A Scientific Exploration of Scenario Planning, Thinking, and Cognitive Biases

By Megan M. Crawford

A Dissertation Submitted to

The Faculty of
The Department of Management Science of the
Business School in the University of Strathclyde
In partial fulfillment of the requirements for the
Doctor of Philosophy degree in
Management Science

July 2021

Dissertation directed by

George Wright
Professor of Management Science
Ron Bradfield
Director SBS In The Gulf - Management Science

Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

One experimental study was partially facilitated by two internal faculty members from the Strathclyde Business School, collectively referred to as "the team". Their support amounted to ensuring a sample of participants within the UAE received their assigned experimental materials. The materials were in digital format, accessed through an online survey platform. The faculty members did not develop nor alter any of the materials. They served only as a temporary proxy in the author's place due to travel restrictions placed on the author at the time.

Signed:

Date: 01 July 2021

Acknowledgements

I absolutely must thank my family for enduring this journey with me, supporting me along the way, and just being the best bunch of people I could ever hope to know. A special heartfelt appreciation to my kids, Jamie and Blake, for our long conversations and odd hours, and to my mom, who supported me all the way to the end. I want to thank my supervisor, Prof. George Wright, for his unending support, at all levels, throughout my investigation. His incredible insights on the practice and business of scenario planning, and personal encouragement to explore the unknown helped me reach beyond all expected boundaries. My supervisor, Dr. Ron Bradfield, deserves all my gratitude for his support with my empirical studies over the years, and his illuminating discussions about the vital importance of facilitators. I also must thank my mentor, Dr. Miguel Moreno, for helping me with the psychological aspects of the theory, statistical methods, and not least of all, his support and encouragement during my toughest times. A special thank you to Anna Sehn for being my first research assistant, and a brilliant one at that. A round of thank-yous to all the teaching and support staff in the Department of Management Science who helped me with the daily operations of my laboratory-based experiments. My compadre, Tom, and our long discussions with Côtes du Rhône always managed to keep my perspectives grounded, and for that deserves my unending gratitude. Finally, I am indebted to the hundreds of volunteers, across the globe, who graciously agreed to be my scenario thinkers. Their participation not only made this investigation possible, but their willingness to share their thoughts and opinions has brought greater knowledge to the fields of scenario planning and foresight.

Abstract

Scenario planning, as a recognised practice, is approaching the better part of a century. In this time it has experienced broad application across various industries and, as of late, growing popularity as an academic discipline. In stark contrast to its prolific use in the field and academia, is the lack in scholarly work that brings verifiable and robust knowledge regarding the efficacy of the practice. In order to understand the impact of scenario planning interventions, it is first necessary to understand scenario thinking. The importance of investigating scenario thinking lies in the notion that scenario planning has less to do with forecasting (i.e. aiming for facts) and more to do with futures-thinking (i.e. working with perceptions). The mental models, experiences, and abilities of scenario teams largely dictate the efficacy of a scenario planning intervention. At this time, however, scenario thinking remains a black box. The present investigation, first, provides a discussion on how to understand scenario thinking. A gestalt perspective is offered, where discrete cognitive features are defined, which comprise the structure of scenario thinking. The motivation to this discussion is understanding the level(s) of influence scenario thinking may succumb to, in the face of changes to external information. Next, three higher-order cognitions (creative, causal, and evaluative thinking) are explored, in depth, and tested against the Intuitive Logics model of scenario planning to help determine i) the robustness of scenario planning against ii) the influence of the cognitive experience. A multi-attribute approach is taken, borrowing methods from cognitive psychology, behavioural economics, and management science. A form of the traditional framing manipulation is used to measure for biases in scenario thinking. Results suggest that even the smallest change in information can lead to several biasing effects across the tested cognitive features of scenario thinking. Understanding the nature of influences on scenario thinking helps reveal the efficacy of scenario planning for management and organisations.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

APA	-	American Psychological Association
CC	-	Causal Connections
CIA	-	Cross-Impact Analysis
CQEC	-	Conversation Quality And Engagement Checklist
CS	-	Consensus
DE	-	Decision Explorer
DF	-	Driving Forces
DI	-	Dialectical Inquiry
DLOQ	-	Dimensions Of The Learning Organization Questionnaire

DV	-	Dependent Variable
ESPAS	-	European Strategy and Policy Analysis System
FEMA	-	The Federal Emergency Management Agency
GDMS	-	General Decision-Making Style Survey
HI	-	High Impact
HP	-	High Predictability
I/P Matrix	-	Impact/Predictability Matrix
IL	-	Intuitive Logics
IV	-	Independent Variable
LI	-	Low Impact
LP	-	Low Predictability
MMSS	-	Mental Model Style Survey
OSF	-	Open Science Framework
PMT	-	Probabilistic Modified Trends
Q12	-	Gallup Workplace Audit
SBS	-	Strathclyde Business School
SI	-	Surprise Index
SP	-	Scenario Planning
ST	-	Scenario Thinking
STIRDEEPER	-	Society, Technology, Industry, Resources, Demographics, Economics, Environment, Politics, Energy, Religion
TAMUCC	-	Texas A&M University-Corpus Christi
TCC	-	Total Number of Causal Connections
TIA	-	Trend-Impact Analysis
UEC	-	University Ethics Committee

Section I – An Introductory Overview

"It is an art *and* a science."
Ron Bradfield, 2015

Chapter 1. **The Overture**

This thesis presents the outcomes of a six-year investigation into scenario planning (SP) and scenario thinking (ST). The information reported here is divided into three overarching sections. The first section presents the justification for the chosen path of investigation, articulates the main objectives of the thesis, provides an historical account of the field of practice, justifies methodological choices, and develops four guiding research questions. The second section reports the outcomes of seven empirical studies developed to help answer the research questions. The third section addresses the research questions through a comprehensive discussion on implications that follow from the empirical studies for both furthering academic research and improving real-world applications.

The greatest strength of SP is arguably the broad agreement in its success. The greatest weakness of SP is definitively proving its successes (Chermack, 2002; Spaniol & Rowland, 2018b). Possibly, more to the point, few can agree what success should even look like. Is SP successful when the mental models of managerial practitioners are different at the end of an intervention compared to the start, as Wack, Schoemaker, Chermack, and others suggest? Or is SP's success found in the outcomes of strategic action, as Aldabbagh, Balarezo, and many private firms suggest? Science considers this a dilemma; Industry capitalises on it; Researchers declare methodological chaos. How to define SP? The number of ways to view SP are so numerous that this simplistic question proved to be somewhat problematic.

The investigation started by gaining a picture of how the private sector presents their scenario work. Private organisations (e.g. consulting firms) were chosen because SP is an applied approach to strategy praxis. SP supports the active development of strategy by being part of “the various activities involved in the deliberate formulation and implementation of strategy” (Whittington, 2006, p. 619). SP is part of the intra-organisational work required in developing and executing an organisation’s strategy, and therefore finds its value in its relationship to strategy.

Four major themes emerged from this explorative field search. First, firms present a fairly homogenous picture of the future. The future is almost entirely opaque (i.e. “faced with an uncertain future”[1]), and firms largely employ fear-related language to describe this uncertain future (e.g. “global economic crisis that is still not fully resolved”, “turbulence on the rise”, and “ever-shifting threat landscape may adversely impact its business, clients, guests, and families”[2]). Second, SP is the ultimate choice in strategy tools (e.g. “One thing certain is that the future is uncertain. Scenario Planning can help” and “Scenario Planning is arguably the most important step in your strategy planning process” [3]). This is understandable considering firms have commodified SP and therefore need to present it as attractively as possible so clients feel compelled to ‘purchase’ the tool. However, this bias offers little in the way understanding *what* SP is. Third, industry descriptions are pre-occupied with presenting outcomes, rather than the process, and employ a high-rate of certainty in their language (e.g. “driving better strategic thinking across top and middle management”, “Our method is a road-tested solution that prepares you for numerous situations”, and “Better decisions today for an uncertain tomorrow”[1-3]). Fourth, profiled organisations claiming to use SP regularly and benefit from the practice are mostly Fortune 500 companies and similar (e.g. Coca Cola, Microsoft, IBM, and VISA). Understandably, brand awareness makes promoting these companies to future clients more attractive than, say, the local neighbourhood shop. However, the bias in promotion goes deeper than that. SP is a resource-intensive process, and as such, is not as attractive to the culture of SMEs, who have far more

---

[1] PA Consulting (https://www.paconsulting.com)

[2] Boston Consulting Group (https://www.bcg.com), Guidepost Solutions (https://www.guidepostsolutions.com), StratForma (https://www.stratforma.com)

[3] Fuld + Company (https://www.fuld.com), Arcus Consulting Group (https://arcusgroup.ca)

limited (financial) resources than Fortune 500 companies. In short, industry knowledge offers a number of insights, but little in the way of gaining much traction in defining SP.

In academia, a number of published reviews help work through the “chaos” so many claim to be endemic in the field (Spaniol & Rowland, 2018b). Chermack & Lynham (2002) offer a representative sample of definitions in the extant literature and frame these against the outcome variables the respective authors associate with their chosen SP definition. Starting with their representative sample, a list of examples is developed that highlight the broader qualities discussed in SP. From the selection of reviews, some epistemological divisions begin to emerge (*see* Amer, Daim, & Jetter, 2013; Bradfield, Derbyshire, & Wright, 2016; Chermack, 2018; Huss & Honton, 1987). Each publication provides a main focus that appears to fall within one of three possible categories: *content*, *action*, or *mental state*. What emerges from the initial search is a categorical division of focus.

Some authors focus on the *content* of SP:

- “An internally consistent view of what the future might turn out to be—not a forecast, but one possible future outcome.” (Porter, 1985, p. 63)
- “A description of a possible or probable future” (Bloom & Menefee, 1994)
- “External scenarios are ‘internally consistent and challenging descriptions of possible futures.’ An internal scenario is ‘a causal line of argument, linking an action option with a goal’.” (Van der Heijden, 1997, p. 5)
- “Scenarios are descriptive narratives of plausible alternative projections of a specific part of the future.” (Fahey & Randall, 1998, p. 6)
- “Scenario planning is an effective futuring tool that enables planners to examine what is likely and what is unlikely to happen, knowing well that unlikely elements in an organization are those that can determine its relative success.” (Alexander & Serfass, 1998, p. 35)
- “A series of imaginative but plausible and well-focused stories of the future.” (Kahane, 1999, p. 511)

- “Scenarios are literally stories about the future that are plausible and based on analysis of the interaction of a number of environmental variables.” (Kloss, 1999, p. 73)
- “A scenario is simply a means to represent a future reality in order to shed light on current action in view of possible and desirable futures.” (Godet, 2001, p. 63)

The authors who speak of SP in terms of *content*, appear to be defining just the “scenario” portion of SP. Scenarios are artefacts of the process, tool, or technique. This is little more than a categorical error. Scenarios are not SP and vice versa. Scenarios are the final product in SP. They distill all the shared knowledge from the workshop into a selection of key storylines about the future.

Some authors focus on the *action* of SP:

- The result of systematic attempts to develop complex statements about future conditions relevant to your company (Linneman & Klein, 1979, p. 84)
- “The process of constructing alternate futures of a business’ external environment.” (Simpson, 1992, p. 10)
- “A disciplined methodology for imagining possible futures in which organizational decisions may be played out.” (Schoemaker, 1995, p. 25)
- “That part of strategic planning which relates to the tools and technologies for managing the uncertainties of the future.” (Ringland, 1998, p. 83)

Scenarios are populated with the details, outcomes, implications, and foundational structures that develop from the wide variety of SP *actions*. The authors who speak of the action, more often than not, take a more wholistic view of SP. These authors present what appears to be the most comprehensive, collective view of SP. They discuss the elements of the workshop or intervention methods, such as brainstorming sessions, clustering exercise, and team interactions. Many of these action-focused papers also discuss the variety of techniques used to eventually construct the scenarios: computer modelling, high vs low outcomes, team construction, intuitive exercises, and validation efforts.

Finally, the remaining group of authors appear to focus on the *mental state* through SP:

- "Hypothetical sequences of events constructed for the purpose of focusing attention on causal processes and decision points." (Kahn & Weiner, 1967, p. 6)
- "A tool for ordering one's perceptions about alternative future environments in which one's decisions might be played out" (Schwartz, 1991, p. 45)
- "An imaginative leap into the future." (Collyns, 1994, p. 275)
- "The present Chairman of Royal Dutch/Shell, C. A. J. Herkstroter, noted that the principal purpose of scenarios may be to "offer a way to decipher the overwhelming and often confusing information of the present." (Duncan & Wack, 1994, p. 21)
- "Scenario planning is inherently a learning process that challenges the comfortable conventional wisdoms of the organization by focusing attention on how the future may be different from the present." (Thomas, 1994, p. 6)
- "Although scenarios can free our thinking, they can still be affected by biases." (Schoemaker, 1995, p. 38)
- "Tools for foresight-discussions and documents whose purpose is not a prediction or a plan, but a change in the mindset of the people who use them." (De Geus, 1997, p. 46)
- "…one path through a person's cognitive map." (Van der Heijden, 1997, p. 5)
- "Scenario planning is one of a number of foresighting techniques used in the strategic development of organizations, which exploit the remarkable capacity of humans to both imagine and to learn from what is imagined." (Bawden, 1998, p. 7)
- "Scenarios are a management tool used to improve the quality of executive decision making and help executives make better, more resilient strategic decisions." (Wilson, 2000, p. 24)
- "…a number of cognitive barriers that appear to limit learning in the scenario development process." (Bradfield, 2008)

- "…provide a modelling framework that mitigates the cognitive biases that would likely to be associated with unstructured strategic choice." (Cairns, Goodwin, & Wright, 2015, p. 11)

Unlike the content and actions of SP, the authors who speak of *mental states* appear to define "scenario thinking", rather than SP. ST refers to the internal processes and cognitive experiences that inform and are influenced by SP (Cairns & Wright, 2018b; Chermack, 2011). ST has more recently come into popularity and is gaining ground in scenario and strategy literature, in what Hodgkinson & Clarke (2007, p. 243) refer to as "the re-humanization of strategy research". The mental processes and features that define ST are still in the early stages of exploration, but what is known (or at least discussed) at this time reveals a near-boundless realm of knowledge, reasoning, judgments, and decision-making. The capacity for ST is as yet fully understood, but cognitive and neuropsychological literature paint a picture of a seemingly endless conceptual cognitive landscape (*see for examples* Kiely, 2014; Lintern, 2007; Runco, 2014b).

By exploring this division in focal definitions within the extant literature, some of the motivations behind declarations of methodological chaos become a little clearer. After nearly 70 years in development and practice, though, many experts feel any attempts to quell the chaos are untenable, "the problem remains that a clear, consistent and agreed upon definition of scenario planning is yet to be achieved, if it is even possible at all." (Bradfield, et al., 2005; Chermack, 2018, p. 47; Martelli, 2001).

1.1 **Statement of the Problem**

From an empirical standpoint, the path towards a unifying theory is possible, even if perhaps largely unknown in the early stages of the discipline. Such developments, after all, have been the case with most verifiable disciplines in their early stages; progressing from splintered, disordered efforts and discoveries towards greater synthesis of knowledge as time and exploration progresses. The evolutionary stage of SP at this time is in its infancy. As Harley, et al. (2010) state, a high degree of theoretical and conceptual activity in a discipline can indicate a relatively immature stage in development. Such a state of development allows for incredible explorative potential. “Scenario planning has grown out of practice” (Chermack, 2018, p. 45). However, with such plurality in the narrative, how best to approach the issue of ordering the chaos?

One option is to explore the scenarios themselves, from a content perspective. This method requires a bottom-up approach, where the last development in the process is investigated. It is clear, though, throughout the literature, that scenarios are only as valuable as the active process. As the discipline began to spread beyond the uses of military wargames and the high walls of Royal Dutch Shell, authors, such as Klein & Linneman (1981, p. 77), began to understand the variability in expression, value, and use of the final scenario products, “Scenario writing techniques are not used to generate a predicted future, rather for the identification of plausible boundaries of environmental expectations within which corporate activities take place.” Schwartz (1991, p. 9) holds that scenarios are to be used more as end-state techniques for adjusting perceptions, "the end result, however, is not an accurate picture of tomorrow, but better decisions about the future." An even more extreme view from some authors is that the scenarios, themselves, are almost valueless. They serve only to promote a different effort in thinking which leads to more effective strategic actions. Possibly most damning to understanding the value of scenarios is found in the external forces that affect the process. For example, Volkery & Riberio (2009, p. 1199) recognise that, “Even well-constructed, thoroughly analysed scenarios can be of little use and relevance, if the organisational capacity to absorb them is poor, if there is no political backing or if relevant specifics of the policy-making process have not been taken into account.”

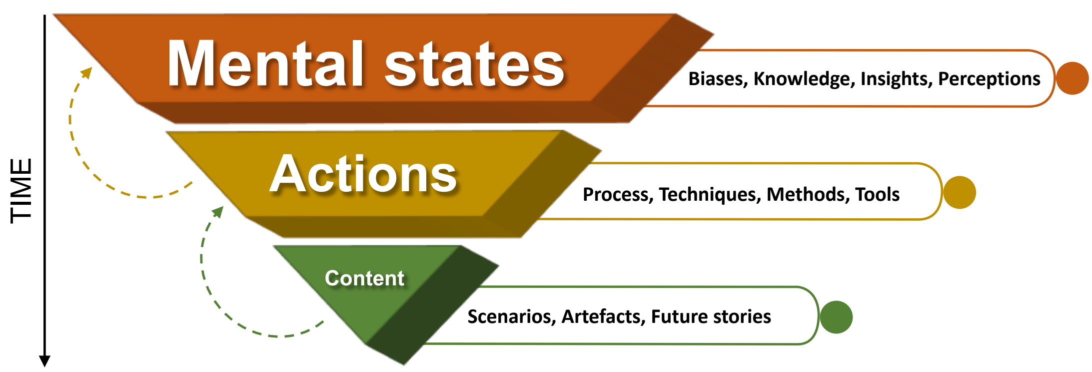
If scenarios are the products of complex actions, whose value depends on a host of other external factors, and content may not even serve a further purpose beyond the workshop setting, then perhaps it would be more appropriate to explore SP, the action. The question quickly arose, *Which SP model do I explore?* Herman Kahn, Pierre Wack, and Ted Newland pioneered the method – and nomenclature – for modern times, beginning in the mid 20th century. Several models spawned from their initial work, the most popular being the Intuitive Logics (IL) model, followed by the more quantitative methods of the Probabilistic Modified Trends (PMT), and then later, Gaston Berger's French model of *La Prospective.* The earliest published, full SP method, though, comes from Vanston Jr, et al. (1977) and lays out their 12-steps to scenario development. The authors describe largely an IL model, but include the option for quantitative features (e.g. computer simulations) to help determine the probability of scenarios. Most of the literature that describes a SP process, largely reflects versions of the IL model. The consistency in methods, however, ends there. For every publication that exists, a new step, stage, or technique is added and/or removed from existing methods, creating further confusion in research. The inconsistencies in models and methods are a reflection of a pragmatic discipline. Chermack (2018, p. 50) recognises the dilemma of pragmatism in a world of Popperian science, "it is generally difficult to engage in intervention research when the intervention (SP) is usually customized, methods are varied as well as the timespans and contexts of different scenario projects."

Just like the scenarios, the action can also prove to be inert, given a number of higher-order variables. Wack (1985a, pp. 139-140) is clear in his assessment of SP's value.

> Scenarios deal with two worlds: the world of facts and the world of perceptions. They explore for facts but they aim at perceptions inside the heads of decision makers. Their purpose is to gather and transform information of strategic significance into fresh perceptions. This transformation process is not trivial—more often than not it does not happen. When it works, it is a creative experience that generates a heartfelt "aha!" from your managers and leads to strategic insights beyond the mind's previous reach. I have found that getting to that management "aha!" is the real challenge of scenario analysis.

The reality that began to emerge from the search was that perhaps the best approach to understanding the discipline was not to take the same path as others in the field, but approach this from a top-down perspective. If the value of SP is based on the effectiveness of the active process, and effectiveness is determined by the mindset of managers (i.e. mental models, biases, knowledge, and perspectives), then perhaps the most beneficial answers would come from investigating the mental states of ST? Figure 1.1 illustrates a hierarchical interpretation of SP research perspectives. The start of any SP intervention begins with the mental states of scenario practitioners and what they bring to the table. Their mental states collectively offer the broadest parameters, with seemingly endless insights, knowledge, experience, and perceptions. Mental states also come with inherent limitations such as short-cut heuristics that can lead to specific reasoning, judgment, and decision-making biases. The knowledge practitioners bring to the table determine the quality of information that is generated and shared during the active participation of a SP intervention or workshop. In return, these knowledge sharing exercises further influence what and how practitioners perceive the issues at hand (i.e. the dotted yellow line). The shared information is by necessity a pared down, focused subset of the broader pool of information that is supplied by the practitioners. The content of each scenario is the distilled information that explores key information about a few plausible futures, and are the final artefacts of the effort. As content is actively generated and decided upon, it informs and alters the kinds of actions and thinking the practitioners continue to develop in the final stages of SP (i.e. green dotted line). The scenario artefacts are the final output of the full SP intervention.

