



The value of experiments in futures and foresight science: A reply

Journal:	<i>Futures & Foresight Science</i>
Manuscript ID	FFS-23-0017
Wiley - Manuscript type:	Commentary
Date Submitted by the Author:	15-Apr-2023
Complete List of Authors:	Derbyshire, James; University of Chester Chester Business School Dhami, Dhami; Middlesex University London Department of Psychology Belton, Ian; Uiversity of Strathclyde Business School Onkal, Dilek; Northumbria University Newcastle Business School
Keywords:	experimental methods, scenario planning, Critical Realism, mixed methods

SCHOLARONE™
Manuscripts

The value of experiments in futures and foresight science: A reply

James Derbyshire¹, Mandeep Dhama², Ian Belton³, Dilek Önkal⁴

1 Introduction

We thank those responding to our paper ‘The value of experiments in futures and foresight science as illustrated by the case of scenario planning’ (Derbyshire et al., 2022a). The research on which the paper was based was funded by a Research and Capacity Building Grant from the Society for the Advancement of Management Studies (SAMS) and the British Academy of Management (BAM). In addition to writing the focal paper, the project funded by the grant also saw the authors conduct three controlled experiments related to scenario planning, findings from which are to be reported in two further papers that are currently under review. We thank SAMS and BAM for their support in developing an area of research which is rapidly increasing in importance.

In the UK, scenarios featured centrally within official government decision-making processes during the coronavirus pandemic and were referred to in the media daily. The pandemic confirmed that, if it were ever possible to deny fundamental uncertainty or that new tools and methods are needed to tackle it, that is no longer the case. The dangers associated with closing the future to make it amenable to standard forecasting tools and probability-based decision aids are clear. Closing the future to make it amenable to these standard tools leads to blindsiding by surprises. Yet, recognising the dangers associated with tools that are presently more mainstream is one thing and evidencing the efficacy of alternative tools quite another. Experiments should be a tool in the methodological toolbox of those searching for this evidence for the reasons we outlined in our paper. Below we provide some reflections on four expert responses to our paper.

2 Valuing experiments in futures and foresight science: Reflections on four expert responses

Rowland and Spaniol (2022) highlight the potential for increased emphasis on experimentation to fragment and divide the field of FFS. But is this so if the underlying ontology adopted is one of critical realism rather than naïve empirical/logical positivism? There would indeed be potential for experiments underpinned by empirical/logical positivism to lead to divisions. Under the strict Popperian falsification-based view associated with empirical/logical positivism, a single experiment refuting a particular finding would falsify it. Yet, as reflected in the replication crisis facing the social and behavioural sciences, there are so few decisive tests, and as we noted, it is exceedingly difficult in an organisational setting to replicate exactly a previous experiment. Moreover, what really matters is the cumulative body of evidence and not the outcome of any one experiment. This is why we advocated the use of conceptual rather than direct replication—which, we believe, is less likely to lead to disagreements about the relative merits of others’ experiments that Rowland and Spaniol (2022) envisage.

From another perspective, we believe that divisiveness is useful if it facilitates discussion of disagreements and efforts to seek solutions. Much more problematic is a field in which everyone agrees. That said, we agree with Merton (1942) that scientific fields do not develop in the ‘rational’ way those adopting the empirical/logical positivist ontology think they do. As noted in Derbyshire (2017) by reference to Kuhn (1962), scientific fields tend to go through alternating periods of stability and disruption, which are reflected in periods of agreement and disagreement on theory and methods. For a new paradigm to emerge a field may need to undergo a period of disruption and disagreement.

