

# The value of experiments in futures and foresight science as illustrated by the case of scenario planning

James Derbyshire<sup>1</sup>  | Mandeep Dhami<sup>1</sup> | Ian Belton<sup>1</sup>  | Dilek Önkal<sup>2</sup>

<sup>1</sup>Middlesex University, London, UK

<sup>2</sup>Northumbria University, Newcastle upon Tyne, UK

## Correspondence

James Derbyshire, Middlesex University, The Burroughs, Hendon, London NW4 4BT, UK.  
Email: [J.Derbyshire@mdx.ac.uk](mailto:J.Derbyshire@mdx.ac.uk)

## Funding information

This research was undertaken as part of a project titled "The effect of scenario planning on individuals' perceptions of uncertain futures," which was funded by the SAMS/BAM Research and Capacity Building Grant Scheme. It was thus jointly funded by the Society for the Advancement of Management Studies (SAMS) and the British Academy of Management (BAM).

## Abstract

An already pressing need to evidence the effectiveness of futures and foresight tools has been further amplified by the coronavirus pandemic, which highlighted more mainstream tools' difficulty with uncertainty. In light of this, the recent discussion in this journal on providing futures and foresight science with a stronger scientific basis is welcome. In this discussion critical realism has been proffered as a useful philosophical foundation and experiments a useful method for improving this field's scientific basis. Yet, experiments seek to isolate specific causal effects through closure (i.e., by controlling for all extraneous factors) and this may cause it to jar with critical realism's emphasis on uncertainty and openness. We therefore extend the recent discussion on improving the scientific basis of futures and foresight science by doing three things. First, we elaborate on critical realism and why the experimental method may jar with it. Second, we explain why the distinction between a conceptual and a direct replication can help overcome this jarring, meaning experiments can still be a valuable research tool for a futures and foresight science underpinned by critical realism. Third, we consider the appropriate unit of analysis for experiments on futures and foresight tools. In so doing, we situate the recent discussion on improving the scientific basis of futures and foresight science within the much longer running one on improving the scientific basis of business, management and strategy research more broadly. We use the case of scenario planning to illustrate our argument in relation to futures and foresight science.

## KEYWORDS

critical realism, experimental methods, scenario planning

## 1 | INTRODUCTION

Providing futures and foresight science with a stronger scientific basis has been the topic of several recent papers, commentaries and responses in this journal (Cairns, 2021; Fergnani & Chermack, 2021a, 2021b; Hodgkinson, 2021; Kishita et al., 2021;

Mandel, 2021; Minkinen, 2021; Münch & von der Gracht, 2021; Phadnis, 2021; Rowland & Spaniol, 2021a, 2021b; Salo, 2021; Schoemaker, 2021). The increased focus on this subject in this field is welcome. An already pressing need to evidence the effectiveness of futures and foresight tools has been further amplified by the coronavirus pandemic, which highlighted more mainstream tools'

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Futures & Foresight Science* published by John Wiley & Sons Ltd.

difficulty with uncertainty. In this paper, we use the example of scenario planning to argue for increased use of experiments as a specific means by which to improve the scientific basis of futures and foresight science.

Scenario-planning research is only a part of futures and foresight science and not the field in its entirety. Nevertheless, it is an excellent example around which to frame our argument. Scenario planning may be among the futures and foresight tools most in need of an improved empirical basis. At the last count there were twenty-three distinct approaches for creating scenarios (Bishop et al., 2007; Phadnis et al., 2014) and few have been tested for their effectiveness (Phadnis et al., 2015). Scenario planning has been proffered as an alternative that overcomes more mainstream tools' problems in dealing with uncertainty (Kay & King, 2020), but this requires that its effectiveness in this regard be evidenced. Among other things, in this paper we elaborate on what might be meant by scenario planning's "effectiveness" by drawing on a critical realist understanding of uncertainty.

Critical realism recognizes the openness, complexity and emergence that lead to uncertainty. Partly for that reason, it has been suggested as an appropriate foundation for research on scenario planning, as-well-as for futures and foresight science more broadly (Frith & Tapinos, 2020; Hodgkinson, 2021; Patomäki, 2006). Yet, exactly because critical realism recognizes openness, complexity and emergence, the experimental method may jar with it, as it seeks closure by controlling for all extraneous factors in order to isolate specific causal effects. The jarring with critical realism this might cause is evidenced by Byrne and Callaghan (2013), whose ontology of "complex realism" (Reed & Harvey, 1992) draws widely on critical realism:

Given the nature of complex systems, causal processes relevant to them are seldom if ever singular. Interaction implies, indeed requires, multiplicity. Moreover, we have to abandon what medicine calls "specific aetiology" which asserts that each and every outcome (disease) has one, and only one, cause. The notion of single cause has profound methodological implications. It is the justification for the primacy attached to randomized controlled trials (Byrne & Callaghan, 2013; p.185).

The implication is that attempting to isolate specific causes through experiments is misguided. Long-standing skepticism towards experimental methods in business, management and strategy research draws on similar logic. For example, Schwenk (1982) quotes Mintzberg's (1977) similar objection to experiments thus:

the very complexity of phenomena determines the organization's behaviour. In other words, processes such as strategy formulation are characterized by the inherent complexity and dynamic nature of the environments in which they operate; recreating these

processes in the artificially simplified environments in the laboratory eliminates the very characteristics that determine the organization's responses. Management Policy researchers have [therefore] been forced to study real behaviour in real organizations, amidst their complexity (Mintzberg, 1977; p.93).

The clear implication is that experiments lack realism. Bolinger et al. (2022) suggest similar logic still limits use of experiments in these fields today. We suggest similar logic to hinder their greater use in futures and foresight science too. For example, Cairns (2021) has questioned whether the search for relationships between variables (as might be undertaken through an experiment) is a universally shared understanding of a scientific approach in futures and foresight science. The sheer paucity of experiments in futures and foresight science implies skepticism towards their value. If we wish to advocate for increased use of experiments in this field, we must address these reservations. In this paper, we therefore do the following:

- i) Elaborate on critical realism and why the experimental method may jar with it;
- ii) Explain why the distinction between a conceptual and a direct replication can help overcome this jarring and why experiments can therefore make a valuable contribution to a futures and foresight science underpinned by critical realism;
- iii) Consider the appropriate unit of analysis for experiments on futures and foresight tools.

