios considerably less believable to subjects than the ±5% scenarios, resulting in narrower rather than wider 50% and 90% subjective confidence ranges.

3 | COMPARISONS WITH OTHER STUDIES

Although my second experiment clearly supports the general claim that scenario construction often alters beliefs, the shift is not necessarily in the direction typically predicted in the reason generation literature to reduce overconfidence. When subjects are asked to generate more extreme scenarios, the availability effect may be countered by a lack of believability concern. Credibility may be low due to either low perceived probabilities associated with any single cause listed (such as a terrorist strike against US athletes in the Olympic camp) or unlikely conjunctions of events necessary to produce more extreme outcomes. The data of Experiment II clearly show that more extreme scenarios may in some cases reduce rather than stretch people's subjective confidence ranges.[3]

An important subsequent study by Kuhn and Sniezek (1996), which the target article failed to mention, likewise produced mixed results. Their careful experiments measured confidence in predictions made on a nine-point scale rather than via statistical confidence ranges expressed in probability terms as I did. Also, their definitions of uncertainty reduction and scenario confidence differed from mine. Fortunately, Kuhn and Sniezek provided an incisive psychological discussion about these key nuances, highlighting the importance of which metrics, methods, stimuli, subjects and conditions are tested in seemingly comparable controlled experiments. Since their study measured *confidence* in the predictions or hypotheses made rather than *perceived uncertainty*, they wrote that "these two studies may not be as contradictory as they appear."

For example, my experiment II asked subjects for a best guess first about some unknown future quantity and then to write one short scenario in which this quantity would be considerably higher than their estimates plus another scenario where it would be much lower. Following that, they were asked to provide a 50% as well as 90% confidence interval for their original estimates. This design entails at least two possible effects: (1) increased confidence in each scenario alone due to possible conjunction effects or other plausibility drivers and (2) a deeper realization about the uncertainty of their best guess due to now having two plausible but opposite scenarios at hand. Kuhn and Sniezek (1996) asked subjects to assess (3) their confidence in judgments made using a nine-point scale and (4) measured uncertainty reduction for five different scenario set ups (from single to multiple variations). Their more complex factorial experimental design was administered to 186 undergraduate students taking an introductory psychology course. This large sample size allowed multiple scenario combinations to be tested (up to five decades out), more complex stimuli designs, and very specific instructions that make a direct (apples-to-apples) comparison of results nearly impossible.

A key point here is that claims about follow-on experiments, as either confirming or refuting earlier findings, need to include appropriate caveats about ways in which the studies differed. Seldom are research extensions or attempts at replications exact replicas of the original study and nor do they need to be since additional issues or questions may have been addressed. The target article did not acknowledge the pertinent follow-on experiments by Kuhn and Sniezek (1996) and falsely claimed that Phadnis et al. (2015) reported opposite results from Schoemaker (1993). Derbyshire et al. (2022) further claimed that one of their own studies found contrary results as well, but without providing any citations at all. Noteworthy here is

that the recent commentary written by Shardul Phadnis about the target paper makes clear that their 2015 article did not seek to replicate my 1993 experiments and did not contradict its findings. Carefully evaluating study designs and internal validities becomes especially important when comparing findings *across* experiments, since fine-grained research results can easily be misrepresented when cast as broad-brush stylized findings without proper caveats.

4 | MY OTHER TWO EXPERIMENTS

I shall not discuss Experiments III and IV here in as much detail since they were not specifically criticized by the authors who called out just one of the prediction questions used in Experiment II (about Olympic medals). Neither experiment I or II tested explicitly for conjunction fallacies (Tversky & Kahneman, 1983) since this was not part of the hypotheses postulated above. The target article confused my broader discussion at the end of my 1993 article with what was hypothesized and tested specifically. Still, experiments III and IV do touch on aspects related to conjunctive events since scenarios usually tell causal stories that link specific outcomes of key uncertainties over time. The two hypotheses I tested about conjunctive events, in sequence, were as follows:

H3: Conjunction fallacies also occur when well-educated and informed business students examine typical political and economic uncertainties.

H4: Inconsistencies in subjects' belief systems transcend pairwise events and produce systematic intransitivities (over pairs of events).