Figure 1.1. Scenario planning hierarchy of topical focus



The main position in the extant literature, however, views the process in reverse. Practitioners are treated like a homogeneous actor in the narrative, where SP is assumed to affect ST. The few empirical studies that exist take a bottom-up approach to their investigations, reflecting this same reverse-influence assumption. Practitioners are assumed to enter with a global ‘limited thinking’ bias, against which SP is tested to determine how much it alters this bias. Theoretical and empirical positions in the field claim that SP can (and should) “stretch” thinking (Van Notten, et al., 2003), be a learning experience (Cairns & Wright, 2018b), change managerial mental models (Chermack, 2011), and modes of thinking (Franco, Meadows, & Armstrong, 2013; Schoemaker, 1995; Van der Heijden, 1997; Van der Heijden, et al., 2002; Wack, 1985a, 1985b; Wilson, 2000; Wright, 2014). The main take-away is that SP should alter ST in some beneficial way through un-biasing mechanisms. This is a lofty claim, particularly given the lack of definitive evidence in support. Some studies present evidence of cognitive changes (Bodin, Chermack, & Coons, 2016; Meissner & Wulf, 2013), while others report no changes in various qualities of decision-making (Chermack & Nimon, 2008; Schnaars & Topol, 1987), and none offer definitive evidence of beneficial, de-biasing changes.

Some key issues arose from the literature review. More often than not, authors do not make explicit the specific SP method discussed in their paper. This has left a major gap in our understanding of efficacy. Most of the methods borrow from the IL model, though all are different in technique, and all empirical investigations measure decision-making (e.g. as opposed to reasoning, judgment, or deliberation). To better understand efficacy, authors need to employ more transparency in their writing. Closely linked to lack of method clarity is environment awareness. Particularly with the empirical studies, authors make little effort to develop a dialogue around environmental cues and the influential differences one workshop setting can have on individuals and the team as a whole, compared to other settings. The field would benefit from greater inclusion of environmental cues in the dialogue. Another key issue that stood out is mentioned in the previous paragraph. Participants and practitioners are not treated like individuals informed by unique histories and education. They are, instead, treated like some generalized exemplar of an executive who brings the same cognitive state to any given SP session. A century of psychological studies and a millennia of philosophical inquiry show this assumption to be demonstrably untrue. Each scenario practitioner, CEO, executive, manager, and stakeholder is not only a unique actor in and of themselves, but can even be said to be unique from one moment to the next. Heraclitus believed the concept of 'becoming' is built on the act of constant experience in an impermanent environment. This is presented in contrast to 'being' which reflects a static state of self. The impermanence of a single self, and the transitivity of knowledge is further discussed in section 1.4.

Then a study stood out that features probably the biggest obstacle to determining the efficacy of SP. Through a fully immersive workshop, mirroring field experiences with SP, Bradfield (2008) reports a pervasive *cognitive inertia* from participants, where the process of SP not only failed to inspire participants away from their existing mental models, but also beyond their contemporary influences. In this light, the interpretation behind several empirical SP studies came into question. What if the authors were actually measuring various forms of cognitive inertia, but not necessarily recognising it as such? In order to understand the efficacy of SP, we must understand ST and what this means to practitioners (both as individuals and as members of management teams). Gaining clarity and deeper understandings of the function of ST can improve "the interweaving of theory and practice" (MacKay & Tambeau, 2013, p. 674), which can bring greater value to the "artful and improvisational" SP praxis (Whittington, 2006, p. 620). How, then, to use the language and tools of empirically-based science to explain the fuzzy, intuitive, even chaotic, phenomenon of ST?

1.2 **Aim, Objectives, and Central Question**

The central aim to this investigation is to supply empirical discoveries that help improve theoretical and practical knowledge of SP by way of ST. This aim will be achieved through the following objectives.

(1) Conduct comprehensive reviews of both SP and ST literature, including a special focus on empirical SP studies and experiments
(2) Interview SP practitioners to gain a snapshot of contemporary praxis
(3) Synthesise insights from reviews with data from both the empirical review and interviews to identify gaps in knowledge
(4) Design and test empirical studies that address some of the identified gaps
(5) Provide data-informed recommendations for future SP research
(6) Provide data-informed recommendations for practitioner-driven SP techniques

The efficacy of SP will be described by framing investigations in terms of ST. The elephant in the room, decade after decade, is the discipline's lack of robust evidence. The basic, but lofty desired result, according to Wack and others, is to alter the mental states of managers in such a way that leads to novel, strategic insights which, in turn, foster better strategic decisions in the future. By applying the scientific method to an intuitively-based practice, this investigation aims to reveal novel methods for understanding the cognitive journey that develops ST, the practice, and specific techniques. In this way, practitioners can use novel methods to develop and improve their own SP interventions. Further to this, academics will be provided with a collection of inter-disciplinary approaches that enrich the field of study. The proposed empirical approach is to narrow "the focus of the study but leave open the questioning" (Creswell, 2009, p. 130). The central question to this thesis asks:

How do cognitive biases in scenario thinking, a collection of higher-order cognitive functions, effect the content of scenario planning?

1.3 **Statement of Potential Significance**

This investigation is important for both scholarly and practical advancement. It contributes to the growing body of research literature in SP by providing new insights into the structure of ST which can bring clarity to existing discoveries in the scholarship. There are too few empirically-based studies for a field that is approaching the better part of a century, and too little agreement across the existing genre. The methods are rooted in the IL mo, as developed by Van der Heijden, et al. (2002) and formalised by Cairns, & Wright (2018a). The theory is anchored in Chermack's *Six Domains of Scenario Planning Theory* (2011) with a view through constructivism. The structure of ST builds from Cairns, & Wright's work against cognitive psychological theories.

This investigation's value to the field of practice is as great as its value to scholarly advancement. Popularity in SP consultancy is on the rise. Unfortunately, as stated in the opening of this thesis, the private sector speaks almost dogmatically of the value, using language that expresses opinions as if they are unwavering facts about SP (e.g. "Our Scenario Planning method is a road-tested solution"[3], "Scenario planning is an excellent tool"[4], "Scenario Planning is the best method to use"[5], "proves that our consulting services are successful"[6]) without definitive evidence to back such claims. Therefore, the following studies, which encompass ecological validity and employ real-world methods, aim to create a comprehensive approach to understanding ST and how facilitators can apply such knowledge to SP workshops. An outcome, of which, is to show the true value of SP and facilitate greater efforts in mainstreaming the praxis.

1.4 Theoretical Approach

SP literature breaks strongly from traditional approaches in strategy and management research. Many presuppose a static, singular environment for an organisation, easily understood, and more importantly, an environment set apart from the practitioner and the organisation. Guidance promotes the use of hierarchical taxonomies (e.g. PEST) to create easily analyzable, homogeneous groups of external variables (Burt, et al., 2006). The assumption is that a one-time analysis of a current situation is sufficient for generalisation. Such a perspective employs a critical realist view, assuming an

---

[4] Lyons-Newman Consulting (https://www.lyonsnewman.com/)

[5] Flevy (https://flevy.com/)

[6] CMG (https://www.cmgconsulting.com/)

objective truth, and that by identifying the structures that are already in place, the players can discover "the truth" (Archer, et al., 1998). There is a comfort in adopting the critical realist view in futures and strategy research. If there is an objective truth that reflects the external environment, then we can assume a unified goal to aim our independent research developments towards. However, SP is not a practice of objective truths, nor does it assume an objective environment that transcends the individuality of the organisation and the actors within.

Cummings and Worely (2001) see organisations as open systems that must find the best fit within their external environment. Emery & Trist (1965, p. 8) discuss the open system in terms of biological fit and survival.

> [A]ny living entity survives by importing into itself certain types of material from its environment, transforming these in accordance with its own system characteristics, and exporting other types back into the environment. By this process the organism obtains the additional energy that renders it ''negentropic''; it becomes capable of attaining stability in a time-independent steady state – a necessary condition of adaptability to environmental variance.

Each stage of the scenario process requires the scenario team to focus on various external factors, determine which factors could plausibly affect their organisation, use this knowledge to affect internal actions, then assess how these new actions could, in turn, further influence various external factors and mechanisms. An assumption of the success of SP is that praxis requires active participation by each team member (i.e. practitioner), which in turn helps them collectively discover their organisation's best fit for long-term survival (Van der Heijden, et al., 2002). Further, it is through this process that management teams are able to view their organisation as an open system, continuously adjusting against and within its contextual environment, rather than a static system with one true path forward.

Breaking from the traditional strategic planning view – often implicitly – constructivist views are adopted fairly regularly in SP literature. Some of the scholarship discusses an external business environment (an extension of the social environment) as one that is socially constructed, where practitioners reflexively build understanding of their experiences, through subjective meanings (*see* Kahn & Wiener, 1967; Mir & Watson, 2000; Schoemaker, 1993). Van der Heijden, et al. (2002) describe SP as an organisationally based social-reasoning process built on shared dialogue. It is through this active dialogue that practitioners engage in sense-making of their perceptions with the purpose of building cohesive stories of plausible external business environments. Smircich and Stubbart (1985, p. 724) state that, "environments are enacted through the social construction and interaction processes of organized actors". Actors (e.g. stakeholders and practitioners) influence and are influenced by their environments. This process creates an emerging reality that is reflexive, but certainly not static.

Such views reflect a perdurantism position that actors have both temporal and spatial parts (Effingham, 2009). It is intuitively easier to understand one's spacial parts; These are the physical spaces an actor physically takes up in their environment (e.g. their presence in a SP workshop and their presence in the organisational environment). An actor's temporal parts, though more conceptually discussed, are seen to similarly take up 'temporal space', which results in multiple versions of the self, such as you-yesterday, you-today and you-tomorrow. Perdurantism explains how we *persist* through time, while changing as we progress (Wasserman, 2016). Perdurance theory suggests that a practitioner will enter a SP workshop with one identity (i.e. their personal set of knowledge and awareness skills), gain different identities throughout the process of the workshop from interacting with and learning from other practitioners, and leave the workshop with a new identity (i.e. new knowledge and awareness skills). Through this lens, it can be explained how an actor was a novice in the past, who becomes ever-more an expert in the emerging future. Put in SP terms, practitioners enter with one set of mental models and leave with a different set of mental models.

In contrast to perdurantism is endurance theory (endurantism; Lowe, 2006). Endurantism views both people and objects as *enduring* through time. We maintain our identity throughout time and space. As a consequence of the theory, endurantism makes it difficult to even declare what temporal parts are supposed to be. A person temporally lies in the past, present, and future, with a consistency to self. The two theories ontologically disagree about temporal existence, and by extension, identity, which includes knowledge, awareness, and learning. Endurance theory does not support a constructivist view of SP, nor ST – as well as perdurantism – since it negates the kinds of learning and cognitive changes (e.g. mental models) inherent in the SP scholarship. This investigation borrows from the perdurance theory of temporal and spacial persistence of the self, both as a physical actor and a cognitive being.

From a design science perspective, this process is seen as a "voyage of discovery," where solutions can emerge from the process (Hodgkinson & Healy, 2008, p. 436). Design efforts, whether physical, dialectic, or methodological, are trialed and workable solutions are determined from their results, independent of any fully formed theoretical understanding of the systems being designed. In this manner, insights and knowledge develop reflexively through active field or case studies which, in turn, provide insights to inform future designs (Boland & Collopy, 2004).

According to Burt, et al., (2006) a constructivist approach to scenario research must necessarily focus on *how* and *what* practitioners define as their business environment in social settings. Definitions are dependent on a process where practitioners become aware of environmental signals and cues, then extract and attribute meaning (i.e. factors; Kiesler & Sproull, 1982; Weick, 1995). These meanings are internally derived based on the practitioner's experiences and knowledge, as an individual, a social actor and force within their environment, as well as a member of the scenario team (Kiesler & Sproull, 1982). Therefore, according to Burt, et al., research into SP should study practitioners (i.e. decision-makers) "as *they* define *their* own environment" (p. 59, authors' emphasis). This moves the research perspective directly from a socially constructed environment to a psychologically constructed perspective.

A strong assumption in behavioural science is that human thinking and knowledge is bounded, or limited in scope (Bradfield, 2008; Kahneman, 2011; Kahneman & Tversky, 1979; Simon, 1972; Wright & Cairns, 2011). Each practitioner's unique set of limitations, in isolation, is assumed to be insufficient in constructing a successful view of an organisation's social environment for the purposes of achieving long-term survival. Supporting the constructivist position, SP pushes practitioners to understand their business environment with a plurality of viewpoints (e.g. the other members of the scenario team). There is value in the personal perspective and equally there is value from negotiated interactions between multiple perspectives (Creswell, 2009). For example, the same signal or cue may be interpreted differently by different practitioners. SP equally values all interpretations as much as the identified signal. Reflecting the philosophy of constructivism, SP literature discusses a general process where a practitioner's background, history, social and cultural norms, and interactions with other members of the scenario team develops the kind cues and signals that are eventually identified throughout a workshop or session, as well as any meanings and interpretations attributed to them.

Practitioners are required to (a) identify goals and a timeline, (b) scope the business environment, and (c) construct multiple plausible futures, given (a) and (b), in order to help organisations increase their chances of long-term survival (Van der Heijden, 1997; Van der Heijden, et al., 2002). SP is a practice in identifying would *could be*, not what *is.* A scenario process is not one of determining a singular end result, nor is it one of proving objective truths. Therefore, SP research must employ a global method of investigation that accounts for multiple plausible realities and contextual interpretations, while affording validly measurable results with generalisable applications.

To respect the constructivist view of SP, the approach of the present investigation is one of pragmatism. Adopting a pragmatic scientific approach to the investigation allows the acknowledgement that ‘truth’ can be a constructed reality (Peirce, 1935), is temporally bound (Creswell, 2009), and is “what works out most effectively in practice” (Honderich, 2005, p. 747). This approach frames the present research within social, historical, and political contexts – amongst others. A pragmatic approach to investigating the cognitive efforts of practitioners as they navigate the stages of SP allows the investigation to acknowledge the subjective modality of ‘possibilities’ in decision-making. This is a key component to SP – the mode of possibility. Peirce claims there is a real ‘possibility’ when there is more than one state of things that no knowledge excludes (i.e. you cannot claim it is false), which a necessary effort in scenario perspectives, given the practitioner’s focus on unknown future states. As Wack (1985b, p. 73) claims, “the future is no longer stable, it has become a moving target. No single ‘right’ projection can be deduced from past behavior. Therefore, we need to accept uncertainty, try to understand it, and make it part of our reasoning”. Peirce’s necessary inclusion of ‘possibilities’ and Wack’s acknowledgement of ‘uncertainties’ are synonymous in this case. Dewey (1991) adds to this by claiming that our experience of the world is based on our interrelationship with it, and the importance of the practicality of that relationship. An experimental approach and critical examination are necessary to our understanding of our environment. By incorporating Peirce’s and Dewey’s pragmatism with Wack’s inclusion of uncertainty, a method of inquiry is used, guided by empirical research. The chosen approach to this investigation is based in cognitive psychology, borrowing from sociological theory. A selection of mixed methods, using psychometrics and econometrics, are applied to operationally define, test, and measure qualities of ST, from a practitioner’s perspective, against an exemplar of SP. It is through a pragmatic, mixed methods approach that the present investigation aims to help determine *how* and *what* practitioners’ define as their business environment in contextually dependent social settings.

Most of the existing empirical research into SP is concerned with the artefacts (i.e. scenarios), either produced from their sample of participants or used as a manipulation in their design. Some studies are void of the process entirely (Kuhn & Sniezek, 1996; Phadnis, et al., 2014). Other studies use known, final-outcome performance measures to determine success or failure of the method (Schnaars & Topol, 1987). Overall, many empirical studies treat the external environment as a static state of events, the participants as cognitively similar, the external environment as containing objective truths capable of being identified, and the results as contextually unbounded. However, these methods do not fit with a constructivist view, and therefore, fail to report necessary qualities of SP, either as a process or a tool.

Though many of the traditional methods do not lend themselves, as they stand, to a pragmatic approach, designed around a constructivist view of the active practitioner and their external environment, some of the tools can be useful in the present investigation. For example, taxonomies such as PEST or STIRDEEPER are popular tools for measurement in strategy and management research. Relying only on these categories to define an external business environment, though, limits the scope of cognitive exploration, because it implies that any given environment can and should be limited to these discrete, immutable categories. However, taxonomies are not entirely incompatible with a reflexive practice. The concept may prove beneficial if used through a modified method. A modification of a taxonomic application within an experimental design has the potential to allow practitioners to view their environment differently, more contextually, even more organically, and leave the potential open for insights that expand beyond present taxonomies. Reversing the method is one option. Rather than present the taxonomy first, and ask for brainstorming second, a facilitator could ask for brainstorming first, then create a taxonomy second, from their shared insights. That is, once practitioners have exhausted their ideas, they are then asked to categorise their identified factors by relation. By employing an exploratory design, we accept a constructivist view that there is no absolute ‘right’ category, the pragmatic approach that all identified factors and categorisations are equally valuable, there could be any number of potential related factors, each experience could be unique for each practitioner, and leave the door open for potentially new categories to be discovered for future taxonomical use. This approach allows the researcher to scrutinise the behavioural conditions that might enable scenario-based techniques to yield beneficial outcomes, while also leaving it possible to delve further into the causal mechanisms that motivate a practitioner's reasoning (Hodgkinson & Healy, 2008).

Using a pragmatic approach also accommodates mixed method designs. Quantitative measures are the standard in cognitive psychology. A variable is operationally defined, a hypothesis is stated, A/B testing administers the variable, behaviours are measured, and results are interpreted against the values of both the variable and behavioural outputs. However, quantitative interpretations of the black box that is *cognition* is only one way of understanding the scenario process, and offers an understandably partial story to the investigation. Qualitative methods, such as narratives, case studies, and discourse analysis, allow for richer understandings of the social process, and compliment the partial story that quantitative methods deliver. By incorporating both approaches into the investigation, a more comprehensive view of the social process with which practitioners engage during SP will be afforded.

Employing a pragmatic scientific approach will require the researcher to account for their own flexibility throughout the investigation. Accounts of data analyses against hypothesised expectations of individual studies will understandably reflect the researcher's own belief structures. The investigation will not only be a reporting of the path of discovery by practitioners engaging in SP, but the researcher's own evolution in understanding and adjustments, as a scientist.

1.5 **Definition of Terms**

Bias – A systematic pattern of deviation from norm or rationality in judgment.
(Tversky & Kahneman, 1974)

Confidence – An individual's belief and trust in their own decision.
(Koriat, Lichtenstein, & Fischhoff, 1980)

Cognition – "the activities of thinking, knowing, and processing information."
(Franco, Meadows, & Armstrong, 2013, p. 725)

Cognitive style – "individual differences in peoples' preferred way of processing (perceiving, organising and analysing) information using cognitive brain-based mechanisms and structures.
(Franco, Meadows, & Armstrong, 2013, p. 726)

Decision-making – A complex process involving directed attention, discovery, judgment, deliberation, determining a course of action, evaluating alternatives, and choosing among them.
(Simon, 1972)

Deep processing – Deliberative cognitive efforts that require attention and are affected by shallow processing.
(Kahneman, 2011)

Dependent variable – The output that is measured from participant performance.

Driving forces – "Those fundamental forces that bring about change or movement in the patterns and trends that we identify as underpinning observable events in the world."
(Van der Heijden, et al., p. 282)

Environment – Relevant physical and social factors outside the boundary of an organization that are taken into consideration during ST. Also referred to as "business environment", "organisational environment", and "external environment".
(Van der Heijden, et al., 2002)

Facilitator – AKA "scenario planner", "is the person (or group of people) involved in promoting and facilitating the learning process."
(Van der Heijden, 1997, p. 136)

Framing – How an issue or situation is represented.
(Tversky & Kahneman, 1974)

Heuristic – Mental shortcuts that ease the burden of judgment and decision-making. (Tversky & Kahneman, 1974)

Independent variable – A variable that remains constant and represents the manipulation within an experimental design. No other factors change its quality.

Participants – The managers and students who volunteered for one of the experimental studies reported in Chapters 5-9.