In so doing, we situate the recent discussion on improving the scientific basis of futures and foresight science within that on improving the scientific basis of business, management and strategy research more broadly. In all instances we use the example of scenario planning to illustrate our argument in relation to futures and foresight science more broadly.

## 2 | THE NATURE OF THE WORLD AND ITS UNCERTAINTY

The longevity of its popularity suggests there is more to the need for scenario planning than mere managerial performativity. Managers cannot pretend that what has been called "large-world uncertainty" (Feduzi et al., 2022) does not exist as more mainstream tools that reduce it to probabilistic risk do (Kay & King, 2020). Feduzi et al. (2022) illustrate "large-world uncertainty" by contrasting it with the "small-world uncertainty" intended as the realm of subjective probability (Savage, 1954). Where a system of interest is characterized by small-world uncertainty, it can be "closed" because its possibilities can be completely listed. Uncertainty exists only in relation to which of these completely known possibilities occurs. Only by achieving such closure can probabilities be legitimately attributed. Scenario planning, by contrast, is designed for the eternally open "large world" with all of its complexity (Wilkinson

et al., 2013), its emergence leading to novelty and what are presently unknowns, and its non-determinism.

In these distinctions between openness and closure, and large- and small-world uncertainty, are hints of Mintzberg's (1977) contrast between the artificially simplified laboratory and the complex real world, which led him to reject experiments. Yet, while we also accept these important distinctions, and recognize that openness is a central if not predominant feature of reality, we do *not* agree with the implication that Mintzberg (1977) draws from this, which is that experiments inherently lack realism. In this section and the next, we build an argument describing why experiments can still be valuable to those who recognize openness, complexity and emergence, and who, because of this recognition, accept the central tenets of critical realism, as many scholars of futures and foresight science do. This argument requires us first to elaborate on the nature of the world and its uncertainty according to critical realism, a task we have already begun by reference to large-world uncertainty above, and one we continue below. As noted by Fleetwood (2005), the way we consider the world to be (ontology) influences what we think can be known about it (epistemology) and how we think it should be investigated (methodology and research techniques). We must therefore clarify our ontology, before discussing epistemology and methodology in terms of why we think experiments can be valuable to futures and foresight science.

Critical realism (Bhaskar, 1978, 2008, 2009, 2010) is not a general theory, but provides a distinctive philosophical standpoint on the nature of the world that has major implications for theory construction, testing and application (Jessop, 2015). Critical realism offers an alternative to both the scientism of logical positivism and the damaging relativism of the reaction to it that is postmodernism (Sayer, 2004). It combines modest versions of empirical realism and social constructivism (Jessop, 2015), which contrasts with the extreme forms of these adhered to by logical positivists and postmodernists, respectively. Empirical realism in its extreme form is naïve as it posits a directly observable reality; social constructivism in its extreme form is misleading as it posits that ideas and social practices are merely constitutive of social relations. Social constructivism in isolation therefore commits the "epistemic fallacy" of assuming that reality corresponds to our knowledge of it (Jessop, 2015). And empirical realism in isolation fails to acknowledge the reflexivity inherent in the production of any form of knowledge. Only together can they provide a realistic representation of reality by each tempering the extremity of the other in isolation.

Empirical realism and social constructivism are melded in critical realism through its layered ontology, which implies that the world is stratified into different strata, requiring different concepts, assumptions and explanatory principles corresponding to their respective emergent properties (Jessop, 2015). There is the intransitive world, the real world of causal mechanisms, actual events and processes (the actual), and empirical observations (the empirical). Note in particular that the real and the empirical are distinct. Actual causal mechanisms, events and processes are not always manifest in empirical observations. This is why critical realists reject a purely correlational or

regularity-based view of causation (Jessop, 2015). The actual comprises patterns of events and processes that result from the interaction of a plurality of mechanisms, tendencies and counter-tendencies, which operate under specific conditions only (Jessop, 2015). The empirical represents traces of the actual, in terms of the selection of potential causes that are actualized in a particular instance (Jessop, 2015). The actual and the empirical rarely if ever correspond in their entirety. For this reason, one of critical realism's central concepts is that of the "demi-regularity," which is a cause that is sometimes actualized and sometimes not (Fletcher, 2017). Yet, at the same time, and despite the uncertainty this inevitably gives rise to, under critical realism an entity can only be considered real if it has causal efficacy or an effect on behaviour in some way— that is, if it "makes a difference" (Fleetwood, 2005; p.199). Note that "realness" of effect does not necessarily imply material realness. For example, God may or may not exist, but belief in God makes a difference to people's actions and so is real (Fleetwood, 2005).

As reflected in its layered ontology, the world according to critical realism is open. Yet, institutions shape the world and produce partial closures (Downward et al., 2002), meaning focal systems of interest will therefore lie on a spectrum from open to closed and from deterministic to nondeterministic (Derbyshire, 2020). This means demi-regularities will be more consistently actualized in some systems of interest than in others, blurring the lines between "large-world uncertainty" and "small-world uncertainty." Moreover, there can be a temporal aspect to closure (Downward et al., 2002) because the emergent outcome of the interaction between a multiplicity of sometimes actualized and sometimes non-actualized causes can take time to produce change. The short term is therefore more closable than the long term. For this reason, for many systems of interest short-term forecasting is more possible than is long-term forecasting. The range of possibilities is much smaller over a shorter time horizon.

Openness can be understood by contrasting it with its opposite: closedness. A closed system is one in which there is no possibility for qualitative changes that are either internally or externally driven. Internally driven qualitative changes emerge from within the system itself as the result of, for example, reflexive changes to the behavior of agents operating within it. An externally driven change may be brought about by forces emanating from other systems with which a focal system has a boundary. If neither of these sources of qualitative change are present, a system can be considered closed. Critical realism characterizes the world as being predominantly open, and many of its systems as comprising diverse causal factors— both internal and external—all of which impinge on the system to different degrees simultaneously. Some are invariant but many others are sometimes actualized and sometimes not. There is positive feedback and nonlinear interaction between these diverse causal factors, meaning they may combine to create ampliative and extreme effects and outcomes, which may extend far beyond what has occurred previously as captured by a measure such as variance, or as embedded in Bayesian priors (Derbyshire & Morgan, 2022).