¹ Chester Business School, University of Chester, corresponding author, email: j.derbyshire@chester.ac.uk

² Dept. of Psychology, Middlesex University

³ Strathclyde Business School, Strathclyde University

⁴ Newcastle Business School, Northumbria University

1
2
3 This leads us to ask several questions: In which part of this process of evolution is the field of
4 scenario planning presently? And, similarly, what is the current evolutionary status of the field of FFS
5 that subsumes scenario planning? Is there currently too much agreement rather than too little? Might a
6 period of divisiveness not be welcome? If academia has a value that the domain of practice does not
7 have, it is in its ability to generate paradigm shifts through disagreement, questioning and dialectics.
8 Should these not be seen as a welcome by-product of increased emphasis on experimentation rather
9 than as something undesirable? Perhaps greater divisiveness is just what is needed. Which brings us to
10 Schoemaker's (2023) response to our paper.
11

12 We agree that Schoemaker's (1993) theoretical suggestion—i.e., that scenario planning's effect
13 may be due to the conjunction fallacy—was not directly tested for in his 1993 paper. Yet, this theoretical
14 suggestion emerged from the empirical tests that were conducted therein. We accepted the theoretical
15 suggestion and have tested it in a fully controlled experiment conducted as part of our broader project,
16 with the results to be outlined in a paper currently under peer review. Schoemaker's (2023) response is
17 written as if our paper (Derbyshire et al., 2022a) is one that directly responded to and refuted
18 Schoemaker (1993). Schoemaker (2023) gives the impression that our intention was simply to assert
19 that the conjunction fallacy is not the basis for scenario planning's effect, without ever publishing the
20 evidence underpinning that assertion. That was not the purpose of our paper. We should add that, it is
21 the cumulative body of evidence that matters, not the findings from any one study (no matter how well
22 it is conducted). We did not intend to suggest by briefly mentioning our forthcoming experimental
23 findings that they are sufficient alone to falsify Schoemaker's (1993) theoretical suggestion about the
24 conjunction fallacy. We were simply using our findings in relation to that suggestion to illustrate how
25 experimental findings may provide incremental improvements to the 'weight of evidence' (Derbyshire
26 et al., 2022b) and the direction in which it leans, which cumulatively (e.g., when added to those of
27 Phadnis et al., 2015) determine what is currently thought to be true.
28

29 Schoemaker (2023) suggests that in Derbyshire et al. (2022a) we 'did not acknowledge the
30 important follow-on experiments by Kuhn and Sniezek (1996) and falsely claimed that Phadnis et al.
31 (2015) reported opposite results from Schoemaker (1993)' (Schoemaker, 2023, p.7). We did not
32 acknowledge the important work of Kuhn and Sniezek (1996), of which we are very much aware,
33 because, as we have stated, the purpose of the focal paper was not to write up our experimental findings.
34 Kuhn and Sniezek (1996) are fully acknowledged in the paper that reports our experiment on the
35 conjunction fallacy. More importantly, Schoemaker (2023), in suggesting here that we made a 'false
36 claim', has made one of his own in relation to what we said about Phadnis et al. (2015). Firstly, we
37 noted in our paper that Phadnis et al. (2015) did not use the same scenario process as Schoemaker
38 (1993). Secondly, rather than claiming that Phadnis et al. (2015) 'reported opposite results', we stated
39 that Phadnis et al. (2015) 'did not replicate' Schoemaker (1993).
40

41 This usefully highlights the difference between a conceptual replication—or 'conceptual
42 extension' as Twang and Kwan (1999) alternatively call it (more on that subject later)—and a direct
43 replication. When we said that Phadnis et al. (2015) 'did not replicate' Schoemaker (1993), we were
44 referring to Phadnis et al. (2015) as a conceptual replication/extension of Schoemaker (1993), which it
45 indeed was as acknowledged in Phadnis' (2022) response to our paper. Even if Phadnis et al. (2015)
46 did not seek directly to replicate Schoemaker (1993), it was a conceptual replication/extension of it. If
47 the conjunction fallacy really is the basis of an effect from scenario planning, we might expect a
48 conceptual replication/extension such as that of Phadnis et al. (2015) to have shown similar results to
49 those in Schoemaker (1993) that prompted this theoretical suggestion in the first place. This illustrates
50 very nicely how conceptual replication/extension can add to the weight of evidence, which was the
51 point we were trying to make by briefly referencing Schoemaker (1993) and Phadnis et al. (2015) in
52 our paper. Whether we made that point clumsily or not is for others to judge, but we certainly did not
53 make the 'false claim' suggested.
54