Practitioners[7] – The managers, executives, directors, CEOs, and stakeholders who participate in a workshop. Sometimes referred to as the "management team". (Van der Heijden, et al., 2002)

Scenarios - "have a temporal property rooted in the future and reference external forces in that context; scenarios should be possible and plausible while taking the proper form of a story or narrative description; and that scenarios exist in sets that are systematically prepared to coexist as meaningful alternatives to one another" (Spaniol & Rowland, 2018a, p. 1)

Scenario planning – A multi-stage, group-based approach that "always aims for the invention of strategy and testing of related organisational characteristics against

[7] Traditional business literature uses the title "practitioner" to identify SP facilitators and workshop leaders, and "participant" to identify those managers, executives, etc who participate in the workshops. These traditional identity distinctions within business literature are not employed in this thesis for the sake of preserving traditional human-based, scientific research vernacular. Volunteers within experimental studies are always designated either "subject" (pre-1990) or "participant" (post-1990), in accordance with American Psychological Association standards. In order to maintain a distinction between those *participating* within a workshop and those *leading* the workshop, "practitioner" and "facilitator" are employed for this thesis, respectively.

multiple representations of the future business environment."
(Van der Heijden, 1997, p. 131)

Scenario planning model – Four "main schools of techniques", or scenario methodologies, as recognized through the history of SP. Intuitive Logics, Cross-impact analysis, Trend-Impact Analysis, and *La Prospective*
(Amer, Daim, & Jetter, 2013; Bradfield, et al., 2005; Huss & Honton, 1987)

Scenario thinking – "contains the key components to promote effective exchange of opinions and beliefs among individuals with a shared interest in some critical issue. The construction of multiple futures holds open airtime for differing opinions about the nature of the future and provides a forum for the debates, questioning, and synthesis of complementary, contrasting, and conflicting viewpoints."
(Cairns & Wright, 2018a, p. 9)

Shallow processing – Automatic, involuntary cognitive efforts that develop learned associations and rely on heuristics for quick judgments.
(Kahneman, 2011)

Stages – The discrete tasks within a SP workshop.
(Cairns & Wright, 2018a)

Threat – Environmental factors that act upon the organisation.
(Van der Heijden, 1997)

Vignette – A short summary that includes pertinent information.
(Abelson, 1976)

1.6 Organisation of Chapters

This thesis is divided into three sections that categorise eleven chapters. The division and flow of the thesis is illustrated in Figure 1.2 at the end of this chapter. Section I presents the theoretical and methodological backgrounds. Section II presents eight empirical studies. Section III offers closing discussions.

The methodology developed throughout the investigation is emergent and attempts to introduce novel collaborations through an interdisciplinary approach. Knowledge-sharing from other disciplines can help reframe perspectives around SP and ST. The purpose to developing such a methodology is to introduce greater insights to the field and further develop a grounded theory. The resulting dialogue is information-rich and touches upon several unique categories. Therefore, the methodology is developed incrementally across the first five chapters of Section I to allow sufficient attention to its development. In conjunction with the emergent writing style, each chapter of Section II introduces an additional focused literature review that builds on the philosophical position and knowledge of Section I.

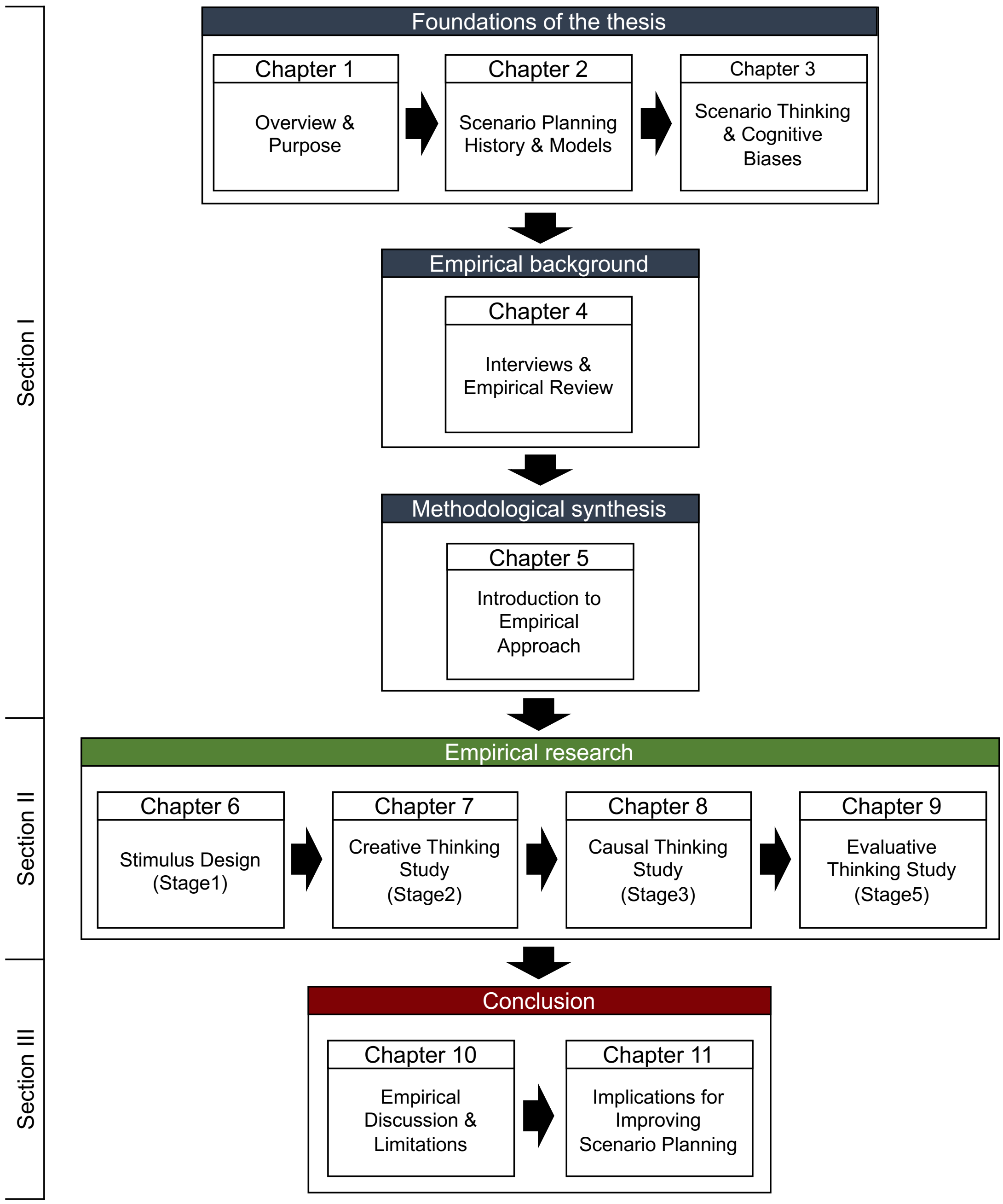
Section I presents the objectives, justification, literature review, and methodology for the investigation. The foundation of this thesis is divided across the first three chapters. Chapter 1 introduced the main problem upon which this thesis is focused. The study and practice of SP is still in its infancy, and as such remains in a somewhat chaotic stage of development. The purpose of this investigation is to bring some order to the chaos by attempting to understand SP at a more fundamental level. This is achieved by exploring the functions of ST and how the cognitive experience can affect the behavioural actions and final content. ST is viewed through constructivist theory, where a practitioner's knowledge is seen as constructed from their individual and group experiences, personal history, and biases, with on-going learning and changing. A pragmatic methodology is adopted to help conform to a constructivist approach, where the researcher's own learning through the investigation will inform mix-methods designs. Chapter 2 presents an historical review of SP with a discussion on the prevailing models. Justification is given for framing the investigation specifically upon the IL model, which provides a map against which the functions of

ST can be delineated. Chapter 3 reviews the present definitions of ST. Cognitive psychological theory is used to categorise the various higher-order functions of ST. Using the chosen SP model identified in Chapter 2, a roadmap is developed where dominating cognitions of ST are plotted against the chosen SP model, to create the Intuitive Logics/Scenario Thinking (IL/ST) framework. This framework is used to develop the structure of research conducted in Section II. Chapter 4 reports three interviews with industry executives and reviews the existing empirical literature. Discoveries from this chapter justify the focused development for the empirical perspectives. Chapter 5 ties together the full methodological design: key cognitive biases and barriers, confidence measures, and the four research questions.

Section II presents the eight empirical studies which reflect the bulk of the investigation. Chapter 6 develops the stimuli to be used as the manipulation across the remainder of the studies. Through priming and validation testing, a novel form of testing ST is created using business vignettes to deliver the independent variable (i.e. framing manipulation). Chapter 7 explores creativity in ST during Stage 2 of the IL/ST framework. Chapter 8 explores causal mechanisms in ST during Stage 3. Chapter 9 explores evaluative thinking in ST, which is largely employed in Stage 5.

Section III presents concluding remarks, discusses extensive implications, and explores further questions. Chapter 10 presents a comprehensive discussion of the empirical results and offers practical recommendations for future research methods. Finally, Chapter 11 concludes this thesis with a discussion on several implications for improving SP practice.

Figure 1.2. Thesis flow chart outline



1.7 **Writing Perspective**

This thesis uses a gender-neutral narrative. Personal titles are used when possible, "participant" is used in place of "man" or "woman", and "they/their/them" are used instead of gendered pronouns. This is to preserve unbiased language where possible, to fully represent the collection of participants across the studies, and conform to contemporary academic publishing standards.

Chapter 2. **Scenario Planning**

> ". . . serious students of human psychology will expect as much madness as wisdom from large groups of people."
> Niall Ferguson, *The Ascent of Money*

In the first chapter, the aim of the investigation was outlined; To help determine, in part, how ST effects the actions and content of SP. In this way, outcomes are sought to improve our collective understanding of the efficacy and value of SP. To date, however, there is no agreed upon single or universal SP method. The discipline has evolved for decades from numerous sources and is now an umbrella term for a plethora of techniques and prevailing models. Faced with this plurality in choice, Chapter 2 takes an historical walk through the evolution of SP, largely focused on the last 50 years of renewed popularity. The aim of this review is to more clearly define SP. A working definition of SP will guide the process of identifying a model to serve as the framework for the investigation. The model will provide a map against which the stages of SP and the functions of ST can be delineated. This unified roadmap will then inform the methods developed for the empirical studies. Understanding the etiology of SP is necessary to ultimately developing clear empirical investigation.

2.1 **History of Scenario Planning**

Like many branches of science, art, and mathematics, the development and popularizing of SP as a recognised organisational intervention, has many simultaneous fronts. Just as Leibniz and Newton both worked on differing forms of calculus almost simultaneously, or as the Theory of Evolution developed in progressive formation across Lamarck, Darwin, and Wallace in short succession, so too does SP reveal its multi-parentage. Imagining scenarios of the future, particularly for the purpose of strategic advantage, and our role within that future, is about as old as living history. The Zhou Dynasty (1046-256 BCE) produced some of the longest lived texts on futures-thinking and military strategy. The infamy of ancient Greece's art of storytelling and philosophy as means for wargaming plausible realities is well known. Then there's ancient Rome, famous for thriving under military and political strategies through successful foresight, In short, the techniques used in SP today, have been employed for strategic advantage for centuries.

The path that brings SP from early human history to our contemporary feet also shares similar developments with science and mathematics. Observations and early interest started the dialogue, but, following a century's-long silence, emerged with accelerated force in modern times, to find a well-supported perch in the 20th century.

SP has been in practice in the west since the post-war period, largely due to the pioneering efforts of Herman Kahn at the RAND Corporation (1948-1961) and the collaborations between Ted Newland (1965-1981) and Pierre Wack (1971-1981) at Royal Dutch Shell. The earliest definitions of SP and scenarios is a source of small dispute. Chermack & Lynham (2002) state the first available definition for SP appears in 1985, possibly simultaneously by Porter (1985) and Wack (1985a). However, earlier publications have been found that also offer their own definitions of SP. In their review of various scenario generation procedures, Mitchell, et al. (1979, pp. 409-410) describe scenario construction (or generation) as "an explicit framework for investigating possible futures" that provides "the decision maker with coherent and definite pictures of likely or possible developments". Vanston, et al. (1977, p. 159) describe SP as an uncertain art, where scenario development follows a multi-step technique that leads to multiple alternative scenarios that are "relevant, reasonable, and logically interrelated". Even earlier, Vanston, et al. (1975) suggest that even though there are any number of possible methods for generating scenarios (depending on the their intended use, the organization, and the group), all scenarios should have at least four basic features (p. 5):

(1) Be plausible
(2) Be self-consistent
(3) Include all critical, relevant factors
(4) Roughly parallel other scenarios in form and scope

More than 40 years later, Spaniol & Rowland (2018) add to this list with the following necessary features:

(5) Future oriented
(6) About external context
(7) Narrative description
(8) All scenarios are comparatively different from one another

The earliest published definition, however, appears to come from Kahn & Weiner's book *The Year 2000: A Framework for Speculation on the Next Thirty-Three Years* (1967, p. 6).

> Scenarios are hypothetical sequences of events constructed for the purpose of focusing attention on causal processes and decision-points. They answer two kinds of questions: 1) Precisely how might some hypothetical situation come about, step by step? and 2) What alternatives exist, for each actor, at each step, for preventing, diverting, or facilitating the process?

As time progresses, scenario methods continue to develop along pragmatic lines, with an almost 'trial by fire' effort. With SP appreciating the better part of a century in use, we should take this time to assess what has emerged – from both academic inquiry and field applications – and determine where we can go. The field is rich with typologies, models, theoretical development, and an almost anatomical map of scenario characteristics. However, as true today as in decades past, are the numerous and ever-emerging methods and models of SP. These differences are as responsible for expanding the field of SP as they are for introducing obstacles to that same field. Some feel that SP is forever consigned to the realm of chaos, incapable of being clearly defined in any manner. However, just as likely is the possibility that the field of SP is a collective of experiences and knowledge that play upon a theme, where realities have been revealed in absence of pre-defined boundaries.

2.1.1 *Frequency & Topic of Publications*

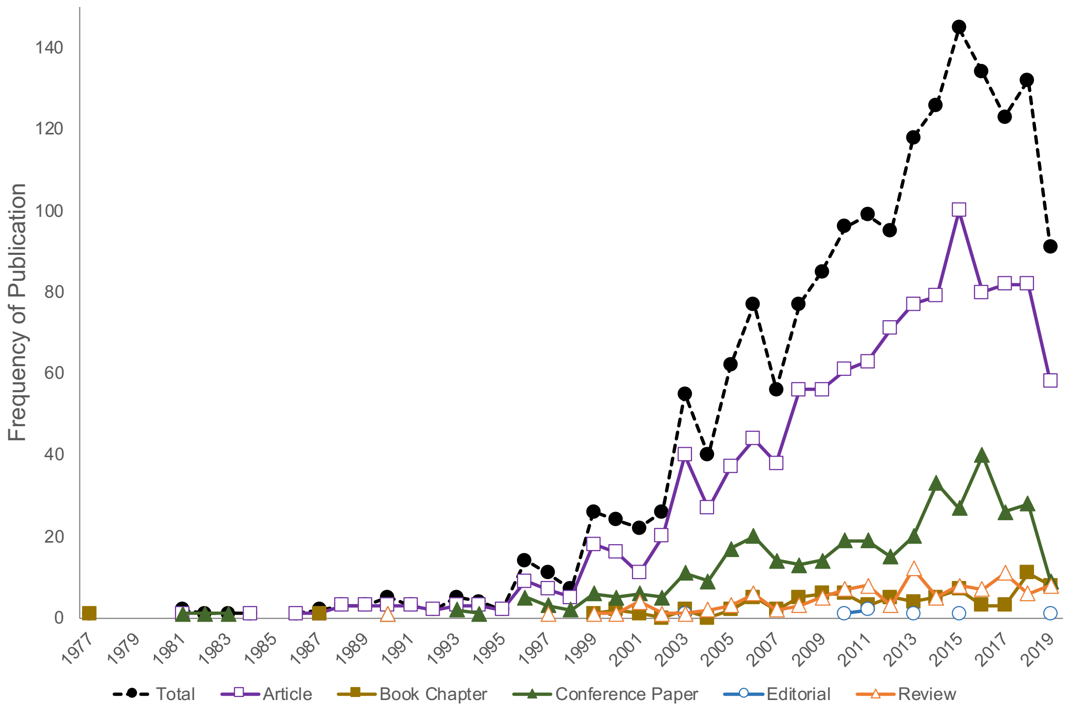
To better understand SP's historical popularity, a full search of publications is conducted.[8] An historical search will bring context to the work conducted in this thesis. By gaining a general understanding of publication behaviours and topics, gaps and focal points of the field may become easier to identify.

A search of the phrases "scenario planning", "scenario thinking", and "scenario building" in all areas of a publications (e.g. abstract, title, key words) was conducted in Scopus , Web of Science (WOS), and Strathclyde University's Andersonian Library. The search criteria included an open date range to 6 September 2019, all document types, open access and traditional, all subjects, and all languages. Scopus returned 2,477 results, WOS returned 1,693 results, and the University of Strathclyde Library returned 42 results. Using R software, the documents were compared by "Title", "Year", and "Document Type", to locate duplicates. After several iterations of different comparison protocols (e.g. "Title" only, "Affiliation publication" only, "Title" and "Year" together), the grouped selection criteria of "Title", "Year", and "Document Type" proved to be the best protocol. No duplicates were left in the master list and no documents were accidentally removed due to similar selection criteria. Duplicates were removed, which resulted in 2,407 unique documents. The different document types that remain are article, review, conference paper & review, note, book & chapter, letter, article in press, editorial, erratum, and short survey. Publication dates range 1967 to 2019. Figure 2.1 shows the frequency of publications (y-axis) per year (x-axis). Between the years 1967 and 1986 there was an average of less than one publication a year. As time progressed, overall activity sharply increases in popularity, then a surprisingly steady drop over the past three years commences.

---

[8] Kobes & Loy (2020) pre-published their review one month before handing in my thesis, that explores the same open-source database as I, to report on nearly the same question "What is the publication history of scenario planning?" Many of their analyses are near identical to my own prior discoveries. To honour the original intent and unique work I conducted prior to the authors' paper, I have not integrated their work into this thesis.

Figure 2.1. All scenario planning publications from 1967 To 2019

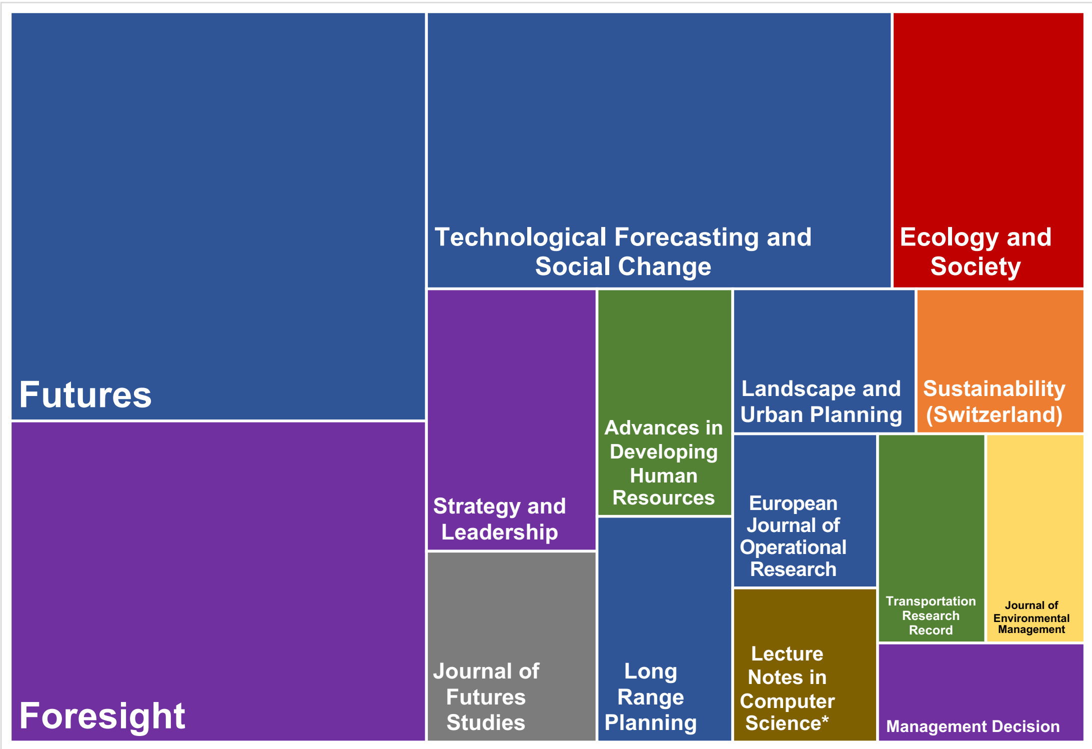


There are 987 different publications that include SP documents. The division of popularity across publications is not uniform. The top 1.5% of journals (≥10 articles) are illustrated in a tree map in Figure 2.2.[9] *Futures* is the most prolific, followed by *Foresight* and *Technological Forecasting and Social Change*. All three are peer-reviewed journals. The popularity of these journals is expected. All three specialise in promoting and disseminating information about the future. The full range of specializations across the journals includes future studies as they pertain to cultures, management theory and practice, societies, science, technology, economics, politics,

[9] It is important to note that even though the bulk of the analyses in this chapter focus on journal articles, published books on the topic of scenario planning have also been strongly influential in disseminating the scholarship. The most notable books on the discipline span 52 years and are referenced throughout this thesis (e.g. Cairns & Wright, 2018; Chermack, 2003, 2011; De Geus, 1997; Kahn & Weiner, 1967; Lindgren & Bandhold, 2003, 2009; Ringland, 1998; Van der Heijden, 1996; Van der Heijden, et al., 2002).

environment and the planet, individuals and humanity, planning tools, life-supporting ecosystems on which human wellbeing depends, computer science, transportation industries and policy, and environmental systems and quality.