This openness results in uncertainty not least because it makes antecedent conditions a poor guide to the future. Yet, perhaps paradoxically, the resulting non-ergodic nature of the world and its systems (Peters, 2019) renders history of great importance. How a system has come to be as it is presently will affect how it develops in the future, but not in the way implied by the stationarity assumed by econometric forecasting. It is not a simple matter of continued trends, but how historic processes lead to the emergence of qualitative changes over time. Such qualitative changes lead to disjuncture and disruption, which negates prevailing strategies and makes long-term forecasting an impossibility. This is highly problematic for business strategy-making because managers suffer from what Audia et al. (2000) call “the paradox of success.” Over time their mental model of the external environment becomes framed by strategies that have led to past success. This leads to “strategic persistence” (Audia et al., 2000) in the face of disruptive changes, which we might also call “strategic inertia.” Strategic inertia caused by past success appears to be quite widespread in strategic management (Croson et al., 2007; Hodgkinson et al., 1999).

Herein is the very *raison d'être* of scenario planning, which is designed to aid strategy-makers to change their mental model of the external environment so that they do not assume future changes to it will be consistent with business-as-usual expectations. Scenario planning is designed to ensure they instead take account of highly uncertain and potentially highly impactful (to current strategy) changes that may occur (Cairns & Wright, 2018). Critical realism helps explain why a tool such as scenario planning is needed to assist with the “unfreezing [of] mental models” (Hodgkinson & Rousseau, 2009; p.539). Critical realism helps in understanding why strategy-makers cannot simply compare their view of the external environment with empirical reality and realize they must change course. In its incorporation of a modest form of social constructivism, critical realism recognizes the human reflexivity involved in the production of any form of knowledge (Van de Ven, 2007), including that related to an organization's external environment and strategy.

The combination of empirical realism and social constructivism in critical realism means that it acknowledges there is a real world “out there” against which we can check our empirical claims (Van de Ven, 2007), but these attempts at verification will be inhibited by attempts to manipulate others' perspectives to advantage (“reflexivity”), which will depend to a degree on having the power to impose a particular interpretation. Individuals try to interpret reality accurately, but also try to manipulate others' interpretation of it. Their attempts to manipulate others' interpretation of it confounds their ability to interpret reality accurately. This is why strategy-makers cannot simply compare their view of the external environment with empirical reality and realize they must change course. Such reflexivity in commercial and central banks' boardrooms contributed to the subprime crisis, which led to the highly disruptive change to the external environment that was the credit crunch (Soros, 2009, 2013). Those in charge had an interest in perpetuating the illusion of low risk associated with collateralized debt—an illusion that could no longer be sustained when changes to interest

rates showed these risks to be something else altogether: incalculable uncertainty (Kay & King, 2020).

In all of the above ways, critical realism helps explain the reasons for uncertainty and the human responses made in its face, which can compound it. Yet, exactly because it emphasizes openness, emergence, complexity and therefore uncertainty, critical realism may be interpreted as jarring with the experimental method. Recall how Byrne and Callaghan (2013) question medicine's “specific aetiology”—that is, its search for singular causes—from a critical realist and complexity-based angle. Byrne and Callaghan (2013) might consider the credit crunch an excellent example of the dangers of just such specific aetiology. Oatley (2019) describes the credit crunch as an emergent outcome of multiple simultaneous developments within the global banking, finance, regulatory and social environment, implying that placing the finger of blame on any one of these causes in isolation would be misleading.

This issue of causal “reductionism” is why experimental methods might jar with the nature of reality implied by critical realism. The perception that experiments lead to causal reductionism is among the factors inhibiting greater use of experimental methods in business, management and strategy research (Bolinger et al., 2022; Schwenk, 1982). In these fields, discussions on this subject have been couched in doubts about experiments' “generalisability” and “external validity” (Bolinger et al., 2022), but suspicions about causal reductionism are what lie behind such doubts. By not representing reality in all of its causal complexity and openness, and by seeking to isolate specific causes, experiments can only say something about causation under the specific conditions of the particular instance of their implementation. That is the logic to which this perception of causal reductionism gives rise. In the next section we counter this objection to the experimental method by drawing on Van de Ven's (2007) discussion of critical realism.

### 3 | CRITICAL REALISM AS A FOUNDATION FOR EXPERIMENTAL RESEARCH ON SCENARIO PLANNING

A futures and foresight tool such as scenario planning is not only implemented within and designed to grapple with a causally complex external environment (Wilkinson et al., 2013), it is also implemented in the complex internal setting of a particular business or organization, and it is itself a complex *process*. Rather than meaning that its process is complicated, we mean by it being a “complex process” that any effect it has emerges from the process' many steps and procedures, the specific environment in which it is applied, and the social interactions of those engaged in the scenario exercise. No one step or part of this complex process can be expected to have the intended emergent effect in isolation.

Yet, this does *not* mean we should avoid testing particular parts of this process for a specific effect that might contribute to its emergent overall effect. Even Byrne and Callaghan (2013), who we have noted to be skeptics of specific aetiology, note the following:

Note that the idea of causal emergence is not inherently holistic. That is to say it is not the simple opposite of reductionism where we explain the properties of the whole in terms of the properties of the components of the whole. We cannot turn this around and pay attention only to wholes (Byrne & Callaghan, 2013; p.21).

We interpret this as meaning we should not fail to take account of the component parts of a process and their individual effects for fear of causal reductionism. Nor, for that matter, should we avoid seeking to identify traces of the emergent effect of the process as a whole, even if we should be skeptical about capturing it in its entirety using any one method. There is a fundamental difference between a reductionism (or, for that matter, an “atomism”) based on breaking up and analyzing the component parts of irreducibly complex processes, on the one hand, and trying to understand the contribution of different parts of a process to its emergent effect, or attempting to uncover traces of the emergent effect itself, on the other.

Consider that many diseases are a function of multiple generative mechanisms. These can include social, economic, biological and other diverse causes. Treatments that address several of these causes simultaneously might work better than any one treatment does in isolation—a logic that has given rise to social prescribing alongside more traditional pharmaceutical treatments. Yet, this does not mean that testing the component treatments for their individual effect is invalid. Both the social prescribing and the pharmaceutical treatments would be tested for their individual effectiveness, even if combining the two is expected to create a non-additive effect that is irreducible to either in isolation. Moreover, their combination would be tested for traces of its emergent and non-additive effect too. The number of cases who receive both treatments who then recover might be compared statistically to the number who recover after being prescribed one or other treatment in isolation, but not both. A non-additive effect might thus be identified that is suggestive of both treatments working in unison to create an effect irreducible to either in isolation. This might be done alongside research that traces the emergent process of qualitative change towards recovery brought about by the treatments' complementarity. This approach would be highly congruent with critical realism.