The data reported in Schoemaker (1993) revealed conclusive evidence of conjunction fallacies for 25% of the cases tested in *H3* and were suggestive for the remaining cases without being definitive in Experiment III. As regards *H4*, subjects were asked to produce subjective correlation matrices for uncertainties faced by Apple Inc in the PC industry. The majority submitted correlation matrices that are mathematically impossible, in the sense that they were not *positive-definite* matrices after testing for negative eigenvalues (also known as characteristic roots). The broader purpose of these last two experiments was to test well-established cognitive biases in behavioral decision theory within the context of scenario thinking, such as coherence of beliefs about how key uncertainties are associated. Within this context, I view controlled experiments to be a useful tool to bring more rigor and deeper insight to complex methodologies that organizations or consultants deploy in field settings.

Although I never explicitly tested in my four experiments whether conjunction fallacies are either necessary or sufficient to produce widening of subjective confidence intervals, I did examine some of the building blocks needed to explore this complex issue. However, as the authors properly emphasize, concerns about the external validity of the tasks studied—as well as selecting the right experimental subjects—remain important. This also underscores the inherent limitations of using controlled experiments for testing the efficacy of complex organizational interventions. If the experiments are conducted over extended periods of time, or across diverse people, functions and departments, external validity becomes even harder to achieve. Just as we should not test rat intelligence only in maze learning settings, but also in their natural environments, the same applies to scenario planning.

Egon Brunswick (1952, 1956) and others interested in experimental design emphasized the importance of studying animals and humans as much in their natural ecologies as possible (Hammond, 1993). This research perspective when applied to human judgment and choice is known as naturalistic decision making (Klein, 2015). In general, we should try to deploy diverse scientific methods to achieve both convergent as well as discriminant validity. For organizational practices like scenario planning, this means conducting experiments, field-studies as well as deep case analyses. To develop mid-range theories as defined by Merton (1968), we shall need "thick descriptions" as advocated by Geertz (2008). In addition, we should recognize that organizational interventions like scenario planning entail artistic and political aspects that help bestow legitimacy in quarters other than academia. This raises broader issues for foresight methods in general about the kind of legitimation criteria (Habermas, 1975) to use when dealing with policy makers, consultants, and scholars.

5 | CLOSING THOUGHTS

Conducting controlled experiments about cognitive aspects of scenario planning is a valid and promising scientific tool to deepen our understandings about various mental processes at play. However, such studies can never fully capture all that happens inside organizations when deploying scenario methodologies. On rare occasions, it may be possible to conduct statistically controlled field experiments ex post, without any actual experimental manipulation ex ante, known as quasi-experimental design analysis (Campbell & Stanley, 1963). This happened, for example, when social security started to be introduced in the United States in 1940. Since it took years before various states participated, researchers could examine the effect of forced savings (via payroll withholdings) on people's discretionary savings (Katona, 1965). While economists predicted a substitution effect between forced and discretionary savings, the data revealed an increase in overall personal savings since many people became more pension minded (akin to a goal gradient effect).

Additional controlled field experiments were conducted when governments started to introduce nudge strategies to encourage healthier eating, more savings and other desirable behaviors societally without mandating them (Congiu & Moscati, 2022). Experimental field studies also enjoy wide acceptance in sociology to test mid-range theory in realistic and unobtrusive ways, with suitable controls (Baldassarri & Abascal, 2017). However, such field research approaches may not present themselves so readily for scenario planning overall, thus elevating controlled experimentation further as an important avenue to assess validity and improve practice.

I hope that additional scenario studies will be conducted in the classic treatment-control mode, while meeting the gold standards of experimental design laid out by Campbell and Fiske (1959). This will let us further examine critical issues related to the design, methods, administration, and impact of scenario planning in practice. Once enough such building blocks have been collected, across researchers, tasks, subjects, and contexts, we should be able to further validate and improve scenario planning in practice. Still, this flexible methodology will likely remain a mix of art, science, and politics by its very nature, making it difficult to get air-tight measures and proof of its value to organizations. The same complexity has bedeviled nearly all organizational inventions that businesses embraced in the past, with many having fallen by the wayside due to lack of sufficient clear evidence.