Figure 2.2. Popularity of journals



Note: Journals are ordered by frequency of published SP documents and colour coded by publisher. In order of prolificness: Elsevier (blue), Emerald (purple), SAGE (green), Resilience Alliance (red), Tamkang University (grey), Multidisciplinary Digital Publishing Institute (orange), Springer (brown), and Academic Press (yellow).

Using R software (R Core Team, 2019), authors' keywords were separated and ranked from most to least popular, by journal.[10] Phrases were preserved where entered as keywords (e.g. "scenario planning", "design thinking"). Keywords can

[10] *Long Range Planning* did not offer authors keywords in the databases. Each article's index keywords were analysed in place of author keywords.

serve as a proxy for the focus of a journal. Reporting the popularity of keywords gives the reader a glimpse into the interests of both the journals and the dialogue around SP. The method gives us a quick view into the applications of SP and helps determine the breadth of popularity.

Data were cleaned to improve accuracy of analysis. “NA” is a placeholder R software substitutes for blank spaces. All “NA” entries were removed. Typos were corrected (e.g. “senario” changed to “scenario”), followed by general spelling and grammatical corrections (e.g. “scenarios planning” changed to “scenario planning”). Articles and conjunctions (e.g. a, the, and) were removed. The default setting to the R software, tidyverse, converts all words to lowercase.

To calculate the popularity (Pop) of keywords by journal, a simple weighting formula was used. The number of terms that comprise a single keyword is represented as (a). The frequency of a keyword (k) was summed $(\sum k)$. The proportion of the use of k was then calculated from the total of all keywords $P_1 = (\Sigma k_a / \Sigma k_n)$. Next, the frequency of each term $(t_i = t_1 + t_2 + t_3 \ldots)$ across all keywords was summed (Σt_i). The proportion of t_i amongst all terms was calculated $P_2 = (\Sigma t_i / \Sigma t_n)$. For keywords that were single words, $k_{i1} = t_i$. Each keyword proportion was then weighted by their individual term proportions to give a more granular and accurate representation of keyword popularity by journal. The formula gives a clearer report of keyword popularity than standard, simplified frequency counts of individual terms. It controls for bias in weighting due to single- vs multiple-term frequencies. As well, the formula preserves contextual meaning by accounting for phrases, instead of context-free single term frequency. The formula is presented in Equation 2.1 with a following example.

Equation 2.1. Pop score

$Pop(k_i) = (\Sigma k_a / \Sigma k_n) * (\Sigma t_i / \Sigma t_n)$

or

$Pop(k_i) = P_2(P_1)$

To illustrate the process, the keyword "futures studies" will be analysed from the journal *Futures*. The keyword "futures studies" is a single keyword comprised of two individual terms, "futures" and "studies" ($a = 2$).

There are no variations of "futures studies" in the keyword list (e.g. "future studies" or "futures study"). The frequency for this keyword is ($k_1 = 2$). There are 150 separate keywords mentioned 204 times across the keyword lists for 85 articles. The proportion of mentions for "futures studies" is calculated ($P_1 = 0.0098$). Next, the proportion of each term is calculated. The terms "futures" and "future" both appear in the term list, therefore the frequency of both terms is combined to represent the use of the concept *futures*, ($t_1 = 11$). There is only one version of the term "studies", therefore only the occurrence of this spelling is calculated ($t_2 = 2$). The sum of individual terms is then calculated ($t_a = 13$). After data cleaning, there are 199 unique terms used 264 times. The proportion of mentions for "future(s)" and "studies" is calculated ($P_2 = 0.0492$). Finally the keyword proportion is weighted by the individual terms' proportion to determine the keywords popularity within *Futures* ($Pop(k_i) = 0.0005$). The range of Pop scores is ($1.857E$-5 – 0.006). When ranked by Pop score, "futures studies" becomes the 4th most popular keyword for all SP articles within *Futures*.

Table 2.1. presents the 20 most popular keywords associated with the 15 journals. Journals are ordered left to right by popularity. Keywords are ordered top to bottom by Pop scores. The shading represents the most commonly shared keywords, weighted by term popularity, across journals. The darker the shading, the higher the proportional sharing of the keywords and terms across journals. The keywords in white are those almost fully or fully unique to the journal.

For example, the keyword "futures studies" is only popular in one journal, *Futures*. However, the terms "futures" and "studies", individually, are featured several times within the full repository of keywords, and are highly popular across three journals. Therefore the keyword "future studies" is represented with a mid-range shading.

Table 2.1. Most popular keywords across journals

Futures	Foresight	Technological Forecasting and Social Change	Ecology and Society	Strategy and Leadership
scenario planning	scenario planning	scenario planning	scenario planning	scenario planning
scenario	strategic planning	scenario	ecosystem services	strategic planning
future	sustainable development	scenario development	participatory scenario planning	business planning
futures studies	research methods	scenario method	social-ecological system (ses)	strategic management
future scenarios	delphi method	decision making	scenario	competitive strategy
world futures studies federation	forecasting	scenario discovery	doxana social-ecological system	corporate strategy
scenario analysis	government policy	scenario use	marine social-ecological systems	strategy
future uncertainty	methodology	strategic foresight	social-ecological model	scenario development team
future workshop	strategic management	robust decision making	ecosystem-based management	management strategy
futures education	strategy	scenario-based decision making	multiscale scenario	economic scenarios
oxford futures forum	foresight	strategic decision making	ecosystem service bundles	scenario development
design thinking	environmental management	participatory scenario planning	landscape ecology	brazil scenarios
australia's future	uncertainty management	strategic planning and forecasting	programme on ecosystem change & society	scenario learning
future of work	schools of scenario planning	foresight	ecosystem service	electricity transmission scenarios
futures literacy	futures method	deductive scenario method	ecosystem services:	political scenarios
scenarios and strategy	generation and dissemination of information	scenario methods	millennium ecosystem assessment	potential strategic responses
scenario thinking	public health policy	scenario-based roadmapping	value-based scenarios	risk management
scenario workshop	science and technology policy	2x2 scenario matrix	ecological monitoring	change management
scenario planning and training	complexity theory	strategic planning	multi-scale scenarios	uncertainty management
science fiction prototypes	technology	decision support	system dynamics model	knowledge management

Table 2.1. *(continued)*

Journal of Futures Studies	Advances in Developing Human Resources	Long Range Planning	Landscape and Urban Planning	Sustainability (Switzerland)
scenario planning	scenario planning	scenario planning	scenario planning	scenario planning
scenarios	integrated scenario planning model	strategic planning	land use planning	adaptive planning
scenario planning outcomes	scenario planning models	planning method	participatory scenario planning	energy planning
scenario planning research	challenges in scenario planning	strategic approach	land use change	environmental planning
causal layered analysis	scenario planning history	environmental planning	climate change scenarios	planning support systems
scenario network mapping	leadership development	risk management	land use modelling	policy portfolio planning
alternative futures	scenario-based training	enterprise resource planning	land-use mapping	r&d planning
futures education	strategic planning	real option planning	scenarios	plan making
futures methods	learning and change strategies	boundary conditions	planning support system	environmental sustainability policy
preferred future	scenarios	planning implication	regional planning	planning
strategic planning	best practice	planning theory	climate change adaptation	green supply chain management
the futures triangle	strategic hrd	planning process	climate change-induced flooding	supply chain management
urban futures	learning in multiple levels	management practice	forest planning	energy scenarios
futures	organization development	planning	transportation planning	hybrid renewable energies
long-range planning	systematic training design	managers	watershed planning	sustainable smart grid technology
scenario building	technology development	scenario analysis	private land conservation	sustainability
strategic actions	change curve	economic conditions	land cover modelling	electric vehicle charging technology
anticipatory action learning	community of practice	organizational performance	scenario downscaling	environmental market failure
strategic thinking	environmental outreach	trend-impact analysis	adaptive management	renewable energies
systemic risk analysis	group potency	quantitative analysis	analytic network process	adaptation tracking

Table 2.1. *(continued)*

European Journal of Operational Research	Lecture Notes in Computer Science*	Transportation Research Record	Journal of Environmental Management	Management Decision
scenario planning	urban regeneration	transportation	scenario planning	strategic planning
strategic planning	scenario optimization	scenario planning	multi-criteria decision analysis	strategy
multiple criteria decision analysis	policy analysis	transportation planning	uncertainty	strategic management
multi-stakeholder decision-making	web 2.0	motor transportation	urban water management	scenario planning
supply chain network design	data-driven design	transportation system	planning support system	management strategy
strategic planning	marketing scenario	strategic planning	environmental management	corporate strategy
supply chain network	dynamic simulation	regional strategic transportation planning	participatory scenario-planning	decision making
decision analysis	knowledge processing	transportation planning boards	adaptive management	corporate planning
decision support systems	finance	transport model	environmental planning	process planning
decision making.	scenario planning and training	integrated transportation systems	regional environmental planning	uncertainty management
modelling systems and languages	risk assessment	intelligent transportation systems	geographic information systems (gis)	economic change
supply chain disruptions	virtual reality	transportation scenario	nutrient management	organizational change
decision processes	big data	sustainable transportation	river management	experiential learning
location models	causal mapping	transportation sustainability	stormwater management	organizational development
soft or methods	scenario analysis	travel demand models	material flow analysis	performance measurement (quality)
uncertainty modeling	security	regional transportation	planning	world economy
water infrastructure planning	decision making-process	scenario planning methodology	societal decision support	balanced scorecard
scenarios	e/mlearning	michigan department of transportations	stakeholder analysis	business failure
capacity models	dynamic behavior	federal transportation	water framework directive	financial markets
multi-attribute utility theory	exposure reduction	transportation infrastructures	multi-attribute utility theory	footwear industry

The shading of Table 2.1 reveals an initial pattern in keyword popularity. Unsurprisingly, “scenario planning” is the most popular across the majority of journals, but becomes less popular in the more niche publications, *Transportation Research Record* and the *Lecture Notes* collection. “Scenario planning” is closely followed by management, analysis, model, and strategy related keywords. The journals *Technological Forecasting and Social Change* and *Strategy and Leadership* have the highest crossover of the most popular keywords and terms of all the journals. Neither journal is the most prolific in SP publications. Together they published just over 50% of the total articles published by the top two journals. This may indicate that these journals publish more position pieces on SP than the other journals. Theory and position papers are less likely to focus on a specific topic than empirical and case studies. As well, neither of these journals are dedicated to a niche topic in industry, science, or business.

Surprisingly, the terms “risk” and “uncertainty” are not heavily featured in the keywords. If keywords serve as a heuristic to understanding the main focus of an article, it is surprising to see very little focus on either of these topics, given that one of the main purposes to SP is to learn how to navigate uncertain futures and manage risks (Cairns & Wright, 2018; Chermack, 2003; Van der Heijden, *et al*., 2002).

It is equally interesting to note that only one journal has popularised the keyword “scenario thinking”. The term appears for the first time in Sandiford, et al., 1991 uncited article “Why Nicaraguan children survive: Moving beyond scenario thinking.” The article, however, does not discuss SP. The concept of ST gains popularity just after 2010, but still largely remains on the fringe of SP literature. ST is discussed in more detail in Chapter 3.

Table 2.1 also helps reveal the variety of techniques associated with SP publications. Balarezo & Nielsen’s (2017, p. 14) recent review of the literature reports a growing trend of “combining SP with more structured quantitative tools”. The text mining analysis reveals that Delphi, causal layered analysis, multiple criteria decision analysis, and various modelling techniques are used in conjunction with SP. This helps support the dialogue that SP is a method that works in conjunction with others to support strategy and planning, as opposed to a method that works in isolation.

2.2 Scenario Planning Literature Review

The historical review illustrates a growing popularity in SP over the better part of a century, concentrated largely in the last 25 years. This popularity is bolstered by the expansion of SP into a greater variety of fields and disciplines – from military, to oil and gas, climate science, urban design, and more. But how does the literature discuss, or frame, SP? With wide applicability, it is important to determine whether SP is truly an umbrella term for disparate practices and techniques or whether a thread of unity weaves throughout. One way to begin understanding the quagmire that is "Scenario Planning", as a field and as a discipline, is to understand how the literature discusses the function of SP. This section's literature review is not exhaustive, but presents the two overarching thematic functions that dominate the scholarship. Authors overwhelmingly speak of SP as either a 'tool' or a 'process'. This linguistic difference is used to further develop the thesis research position in the proceeding chapters.

2.2.1 *As a Tool*

Schoemaker (1995) discusses SP as a tool that formalises the relationship between elements, so further quantitative models can be developed. SP is seen as a type of corporate map, where the "avalanche" of information is distilled into stories of interacting elements (Schoemaker & Van der Heijden, 1992, p. 815). From these elements, quantitative models can be developed. Van der Heijden (1994) and Schwartz (1991) both describe SP as a tool for ordering perceptions about plausible futures. Schwartz emphasises that the tool, however, is only as important as the people who use it. Kahn & Weiner (1967) emphasise in a number passages that SP is a pedagogical tool for support. Heugens & Van Oosterhout (2001, p. 864) see SP as an effective tool for evaluating the robustness of an organisation's strategies, "making it more suitable for stimulating organizational interpretation and sense making processes," by comparing them against several plausible futures. Schnaars (1987) discusses SP exclusively as a forecasting tool, where Ducot & Lubben (1980,

p. 54) echo this sentiment that certain methods will make SP a “credible tool in the art of forecasting”.

Arguably, the bulk of the literature that discuss SP as a tool, are actually focusing on the scenario as an artefact of the process. McKiernan (2017) describes scenarios as “filters” in the process of sense making, giving and receiving. Cornelius, *et al*. (2005) explore how SP provides a “useful tool”, by way of the individual scenarios, for understanding uncertainties at Royal Dutch/Shell Corp. Gordon (1994a, 1994b) and Pagani (2009) both discuss scenarios as tools for developing strategies by using more quantitative methods. Becker (1983, p. 95) specifically emphasises that, as a tool for organisations, “scenarios should not be viewed as forecast[s] of what will be,” but rather used to minimize risk. Durance and Godet (2010, p. 1489) attempt a strong distinction between scenarios and foresight, stating that, “a scenario is not an end in itself.”

2.2.2 ***As a Process***

A number of typologies reflect on the process of SP. The Comprehensive Scenario Intervention (CSI) typology presents four overarching themes to any given SP intervention (Crawford, 2019). The second theme, *Process Design*, makes it clear that SP is seen as an ongoing development of data generation, collection, and information sharing. The CSI typology frames the efficacy of SP through a process that challenges practitioners’ perspectives, by increasing flexibility in thinking, to ultimately ensure the survival and success of the organisation (Schoemaker P. J., 1995; Van der Heijden, *et al.* 2002). The second theme originates from the Van Notten, et al. (2003) typology, which presents the efforts of a SP intervention on a spectrum from intuitive to formal processes. The intuitive process takes a flexible approach to scenario development. As an intervention moves towards a more formal process, it exchanges qualitative efforts for more quantitative measures. The process resembles a rational and analytical exercise, where simulations and dynamic modelling inform scenario development. The Börjeson, et al. (2006) typology considers the facilitation of finding common understandings between heterogeneous groups of practitioners an inherent strength in the process, much in the same manner as the Delphi process finds a common ground across participants. The Heugens & Van Oosterhout (2001) typology speaks of process as components within SP and ST: construction, sense making, and creativity. Wack (1985b, p. 149) takes this a step farther, equating the energy spent throughout the process as multiplicative, “The scenario process of converting information into fresh perceptions has something of a “breeder effect”: it generates energy, much more energy than has been consumed in time and effort during the process.”

The literature search shows that perspectives of SP as either a tool or a process are not temporally bound, either. Unlike other sciences and more established disciplines, the perspective that SP is primarily one type or another, spans both early and contemporary publications.

2.2.3 *As Either or Both*

Many authors use the two concepts interchangeably. Some are clear when they switch between speaking from a content/tool position to a holistic action/process position. Pulver & Van Deveer (2007, pp. 1, 3, 5, emphasis added), for example, speak of scenarios (the artefacts) in terms of tool-based language, and the experience of SP as a process.

> First, we examine scenarios as decision-support *tools* and call into question the automatic presumption that scenarios support or influence policy-making… One common understanding of scenarios is as *tools* to support decision-making… Second, we step back and focus on the *process* aspects of scenarios, investigating them as social objects that link social worlds and influence political and social spheres… Scenario exercises are social processes.

Others are not so clear. Franco, et al. (2013) interchange the terms when describing SP. In the first two sentences alone, they blend the concepts (p. 723, emphasis added).

> …scenario planning has enjoyed increasing acceptance among practitioners and academics as a *tool* for supporting strategy formulation in organisations. In simple terms, a scenario-driven strategy *process* involves building a set of challenging but plausible futures that are used as 'wind tunnels' to test whether the organisation's strategies can withstand the turbulence of an uncertain environment

Chermack (2003, 2005, 2011, 2018; Chermack & Lynham, 2002) freely moves between the two concepts for both the action and the artefact. For example, a selection of passages from his book *A Theory of Scenario Planning* (2003, pp. 2, 3, 4, 27, 93, emphases added) reveals the kind of interchangeability employed by the author,

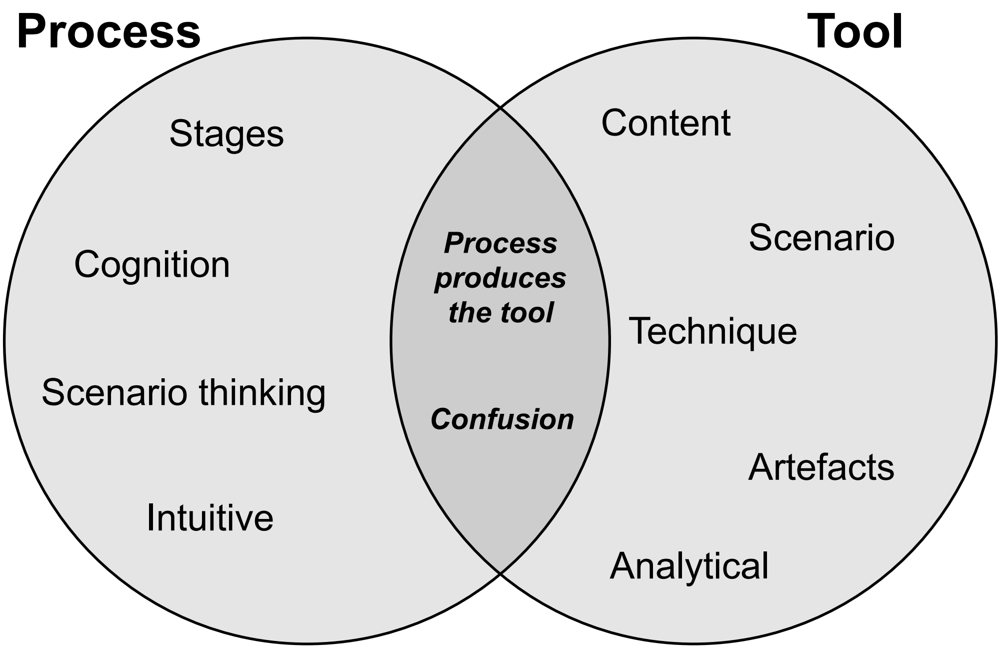
> The popular application of scenarios has resulted in a variety of approaches and methods for conducting the scenario building *process*… scholars and practitioners will realize the value of scenario planning as a *tool*… Scenario planning has emerged as a *tool* for considering uncertainty in the planning *process*… HRD professionals ought to be very much in favor of a *process* such as scenario planning… Although it was originally developed as a *tool* for strategic decision-making, scenario planning is increasingly noted as an important *tool* for learning… We can therefore randomly assign subjects to a treatment condition in which the subjects will receive extensive exposure to scenario planning as a specific planning *process*, or to a control condition in which the subjects receive no planning *process*".

What is most interesting about the last passage is that Chermack is discussing his agenda for testing the theory of SP. However, in his later quasi-experimental papers, he primarily discusses using SP as a tool for affecting cognitions (Bodin, Chermack, & Coons, 2016; Chermack & Nimon, 2008; Chermack, Van der Merwe, & Lynham, 2006; Glick, *et al.,* 2012; Haeffner, *et al.,* 2012; Hawkins & Chermack, 2014).

Bishop, et al. (2007) recognise the confusion in terminology across the SP literature and attempt to create ground rules, though, it is unclear how many have taken their suggestions. Balarezo & Nielson (2017) present a methodological outline for exploring the SP intervention framework by dividing the intervention across antecedents, processes, outcomes, and variables. This sets the perspective of SP as an intervention process apart from SP *techniques*, which are often interpreted as tools. Van de Riet, et al. (2008) also offer a helpful clarification between the two perspectives. Their typology suggests when the purpose is for development or evaluation, these are *content*-focused scenarios, and are seen as instrumental artefacts of the intervention (i.e. tools). This appears to be clearly reflected in the literature that discusses the tool SP in reference to forecasting. When the purposes are to learn, train, achieve consensus, or change mental models in some way, then these are *process*-focused scenarios. The contents are less important than the journey. In the same manner as the 'SP tool' literature, we see Van de Riet, et al.'s division supported by the 'SP process' literature, which primarily discuss cognitive and conversational effects. Figure 2.3 shows a Venn diagram of the main language and foci the scholarship takes when discussing SP as a tool, process, or both.