55 While our advocacy of conceptual replication may have lowered the bar in terms of exactly
56 replicating what was done in an original study, it is still key to establishing a cumulative body of
57
58

1
2
3 evidence through experimentation that experiments be adequately reported. Their reporting should
4 contain enough information for comparisons, caveats, and distinctions between studies to be drawn.
5 This adequate reporting of experiments is needed because, as Schoemaker (2023, p.7) correctly notes,
6 'claims about follow-on experiments either confirming or refuting early findings need to include
7 appropriate caveats about ways in which the studies differed'. We agree wholeheartedly and therefore
8 welcome the clarifications about Schoemaker (1993) in Schoemaker (2023). However, we suggest it
9 might have been better to include these clarifications in Schoemaker (1993) in order to allow for the
10 caveats to which Schomaker (2023) alludes. Claims about follow-on experiments either confirming or
11 refuting early findings need to include appropriate caveats about ways in which the studies differed,
12 exactly as Schoemaker (2023) states, but this can only happen if the original study (e.g., Schoemaker,
13 1993) provides the detail needed to make the caveats.
14
15

16 That said, we agree that many a managerial tool has waxed and waned down the years due to
17 lack of empirical support. We agree that experiments are one way to defend a tool such as scenario
18 planning against the criticism that it lacks empirical support, but we add two caveats in this regard.
19 Firstly, the continued demand for scenario planning, which now spans decades, and the fact that it is a
20 primary tool used by government for decision-making under uncertainty, suggests there is more to it
21 than a mere managerial fad. There are very few tools for considering the future that do not require the
22 future to be closed and scenario planning is one, which distinguishes it from others. Secondly, under
23 the critical realist ontology we set out in Derbyshire et al. (2022a), to which many in FFS would adhere,
24 experiments are just one method for evidencing the efficacy of a tool such as scenario planning. It is
25 useful when used in conjunction with other tools such as case study, as we implied in our paper. We
26 agree with Schoemaker (2023) about the need for thick descriptions through case studies, which should
27 be carried out alongside experiments.
28
29

30 Moving onto the two remaining responses to our paper, Salo (2023) elaborates on and
31 summarises very well the many difficulties associated with conducting experiments, and even more so
32 replications of experiments, on a tool such as scenario planning that is used in highly idiosyncratic
33 organisational contexts. An important danger highlighted by Salo (2023) is that of increased emphasis
34 on experimentation giving rise to experiments that have limited correspondence with the settings in
35 which scenario planning might naturally occur. In other words, there is a danger of conducting scenario-
36 planning exercises under circumstances in which they would not otherwise occur, and which therefore
37 badly replicate the real-world context in which they might happen, just for the sake of increased
38 experimentation. Such artificially motivated scenario planning conducted simply to increase
39 experimentation would lack external validity. It would certainly fall foul of the principles for designing
40 field experiments highlighted by Phadnis (2022). It is for all these reasons that we emphasised the
41 importance of representative design in our paper. See Dhami et al. (2004) for a broader discussion on
42 the importance of representative design in relation to the experimental method.
43
44

45 The tensions at play here come starkly into focus when we consider that Salo (2023) quite
46 rightly also highlights the level of granularity (and, therefore, the increasing sample size and effort)
47 needed to test anything other than a quite general and high-level hypothesis, such as that related to the
48 conjunction fallacy. Once one starts trying to take account of idiosyncratic contexts and their varying
49 effect on the outcome from scenario planning, sample size and other considerations can become very
50 taxing. In other words, once one attempts to isolate the specific procedures, contexts and circumstances
51 associated with desirable effects by having a more elaborate experimental set up involving many
52 dimensions of analysis, the effort needed to conduct experiments increases rapidly, and concomitantly,
53 the feasibility of implementing it in a sufficient number of real settings decreases rapidly.
54
55