Van de Ven (2007) contrasts two types of research from the perspective of critical realism: research that looks for relationships between variables and research that examines processes of change. As noted earlier, scenario planning is designed to change participants' perception of the uncertainty of the external environment, which can be influenced by past success, leading to strategic inertia. Scenario planning is therefore a process-based tool designed to bring about change. Studying processes of change requires research tools such as case studies that can uncover emergent effects by tracing what unfolds over time (Van de Ven, 2007). Understanding how particular patterns emerge from a process unfolding over time requires a broad causal perspective that experiments cannot deliver. This raises an important question: if a tool such as scenario planning is supposed to

initiate a process of change, why should we not exclusively prefer case studies as an empirical tool for researching scenario planning?

Byrne and Callaghan's (2013) recognition that wholes and parts should not be mutually exclusive foci of research provides the initial hint of an answer. Van de Ven (2007) makes the reasons more explicit by emphasizing the need for both variance-based and process-based research in terms of them providing complementary answers to different questions: “what” and “how.” Behind every “what” question is a hypothesized causal mechanism or process—a “how.” Whether implicit or otherwise, the logic behind a “what” answer that is provided by a variance model is a process story about how a sequence of events unfolds (Van de Ven, 2007). The implication is that only by answering both types of question—“what” and “how”—can we hope to have something approaching a satisfactory answer. This implies the complementarity of experimental and more qualitative research approaches such as case studies. Neither can be considered to provide a complete answer when used in isolation.

Returning to the example of strategic inertia, Audia et al. (2000) carry out longitudinal archival research and paint a rich and qualitative picture of the “paradox of success.” Having painted this rich picture qualitatively, Audia et al. (2000) then conduct a fully randomized experiment that allows them to isolate a specific causal mechanism contributing to strategic inertia—that related to the organization's past performance. Their experimental task is exactly the same between two groups of participants, but for one group data on the past performance of the organization shows it to be positive, whereas for the other it is negative. This allows Audia et al. (2000) to confirm the hypothesis they created through qualitative research: past success blinds managers to the need to change strategy in light of radical changes to the external environment. Herein we have both a “what” and a “how.” This is an example of a combined approach that could be used by researchers of scenario planning, who can draw on a plethora of case studies to create hypotheses followed by experiments that test an aspect of the processual effect captured by the case study. Generalizability and external validity, which we discuss further in the next section, would be achieved in combination. That case studies can contribute to generalization has been convincingly argued by Flyvbjerg (2006).

Van de Ven (2007) provides much to support use of experimental methods underpinned by critical realism, but also highlights experiments' inherent limitations. While promoting greater use of experiments, it is important not to lose sight of these limitations by placing the experimental method on a scientific pedestal. Nor, relatedly, is it wise to place scientific knowledge on a pedestal above that of other types of knowledge (Funtowicz & Ravetz, 1993; Ravetz, 2022), especially in a practice-led domain such as futures and foresight science. Hodgkinson (2021) relays the chastening experience of trying to persuade an audience of prominent practitioners to adopt scientific methods in relation to futures and foresight tools. The audience did not quite throw rotten tomatoes, but it appeared akin to that. Hodgkinson (2021) suggests that appealing to the virtues of scientific theory (and method) is a necessary but insufficient part of arguing for the adoption of more scientific methods. On the one

hand, developing scenario planning and futures and foresight science as scientific fields of research requires more than just “anecdotal evidence” from practitioners. On the other hand, it also requires more than just being scornful of what may seem to academics to be the unscientific basis of practice. Truly “engaged scholarship” (Van de Ven, 2007) requires the bridging of any divide between practitioners and scholars in these fields.

Critical realism recognizes that all research methods have limitations (Hodgkinson & Rousseau, 2009). Rather than elevating one method over another, critical realism makes such elevation unnecessary by prizing mixed methods (Hodgkinson & Rousseau, 2009). In so doing, it assists in overcoming tensions between rigor and relevance in academic research (Hodgkinson & Starkey, 2011). Moreover, critical realism can assist in overcoming the tribal divisions between “deductivists” and “constructivists” that Wilkinson (2009) has highlighted in the scenario planning field. Both tribes are accommodated within critical realism, which is a meld of empirical realism and social constructivism, with each tempering the extremity of the other in isolation. Under critical realism, just because reality is in part constructed does not make it any less real. It is approaches to research that do not recognize this partial construction that lack realism.

Critical realism implies a pluralist approach to research in futures and foresight science. Both experiments and case studies are an important part of this pluralism, but both are still only one part of it. Case study research would not somehow be diminished by increased use of experiments. Case studies would gain in credibility if combined with experiments, with the two together providing a “what” and “how” answer. Experiments are just one way to construct and then test models of the world. We advocate for them in this paper because we believe in a pluralist approach, but do not believe that such pluralism exists presently as experiments are extremely rare in research on scenario planning and in futures and foresight science more broadly. A central implication of critical realism for improving the scientific basis of research in these fields is more balance in this regard.

#### 4 | THE DISTINCTION BETWEEN CONCEPTUAL AND DIRECT REPLICATION

Psychology is a field from which futures and foresight is likely to draw if experiments are to be used more in it. Derksen and Morawski (2022) suggest psychology is going through a period of turmoil due to its “replication crisis” (Pashler & Harris, 2012). Failed replications are perceived as having dented trust in the field’s findings and its scientific standing. However, Derksen and Morawski (2022) examine this “crisis” through the lens of enactment theory, which recognizes the multiplicity of causes and the dynamic nature of reality. If we accept this multiplicity and this dynamism, which are also emphasized by critical realism, then we must also accept that the failure of an experiment to replicate is not necessarily reflective of flaws in its design or implementation. Moreover, we must accept that direct replication is often unrealistic, and that conceptual replication is more

realistic, especially in social settings. This is an important distinction for experiments conducted on futures and foresight tools such as scenario planning, which have a social component.