Ghemawat (2016) documented the rise and fall of more than 80 management innovations since the 1960s, while Cummings and Daellenbach (2009) examined the numerous fads and fashions in business improvements over those decades as well. Some had much staying power, such as decision analysis which during the 1960s became well-grounded in scholarship. Scenario planning, however, remained less well moored academically as a business tool even though it gained organizational traction as the outside world grew more uncertain. To continue this broad market acceptance, the field of scenario planning would do well to embrace the rigors of controlled experimentation as a powerful arrow in its academic quiver. Indeed, it may be the only arrow that practitioners and scholars can marshal scientifically to fend off attacks from skeptics and competing methods.

ACKNOWLEDGMENTS

My thanks to Shardul Phadnis, Edward J. Russo, and George Wright for their helpful feedback on earlier drafts.

DATA AVAILABILITY STATEMENT

Data sharing not applicable—no new data generated, or the article describes entirely theoretical research.

ORCID

Paul J. H. Schoemaker http://orcid.org/0000-0002-3619-5335

ENDNOTES

[1] Gigerenzer and Gaissmaier (2011) especially challenged the external validity of Kahneman and Tversky's experiments and some of my own joint research has challenged some internal validities as well (Hershey and Schoemaker, 1980).

[2] As noted by Phadnis et al. (2015), my articulation of these factors provided an essential foundation for their own study and presumably for Kuhn and Sniezek (1996) as well earlier.

[3] There were no significant differences in the accuracy of best guesses among the various conditions tested for my three trivia questions, whose answers were as follows. The USA won 94 Olympic medals in South Korea in 1988 versus 132 for the USSR, yielding a ratio of 0.712. The Dow Jones average closed at 1978 on December 31, 1988. George Bush won the 1988 election with 53.9% of the nearly 90.1 million popular votes cast.

REFERENCES

Baldassarri, D., & Abascal, M. (2017). Field experiments across the social sciences. *Annual Review of Sociology*, *43*, 41–73.

Brunswick, E. (1952). *The conceptual framework of psychology*. University of Chicago Press.

Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Campbell, D. T., & Fiske, D. W (1959). Convergent and discriminant validity by multitrait–multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally.

Congiu, L., & Moscati, I. (2022). A review of nudges: Definitions, justifications, effectiveness. *Journal of Economic Surveys*, *36*, 188–213.

Cummings, S., & Daellenbach, U. (2009). A guide to the future of strategy? *Long Range Planning*, *42*, 234–263.

Deighton, J., Romer, D., & McQueen, J. (1989). 'Using drama to persuade'. *Journal of Consumer Research*, *16*, 335–343.

Derbyshire, J., Dhami, M., Belton, I., & Önkal, D. (2022). The value of experiments in futures and foresight science as illustrated by the case of scenario planning. *Futures & Foresight Science*, e146.

Geertz, C. (2008). Thick description: Toward an interpretive theory of culture, *The Cultural Geography Reader* (pp. 41–51). Routledge.

Ghemawat, P. (2016). Evolving ideas about business strategy. *Business History Review*, *90*, 727–749.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482.

Habermas, J. (1975). *Legitimation crisis*. Beacon Press.

Hammond, K. R. (1993). Naturalistic decision making from a Brunswikian viewpoint: Its past, present, future. *Decision Making in Action: Models and Methods*, 205–227.

Hershey, J. C., & Schoemaker, P. J. H. (1980). Prospect theory's reflection hypothesis: A critical examination. *Organizational Behavior and Human Performance*, *25*, 395–418.

Hogarth, R. M., & Reder, M. W. (1986). Editors' comments: Perspectives from economics and psychology. *Journal of Business*, *Oct 1*, S185–S207.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–350.

Katona, G. (1965). *Private pensions and individual saving*. Survey Research Center, Institute for Social Research, University of Michigan.

Klein, G. (2015). A naturalistic decision-making perspective on studying intuitive decision making. *Journal of Applied Research in Memory and Cognition*, *4*, 164–168.

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

Kuhn, K. M., & Sniezek, J. A. (1996). Confidence and uncertainty in judgmental forecasting: Differential effects of scenario presentation. *Journal of Behavioral Decision Making*, *9*(4), 231–247.