The body of research conducted in this thesis takes the view that SP is a *process*, defined by a number of necessary and sufficient elements, where value lies primarily in the change that manifests in the perspectives, cognition, and willingness of practitioners as they both envision and strategise for the future. To consider SP a tool is to limit the scope of abilities the method brings to the table. Tools have well defined functions. They are designed for a specific task, to solve a specific problem, in an optimal way. However, SP is a method of pragmatism, where flexibility in application, process, and components reveals a broad spectrum of expression.

Figure 2.3. Division of perspectives in the literature



2.2.4 ***Scenario Planning Models***

Framing SP from a process perspective focuses the investigation on the models of practice that guide the active process (as opposed to content-focused scenario research). Huss & Honton (1987) identify three major models of SP (also referred to as "schools"):

- IL (Ogilvy & Mandel, 1984)
- Cross-impact analysis (CIA; Gordon, 1994a)
- Trend-Impact Analysis (TIA; Gordon, 1994b)

Sometimes the IL approach is referred to as the 'Shell approach' due to its early popularity with Pierre Wack when employed at Royal Dutch Shell (Wack, 1985a, 1985b) or the 'SRI school' due to the pioneering days at the Stanford Research Institute (International, 2019).

Added to the list the French Model of *La Prospective*, developed by Gaston Berger, and expanded by Godet (1987). Bradfield, et al., later combine CIA and TIA under the single umbrella of PMT in their 2005 paper. TIA and CIA were spearheaded by Theodore J. Gordon and John Stover in the 1970s at The Futures Group in

Connecticut. Bradfield, et al., along with Amer, et al. (2013) offer comprehensive reviews of the SP models and describe the three main scenario approaches as plausibility based (IL), probability based (PMT), and preference based (*La Prospective*). Chermack (2011) identifies 10 models of SP, though several are modifications of the main three.

2.2.4.1 Intuitive Logics

The IL model is defined by eight stages in the "standard" or "SRI" approach (Bradfield, Derbyshire, & Wright, 2016; Bradfield, *et al.*, 2005; Cairns & Wright, 2018; Duncan & Wack, 1994; Huss & Honton, 1987; Schoemaker, 1991, 1995; Van der Heijden, 1997; Van der Heijden, *et al.,* 2002; Wright, Bradfield, & Cairns, 2013; Schwartz, 1991; Chermack, Van der Merwe, & Lynham, 2006).

Stage 1 is used to set the agenda for the SP intervention, by the practitioners and facilitators. Interviews and research (e.g. historical, market, internal, and external) are conducted to help define the scope of the intervention. Huss & Honton (1987) suggest the narrower the scope of the strategy, the easier the SP intervention will be, to carry out. Duncan & Wack (1994, p. 20), however, warn against this technique because it increases the practitioners' chances of "missing key determinants of future conditions or events." Goals are made explicit and agreed upon (e.g. "We want to increase our product prices"). Timelines are determined (e.g. "What will the future look like in 2, 5, and 10 years?"). SP practitioners are identified and invited to participate in the workshops. Length of the session(s), locations, and other logistics are determined. The product, of which, is a rich library of information regarding past behaviours, present issues, and future desires/focus.

After familiarizing with the organisation's information and agenda, practitioners are tasked, in Stage 2, with identifying as many external driving forces (DF) as possible that could have an impact on the organisation and their ability to reach their goals within the designated timeline. "Driving forces are those underlying and impacting factors that set the pattern of events and determine outcomes in the business environment and timescale being considered – the forces that make things happen"

(Van der Heijden, *et al.,* 2002, p. 202). All authors recommend that scenario teams consider the widest range of DF that could impact the organisation and their path towards achieving their goals. The information searched for and discovered (two different features altogether) in this stage form the "foundation stone" for all subsequent stages (Bradfield, 2008, p. 207). Different techniques are used to aid the scoping process, most notably the PESTEL and STIRDEEPER structured frameworks.[11] DF are divided into two overarching categories: critical uncertainties and predetermined. Critical uncertainties are those DF that exist in the present and potential future organisation's environment, but do not have a predictable path of development (Kahn & Weiner, 1967; Van der Heijden, *et al.,* 2002). Predetermined variables are those DF and trends that will have an impact, but will eventually resolve themselves in time. These are "already in the pipeline" and offer a certain level of predictability (Van der Heijden, *et al.,* 2002, p. 269).

Moving from Stage 2 to 3 is an issue of some confusion within the literature. Whether a pre-determined stopping point or threshold (e.g. no new DF are generated) serves as the cue, practitioners move from identifying external factors to thinking about the relationships between DF that now exist for the SP group. DF are grouped together by causal relations through a task identified as clustering. Clustering is an effort in sense-making by reducing a large number of individual factors – which can be in the hundreds – into a "small and manageable number of higher-level concepts," though clusters will usually vary in size and complexity (Van der Heijden, *et al.,* 2002, p. 205). Sometimes it's discovered in this stage that two DF have a plausible, but weak causal relationship. In these cases, practitioners are encouraged to identify more DF to enrichen and strengthen the relationship between the two, more distant DF. The IL process regularly encourages practitioners to revisit the efforts of previous stages, through an iterative process, whenever further information is needed

---

[11] PESTEL: Political, Economic, Social, Technological, Ecological, Legal
STIRDEEPER: Society, Technology, Industry, Resources, Demographics, Economics, Environment, Politics, Energy, Religion

to improve the plausibility and believability of the scenario content. This stage is completed when all DF are causally linked together in their own internally consistent, yet independent clusters, that each summarize a theme of uncertainty.

Once DF are clustered around associated themes of causality, each cluster is assessed for plausible outcomes in Stage 4. To help maximise resources (e.g. time, temperament, cognitive load), two extreme, yet highly plausible outcomes are defined, which span the scenario timeline (Cairns & Wright, 2018). Outcomes should be different enough from the other to develop new storylines of final causality. Outcomes do not, however, necessarily need to give a "good vs bad" storyline. Cluster outcomes may be complex and as such, develop along unstable paths. This allows for any number of outcome dyads (e.g. collapse A vs collapse B, progressive vs moderate, or disciplined vs transformational). In many popular IL techniques, this task is actually integrated into Stage 6. Two main purposes are served by diverting this task to a later stage. First, it can streamline the process by requiring the scenario team to focus on only two clusters, instead of all the clusters. This is a beneficial technique when working with time-poor professionals. Second, it can help avoid biasing Stage 5 by keeping clusters 'neutral' (i.e. neither positive nor negative in outcomes), therefore allowing practitioners to judge each cluster on its content, rather than expected/preferred outcomes.

Stage 5 requires practitioners to determine how impactful and predictable each cluster is projected to be. To do this, they plot their clusters on an Impact/Predictability (I/P) Matrix. The relative degree of impact each cluster is projected to have on the goals or focal issues, given the timeline, is determined on a scale from "low" to "high" impact. Predictability – or uncertainty, depending on the preferred method – is also determined on a relative scale from "low" to "high". Specifically, predictability is focused on what the impact will be, not whether there will be an impact at all (Cairns & Wright, 2018).

Stage 6 tasks the scenario team with focusing their perspectives on a subset of information. Practitioners choose the two most highly impactful and least predictable (i.e. most uncertain) clusters from their I/P Matrix. The motivation behind choosing only two clusters is to help eliminate the excessive work that comes from scenario development of every possible cluster combination. By focusing on those events which are projected to have the highest impact on the organisation's ability to reach their goals, while remaining the least predictable on that path helps focus, practitioners take a risk-averse approach to SP. If Stage 4 was developed, in full, then the corresponding cluster outcomes are combined in a 2x2 method, to result in a skeletal framework of four different scenarios. If Stage 4 was skipped, then two extreme, yet plausible outcomes are defined, which span the scenario timeline, and combined to create four different scenarios.

A broad range of descriptors are developed in Stage 7, to build a logical structure towards the four plausible futures. To develop descriptors, practitioners dive into the details of the future society by questioning the states of various frameworks and their implications (Cairns & Wright, 2018; Chermack, 2011; Huss & Honton, 1987). For example, the PESTLE technique can be used to ask such questions as "what would be the state national politics/society/the legal system/economy at time X?" which helps them generate implications for the outcomes, then end states for each scenario.

Stage 8 is the final effort in the process, where chronological storylines are developed. Details are filled in to create a logical and coherent flow of information, including identification of signposts and leading indicators. The efforts ultimately result in a picture of a plausible future for each scenario. Practitioners are then tasked with reviewing the scenario implications for the organisation, "what does this scenario mean for the client in terms of its future strategy?" (Wright, 2014, p. 34). Schwartz (1991, pp. 247, original emphasis) believes a full scenario process that follows these basic eight stages will result in a successful intervention. Table 2.3 summarises the IL process.

> If the scenarios have been built according to the previous steps, then the scenarios will be able to translate movements of a few key indicators into an orderly set of industry-specific implications. The logical

coherence that was *built into* the scenarios will allow logical implications of leading indicators to be *drawn out* of the scenarios.

Table 2.2. Intuitive Logics model

Stage	Task
1	Setting the agenda – background research, define goals and timeline
2	Determining the DF – critical uncertainties and predetermines
3	Clustering the DF – causal associations between DF
4*	Defining the cluster outcomes – extreme plausible outcomes of clusters
5	Impact/Predictability matrix – determine most impactful and uncertain clusters
6*	Framing scenarios – combine outcomes of four scenarios
7	Scoping scenarios – broad descriptors and implications
8	Developing full scenarios – write narratives and test for plausibility

*Note: Stage 4 is optionally integrated at the end of Stage 6 in some popular practices.

2.2.4.2 Probabilistic Modified Trends

PMT encompasses both trend-impact and cross-impact analysis. Both methods are probability based. They attempt to expand beyond existing forecasting methodologies by including model-based modifications that account for practitioners' expectations about the future (Gordon, 1994a). The PMT model is, therefore, largely normative in development.

Though Gordon's 1994 paper does not make explicit reference to discrete process stages, Huss & Honton (1987) break down the TIA method into eight stages.

The first stage is the same across all scenario models: define the topic. Two to three key drivers, relevant to the topic, are identified. Each driver is divided into different states and set into a matrix to provide the initial boundaries of the probabilities and alternative scenarios. "Surprise-free extrapolation is the first step" (Gordon, 1994b, p 3). To generate a surprise-free forecast, different boundaries are fit together that reflect the historical data.

A subset of the alternative scenarios is selected in Stage 2. The quantity and type of alternatives are often chosen based group consensus and goals of the project. The alternative scenarios should be plausible, potentially powerful in impact, and verifiable in retrospect. From these efforts, a scenario space is created.

Judgment and imagination are considered crucial in Stage 3 (Gordon, 1994b). The purpose is to create departures from the surprise-free forecast. Insights are gathered through a number of sources: literature search, Delphi, and face-to-face workshops.

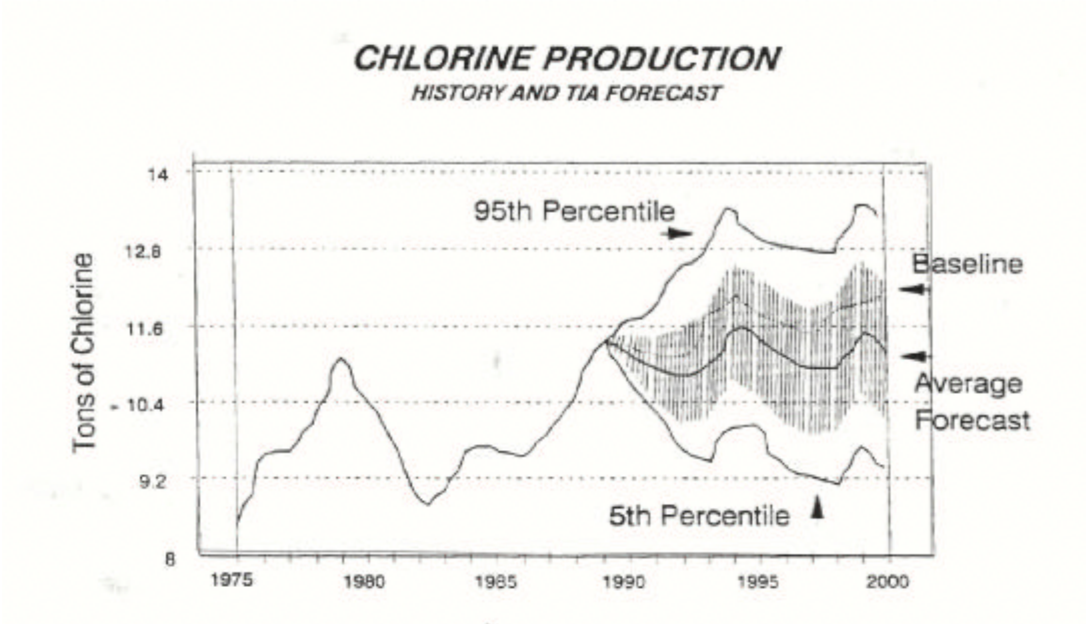
Practitioners take the expanded list of plausible events and extrapolate intitial trends using standard time series tecniques in Stage 4.

A list of impacting events is developed at the start of Stage 5. Similar techniques, such as Delphi, are used to generate this information.

Estimates are made of the probability of occurrence, for each event, as a function of time. Stages 5 and 6 are similar to the I/P matrix exercise in the IL model.

In Stage 7, all quantitative data (impact and probability) are combined and analysed using TIA-modelled software. The result is an adjusted extrapolation of the impacting trends, bracketed by upper and lower quartile limits of probabilies. The probability limits can also reflect other levels, as decided by the scenario team. The simplest approach assumes independence of events, and therefore the magnitude and probability of impact as the future progresses are isolated. Figure 2.4 is an example of TIA analysis regarding U.S. chlorine demand, as illustrated in Gordon's 1994 TIA paper (p 5). The years of production are given across the x-axis and the production of chlorine, in tons, is given on the y-axis. The single line from 1975-1990 represents the historical data collected during the research phase of the process. The projected data is given from 1990 to 2000. The upper solid line represents the 95th percentile and the lower solid line represents the 5th percentile; the space between these lines represents 90% of the forecasted chloroine production at every point in the future. The solid middle line is generated from the surprise free forecast in Stage 1 and an averaging of the alternative scenarios, weighted by their probabilities and impact factors.

Figure 2.4. Trend-Impact Analysis example: US chlorine forecasts



Note: Reproduced from Gordon, 1994a, p 5, Figure 3.

Stage 8 closes off the process by tasking the scenario team to develop narratives for each scenario, based on the analyses of Stage 7. Table 2.4 summarises the process TIA process.

Table 2.3. Trend-Impact Analysis model

Stage	Task
1	Select topic and identify key scenario drivers.
2	Create scenario space
3	Identify important trends and collect time series data
4	Prepare a naïve extrapolation
5	Establish a list of impacting events
6	Establish probabilities over time
7	Modify extrapolations
8	Write narratives

CIA is a computer-assisted interactive approach to communication and decision-making. In contrast to the independence of the TIA method, CIA assumes all events are interrelated, in some way. The stages and techniques are similar to the TIA method. Information is generated from desk research, Delphi techniques, and face-to-face workshops. The main variation is in the final analysis method.

The process begins by defining the events to be included in the study. This is arguably the most important effort in the process. If factors and events are missed, then the final scenarios will be blind to potential future realities; However, inclusion of events that are not pertinent can complicate the analysis.[12] An initial set of events is compiled, then refined and pared down (generally 10-40 events). This creates the scenario space for the remainder of the process.

Next, initial estimates of event probabilities are generated, as if each event occures independently; However, there is an assumption with some approaches that the

---

[12] For example, a ten-by-ten matrix requires 90 conditional probability judgments be made, and a 40-by-40 requires that 1,560 judgments be made.

probability of one event has already taken into account the probability of the other events, thus employing the CIA method at the start, even if implicitly.

Using a matrix to plot events, initial probabilities are adjusted to account for the cross-impact probabilities of other events. Such estimates are often reached using the same techniques as TIA: : literature search, Delphi, and face-to-face workshops. By combinging probabilities, the "learning process that occurs while the cross-impact matrix is being estimated is one of the major benefits of performing a cross-impact analysis" (Gordon, 1994b). The following formula (Equation 2.2) is applied to the range of conditional probabilities to satisfy consistency requirements.

Equation 2.2. Cross-Impact Analysis consistency formula

$$P(1) = P(2) \times \ P(1/2) + P(2c) \times \ P(1/2\,c)(1)$$

where:

P(1) = probability that event 1 will occur;

P(2) = probability that event 2 will occur;

P(1/2) = probability of event 1 given the occurrence of event 2;

P(2c) = probability that event 2 will not occur; and

P(1/2c) = probability of event 1 given the nonoccurrence of event 2.

Substituting the maximum and minimum probabilities for *P*(1/2) and *P*(1/2c), the limits can be calculated. After establishing all probabilities across the matrix, a calibration run of the matrix is calculated. The calibration is similar to post-hoc methods. The calibration ends with determining the impacts using odds ratios and determines the odds of event_n given the occurrence/non-occurrence of $\text{event}_{n\pm 1}$. Figure 2.5 is an example of a CIA matrix regarding the future of the chemical industry, as illustrated in Gordon's 1994 CIA paper (p 14).

Figure 2.5. Cross-Impact Analysis example: chemical industry

		The Probability of This Event Becomes				
If This Event Occurs	Initial Probability by 1985	1	2	3	4	5
1. Use of plastics expands six fold	.15		.30	.25	.10	.15
2. Increased governmental intervention in innovation	.20	.10		.35	.07	.40
3. Chemical research performed on computers	.25	.15	.20		.15	.05
4. Chemical industry expands into textiles	.10	.15	.25	.25		.15
5. Declining returns on conventional research	.20	.25	.15	.50	.20	

Note: Reproduced from Gordon, 1994b, p 2, Figure 4.

At this stage, the matrix is ready for sensitivity testing or policy analysis. If the initial event probabilities were estimated in isolation (e.g. TIA method), then CIA can be used as a sensitivity test of new event impacts. To test sensitivity, each uncertainty (event) is assumed to exist. If the assumption that event$_n$ will occur results in significant differences in the matrix output, then CIA method considers this event and associated judgments important and worthy of further exploration. Policy testing begins by defining a policy that would have an effect on the events in the matrix. Probabilities and odds are then adjusted to reflect the impact of the policy. Any differences are considered the impact of the policy. Changes can be traced back through the matrix to establish causality chains. Table 2.5 summarises the process CIA process.

Table 2.4. Cross-Impact Analysis model

Stage	Task
1	Defining the events
2	Generate probability estimates
3	Construct cross-impact probabilities matrix
4	Calibration to determine impact odds ratio
5	Sensitivity testing and policy analysis

2.2.4.3 La Prospective

The model is described as pre-active and pro-active (Godet & Roubelat, 1996). There is no single scenario method within the *La Prospective* model, but rather several methods that revolve around a unifying structure. All methods largely employ a mathematical and computer-based probabilistic approach to scenario development. Unlike the previous two models, *La Prospective* uses an integrated approach (Bradfield, *et al.,* 2005, p. 803).

> …mixed systems analysis tools and procedures', including morphological analysis for scenario building, Micmac for identifying key variables, Mactor for analysis of actors' strategies and Smic-Prob-Expert for determining the probability of scenarios (Micmac, Mactor and Smic are all acronyms for specific computer programmes developed by Godet).

The *La Prospective* model incorporates features of the IL model, but is more elaborate, complex and more mechanistic, relying on computer-based modelling with roots in TIA and CIA. Hypotheses are developed which serve as the foundation for workshop scenario development, using similar methods to the IL model. The hypotheses are then assigned single and crossed conditional subjective probabilities and analysed using one of the popular software packages (e.g. Micmac, Mactor and Smic). Subjective probability analysis is used as a decision criterium, and scenarios are ranked accordingly. Those that appear to break mental maps or indicate new possible directions are prioritised. No table is offered for the *La Prospective* model for two reasons. It is an amalgamation of the previous models and therefore a new table would be little more than a repeat of Tables 2.3-5. As well, the title is more of an umbrella term for several methods that follow a basic intuitive-plus-computer-based structure.

2.3 Scenario Planning Summary

A number of methods and insights can be applied to the other SP models. The three main models discussed in this chapter are the IL, PMT, and *La Prospective.* This investigation takes a closer look at the general IL model of SP. The IL model was chosen for several reasons.