A direct replication is “an experiment whose design is identical to an original experiment’s design in all factors that are supposedly causally responsible for the effect” (Romero, 2019; p.2). Direct replications are therefore studies that seek to estimate the same population parameter and test the reliability of an original study (Clemens, 2017). A conceptual replication, by contrast, attempts to establish the same theoretical conclusion as an original study with different experimental manipulations or measures (Gouveia, 2021; Schmidt, 2009). Proponents of conceptual replication (“conceptualists”) argue for the primacy of context because behavior is far from universal, as it is (in part) socially, culturally and historically informed. For conceptualists, this makes direct replication unrealistic and they therefore have a more nuanced understanding of Popper’s ideas about falsification (Popper, 1959), which are central to the present hand-wringing about replicability (Derksen & Morawski, 2022). Proponents of conceptual replication emphasize that science is a collective and a cumulative process that places the goal of theory development above the mechanical objectivity of direct replication (Derksen & Morawski, 2022). Reproducibility of a finding is of less concern than evidence of validity of a theory. As such, conceptual replications might more accurately be called “conceptual extensions,” which test and extend theory (Derksen & Morawski, 2022). This is particularly pertinent to scenario planning in particular, which has been suggested to be theoretically underdeveloped (see Derbyshire, 2017; p.77).

Nosek and Errington (2020) specifically argue that many conceptual replications are generalizations rather than replications (Derksen & Morawski, 2022). The implication is that science does not advance through an impenetrable thicket of amassed empirical findings, but through development, refinement and replacement of theories. Theory development does not rest on any one specific experimental outcome, but on many examples and extensions across contexts. It is through variations on an earlier study, achieved through conceptual replication, that the underlying invariant and stable aspects of reality are brought into view (Derksen & Morawski, 2022).

In an implication that is highly congruent with critical realism, from the perspective of conceptual replication, even a technically identical experiment may not guarantee the same findings if the context changes (Schwarz & Clore, 2016). This might be particularly true of scenario planning, which is implemented to examine varying strategic issues in different organizations, operating in different sectors. This chimes with critical realism’s emphasis on the multiplicity of actualized and non-actualized causes, which implies not only that experiments in different contexts may produce different results, but that the same experiment implemented on another occasion in exactly the same context may also produce a different result, due to changes in the nature of that context over time. One never steps into the same river twice (Crandall & Sherman, 2016; p.94; Derksen & Morawski, 2022; p.1493).

In sum, there is little reason to expect that an experiment aiming to conceptually replicate some findings will yield exactly the same results (effects) as the study it seeks to replicate (Clemens, 2017; Gouveia, 2021). As such, there is no such thing as a “conceptual failure to replicate” (Doyen et al., 2014; p.28; Gouveia, 2021). In this sense, follow-up studies that change the methods or measurements used are better interpreted as extensions rather than replications (Clemens, 2017; Gouveia, 2021). Extensions test the inference made by the original study in alternative contexts or using alternative methods, testing the robustness of the original study (Clemens, 2017). This provides important insights into how experiments would be used to research futures and foresight tools such as scenario planning.

For example, Phadnis et al. (2015) conduct three experiments that test the hypothesis of Schoemaker (1993) that scenario planning works by broadening participants' probabilistic confidence intervals via the mechanism of the conjunction fallacy. However, Phadnis et al. (2015) do not use exactly the same scenario process as that used by Schoemaker (1993). Yet, if the conjunction fallacy is indeed the mechanism by which scenario planning has an effect we may expect this effect still to be found under Phadnis et al.'s (2015) different scenario process, as the key elements of it related to this proposed causal mechanism—the writing of narratives—is still present. If Phadnis et al. (2015) had found a similar conjunction-fallacy effect to that proposed by Schoemaker (1993) as an explanation for his own findings, this would have represented a conceptual replication, but not a direct one. The finding would have been extended to a somewhat different scenario method and would therefore have gained in credibility as a general causal mechanism.

However, as it was, Phadnis et al. (2015) did not replicate this finding. The authors of the present paper have also failed to extend Schoemaker's (1993) findings by replicating them in a similar experiment, as reported elsewhere. This adds to the cumulative body of evidence suggesting the conjunction fallacy may not be the causal mechanism by which scenario planning has an effect. This illustrates the possibility to advance theory by eliminating causal mechanisms—or, at least, by reducing the likelihood that they exist—even without direct replication's stark view on falsification. Conceptual replication, therefore, does not lead to an “anything goes” free-for-all.

Expecting constant, direct replication from experimental methods takes away from the real value they can add (Deaton & Cartwright, 2018). We have argued that external validity is better achieved through use of mixed methods which together combine to make findings more generalizable. We later argue that generalizability can also be added to by carrying out the same experiment in multiple contexts. While we would not expect exact replication in every context, we might expect a genuine causal effect to show up more often than not. The inflation of the perceived probability of a particular outcome (i.e., a conjunction fallacy) as a result of the writing of narratives in scenario planning could be what critical

realism calls a demi-regularity, although evidence is mounting that it is not.

It is the body of evidence as a whole that is the source of knowledge, not one particular study in isolation (Hodgkinson & Rousseau, 2009). Scenario planning, like many futures and foresight tools, has group-based and participatory aspects. There is little reason to expect that an experiment on scenario planning will yield exactly the same effect as another study if it depends on group-based effects and the group undertaking it is different on each occasion. We argue that conceptual replication ought to be the norm in relation to scenario planning because of the manifold contexts in which it is used, its group-based and participatory nature, and the many alternative approaches there are for implementing it. Furthermore, we suggest that such an understanding of replication is highly congruent with critical realism, which has been suggested to be a useful philosophical foundation for research on scenario planning and in futures and foresight science more broadly.

## 5 | WHAT IS THE APPROPRIATE UNIT OF ANALYSIS?

Van de Ven (2007) notes that critical realism views science as a process of constructing models that represent aspects of the world, then comparing and contrasting between the findings from these models (Rescher, 2000). No one model can represent the world—or even a discrete portion of it—in all its causal complexity. A degree of reduction will thus always be necessary, which is why we seek representativity. We must seek to represent the aspect of reality we seek to understand as accurately as possible, while knowing we cannot represent its full complexity. We therefore turn finally to consider the issue of representing the population under study accurately in an experiment. This is a question about the appropriate unit of analysis—one that has important implications for whether the randomness needed for experiments is “doable” in the context of futures and foresight science. To understand better why establishing the appropriate unit of analysis is important it is worth firstly considering the different roles of randomness in experimentation.