Lakoff, G. (1987). *Women, fire, and dangerous things*. University of Chicago Press.

Merton, R. K. (1968). *Social Theory and Social Structure (1968 enlarged ed.)*. Free Press.

Phadnis, S., Caplice, C., Sheffi, Y., & Singh, M. (2015). Effect of scenario planning on field experts' judgment of long-range investment decisions: Effect of Scenario Planning. *Strategic Management Journal*, *36*, 1401–1411.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. A. J. Erlbaum.

Schoemaker, P. J. H. (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic Management Journal*, *14*, 193–213.

Schoemaker, P. J. H. (2022). *Advanced introduction to scenario planning*. Elgar Publishing.

Sunter, C. (1987). *The World and South Africa in the 1990s*. Human & Rousseau.

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

Wack, P. (1985). Scenarios: Uncharted waters ahead. *Harvard Business Review*, *63*(5), 73–89.

How to cite this article: Schoemaker, P. J. H. (2023). Exploring scenario planning through controlled experimentation: Commentary on Derbyshire et al. (2022). *Futures & Foresight Science*, 5, e153. https://doi.org/10.1002/ffo2.153

APPENDIX: FURTHER DETAILS ABOUT EXPERIMENTS I AND II

The subjects were first asked to select strategically important issues from work or home, entailing a time frame anywhere from 1 to 5 years (see the article for the exact instructions and procedures). The issues they generated at the beginning of the experiment included new product development, technological uncertainties, organizational design, divisional performance, competitor behavior, as well career options and personal finances. Each student subject was also asked to specify a few quantitative variables that would measure the important aspects of the particular issue they focused on. Examples of target variables included profit margins, sales, market share, budget size, returns on investment, salary, wealth creation, and so forth.

The confidence ranges that subjects were asked to estimate later in the experiment concerned "strategic variables" they deemed to be of personal or professional interest, due to entailing significant stakes, uncertainty, and complexity. A within-subject design was used consisting of the following three steps, completed over a period of 6 weeks.

(i) Submission of best estimates, as well as 50% and 90% confidence ranges by the subject and a suitable colleague for various target variables of interest.
(ii) Construction of detailed scenarios about factors influencing these variables, with interim feedback from the instructor concerning their completeness, coherence and scope.
(iii) Repeat of the stage (i) measures concerning the means, 50% and 90% confidence ranges of the key variables originally selected by the subject.
(iv) Collect the same from a colleague after having read the subject's scenarios.

Experiment II used a *common* set of external target variables about which subjects first had to make best guesses. Following that, they were asked to rethink their best guesses by contemplating extreme outcomes. Instead of administering a detailed scenario process as in experiment I, subjects were merely asked to write a one-page scenario script, in the spirit of the classic reason generation literature (Koriat et al., 1980), that might bring about a significant change in their original estimate of the target variable. A two-stage between-subject design was used with embedded controls. In stage I, 75 MBA students were *randomly* assigned to one of the following three prediction tasks. The study was conducted during the summer of 1988 and each of the target events would be known by year-end.

1. Ratio of USA to USSR medals obtained during the summer Olympics of 1988 in Seoul, South Korea.
2. Level of the Dow Jones Industrial Average at the end of 1988.
3. The percentage of the popular vote for George Bush during the 1988 presidential election.

Each subject was then asked to develop both a "high" and "low" scenario case for the assigned variable. The "high" scenario had to be such that the target variable would be either 30% (one experimental condition) or 50% (a second experimental condition) above the subject's best guess (in their own view). In the "low" case, it would have to be 30% (or 50%) below expectation. Upon completing the two extreme scenarios (high and low), the subject was also asked to provide both a 50% and 90% subjective confidence range. One week was allowed for the completion of this take-home exercise. Stage II consisted of collecting from all subjects their 50% and 90% confidence ranges regarding all three questions. This was done after the single-question scenarios, point estimates, and ranges had already been submitted. This design permits a cross-sectional contrast between a scenario versus control group as well as between a moderate (±30%) versus more extreme (±50%) condition within the scenario group for two of the questions. For the last question (which was about an upcoming presidential election outcome), moderate extremity was set at (±5%) versus (±10%) for the more extreme scenario condition.