First, the IL model has been reported in numerous reviews to be the prevailing model of SP used by facilitators (Amer, Daim, & Jetter, A review of scenario planning, 2013; Ringland, 1998). The IL model is so commonly used, it's often referred to as the "standard" approach (Bradfield, Derbyshire, & Wright, 2016; Ramírez & Wilkinson, 2014; Wright, Bradfield, & Cairns, 2013). The value of the insights from empirically investigating the techniques used in the IL model has the potential to be greater than investigations into the other SP models. Discoveries could be more applicable, disseminated amongst more practitioners, and facilitate greater ROI for private organisations. It is also important to recognise that no single SP technique, method or model of practice exists in a vacuum. Discoveries made in one branch have applications for the other branches. This is particularly true of the *La Prospective* model, which borrows directly from IL model. Therefore, designing and investigation around a single model's methodology still brings valuable insights to the field, at large.

Second, it lends itself to a quantitatively based investigation the least, compared to either PMT or *La Prospective*. This makes the task both extremely interesting and novel. This thesis proposes to not only present an investigation into the efficacy of IL techniques, but to do this through mixed-method approaches, which attempt to discover whether quantifying an "intuitive" system of reflexive interactions and decision-making can be both valid and robust. The greater the challenge, the greater the value of novel insights that may arise from the investigation.

Closely tied to the previous, and understandably so, the "intuitive" nature of the practice has resulted in very few empirical studies (17 as of December 2018; Chermack, 2018). SP, as a discipline, has not strongly addressed the concept of intuition; How the different methods and techniques may inspire it; How practitioners integrate it into their experience of the process; How it generates, bolsters, or inhibits decisions. If SP is a practitioner driven method (Spaniol & Rowland, 2018b) or even a practitioner's art (Bradfield, 2008), then understanding a practitioner's mental processes will be integral to any further studies of the IL model. The practitioner-driven argument reiterates an earlier research statement, that this investigation will need to explore ST, as much as the process of SP. How does the method affect cognition? And how does cognition affect the outcomes of the method?

In the next chapter, a literature review of publications that reflect on ST is presented. Cognitive psychological theory is integrated into the review and offers an operational definition of the identified mental phenomenon. Once an overview is established, the discussions will focus on specific ST features that are most relevant to this investigation: heuristic and biases, uncertainty, and individual vs group decision-making.

Chapter 3. **Scenario Thinking**

"The ability to dream is all I have to give.
That is my responsibility; that is my burden."
Harlan Ellison, *Stalking the Nightmare*

The first chapter introduced the theoretical foundation for the investigation. SP is defined with a constructivist approach, where praxis takes on a reflexive nature of knowledge exchange. The second chapter presented a literature review of SP, and a model was identified to develop a foundation for empirical study. The IL model is a multi-stage, iterative process that relies on the subjective input of practitioners to develop stories of different, plausible futures. Given that SP is a practitioner-driven process, it is necessary to gain greater understandings of the mental processes regularly employed, in order to better understand SP. Such mental processes are referred to, in general terms, as "scenario thinking". To understand SP, it is necessary to understand ST, for there are no scenarios without, first, the practitioner's mind. The aims of this chapter are to review literature within the SP scholarship that discusses and explores concepts of ST, then use cognitive psychological theory to develop a conceptual *process of cognitions* that, together, help define ST. Dominant cognitions are plotted against the IL model to further develop an Intuitive Logics/Scenario thinking (IL/ST) framework. Presented against this framework are key barriers that SP was developed to counter. The IL/ST framework and key barriers will serve as a roadmap for the empirical studies reported in Section II.

3.1 **Scenario Thinking Literature Review**

Martelli (2001) considers ST part of the necessary foundation to the holistic process of SP. Cairns & Wright (2018a, p. 19) define ST "as a creative process that involves subjectivity, intuitive and emotion, but which *also* involves rationality and objectivity". Mackay & McKiernan (2018, p. 39) describe how practitioners engage with ST by performing "a cognitive process concerned with imagining how the future might unfold in multiple ways through the analysis and judgement of the effects of the actions and reactions of shaping forces." At this point in the field, the confusion in language that conflates SP with ST must be abandoned for a more refined view that has emerged from both praxis and theory. It is necessary for practitioners and researchers, alike, to be clear in their discussions and declare whether they speak of SP or ST.

Wack, a leader in the discipline of SP, repeatedly discusses his motivations behind the developments. He was not concerned with forecasting or accurately predicting Shell's future. He makes it clear across his corpus that SP will not lead to predictions. Wack was focused on affecting the quality of judgments in managerial types, and the organisation more broadly. Even though Wack largely discusses "scenario planning", the art, his work is truly about "scenario thinking", the science (Wack, 1985a, p. 150).

In times of rapid change, a crisis of perception (that is, the inability to see an emergent novel reality by being locked inside obsolete assumptions), often causes strategic failure, particularly in large, well-run companies. Problems resulted from a crisis of perception rather than from poor strategic reasoning. These decision maker's strategies made sense and indeed were often brilliant – within a context of their limited worldview.

The artifacts (i.e. scenarios) are the tangible results of the process, against which practitioners can map their potential strategies, but are not the only goal of the exercise. "The end result… is not an accurate picture of tomorrow, but better decisions about the future" (Schwartz, 1991, p. 9). Where the scenario method offers multiple views of the future, it engages Boje's (2001, p. 10) notion of the "antenarrative" – giving "attention to the speculative, the ambiguity of sensemaking and guessing as to what is happening in the flow of experience". Such antenarratives provide "sensemaking that is coming into being, but not finished or concluded, in narrative retrospection" (p. 4). The lynchpin in the process's value lies in the learning capacities of the practitioners, "… scenario thinking is a learning experience rather than a desk-based exercise…" (Cairns & Wright, 2018b, p. 13). Schoemaker (1995, p. 27) concludes that, "Above all… scenarios are aimed at challenging the prevailing mind-set."

The aims of an intervention are to challenge practitioners' perspectives of the organisational environment, by increasing flexibility in thinking, to ultimately ensure the survival and success of the organisation (Schoemaker, 1993; Van der Heijden, Bradfield, Burt, Cairns, & Wright, 2002). The key to such a strategic intervention lies in guiding teams to "design scenarios so that managers would question their own model of reality and change it when necessary" (Wack, 1985a, p. 84). Wack takes this concept further when he states, "If the scenario process does not bring out strategic options previously unconsidered by managers, then it has been sterile" (p. 10). Bood & Postma (1997, p. 635) give six main functions to scenarios (and SP), two of which directly address the changes in practitioners' cognition: 1) making managers aware of environmental uncertainties and 2) stretching of managers' mental models.

However, we cannot meaningfully affect what we do not understand. If ST is the vehicle for change, then at this time, our car is driverless. Any changes experienced in managerial or organisational thinking has been little more than random crashes along the way. It was premature, though not necessarily wrong, to claim efficacy of the scenario method so early in the field's development. To advance the dialogue, it is necessary to make explicit the cognitive processes employed in ST and the challenges the extant literature discusses.

Yet, as McKiernan (2017) points out, little is known about the cognitive processes of ST and their contextual conditions. Within the hundreds of publications discussed in Chapter 2, as well as several reviews (*see* Amer, Daim, & Jetter, 2013; Balarezo & Nielsen, 2017; Chermack, 2018; Chermack, Lynham, & Ruona, 2001; Varum & Melo, 2010) less than 20 of those publications involve empirical efforts to understand the cognitive relationship with SP/ST (further discussed in Chapter 4). Most of the dialogue is speculative at best, and the farther away the narrative moves from academia, the more dogmatic it becomes (e.g. see section 1.3). This is a failure many disciplines experience. An aim of the investigation is to take control, in part, of this narrative, by adding scientifically sound knowledge to the on-going dialogue.

3.2 Terminology

The terminology of the field is treated quite loosely at times (as previously evidenced in section 2.2 with the dialogue around "process" vs "tool"), with a fair amount of interchangeability within and between terms. In an effort to separate ST from SP, it is helpful to take a moment and make explicit the meanings behind the language. Clarity is established for the purpose of continuity across later chapters. "Scenarios", as discussed earlier, are the artefacts of that process, where elaborate, plausible, visions of the future exist in story or narrative form (Spaniol & Rowland, Defining scenario, 2018). "Scenario planning" is the wholistic process of thinking about, discussing, developing, and planning for the future through a systematic and iterative process (Bradfield, Cairns, & Wright, 2015; Chermack, 2011; Van der Heijden, 1997; Van der Heijden, *et al.*, 2002). "Scenario thinking" is the complex cognitive process that facilitates SP and determines scenario content (Bradfield, 2004; Mackay & McKiernan, 2018; McKiernan, 2017).

3.3 Motivational Systems

ST is not a single cognition in and of itself, but rather plays host to several discrete cognitions that function through various levels of interrelatedness. The specifics of ST may vary with the requirements of a given SP intervention. At this time, the field has yet to define what a full cognitive system of ST could encompass. "A cognitive system is one that performs the cognitive work of knowing, understanding, planning, deciding, problem solving, analyzing, synthesizing, assessing, and judging as they are fully integrated with perceiving and acting" (Lintern, 2007, p. 398). What is revealed from the review in this section is that ST is a gestalt. Each feature, known or otherwise, provides an integral function to a wholistic ST process. No feature, though, works in isolation. The interrelationship of the cognitive structure leads to a final output that is qualitatively different in value and information than all the features, individually.

This proposes a difficult task for the study of ST. Should empirical research attempt isolated measures of individual cognitions? Would this risk potentially misleading the narrative regarding the efficacy of SP? More fundamental to psychological theory, is it even possible to measure cognition?

The first group of cognitions are those considered "cognitively ubiquitous" (Tooby, Cosmides, & Barrett, 2005, p. 316). These are continuous processes, entering almost all situations, and regulating almost all behaviours. These processes are necessary and fundamental to ST.

Superseding all mental processes are the necessary functions of short- and long-term memory. Memory encoding and retrieval are the *alpha* and *omega* of ST. The scenario process begins with encoding information into memory banks for processing and deliberation (Bradfield, 2008). Information is then recalled from memory, throughout the process, according to the demands of the task and cognitive schemata (Tversky & Kahneman, 1982).

Creativity is discussed to varying degrees throughout the scenario literature. Conceptually, creativity is universally treated as a desired and appealing quality to ST (MacKay & McKiernan, 2010). Practitioners who can exercise greater creative thinking during the scenario process, from early search tasks to scenario development, are considered to benefit SP more than those that think more normatively or rationally (Bood & Postma, 1997; Cairns & Wright, 2018; Van der Heijden, *et al.*, 2002; Yeoman & McMahon-Beattie, 2005). Conversely, if SP can bolster greater creativity in the practitioners as the process unfolds, then SP is considered to be of greater benefit to the practitioners and organisation.

Causal thinking is another fundamental cognitive function that encompasses several elements (e.g. causal reasoning, associative thinking, convergent thinking, and mapping). We assume to live in a causal universe, that there is an order to causal relationships, and assign causality to events based on a number of perceived relationships (Hastie, 2015). ST employs our innate ability to find causality between events and SP methods aim to enhance this cognitive function (Wright, Bradfield, & Cairns, 2013; Wright, Cairns, & Goodwin, 2009). Derbyshire & Wright (2017) address the need for employing different types of causal thinking, by way of Aristotle's theory of causality. They warn that ST emphasises efficient causes over all others.

(1) Material cause, or "that out of which", underpins the emergence of plausible futures in intuitive logics-based scenario planning.
(2) Formal cause, or "the account of what-it-is-to-be", explains the structure, plans, or blueprints that define the object.
(3) Efficient cause, "the primary source of the change or rest", more commonly known as cause-and-effect.
(4) Final cause, "the end, that for the sake of which a thing is done", explains the purpose and motivations of the behaviours.

Evaluative thinking is another way of identifying the valuation processes (Buckley, et al., 2015; Tooby, Cosmides, & Barrett, 2005). Just as we must, by necessity, give some forethought to our actions in order to interact with our environment, so too must we equally and automatically employ value processes regarding our environment and our decisions. Evaluative thinking includes both probability (numerically based) and non-numerical thinking (intuitively based).

The next group of cognitions are those which build from the primary and necessary cognitive structures. These are features that combine multiple cognitive functions for execution.

McKiernan (2017) offers a list of potential cognitions as elements of ST. Prospective thinking is the mental act of creating and contemplating future events, their paths, interactions, and consequences. On a similar note, pre-factual thinking is the active development of if-then propositions about future action/outcome causal relations (Epstude, Scholl, & Roese, 2016; Sanna, 1996). Forethought can provide direction to the multiplicity of plausible or potential paths (Bandura, 2001; Israeli, 1941). Episodic future thinking uses similar cognitive and neurological features for imagining events of the past to imagine future events, but unlike the previous cognitions, these imaginings include the actor in the scenes (Botzung, Denkova, & Manning, 2008; Schacter & Addis, 2009). Counterfactual and episodic counterfactual thinking are the efforts of imagining alternative versions to the scenes, where the latter includes the actor's personal past or future (Bradfield, Derbyshire, & Wright, 2016; Schacter, et al., 2015).

A few publications discuss the integral nature of systems thinking with SP (Schwartz, 1991). This is the ability to "understand the parts that make up the whole and the relationship between the two," (Glick, et al., 2012, p. 495), see the consequences of actions and how things unfold overtime (Checkland, 1981; Chermack, 2002, 2004; Lindgren & Bandhold, 2003; Van der Heijden, *et al.*, 2002).

Role thinking is another commonly discussed, yet little understood, cognitive feature of ST. Also referred to as perspective thinking and psychodramatic techniques, role thinking requires a person to consider the viewpoints of other actors/protagonists (e.g. stakeholders, executives, or clients) in order to anticipate their action and reactions (Brandenburger & Nalebuff, 2011; Green & Armstrong, 2011; Janis & King, 1954). Role thinking and role playing are discussed interchangeably at times, but are two distinct efforts. While role thinking can be performed in isolation, role playing not only requires participatory simulated behaviours, but the also the necessary precursor of role thinking (Green, 2002; Önkal, Sayım, & Lawrence, 2012; Wright, 2002).

Intuition is one of the least understood, and consequently overly defined, cognitive function. Some authors speak of intuition almost as if it were a homunculus, deftly and secretly controlling our thoughts and behaviours (Gladwell, 2005). Others believe intuition to be an executive function that could eventually be understood through systematic investigations into cues, attention, and memory functions (Glöckner & Witteman, 2010; Kahneman, 2011; Kahneman & Klein, 2009; Mulford, 1916; Tversky & Kahneman, 1983). In an attempt to summarise, intuition is a cognitive function that aids in judgment and behavioural choice.

3.4 Key Cognitive Barriers

Cognitive barriers are defined as mental phenomena that skew thinking, reasoning, and decision-making from an optimal point, creates ambiguity, or inhibits desired efforts and outcomes in SP.

3.4.1 *Heuristic & Biased Thinking*

Acting upon the discrete, yet interrelated cognitions of ST are various cognitive heuristics and biases which guide the decision-making process and outcomes. The help develop mental models and schemas, guide attention, and inform interpretations when thinking, reasoning, and making decisions. Heuristics are cognitive functions used for quickly and efficiently processing complex or large amounts of information in order to reach a timely decision (Gilovich, Griffin, & Kahneman, 2013). Heuristics are informed, in part, by past experiences (i.e. long-term memory) and knowledge gained (i.e. working memory). Generally, our cognitive heuristics are necessary to successfully navigate our daily environment (e.g. heuristic thinking makes it possible to simply walk down the pavement without being frozen in endless loops of information processing about every signal and cue in our environment). Biases are the products, if you will, from employing heuristic thinking. A cognitive bias is the systematic deviation from a norm, rationality in judgment, or logical argument (Tversky & Kahneman, 1974). Specifically with regard to SP and strategic planning, where uncertainty is high, biases can affect belief formation, policy, strategic, and economic decisions. There are an unknown number of heuristics and biases and their value to human experience is largely ambiguous. For every beneficial outcome of a bias, there appear to be equally as many negative outcomes. SP is mainly concentrated around the efforts to inhibit negative biased thinking and expand positive un-biased thinking.

These phenomena not only act upon different cognitive functions, but they also act upon the mental models (both exciting and inhibiting) which are developed from variously interacting cognitive functions, informed from the external and internal world. The field of SP categorises mental models as having bounded rationality (Simon, 1972), limited knowledge (Wack, 1985b), and employing heuristics which give rise to biased thinking (Bradfield, 2004, 2008). Mental models, or cognitive styles (Franco, Meadows, & Armstrong, 2013), are unique to each person, but can have shared qualities across groups. SP literature assumes that the process can lead to broader thinking (challenging mental models), increased learning (expanding knowledge), and increased awareness (stepping beyond quick heuristic thinking).

3.4.2 ***Uncertainty***

In all scenario work, uncertainty is an ever-present factor. Much like ST, uncertainty is a term used to represent the ubiquity of various qualities and forms. Uncertainty includes US Secretary Rumsfeld's (2002) now famous "known unknowns" and "unknown unknowns", along with Zizek's (2004) addition of the potentially even more dangerous, "unknown knowns". Uncertainty cannot be eliminated, though it can be diminished, and it cannot be fully understood, though it can altered. Therefore, any investigations into the efficacy of SP, by way of delimiting ST, must acknowledge and work with uncertainty as a factor in practitioner thinking.

3.4.3 ***Individual vs Group Decision-Making***

Cognitive psychology also supports the perspective that individual decision-making can succumb to different types of biased thinking compared to group decision-making (Bennett, 1990; De Bono, 1987; Eden, 1992; Eden & Ackermann, 1998). Much of the extant literature regarding biased thinking, reasoning, and decision-making focuses on individual cognition. Tversky & Kahneman sparked the conversation by investigating individual preferences for subjective probability, risk-averse, risk-seeking, isolation effect, overconfidence, representativeness, base-rate neglect, framing effect, anchoring, availability, and conjunction fallacy (Kahneman & Tversky, 1972, 1979; Tversky & Kahneman, 1973, 1974, 1981).

Biases within group-based decision-making focus on phenomena such as groupthink (Janis, 1982), social loafing (Jones & Roelopsma, 2000), polarization (Sunstein, 2007), and escalation of commitment (Bazerman, Giuliano, & Appelman, 1984; Forsyth, 2010), which can encapsulate features of individual-level biases. For example, groupthink can occur when concurrence-seeking is desired by group members more than critical thinking and results in a “deterioration in mental efficiency, reality testing and moral judgments as a result of group pressures” (Janis, 1971, p. 84). Individuals converge on a single, agreed-upon issue for the sake of ease. The issue can often be the product of a dominant member’s input, around which the other group members rally. If the dominant member’s working mental model suffers from dysfunctional biases (e.g. anchoring, myopia, or hindsight) then the emerging groupthink will also encapsulate the same biases.

Given that SP require both individual and group decision-making, these differences must be explored. To effectively attenuate negative biases, it is important to understand what biases people bring to the table, and what biases may arise from the process. There are, arguably, just over 50 recognised decision-making, rationality, belief, and behavioral biases, with the understanding that an uncountable number of more biases may yet to be discovered. The field of research is still quite young, but popular, which has led to several disagreements on precisely what constitutes as a bias. Rather than start at the beginning of a long and contested list to test for each bias within participant samples of SP practitioners, efforts are first made to limit the list to the most pertinent and popular biases believed to be associated with SP and ST.

3.5 **Intuitive Logics/Scenario Thinking Framework**

SP can be viewed as separate, integrated stages, coordinated together through incremental decision-making efforts. Balarezo & Nielson (2017) recognise that when discussing SP influences, we are actually discussing two directions of potential influence. The research areas are broken down into how SP affects cognition vs how context (i.e. ST) affects SP. This research is interested in exploring the second form of influence. We cannot understand the efficacy of SP until we understand the robustness of the method against necessary and potential influencing factors. This may very well be why different studies reveal different outcomes and different organisations report different successes.

ST is a process of discrete, interrelating cognitions. Each stage of the IL model emphasises the domination of one cognition over others. Cognitive psychological theory helps inform the dominant cognitive processes to be measured at each stage, as well as how to measure them. The systematic process of moving from one dominant cognitive engagement to the next is championed as the efficacious SP method for breaking from cognitively easier normative thinking (which relies heavily on heuristics), and broadening awareness and perspectives. Set against the IL model, eight dominant cognitive processes are revealed at each stage.

The most important effort in Stage 1, regarding the present investigation, is the personal research carried out by the practitioners to familiarise themselves with the agenda (i.e. memory encoding). Not only do practitioners enter with their own mental models which dictate the kind of research carried out, but the discoveries made during the research stage also develop and set their mental models and schemas. Practitioners then use their mental models and schemas to interpret their views of the external business environment and future scenarios. Stage 1, therefore, lends itself to set the groundwork for systematically introducing variables that could help not only reveal the presence of biased thinking throughout the process, but the impact of specific biases.

Stage 2 tasks are often referred to as "brainstorming" (Bradfield, Derbyshire, & Wright, 2016; Bradfield, *et al.*, 2005; Bradfield, Cairns, & Wright, 2015; Davis, Bankes, & Egner, 2007). Brainstorming is a function of creative thinking (Baas, De Drue, & Nijstad, 2008; Lamm & Trommsdorff, 1973).
Stage 2 emphasises greater efforts in creative thinking than other cognitions. Experimental methods are, therefore, developed to detect and measure creativity during this stage.