Achieving external validity (i.e., generalizability to a broader population of interest) requires the sample on which an experiment is conducted to be representative, which requires the application of randomness. While in many contexts it is unrealistic to include every member of the population in an experiment, it is important to ensure that every member has equal chance of being part of the sample on which the experiment is conducted. This requires use of a probabilistic sampling method, which applies randomness to select individuals from the population to be part of the sample. This is known as random sampling. Random assignment is something different from random sampling. The random assignment of individuals from a random sample to “treatment” and “control” groups ensures that these two groups are representative of each other, meaning they have what is known as “internal validity.” Internal validity is needed to ensure the two groups can be compared, which

is the means by which a specific causal or treatment effect is isolated. Random sampling does little to aid internal validity and random assignment does little to aid external validity (Eden, 2017). Only by employing both random sampling and random assignment can an experiment hope to achieve both external and internal validity.

Randomization is therefore essential to the perceived quality of experiments. According to the Maryland Scientific Methods Scale (Farrington et al., 2002), full randomization is the hallmark of a well implemented experiment. Fully randomized experiments therefore achieve a maximum score of 5 on this quality gauge. Experiments that do not have random assignment to a "treatment" and "control" group, such as observational studies, have a lower ranking on this scale, and therefore have a lower scientific value according to this measure of quality. However, for reasons we outlined when discussing the role of randomness above, we suggest that only experiments that involve both random sampling and random assignment should be considered gold standard. We do not consider any experimental study of scenario planning to have ever reached this standard and it may be an unreasonable expectation in relation to that particular futures and foresight tool. If the individuals responsible for strategy in an organization are the unit of analysis, it is difficult to see how a set of managers or executives could be randomly sampled and randomly assigned to "treatment" and "control" groups, with one group undertaking the scenario planning and the other not, all within the same organization and without any contamination between the two groups. For that reason, the few experiments that have been undertaken have tended to have graduate students as participants.

For example, Meissner and Wulf (2013) conduct an experiment on scenario planning that uses three groups of randomly assigned graduate management students. One group undertook a full scenario process, a second undertook a partial scenario process, and the third was a control group. Meissner and Wulf (2013) explicitly claim external validity for their study using the logic of random assignment captured in the above discussion. Yet, limited external validity might still be thought a problem if we consider that (a) the study's participants are graduate management students and not necessarily business executives, managers or strategy-makers, and (b) even if they were business executives, managers or strategy-makers, the experimental stimuli were not representative of those faced by these populations. The first issue (a) refers to an unrepresentative sample and the second (b) to an unrepresentative design (Dhami et al., 2004). Both issues bring the external validity of the study into question. Specifically, the participants may not be representative of the population who would normally undertake scenario planning, to whom the researchers would presumably wish to generalize their findings. Second, the stimuli may not represent the multivariate and interrelated factors that are present in a real scenario-planning setting, and again, this limits the generalizability of the findings. Importantly, Dhami et al. (2004) show that this lack of representativeness is not only a threat to external validity and generalizability, but also to the internal validity of the experiment itself.

The second issue of not accurately representing a real scenario-planning setting is particularly important in relation to the earlier

discussion on critical realism in which the perception that experiments are reductionistic was highlighted. In that discussion we recognized that some abstraction from reality's full complexity is inevitable, but the question is the extent of this and whether it compromises the external validity (and therefore generalizability) of findings. This issue of representative design is very important because psychological processes, including those related to business strategy-making, do not occur in a vacuum. They are adapted to the particular environmental context in which they take place, which the experiment must seek to replicate as accurately as possible to be representative (Dhami et al., 2004). These problems, highlighted in relation to Meissner and Wulf (2013) above, also apply to what is perhaps the most well-known, and certainly the most well-cited, study of an experimental type on scenario planning, which is that by Schoemaker (1993). 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). These problems, however, simply reflect the difficulty of applying randomness in, and providing environmental stimuli that accurately represent, the real contexts in which scenario planning is undertaken. They are inevitable, to an extent, when conducting experiments on scenario planning because of the nature of the tool and the social contexts in which it is applied.

However, as our earlier discussion on conceptual replication suggested, just because the ultimate experimental gold standard may be unobtainable does not mean experiments on scenario planning should be abandoned. It is the accumulated body of knowledge that assists in developing theory, which is not reliant upon any one study. There is considerable opportunity today to conduct experiments on scenario planning with random assignment (if not random sampling) using online platforms such as Mechanical Turk or Prolific. But experiments that employ participants recruited via these platforms will always be subject to questions about their external validity as they can only recruit self-selecting individual participants. Real scenario processes are based on group interaction and social processes.

We do not wish to discourage the use of these platforms as we feel well-executed experiments implemented using them can still offer insights if they are targeted at particular aspects of the scenario process. Such experiments would have high levels of internal validity from random assignment and enable participation by a variety of participant groups that may have input to strategy-making while not necessarily being able to participate in group-based scenario processes (e.g., due to pandemic, war, or other accessibility issues). The ability of online platforms to reach a diverse set of participants is a strong plus given how WEIRD (Western, Educated, Industrial, Rich and Democratic) the backgrounds of many participants in experiments are (Henrich et al., 2010). However, ideally, it would be better to study "real" scenario processes that include a participatory aspect



resulting from group interaction, which cannot be accurately represented by recruiting individual participants via an online platform. Any experiment seeking traces of scenario planning's emergent effect as a full process must seek to replicate that full process, including its group-based aspects, as accurately as possible to have a representative design.

That said, we still consider the unit of analysis for such experiments to be the individual participant rather than the firm. As we noted earlier, scenario planning's effectiveness relates to whether it changes managers' mental models about the external environment and helps them recognize the full extent of its uncertainty. Even if we consider scenario planning to give rise to a group-level effect from the process as a whole, the causes of which are partly the social interactions that take place as part of it, we should still expect this effect to be manifest in the thinking of individual participants, making individual participants within a real scenario setting the appropriate unit of analysis. The same scenario experiment carried out across multiple settings (i.e., multiple implementations of the same scenario exercise in different organizations) would increase external validity. Single studies are always limited in terms of generalizability. It is the body of knowledge as a whole that yields valuable insights and assists in advancing theory (Hodgkinson & Rousseau, 2009).

## 6 | CONCLUSION

Uncertainty is only going to become more pronounced over time as society and its systems become ever more interconnected. The need is therefore to maneuver futures and foresight science into the mainstream as one of the few fields that seeks to grapple with this uncertainty, rather than pretending it does not exist by reducing it to risk. This requires a step-change in the amount and type of empiricism undertaken in this field. To meet this need, we call for a concerted effort to increase the number of experiments in futures and foresight.