56 The tension here is between the need for external validity and the need for an experiment to be
57 practically feasible. We highlighted this tension in relation to our discussion of online platforms as a
58 tool for experimentation. On the one hand, they make experiments more feasible from a practical
59 standpoint. On the other hand, however, experiments conducted using online tools are open to the
60 accusation of lacking external validity when it comes to the testing of tools such as scenario planning

1
2
3 that are typically implemented offline, and which have a group-based social aspect that is difficult to
4 replicate online. Rather than abandoning the experimental method, however, any issues around external
5 validity and practical feasibility simply reinforce the need for a multi-pronged approach to research,
6 which would include experiments conducted in both a real and online context, alongside case studies
7 and action research, and other methods able to provide a deeper dive and assist in uncovering and
8 fleshing out specific causal mechanisms, thereby generalising findings.

9
10 For still further reinforcement of the points made in Derbyshire et al. (2022a), we can draw on
11 the work of Twang and Kwan (1999), who also consider replication and theory development from a
12 critical realist perspective. In our paper we distinguished between direct and conceptual replication.
13 Twang and Kwan (1999) make a still broader distinction between empiricism conducted for the various
14 purposes of: checking of analysis, reanalysis of data, exact replication, conceptual extension (which we
15 called conceptual replication), empirical generalisation, and generalisation and extension. These
16 distinctions highlight the multi-faceted nature of empirical research, including that based on
17 experiments. Twang and Kwan (1999) consider replication to be indispensable to theory development
18 in management science. And as we did in our paper, Twang and Kwan (1999) also characterise
19 knowledge development as a cumulative process in which it is the body of knowledge overall that
20 develops. They recognise that no single study can stand alone as definitive.

21
22
23 Finally, the difficulties associated with conducting experiments on scenario planning again
24 resurface in Phadnis (2022). Phadnis (2022) recognises the possibility for random assignment to
25 treatment and control groups, which can increase what we referred to in our paper as internal validity,
26 but also recognises that the random sampling needed to establish the external validity of an experiment
27 is all but impossible in an organisational context. However, Phadnis (2022) proffers a potential solution
28 to this problem. Namely, field experiments, which are experiments conducted in the natural setting in
29 which the phenomenon under study occurs. Indeed, Phadnis et al. (2015) conducted a field experiment
30 as well as a conceptual extension of Schoemaker (1993). Phadnis (2022) highlights the recognition in
31 our paper of two threats to external validity—unrepresentative sample and unrepresentative design. The
32 former is when a study's participants are not representative of the target population and the latter is
33 when the experimental context is not representative of the real-world decision context. In response,
34 Phadnis (2022) very usefully highlights Harrison and List's (2004) six factors for guiding the design of
35 field experiments, which shaped the design in Phadnis et al. (2015). Field experiments that are cognisant
36 of these six design factors can, we believe, increase the external validity of experiments in which
37 random sampling is impossible. We again reiterate, however, that it is a multipronged approach to
38 research, which would be achieved by those in the field using a range of empirical methods, that will
39 ultimately lead to generalisable findings.

40 41 42 **3 Conclusion**

43
44 We believe the implication of our paper is a relatively uncontroversial and irrefutable one.
45 Namely, that the FFS field—and the scenario planning subfield in particular—requires a step change in
46 the variety of empirical studies conducted by those who research in it. We suggest that experiments are
47 relatively neglected in this field in part because of the difficulty of conducting them in ever changing
48 and idiosyncratic organisational contexts—a difficulty recognised in our paper and reinforced by the
49 responses. Yet, if there is one thing in common across both the focal paper and its responses it is that
50 the obstacles are *not* insurmountable. While it may well be impossible to conduct an experiment with
51 full random assignment and random sampling, and one that fulfils all the other various requirements for
52 representativeness and external validity discussed here and in the focal paper, it is nevertheless possible
53 to conduct experiments that variously and partially meet some of these criteria, depending on the context
54 in which they are conducted. To employ an apt cliché, we should not let the perfect get in the way of
55 the good. In a field that is currently relatively lacking in experimental studies, even an experiment
56 relatively lacking in many of these important regards can still contribute to advancing the field, can still
57 garner many citations, and can still add to the weight of evidence.