Stage 3 shifts practitioners' efforts from creatively identifying DF to clustering the DF together based on their associated causal relationships. Associative processing is a function of causal thinking and is the dominating cognition in Stage 3. For this stage, experimental methods will focus on measuring associative, causal thinking.

Focusing on the opposing or differently developing outcomes reflects a strong effort of parallel thinking in Stage 4. As a group, practitioners must explore alternative outcomes along the same cluster timeline in a disciplined manner. This motivational cognition is not discussed in the SP/ST literature. However, the methods described by Van der Heijden (1997) and Cairns & Wright (2018) reflect such thinking. Experimental designs for this stage will require methods that detect and measure alternative outcomes from parallel thinking.

Stage 5 moves the practitioners into efforts of evaluative thinking. There are, in fact, two tasks in this stage – impact and predictability – but both require evaluating these qualities without the aid of numerical guidance or measures. SP literature treats this stage somewhat lightly. As well, different SP methods, even within the IL model, teach this stage differently. For example, some frame the task as "determine how impactful the cluster will be on the organisation's ability to reach their goals", while others direct practitioners to "determine the cluster's impact on the organisation". This is a small difference in wording, but creates differently defined tasks that will affect how the practitioners evaluate each cluster's impact. Another difference at this stage is that some methods rank-order by impact (Chermack, 2011), while other methods treat each cluster's impact as an independent measure (Wright, 2014). For the purposes of this investigation, the experimental design will make explicit, the impact and predictability instructions for participants.

Stage 6 requires practitioners to switch to a mode of causal reasoning. As dyad outcomes from independent clusters are combined into four new scenario storylines, practitioners must negotiate how the outcomes causally relate to one another while maintaining internal consistency. During the sense-making exercise, practitioners are known to employ more analytic thinking as they validate plausibility and affirm internal consistency. This stage has a higher chance of prompting practitioners to engage in deeper causal reasoning, employing potentially all of Aristotle's four causal mechanisms. Experimental methods will need to account for detecting and measuring *different* forms of causal reasoning.

The complex and deep-dive mode of Stage 7 employs a systems thinking effort (Schwartz, 1991; Ward & Schriefer, 1998). Practitioners are forced to "adopt a systems view of their organizations and the environments in which they operate" (Chermack, 2003, p. 8). Two popular designs for measuring and testing systems thinking are closed- and open-ended methods. Closed-ended methods use typologies or taxonomies to determine boundaries of cognitive performance. Open-ended methods, allow the participants to develop their own boundaries.

Stage 8 development and writing employs a synthesis of creativity, causality, future rehearsals, episodic future thinking, prospective thinking, evaluative, role thinking, political thinking and systems thinking, at the very least. This final stage requires a synthesis of cognitive processes, each taking turn, through internal deliberation and external dialogue to develop the scenarios in full. An experimental design could focus on group interactions or as is common in SP studies, analyse the final scenarios.

Table 3.1 maps the dominant cognitive processes within ST against the existing IL model, to develop the Intuitive Logics/Scenario thinking (IL/ST) framework (Cairns & Wright, 2018; McKiernan, 2017; Van der Heijden, 1997; Van der Heijden, *et al.*, 2002; Wright, 2014; Wright, Bradfield, & Cairns, 2013).

Table 3.1. Intuitive Logics/Scenario thinking framework

Stage	Task	Scenario Thinking
1	Setting the agenda – background research, define goals and timeline	Memory encoding
2	Determining the DF – critical uncertainties and predetermines	Creative thinking- Divergent
3	Clustering the DF – causal associations between DF	Causal thinking- Associative
4	Defining the cluster outcomes – extreme plausible outcomes of clusters	Parallel thinking
5	Impact/Predictability matrix – determine most impactful and uncertain clusters	Evaluative thinking
6	Framing scenarios – combine outcomes of four scenarios	Causal reasoning
7	Scoping scenarios – broad descriptors and implications	Systems thinking
8	Developing full scenarios – write narratives and test for plausibility	Synthesis

The IL/ST framework will inform the development of a series of empirical studies that test and measure dimensions of ST against SP. The next steps in development are to gain a sample view of real-world SP applications and explore existing empirical work on heuristic thinking and biased decision-making in ST. A sample of real-world qualitative data will help provide a snapshot of how practitioners value the process and motivations for using SP. In tandem with the interviews will be a structured review of empirical work that has attempted to measure biased thinking and decision-making within a SP context. The purpose for gathering both real-world and empirical data is to help reveal gaps in the field and issues that are most relevant, which will, in turn, develop the guiding questions for the investigation. A dual approach is used because it is equally important that the present efforts attempt to bring valuable insights to organisations and future practitioners, as much as it is to bring value to a growing theoretical development.

Chapter 4. **Interviews and Systematic Review**

"There is no such thing as a worthless conversation, provided you know what to listen for. And questions are the breath of life for a conversation."
James Nathan Miller, c.1965

"Nothing of what you say displeases me, so keep talking,"
Don Quixote, c.1605

The previous two chapters conducted a broad literature review of SP and ST, respectively. Chapter 4 builds on their discoveries by reporting a sample of real-world SP experiences and reviewing empirical ST studies. Real-world data will offer a snapshot of executives' motivations and values regarding SP. In a sense, conducting interviews with SP practitioners – past and present – will help the researcher gauge their perspectives and gain an intimate view of their thought process not otherwise available in traditional experimental and quasi-experimental studies. Empirical data will help reveal the kind of progress achieved from investigating and measuring ST. Both topics move the thesis narrative from theory to practice and praxis. The aim is to identify important themes and gaps in the field.

4.1 **Scenario Planning Practitioner Interviews**

To gain real-world perspectives around SP, experiential insights are sought. The personal experiences of professionals who participate in SP workshops, within organisations, are valuable sources of knowledge. There exists case studies and interviews of SP practitioners and facilitators. However, these sources are influenced by a number of factors (e.g. interviewer's agenda or focus of the conversation) that limit the amount of available information. Therefore, a series of interviews was conducted. The method for the interviews was unstructured (Fontana & Frey, 2005). An unstructured method is used when the interviewer desires to elicit people's social realities (Zhang & Wildemuth, 2009). An unstructured method was chosen to allow for discovery of potential themes and processes in SP praxis. The method relies on social interaction between the actors (interviewer and interviewee), where neither questions nor answers are predetermined (Minichiello, Aroni, Timewell, & Alexander, 1990). A reflexive approach was taken by the researcher to allow for questions to emerge from all actors in each interview. An unstructured method was determined to be the most beneficial format, in part, because it could allow for discovery of novel insights previously unknown to the researcher (i.e. interviewer). The approach allows the executives (i.e. interviewee) to tell their stories and for the research to learn from them in the process, as opposed to guiding them through pre-developed boundaries to the dialogue. Understandably, an unstructured, reflexive approach leaves the interview more susceptible to interviewer effects and biases. To help control for these potential effects, a constructivist perspective was taken. To understand the interviewee's reality, the interviewer needed to approach the information through their perspectives and terminology and allow adoption of changes as they enter the interviewing space (Denzin, 1989).

4.1.1 *Purpose*

There are two main purposes for conducting interviews. First, a sample of interviews will offer real-world perspectives around SP praxis. Second, discoveries from the interviews have the potential to reveal gaps in both the executives' and researcher's knowledge.

4.1.2 *Participants*

Between September 2014 and February 2015, three unstructured interviews were conducted with executives who either presently worked or had formerly worked with companies that used SP, of which they participated. The years of their experience ranged from one to eight years. Two interviewees worked in finance and marketing firms located within continental Europe. One interviewee presently worked in the Scottish energy sector. All interviews took place in Scotland.

4.1.3 *Interviewer*

The quality of the conversation is influenced by how the researcher presents their self (Zhang & Wildemuth, 2009). To promote a reflexive, constructivist method, the researcher/interviewer presented herself as a learner, one who desires and is willing to understand the interviewee's experiences from their perspectives (Burgess, 1984). A *conversational nature* was adopted to help facilitate the interviewer's own responsiveness to new information, gaps in shared information, novel language use, context, opacity in conceptual knowledge, and any situational changes. The primary aim was to listen, secondary aim to lead and adjust the conversation when needed.

4.1.4 ***Method***

Interviews were held at the University of Strathclyde, a neutral location not affiliated with the interviewees' employers. Sampling was a blend of convenience and purposive methods. Each interviewee (n = 3) was recruited through previous conversations regarding SP. The interviewees self-identified as having a practical history with SP workshops in either their present or former employment. Each interviewee was asked if they would be willing to participate in an interview to discuss their experiences using SP in their workplace. All agreed to an interview.

At the start of each interview, interviewees were informed that the purpose of the interview was to understand the details of their experiences using SP within their respective organisations. Interviewees were informed that their personal and employer's identities would be anonymized. Only their gender, professional fields at the time of SP workshops, and employer's locations would be reported. Participants were then asked their permission for notes to be taken and direct quotes to be recorded. All participants gave their verbal consent. Audio and video recordings were not possible at the time of each interview, therefore, note-taking was the only method of data gathering. Understandably, note-taking can disrupt the natural flow of the conversation and the author can fail to record all of the conversation, hence why audio and video recordings are preferred within this method. To help mitigate interruptions, notes were typed in real-time on a mac laptop. The laptop as an apparatus for notetaking served two benefits. First, typing instead of handwriting increased recorded word capacity, which allowed for more opportunities to record direct quotes. Second, the screen always faced the interviewer, never the interviewee. This proved to be beneficial by allowing the interviewer to have information on the screen without adding distractions to the interviewee's visual field, and therefore preserving a similar *tabula rasa* setting as was presented with the first interviewee. Following suggested interviewer procedures, more detailed notes were immediately written up after each interview (Fontana & Frey, 2005). As the interviews developed, terms and phrases were learned, and a short hand was developed to increase the interviewer's word capacity. Many of the short-hand symbols were borrowed from mathematics and strings borrowed from texting abbreviations. Table 4.1 presents a sample of the interviewing shorthand.

Table 4.1. Interview shorthand

Word or Phrase	Shorthand
"scenario planning"	sp
"scenarios"	scn
"workshop"	wks
"organisation"	org
"change"	Δ
long pause or missed words	…
"you know"	yk
"fluctuation"	fluc
"recommendations"	rec
"project"	proj
"projections"	projn
"department"	dept
"management"	mngm
"information"	info
"meeting"	mtg

The open-ended, reflexive approach resulted in a learning process for all participating in the interviews that brought new insights into each successive dialogue. An *aide memoire* was not used in the first interview, but the information gathered from that interview informed the development of such a tool and was used in a flexible manner through the remaining two interviews (Minichiello, Aroni, Timewell, & Alexander, 1990). The main function of the *aide memoire* was to help with prompting and redirection. There was no order to the content and questions were still not predetermined. The evolution of the *aide memoire* is described in each interview section.

When a new topic emerged, efforts were taken to not stifle the flow of the new information. The interviewee was allowed to share new information at their own pace. Previous topics and unanswered questions were revisited only when gaps in the conversation developed from the interviewee. The gaps were categorised as a "need for prompting" or "redirection", depending on the context of the conversation at that moment.

4.1.5 *Results*

The unstructured nature of the interviews resulted in a time span of 10-60 minutes for a single interview. The questions represent an ending of information by the interviewee and were either a prompt for more information or an effort to redirect. Redirection, however, was only used when a topic was perceived to be exhausted. The format below gives the basic structure of the questions and the text, in full, that the interviewee was able to type in real-time plus edits made after each interview. When possible, the interviewee asked for clarity on a passage just typed, to ensure the interviewee's language was captured as clearly as possible. Efforts were made to preserve each interviewee's vernacular as much as possible. The ellipses represent pauses by the interviewee. The edits included correcting abbreviations. All interviews started with collecting the interviewee's agreement to the interview conditions. Interviews ended when the interviewee felt there was no more information they could offer or had to prematurely end due to other commitments in their schedule. Interviewees are identified by the order of their interview: interviewee 1 (Inv-1), interviewee 2 (Inv-2), interviewee 3 (Inv-3).

4.1.5.1 Interview 1

Inv-1 first asked where to start. Inv-1 was asked to describe their job title or department they worked in and in which country their employer was (is) located. Inv-1 said they worked in the financial department of a consulting firm within the EU for just over two years.

The purpose of making the next question vague was to see what the most salient or important parts of their experience with SP would be. For the purposes of this interview, the first shared details are considered more salient (and potentially more important to Inv-1) than the later shared or prompted details. Therefore, the order of Inv-1's information is preserved in this section.

Researcher	Inv-1
OK, would you be willing to describe how scenario planning was used in your organisation?	*Workshops were held basically quarterly.*
How long did the workshops last?	*We would take a day out for them.*

Inv-1 appeared a little uncomfortable and was giving short answers with long pauses. The tactic in questioning style was changed and next more directed questions were asked, while still trying to preserve enough ambiguity to allow for exploratory conversational development.

Researcher	Inv-1
What kind of information did you cover in the workshops? Not specific details, but rather what was the purpose of holding scenario planning workshops?	*Ah, well, it was always to track the company's performance against the previous year's projections. We had to come up with annual projections of cost and price fluctuations to help… develop plans for the next year. Our workshops were quarterly to time with our fiscal cycles… to help track our progress… determine how far we were from them, whether we needed to adjust our projections… you know, what new information we missed.*
How was the information used?	*Well… my boss would take our recommendations and scenarios to management from other departments and they have their own meetings… all the departments would do similar and management would hold meetings to decide the next steps.*

So were these workshops taken as multiple steps along the management hierarchy?	*Yes, I suppose. Yes… I mean, we would do some scenario planning, come up with projections, then my boss would take our work to higher ups.*
Do you think scenario planning helped the organisation, or maybe just your department?	*Sure! I mean, we were never far off from our projections… and it helped the different branches catch up with each other… you know, coordinate.*
What did you do in the workshop? Your part? Did you ever lead a workshop?	*No, I never led them. We brought in outside guys for that… consultants from, I guess a different firm… they didn't work in our company. The rest of us prepared information for the workshops.*
How long did you usually have to prepare your part?	*Oh… sometimes just a few days… maybe longer at other times.*
You said they took a day? How big were the workshops? What I mean is how many people attended each workshop?	*Just the managers… usually maybe seven to ten.*
Did any other employees attend these department-level workshops?	*No.*
How easy or difficult were these workshops, since you had a day for it all?	*Well, they were really involved, you know… lots of talking and trying to measure up what we knew… we would eat lunch right there, you know! Some people would have to leave… that happened often actually… we were busy people after all, with our own work and deadlines to meet. So sometimes it would be busy and everyone would be talking over each other, then people would leave or come later, and we*

	would be in little groups… always little groups… talking through some specific idea or trying to work through how some projections would make sense.

Inv-1's description of the dynamics of the room seemed like a key point. The next question was informed by a prior conversation the interviewer had with two experts in SP who mentioned their perspectives of common pitfalls in the practice.

Researcher	**Inv-1**
What kind of dynamics do you remember from the workshops, between the attendees and with the facilitators?	*Well… the workshops always started with the facilitators reviewing our previous projections and what we aimed to accomplish that day… maybe accomplish isn't the right word… Everyone would include something, interject where they saw fit, y todo… But it wasn't, like, a space of equality or anything like that.*
Would you be able to elaborate on this?	*I suppose some would say it could get a bit one-sided… no matter what was discussed in the workshop, for example, our boss always had the final say. But it's not like he ever just bullied his way through the workshop, you know? We all had something to add to the scenarios and projections.*

The interview ended at this point because Inv-1 had another appointment. They were thanked for their time. Immediately after the interview notes were reviewed, ambiguously abbreviated text was corrected, and sidenotes to the text were added to preserve the interviewer's thought process where possible.

4.1.5.2 Interview 2

Information from the first interview led to the development of a short *aide memoire* that included the following topics: purpose, frequency, time length, practitioners, and value of SP. These topics were used as a guide for data gathering and not for order of questions.

Inv-2 was in marketing and participated in one SP workshop while working with their previous employer. Inv-2 was employed at the firm for seven years and held different positions within the company, but participated in SP only when they held a higher position in the marketing department. The second interview lasted almost 60 minutes. The length of the interview was extended, in part, due to a phone call Inv-2 received towards the end of our session and to unrelated conversations during the session. Though the flow of the interview was disrupted twice, it is unclear whether these disruptions limited the information that Inv-2 shared.

Researcher	**Inv-2**
What was the focus and purpose of your company using scenario planning?	*The firm recently brought on a new VP of Field Marketing. She wanted to take things in a new direction, I think the firm was dealing with some major disruptions… it's honestly why I eventually left. Our division was trying to find solutions, but… it was pretty chaotic there.*
How long did the workshop take, from beginning to end?	*I think the whole process took about four months. I wasn't involved at every step, but that's about how long it took*

	my division to work through it all and come up with some scenarios and recommendations for the firm.
Was the workshop every day for the four months?	*No. There were early meetings to discuss our plans, gather information from within the firm, OAS statements, some interviews… then there were two workshops, each lasting about a day to two days. Outside of them some of the team leaders consolidated the workshop information and got back with everyone. A packet was sent to each of us, before the first session, things to prepare.*
What does OAS stand for?	*Objective, Advantage, and Scope.*
How long did your division dedicate to preparing for the first workshop?	*Quite a lot, actually. Each one of us was called on to the project. We were looking at several timelines… lots of projects were going on… I had to focus on one aspect of the larger picture, so had to do some market research, things of that quality.*
Did you have time to read through all the materials sent to you before each session?	*Not really, but most of it.*
How much time passed between the workshops?	*A couple months perhaps? More than a month, I'm sure.*
Did you ever forget what your group talked about during the sessions because of the time lag in-between?	*Not really. Maybe some. But a lot of that was taken care of with the prep we had to do before each workshop, specifically.*
How many participated in the workshops?	*There were somewhere around 20 in all, I think. Not everyone made both*

	workshops, though, I managed to make it to both… it would have looked bad if I hadn't. It's very busy work, very tiring. Exhausting actually.
Do you remember who attended? Not necessarily names, but job ranks and titles perhaps?	*We had senior managers there, the VP, of course, and other executives.*
How was the first workshop ordered? What was the process?	*We began with introductions from everyone. We had others facilitating the workshop. The issues our firm was facing were discussed and the goals, objectives, scope of our strategy were all discussed. We wanted to explore potential opportunities that would take the firm in different directions. I remember we were pushed to think "outside the box". I think we did a good job of it. That first session we did a lot of exploratory work. Then we looked at what was important, what did we know versus what didn't we know, and prioritised our thoughts into different focus topics.*
What kind of technology did you use during the workshops?	*We used post-its, PowerPoint, a whiteboard… a lot… and internal software for modelling, transfer and sharing.*
What was your impression of the workshop?	*It was good. Lots of shared insights. Some agreement between the senior and junior members.*
How many scenarios did your group develop?	*Four*

Did this seem like a good number of scenarios?	*Sure. Definitely. We had one that covered utter devastation, growing with new clients, one that discussed our relationship with potential EU member states, and I think the other one was slanted towards technological developments.*

The interview was interrupted at this time by a phone call Inv-2 received. They had to leave shortly thereafter so there was only time for one more question.

Researcher	**Inv-2**
Did you feel that scenario planning was valuable for your firm?	*It was part of the culture. Adopt strategic planning in all its forms. Throw money at it. Whatever works to grow the firm! "Save the firm" in my opinion.*

4.1.5.3 Interview 3

The *aide memoire* was expanded, after the second interview, to include technological aids.

Inv-3 was an engineer at a clean energy firm within Scotland and participated in one SP workshop during their eight-year tenure. The workshop was within the last two months. The third interview lasted 20 minutes. Much like the first interview, Inv-3 began by asking what to talk about first.