We do not do so because we wish to place the experimental method on a pedestal that by implication relegates other methods to a different scientific status. Rather, we do so because we believe in pluralism, but do not consider a field dominated by case studies to be pluralistic. Regardless of how useful they can be, they can only show reality from one level of magnitude. Pluralism implies use of a range of methods, among them the experimental method, which can provide a focus and level of magnitude that other methods cannot. Experiments are particularly useful for zooming in on one particular part of a process and studying how it works.

Using the experimental method is not to deny that a process of change is a process, or that it may have an emergent effect greater than the summed effects of its component parts when studied individually. There is a fundamental difference between a destructive and misleading reductionism or atomism that is based on breaking up and analyzing the component parts of irreducibly complex processes, on the one hand, and trying to understand the contribution of different parts of a process to its emergent effect, or attempting to uncover traces of that emergent effect as a whole, on the other. We

should not throw the experimental method out with the reductionistic bathwater.

We believe that online platforms can accelerate research on scenario planning and should be used more by those researching it. This is in part because we consider the unit of analysis for scenario planning to be the individual not the firm, because our understanding of its effectiveness relates to its ability to change individuals' perspective on the external environment, leading them to recognize the full extent of its uncertainty and the potential for radical changes to it, which disrupt business-as-usual. The change to strategy that may ensue from this process is a firm-level effect, but is unlikely to occur unless the scenario exercise firstly changes the perspective of the individuals who undertake it. It is people who change things, not abstract entities such as "firms," which are merely an administrative and legal construct for groups of people who operate in concert to achieve a particular end.

Nevertheless, for those wanting to pursue a gold standard experimental approach on scenario planning, as one example of a futures and foresight tool, we suggest the ultimate would be an experiment conducted on participants in a real company undertaking a scenario exercise, then conceptually replicated across several other companies, leading to comparison and contrast of the results across contexts. We may not expect exact replication in all instances, but we would expect important causal mechanisms, and the emergent effect of the scenario-planning process as a whole, to leave at least some trace across many of them. We do not believe this approach to be beyond the realms of possibility. Incidentally, it is not beyond what a research council would be willing to fund either.

Scenario planning's famous use by Shell to anticipate the oil crises of the 1970s is a laurel on which we cannot rest an entire field forever. As much as we admire what has been uncovered about scenario planning using presently dominant methods such as case study, we strongly believe that mixed-method approaches would bring added advantages to futures and foresight science work, benefitting both providers and users of scenarios. A central implication of critical realism for improving the scientific basis of any field is pluralism in the form of mixed methods.

## ACKNOWLEDGMENTS

"This research was undertaken as part of a project titled 'The effect of scenario planning on individuals' perceptions of uncertain futures', which was funded by the SAMS/BAM Research and Capacity Building Grant Scheme. It was thus jointly funded by the Society for the Advancement of Management Studies (SAMS) and the British Academy of Management (BAM)."

## DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## ORCID

James Derbyshire  <http://orcid.org/0000-0002-1505-412X>

Ian Belton  <http://orcid.org/0000-0002-2479-6563>

## REFERENCES

- Audia, P. G., Locke, E. A., & Smith, K. G. (2000). The paradox of success: An archival and a laboratory study of strategic persistence following radical environmental change. *Academy of Management Journal*, 43, 837–853.
- Bhaskar, R. (1978). *The possibility of naturalism*. Harvester Press.
- Bhaskar, R. (2008). *A realist theory of science*. Routledge.
- Bhaskar, R. (2009). *Scientific realism and human emancipation*. Routledge.
- Bhaskar, R. (2010). *Reclaiming reality*. Routledge.
- Bishop, P., Hines, A., & Collins, T. (2007). The current state of scenario development: An overview of techniques. *Foresight*, 9, 5–25.
- Bolinger, M. T., Josefy, M. A., Stevenson, R., & Hitt, M. A. (2022). Experiments in strategy research: A critical review and future research opportunities. *Journal of Management Research*, 48, 77–113.
- Byrne, D., & Callaghan, G. (2013). *Complexity theory and the social sciences: The state of the art*. Routledge.
- Cairns, G. (2021). Resistance to hegemony in theorising scenario methods: A manifesto in response to Fergnani and Chermack, 2021. *Futures & Foresight*, 3, e76.
- Cairns, G., & Wright, G. (2018). *Scenario thinking* (2nd ed). Switzerland, Palgrave Macmillan.
- Clemens, M. A. (2017). The meaning of failed replications: A review and proposal. *Journal of Economic Surveys*, 31, 326–342.
- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66, 93–99.
- Crosan, R., Anand, J., & Agarwal, R. (2007). Using experiments in corporate strategy research. *European Management Review*, 4, 173–181.
- Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* (1982), 210, 2–21.
- Derbyshire, J. (2017). Potential surprise theory as a theoretical foundation for scenario planning. *Technological Forecasting & Social Change*, 124, 77–87.
- Derbyshire, J. (2020). Answers to questions on uncertainty in geography: Old lessons and new scenario tools. *Environment & Planning A*, 52, 710–727.
- Derbyshire, J., & Morgan, J. (2022). Is seeking certainty in climate sensitivity measures counterproductive in the context of climate emergency? The case for scenario planning. *Technological Forecasting & Social Change* (in press).
- Derksen, M., & Morawski, J. (2022). Kinds of replication: Examining the meanings of 'conceptual replication' and 'direct replication'. *Perspectives on Psychological Science*, 17, 1490–1505.
- Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, 130, 959–988.
- Downward, P. (2002). Critical realism, empirical methods and inference: A critical discussion. *Cambridge Journal of Economics*, 26, 481–500.
- Doyen, S., Klein, O., Simons, D. J., & Cleeremans, A. (2014). On the other side of the mirror: Priming in cognitive and social psychology. *Social Cognition*, 32, 12–32.
- Eden, D. (2017). Field experiments in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, 4, 91–122.
- Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2002). The Maryland scientific methods scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. Mackenzie (Eds.), *Evidence-based crime prevention* (p. 13Y21). Routledge.
- Feduzi, A., Faulkner, P., Runde, J., Cabantous, L., & Loch, C. H. (2022). Heuristic methods for updating small world representations in strategic situations of Knightian uncertainty. *Academy of Management Review*, 47, 402–424.
- Fergnani, A., & Chermack, T. J. (2021a). The resistance to scientific theory in futures and foresight, and what to do about it. *Futures & Foresight Science*, 3, e61.
- Fergnani, A., & Chermack, T. J. (2021b). Toward a science of futures and foresight: Response to 10 experts' commentaries. *Futures & Foresight Science*, 3, e97.
- Fleetwood, S. (2005). Ontology in organization and management studies: A critical realist perspective. *Organization*, 12, 197–222.
- Fletcher, A. J. (2017). Applying critical realism in qualitative research: Methodology meets method. *International Journal of Social Research Methodology*, 20, 181–194.
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12, 219–245.
- Frith, D., & Tapinos, E. (2020). Opening the 'black box' of scenario planning through realist synthesis. *Technological Forecasting & Social Change*, 151, 119801.
- Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures*, 25, 739–755.
- Gouveia, S. D. P. V. B. (2021) The limits to Randomised Control Trials—Lessons from the replication crisis, *Rerum Causae*. <https://rc.lse.ac.uk/articles/182/>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466, 29.
- Hodgkinson, G. (2021). Why appealing to the virtues of scientific theory (and method) is necessary but insufficient for effecting systemic change: A commentary on Fergnani and Chermack 2021. *Futures & Foresight Science*, 3, e79.
- Hodgkinson, G. P., Bown, N. J., Maule, A. J., Glaister, K. W., & Pearman, A. D. (1999). Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal*, 20, 977–985.
- Hodgkinson, G. P., & Rousseau, D. M. (2009). Bridging the rigour-relevance gap in management research: It's already happening! *Journal of Management Studies*, 46, 534–546.
- Hodgkinson, G. P., & Starkey, K. (2011). Not simply returning to the same answer over and over again: Reframing relevance. *British Journal of Management*, 22, 355–369.
- Jessop, B. (2015). The symptomatology of crises, reading crises and learning from them: Some critical realist reflection. *Journal of Critical Realism*, 14, 238–271.
- Kay, J., & King, M. (2020). *Radical uncertainty: Decision-making for an unknowable future*. Bridge Street Press.
- Kishita, Y., Kusaka, T., Mizuno, Y., & Umeda, Y. (2021). Toward theory development in futures and foresight by drawing on design theory: A commentary on Fergnani and Chermack, 2021. *Futures & Foresight Science*, 3, e91.
- Mandel, D. R. (2021). A positive futures for futures and foresight science needs fierce competition in the marketplace of ideas: A commentary on Fergnani and Chermack, 2021. *Futures & Foresight Science*, 3, e67.
- Meissner, P., & Wulf, T. (2013). Cognitive benefits of scenario planning: Its impact on biases and decision quality. *Technological Forecasting & Social Change*, 80, 801–814.
- Minkinen, M. (2021). Rigor and diversity in the futures field: A commentary on Fergnani and Chermack, 2021. *Futures & Foresight Science*, 3, e69.
- Mintzberg, H. (1977). Policy as a field of management theory. *The Academy of Management Review*, 2, 88–103.
- Münch, C., & von der Gracht, H. (2021). A bibliometric review of scientific theory in futures and foresight: A commentary on Fergnani and Chermack, 2021. *Futures & Foresight Science*, 3, e88.
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, 18, e3000691.
- Oatley, T. (2019). Toward a political economy of complex interdependence. *European Journal of International Relations*, 25, 957–978.

- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7, 531–536.
- Patomäki, H. (2006). Realist ontology for futures studies. *Journal of Critical Realism*, 5, 1–31.
- Peters, O. (2019). The ergodicity problem in economics. *Nature Physics*, 15, 1216–1221.
- Phadnis, S. (2021). Advancing scenario planning theory: A commentary on Fergnani and Chermack, 2021. *Futures & Foresight Science*, 3, e72.
- 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.
- Phadnis, S., Caplice, C., Singh, M., & Sheffi, Y. (2014). Axiomatic foundation and a structured process for developing firm-specific Intuitive Logics scenarios. *Technological Forecasting & Social Change*, 88, 122–139.
- Popper, K. (1959). [2002] *The Logic of Scientific Discovery*, London: Hutchinson. *Republished 2002*. Routledge Classics.
- Ravetz, J. (2022). Science— post-normal perspectives. *Futures*, 140, 102958.
- Reed, M., & Harvey, D. L. (1992). The new science and the old: Complexity and realism in the social sciences. *Journal for the Theory of Social Behaviour*, 22, 353–380.
- Rescher, M. (2000). *Realistic pragmatism: An introduction to pragmatic philosophy*. State University of New York Press.
- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, 14, e12633.
- Rowland, N., & Spaniol, M. (2021a). On inquiry in futures and foresight science. *Futures and Foresight Science*, 3, e37.
- Rowland, N., & Spaniol, M. (2021b). The reception of theory in futures and foresight: A commentary on Fergnani and Chermack. *Futures & Foresight Science*, 3, e66.
- Salo, A. (2021). Developing the needed scientific theory will not be easy: A commentary on Fergnani and Chermack, 2021. *Futures & Foresight Science*, 3, e73.
- Savage, L. J. (1954). *The foundation of statistics*. Wiley.
- Sayer, A. (2004). Why critical realism? In S. Fleetwood, & S. Ackroyd (Eds.), *Critical Realist Applications in Organisation and Management Studies*. Routledge.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13, 90–100.
- Schoemaker, P. J. H. (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic Management Journal*, 14, 193–213.
- Schoemaker, P. J. H. (2021). Theory development in foresight research: Commentary on Fergnani and Chermack, 2021. *Futures & Foresight Science*, 3, e74.
- Schwarz, N., & Clore, G. L. (2016). Evaluating psychological research requires more than attention to the N: A comment on Simonsohn's (2015) 'Small Telescopes'. *Psychological Science*, 27, 1407–1409.
- Schwenk, C. R. (1982). Why sacrifice rigour for relevance? A proposal for combining laboratory and field research in strategic management. *Strategic Management Journal*, 3, 213–225.
- Soros, G. (2009). *The crash of 2008 and what it means*. Public Affairs.
- Soros, G. (2013). Fallibility, reflexivity, and the human uncertainty principle. *Journal of Economic Methodology*, 20, 309–329.
- Van de Ven, A. H. (2007). *Engaged scholarship: A guide for organizational and social research*. Oxford University Press.
- Wilkinson, A. (2009). Scenarios practices: In search of theory. *Journal of Futures Studies*, 13, 107–114.
- Wilkinson, A., Kupers, R. & Mangalagiu, D. (2013) How plausibility-based scenario practices are grappling with complexity to appreciate and address 21st century challenges, *Technological Forecasting & Social Change* 80, 699-710.

**How to cite this article:** 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. <https://doi.org/10.1002/ffo2.146>