1
2
3 And it is the cumulative weight of evidence that we wish to emphasise in conclusion. No one
4 study or one researcher can expect or be expected to produce definitive findings that are forever set in
5 stone. In that respect, we agree with the strict Popperian thesis in that all findings are provisional. The
6 cumulative weight of evidence shifts over time and rests at any one time on many different studies,
7 which employ a whole variety of methods of different types. We reiterate that we did not highlight
8 experiments in our paper because we wish to place them on a scientific pedestal that, by definition,
9 must diminish the usefulness and value of other empirical approaches. We highlighted experiments and
10 made them the focus of our paper because we consider their paucity to be something that needs to be
11 addressed by the field. To end on a positive note, this paucity is a golden opportunity for anyone who
12 endeavours to rectify it.
13
14

15 **References**

- 16
17 Derbyshire, J. (2017) Potential surprise theory as a theoretical foundation for scenario planning,
18 *Technological Forecasting & Social Change* 82, 215-225.
19
20 Derbyshire, J., Dhimi, M., Belton, I. & Önköl, D. (2022a) The value of experiments in futures and
21 foresight science as illustrated by the case of scenario planning, *Futures and Foresight Science* (in
22 press).
23
24 Derbyshire, J., Feduzi, A. & Runde, J. (2022b) Borrowing from Keynes' *A Treatise on probability: A*
25 non-probabilistic measure of uncertainty for scenario planning, *European Management Review* (in
26 press).
27
28 Dhimi, M. K., Hertwig, R. & Hoffrage, U. (2004) The role of representative design in an ecological
29 approach to cognition, *Psychological Bulletin* 130, 959-988.
30
31 Harrison, G & List, J. (2004) Field experiments, *Journal of Economic Literature* 42, 1009-1055.
32
33 Kuhn, T. S. (1962) *The structure of scientific revolutions*, University of Chicago Press, Chicago.
34
35 Kuhn, K. M. & Sniezek, J. A. (1996) Confidence and uncertainty in judgmental forecasting: Differential
36 effects of scenario presentation, *Journal of Behavioral Decision Making* 9, 231-247.
37
38 Merton, R. K. (1942). Science and technology in a democratic order, *Journal of Legal and Political*
39 *Sociology* 1, 115–126.
40
41 Phadnis, S. S. (2022) Considering field factors to enhance external validity of scenario planning
42 experiments: A commentary on Derbyshire et al. 2023, *Futures & Foresight Science* (in press).
43
44 Phadnis, S., Caplice, C., Sheffi, Y. & Singh, M. (2015) Effect of scenario planning on field experts'
45 judgment of long-range investment decisions, *Strategic Management Journal* 36, 1401-1411.
46
47 Popper, K. (1959). [2002] *The Logic of Scientific Discovery*, London: Hutchinson. Republished 2002.
48 Routledge Classics.
49
50 Rowland, N. & Spaniol, M. J. (2022) On replication in science: Commentary on Derbyshire et al.
51 (2022), *Futures & Foresight Science* (in press).
52
53 Salo, A. (2023) On the boundaries of experimental research on scenario planning, *Futures & Foresight*
54 *Science* (in press).
55
56 Schoemaker, P. J. H. (1993) Multiple scenario development: Its conceptual and behavioral foundation,
57 *Strategic Management Journal* 14, 193-213.
58
59 Schoemaker, P. (2023) Exploring scenario planning through controlled experimentation: Commentary
60 on Derbyshire et al. 2023, *Futures & Foresight Science* (in press).

1
2
3 Tsang, E. W. K. & Kwan, K-M. (1999) Replication and theory development in organizational science:
4 A critical realist perspective, *Academy of Management Review* 24, 759-780.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Peer Review Only