Researcher	**Inv-3**
Does your firm hold scenario planning workshops regularly or intermittently?	*Somewhat yes, I suppose. Like I said, I'm new to this, but I've heard of them as a regular part of our strategy.*

	Though I believe they're making strides to integrate this more regularly.
What was the focus and purpose for scenario planning?	*We work in clean energy and we're breaking new ground and into new markets fairly rapidly. We used scenario planning to see what our transition options are.*
What was the style of scenario planning your group used? Did you work from a future point backwards or from the present to different future points?	*A little of both, but if you twisted my arm, we did focus more on the kind of territory and customers we wanted to gain… we looked at the kind of paths that would take us there. What would get in the way. And who.*
How long did the planning take, from beginning to end?	*Probably a year. The amount of work that went into it all was very intensive. And it wasn't just scenario planning we used. Other teams were integrating other strategic analyses. Nothing is small in energy. Too much money involved to mess up.*
What other steps were taken along with the workshops? Like interviews, other meetings?	*Let's see, there was at least one early interview… we were building an understanding of the most important elements for our firm… our past trends… Then fair bit of research, modelling projections, risk analyses, lots of meetings.*
How many workshops did your group hold?	*One… No, two.*
Were you involved throughout the whole planning?	*Mostly just at the start. I came in again at the end.*

How long did you dedicate to preparing for the first workshop?	*Loads. A few weeks before the first workshop. Then there's the meetings with everyone after, from the top down, to develop our plans and how they'll be implemented.*
Did you have time to read through all the materials sent to you before each session?	*Mostly. Not everything involved my input. I read what I needed to.*
Did you ever forget what your group talked about during the sessions because of the time lag in-between?	*Haha. No, I was involved in so much of the work.*
How many participated in the workshops?	*Well, I attended the first workshop. There were upwards of 30, though more if you consider the whole process from beginning to end.*
Do you remember who attended? Not necessarily names, but job ranks and titles perhaps?	*Engineers from my team, senior people, some investors, suppliers, a couple representatives from Scottish Government, as well. At least for some of the workshops.*
How was the first workshop ordered? What was the process?	*Introductions were made. Our previous projections, aims, everything from the interviews, time constraints were all laid out. Most of the time was like a big brainstorming session. Everyone working on linking ideas within and outwith, to other ideas.*
Did you work in one large team or did you split into smaller teams at any time?	*Aye, we started altogether, of course, but as we developed more information, we were split into teams of five or so, based on our experience.*

What kind of technology did you use during the workshops?	*Paper, computers, whiteboards mostly.*
How many scenarios did your group develop?	*We had five in the end.*
Did this seem like a good number of scenarios?	*Yes.*
What was your impression of the workshop?	*Well, it started off with bringing specialists and stakeholders together to break new efforts in our strategy. But in the end, the CEO and Board of Directors call the shots. It becomes very political very fast.*
Did you feel that scenario planning was valuable for your firm?	*We'll see. The workshops were good, we shared a lot of information. Came up with some brilliant projections. Time will tell.*

4.1.6 ***Interview Discussion***

The interviews served to bring real-world insights into SP praxis. A summary table of themes that emerged from the three interviews is provided in Table 4.2. There were several points within all interviews when responses were short or quickly ended. Other times, gaps or unfamiliar information was given. When these moments were encountered, efforts were made to encourage the interviewee to relate a further experience or perspective to the topic (Burgess, 1984). For example, Inv-2 mentioned "OAS statements", of which the interviewee was not familiar and asked for further clarity. Inv-3 mentioned the large number of workshop attendees brainstorming together, which gave the impression that either chaos was part of the process or chaos was controlled through division of labour, which prompted a follow-up question for clarification.

Table 4.2. Summary of interviews

Theme	Inv-1	Inv-2	Inv-3
Frequency	Regularly	Intermittently	Regularly
Role	Practitioner	Practitioner	Practitioner
Length	1 day	4 months	1 year
Motivation	Track performance	Risk avoidance	Exploratory
Group size	7-10	~20	30+
No. of scenarios	-	4	5
Technology	-	Paper, whiteboards, PowerPoint, internal software	Paper, whiteboards, computers
Value	Yes	Yes/Maybe	Yes/Maybe

Note: The "-" is used to indicate a lack of information to the theme from the interview.

All interviewees reported that their organisation held multiple SP interventions, before and/or during their employment, for various lengths of time. All gave the impression that from an organisational perspective, SP has value. Further review of the data revealed an interesting relationship between *frequency* and *motivation*. Inv-2's organisation appeared to use SP intermittently, while organisations for Inv-1 and Inv-3 seemed to have integrated SP as a regular part of their strategic profile. At first glance, it appeared that each interviewee gave a different reason for their organisation's motivation to include SP in their strategic development efforts – track performance, avoid risks, and explore options. However, accounting for further contextual information within each interview, and comparing across the interviews, two main divisions within this theme emerged. Inv-2 gave the impression that their organisation used SP at a time of negative business conditions, where increased external and internal threats played more into the awareness and language of the interviewee (e.g. "major disruptions" and "Save the firm"). The organisations Inv-1 and 13 worked at appeared to use SP in a variety of settings, but largely what could be described as positive business conditions. One organisation was presented as maintaining their projections, with an intuitive suggestion that these were generally positive (i.e. not threats or failures mentioned), and the other organisation was presented as being in a period of expansion. Though the sample size is small, the interviews highlight a feature of SP praxis that is widely discussed in business literature, which is the frequency of SP, given perceptions of threat in the organisational environment. Two issues arise from the thematic analysis of the interviews. First, the study lacks breadth of sampling to understand the prevalence of SP. It may be the case that SP is sought out (i.e. valued) more often under one condition (high-threat or low-threat) over the other. As well, the interviews were ambiguous about the quality of output and strategic support derived from the respective SP workshops.

To address the first issue, it would seem logical to assume that SP – and by extension, other strategic practices – employed in response to different conditions (e.g. high-threat vs low-threat) would be valued differently since the conditions, themselves, present different value systems and potentially different outcomes. The theory of threat-rigidity proposes that perceptions of higher threat diminish risk-seeking behaviours in managers (e.g. engaging in strategic planning and business model adjustment), while perceptions of opportunity (i.e. low-threat) induce more risk-taking attitudes (Dutton & Jackson, 1987; Staw, Sandelands, & Dutton, 1981). However, results have been mixed. Prospect theory has, as of late, served a contrasting view to threat and risk-based strategic planning engagement (see Asgary & Levy, 2009; Fiegenbaum, 1990; Fiegenbaum & Thomas, 1988; Holmes Jr, et al., 2011; Shimizu, 2007). Prospect theory holds that people engage in risk-seeking behaviours when they perceive higher threats, as opposed to more favourable conditions with lower threat levels (Kahneman, 2011; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). In short, the opposite of threat-rigidity theory. Saebi, et al. (2017) reviewed the strategic behaviours of over 1,000 Norwegian companies and compared them against threat-rigidity theory and prospect theory. The authors discovered that companies held higher value for risk-seeking, strategic practices when they perceived external threats (negative), as opposed to seeking new opportunities (positive). March & Shapira (1987) report that managers assign disproportionately greater value to strategic planning and interventions when faced with a failure to meet targets (e.g. high-threat or negative conditions), as opposed to when targets are assumed to be secure (e.g. low-threat or business-as-usual conditions). The 2020 coronavirus pandemic (COVID-19) appears to also have served as a major catalyst to not only the popularity of strategic planning, but SP in particular, on a global scale (Crawford & Wright, unpublished paper).

Using prospect theory to predict managerial behaviours, we should expect SP to be perceived with greater value to an organisation during high-threat conditions. All interviewees, however, indicated that their experiences with SP had some level of positive value to their organisation (Inv-1 "Sure!... it helped the different branches catch up with each other", Inv-2 "It was good. Lots of shared insights.", Inv-3 "The workshops were good, we shared a lot of information."). This is fairly surprising, and more importantly, offers an initial glimpse into ST, if in retrospect. Irrespective of threat-levels, process, and outcomes, all interviewees felt SP brought positive value to their organisations – largely through the process of information sharing. One answer is that the interviewees were correct and that their internal valuations aligned with their organisation's. Another answer is that SP was valuable to each of the interviewees, but one or all of them were mistaken on how valuable it was to their organisation. A third potential answer is that some or all of the interviewees artificially inflated their perceptions of SP's value. Two out of the three potential answers imply biased thinking in some form. A style of thinking that prospect theory, and even threat-rigidity theory, propose. With a growing scientific lexicon of potentially hundreds of cognitive biases in action, it is important to try to understand how such thinking could affect the *process* and *value* of SP.

To address the second issue, no interviewee was able to report the quality of output and true strategic support derived from the respective SP workshops. Does SP lead to more successful business strategies for those that utilise the process regularly during low-threat conditions, or for organisations who intermittently capitalise on the process during high-threat conditions? To recall, Inv-1 stated that their group was "never far off" from their previous projections, indicating new insights were potentially a rarity. This attitude lies in stark contrast to one of the core tenets of SP, which assumes its value lies in altering the mental models of managerial practitioners and inspire new insights (Chermack, 2011; Schoemaker, 1995; Wack, 1985b). Yet Inv-1 felt that maintaining expectations was a valuable outcome to SP. Inv-2, on the other hand, admitted to leaving the organisation for being too disruptive and chaotic, yet felt the workshops were good, even as they were framed as a somewhat last-ditch effort with no real outcomes shared during the interview. And though Inv-3 admits to the potential issues of in-office politics dominating strategic decision-making, while also being personally inexperienced with SP, they still considered the process valuable to their organisation. The interviews show that future scientific and theoretical investigations into SP and ST must find methodological ways to account for value, whether by using proxies or primary sources.

4.1.7 ***Limitations***

By using an unstructured method, it cannot be known how much the role of interviewer influenced the direction, flow, and information sharing of the conversations. For example, it is difficult to determine how much information failed to be gathered from the interviewees. Even though new questions were developed as the interviews progressed (both within and between), it is not possible to know whether an earlier interviewee would have been able to answer any question that was identified later. This same opacity holds for all new questions discovered from later investigations. But the point of the interviews was not to gather complete knowledge of SP, nor complete knowledge about SP praxis from the interviewees. The purpose was to gain knowledge about their experiences in the manner they chose to share.

By not using an *aide memoire* from the start, it was not possible to cover similar topics across all the interviews. The method does not allow for equivalent comparisons of data. However, to reiterate, the method of the interviews was exploratory because the purpose was to gain insights about an unknown process from a demographic of which the researcher had not been a part at that time (SP practitioners).

The novice, untrained background of the researcher may have introduced limitations and bias into the method. There are a few leading questions that were given to probe farther into details of the interviewees' experiences. Though attempts were made to present probing questions with as much ambiguity as possible, the adoption of the researcher's/interviewer's own language by Inv-3 (i.e. "split") shows that some immediate influence was made on the answers from the interviewee. There is also missing information from the participants that could have been gathered by a better trained interviewer with more in-tuned critical listening skills. It is also recognised that the experiences between the three interviewees is different enough that it cannot be ruled out the possibility that the researcher exercised a selective approach in sampling participants. As stated in the Method section, participants were identified during conversations that brought out the fact they had experience in SP with their employer. The three interviewees were not the full population of SP practitioners encountered by the researcher. Therefore, it may be the case that the sampling method was implicitly biased by choosing professionals who exhibited differences from the previous interview(s). The results, however, offer a glimpse into the variety of contextual environments that surround SP in private organisations.

Another limitation can be found in the recording method of the interviews. By typing all notes, instead of audio recording, some information, which could have aided in more direct quotes, was lost. There were concerns that the sound of typing would disturb the interviewees, but no one appeared to respond to this effort, and when asked about their impression, post-interview, all said they were not bothered by the sound of typing. Inv-3 even said they stopped hearing it after the first few minutes.

The sampling method also introduced a bias into the data. Interviewees were selected based on meeting a specific criterion – former experience practicing SP. It is reasonable to assume that practitioners of SP are more likely to value the method more highly than those without experience, if for no other reason than they were part of the process. In general, people prefer the efforts they seek to engage in over those they do not. Without more interviews from a more heterogeneous sample, the skew of bias cannot be known. However, the purpose of the interviews was to gain initial insights into SP praxis, rather than develop an in-depth qualitative study of SP practitioners.

4.1.8 *Interview Conclusions*

The three interviews offer a number of initial insights. First, SP was utilized during times of differing threat-levels (high vs low). The interviewees (i.e. practitioners) found positive value in their SP experiences, irrespective of the conditions and outcomes (indicating potential strong biases in ST). Finally, the interviews give a glimpse into how SP was practiced and the kind of stages the practitioners experienced. With respect to my thesis' central question, "*How does scenario thinking, a collection of higher-order cognitive functions, effect the actions and content of scenario planning?*" the interviews offer a starting focal point – The value of SP as a factor of organisational threat conditions. The presence of threat conditions can be treated as a constant, in that every organisation works within some state of threat condition. An organisation's given condition could be seen lying on a spectrum from low-threat to high-threat. Threat conditions can then be used as primes for practitioners' ST. Since cognitions cannot yet be directly measured, SP performance measures could be included as proxies, which offer quantifiable indicators of ST qualities. The value found in SP, by the practitioners, could be measured against their own performance outcomes to help bring greater understanding to i) the relationship between ST and SP and ii) the kind of values found in SP. **This investigation proposes to explore whether differing external threat conditions have an effect on the qualities of ST, and how this relationship manifests in the practice of SP.**

The interviews offer a focal point for developing my investigation's line of inquiry. Next, it is important to understand how the existing scholarship explores the relationship between ST and SP. Given that a key tenet of SP is that its value lies in altering practitioners' ST in some beneficial way through un-biasing mechanisms, understandably, the bulk of the empirical studies are largely concerned with detecting and measuring biased reasoning and decision-making at some or all stages of SP. A review will help determine what was studied, which methods were used, what conclusions were reached, and what kind of bias(es) could support an empirical investigation into the relationship between ST and SP, using a threat-level focus, to help answer the central question.

4.2 Systematic Review of Empirical Studies Focused on Heuristics and Biases in Scenario Planning

4.2.1 *Rationale*

Little is known of the relationship between ST and SP. The corpus of scholarly literature promotes the assumption that SP alters the practitioner's abilities to perceive the world around them in a way they were unable to achieve before (Aligica, 2005; Balarezo & Nielsen, 2017; Godet, 1987; Kahn & Weiner, 1967; Schoemaker, 1995; Spaniol & Rowland, 2018b; Wack, 1985a, 1985b; Wright & Goodwin, 2009). Generally this assumption is extrapolated towards perceptual changes for the better, and particularly about the organisation's future (Cairns & Wright, 2018; Chermack, 2011; Tetlock, 2005; Van der Heijden, *et al*., 2002). However, as many have admitted, the empirical foundation necessary to support these claims is still building.

The dearth of evidence appears to be the product of two overarching issues. First, there are only a few empirical studies that attempt to measure a causal relationship between ST and SP. Chermack (2018) recently reviewed the SP scholarship, in part, to answer similar questions. Though his article is limited to publications between the years 1995-2016, this review encompasses the bulk of SP literature. Chermack's review identifies 17 peer reviewed articles that use survey research and statistical analyses. In a field that has been active and developing for more than 70 years, where a sampling of the most prolific 20 years results in only 7% of the scholarship reflecting empirically-backed, peer reviewed publications, it is safe to say that the field is lacking sufficient empirical support. For SP to be called a 'science', science must be practiced.

As well, for SP 'theory' to be entertained in the literature, the field must, necessarily, establish a theory. The second reason SP is lacking in evidential support can be traced to the field's emerging theoretical framework (or lack thereof depending on who you are reading). As true today as in decades past, are the numerous and ever-emerging practices and schools of SP. This variety is as responsible for expanding the field of SP as it is for introducing obstacles. Some feel SP lacks sophistication, and possibly forever consigned to the realm of chaos, incapable of being clearly defined in any manner. Godet (1990, p. 199) highlights the practice-based popularity of SP as the key factor to the limited theoretical development, "theoretical research and sophisticated tools have been neglected in favour of multiple applications." Nearly 20 years later, Bradfield (2008) makes a similar claim, attributing the confusion of the field to the lack of a solid, theoretically based foundation underpinning the techniques. Spaniol & Rowland (2018b) recently propose that even the *discussions* of theory and practice in the field are paradoxical. Their argument reasons that to attempt order in chaos, one must first acknowledge the chaos exists, but to acknowledge that chaos exists negates any attempts at sense-making.

Not all agree with this perspective. Others feel the necessary 'chaos' of any new field's embryonic stages is steadily leading towards clearer formulations of the boundaries. Crawford (2019) argues the chaos perspective deserves a reinterpretation, where the field has produced a richness in theory (and method), and where many discuss potential components that could contribute to a unifying foundational theory. The shift in perspective is akin to seeing 100 scattered toothpicks on the floor as a mess vs understanding there is an open box on the counter that can hold 100 toothpicks. Chermack (2011) manages to take these fragments of the field and integrate them into arguably the most comprehensive theoretical foundation for SP, to date. Chermack's *Scenario Planning Theory* is comprised of six domains: dialogue, learning, mental models, decision making, leadership, and organisation performance/change theories. His reasoning is that more than any other domains, these six have the highest repetitious mention and use within the SP literature. Chermack's proposed theory is still quite recent, though, and understandably requires a fair bit of empirical work for support. Opinions are shifting, as of late; The field is rich with theoretical development, but that development is only beginning to find solid ground.

The lack of a strong theoretical base coupled with few investigations into SP efficacy has led to a practice that is, at times, difficult to discuss. In an effort to help build more concrete understandings, the second section of this chapter offers a review of empirical investigations that have attempted to measure the relationship between SP and ST. Specifically, researchers claim that SP helps practitioners avoid the typical pitfalls of biased perspectives that come from employing heuristics while thinking about the future. Though the practice of experimental investigations into the causal relationship between SP and ST is, arguably, still in its infancy, it is important to carry out a review at this time to identify gaps in the literature and help direct future research efforts.

4.2.1.1 Participants

Traditionally, experimental studies are carried out within universities and therefore use students as participants (Arnett, 2008). Often, the students are undergraduates – between 18-25 years old. However, case studies and quasi-experimental models are known to use convenience samples or non-randomized methods for obtaining their participants. These studies can include any number of demographics: experts, managers, CEOs, patients, etc. One argument is that different sample demographics will lead to different outcomes, within the same method. An aggregate of forecasted probabilities from undergraduate students may be significantly different from an expert group’s aggregate probabilities on the same forecasting task. This is the conclusion Wright & Goodwin (2002) reach in their comparison of responses from novice (undergraduate students) and expert (MBA managers) samples on a framed choice task regarding business practices. This review will address the particulars of sample demographics and how they may relate to research findings.

4.2.1.2 Intervention

There are claims that SP can counter certain biased decision-making that would otherwise arise without the method (Balarezo & Nielsen, 2017; Cairns & Wright, 2018; Godet, 1987; Wack, 1985a, 1985b). These claims are divided into two main categories. One category of claims suggests a partial method of SP will counter biased decision-making. The other category of claims suggest the full SP method is required to counter biased decision-making. Further complicating this division, as stated earlier, there are arguably as many versions of SP as there are practitioners (Chermack, 2018). Each study's method will be compared and contrasted.

4.2.1.3 Comparisons

Some experiments compared SP with a different strategic planning method. Though few, this chapter will discuss the outcomes of these comparative studies. Implications on how to carry out comparative experiments against SP will be further discussed in the Discussion section.

4.2.1.4 Outcomes

This is one of the most difficult features to explore. With no unifying theory, and no agreed up on methodological approach, interpreting the possible effects of SP can be, as some claim, impossible (Spaniol & Rowland, 2018b). However, this issue is precisely why a systematic review is necessary. By featuring the use of similar terminology against incongruent methods, this review may shed light on why, with such few scholarly empirical studies, there exists such dramatic differences in certain outcomes. Uniform methods and definitions are used to conduct a systematic review of empirical studies. Due to the nature of the data, a meta-analysis is not possible.

4.2.2 *Method*

Guidance published by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), as determined by the Campbell Collaboration, is used to develop the method for this review (Moher, et al., 2009).

4.2.2.1 Eligibility Criteria

All documents published in a peer reviewed academic journal, as a doctoral thesis, or in pre-print are permitted in this review. Books must be published through an academic publishing company to be included, with referenced empirical studies from peer-reviewed journals. Peer reviewed journal publications preserve the highest industry standard for scientific work. The peer review process is the last stage in the scientific process and acts as a quality control. There are several variables to consider in the peer review process (e.g. double vs single blind, depth of evaluation, number of referees, potential bias in the approval process), but it is generally agreed that the peer review process employs higher standards for ensuring good quality scientific research is published, and arguably more importantly, poor quality research is not. Including doctoral theses allows this review to broaden the search while preserving the higher standards of the peer reviewed process. The inclusion of pre-prints affords the review an option to see manuscripts that may soon be published, but escape the initial, more limited search of existing and listed publications. However, pre-print discoveries are only included in the analyses if they are published by the time of writing.

Publications must use a method of SP that follows one of the three main models: IL, PMT, or *La Prospective* (Bradfield, Wright, Burt, & Van der Heijden, 2005; Godet, 1987; Gordon, 1994a, 1994b; Huss & Honton, 1987). The IL model defines SP as "a device for ordering one's perceptions about alternative environments in which one's decisions might be played out" and follows an eight-step approach (Huss & Honton, 1987, p. 22). The PMT model defines SP as "a forecasting method that permits extrapolations of historical trends to be modified in view of expectations about future events" (Gordon, 1994b, p. 1) and "an analytical approach to the probabilities of an item in a forecasted set" (Gordon, 1994a, p. 3), for the respective methods, with prescribed steps. The *La Prospective* model defines SP as "trying to consider many unknowns in the most objective manner possible… as part of a collective futures-thinking exercise in which structured thoughts and a common language are needed" (Godet, 2000a, p. 6). Publications must also match a basic definition of scenarios. Spaniol & Rowlands (2018, p. 1) offer an operant or synthesized definition of scenarios from the extant literature,

> scenarios have a temporal property rooted in the future and reference external forces in that context; scenarios should be possible and

plausible while taking the proper form of a story or narrative description; and that scenarios exist in sets that are systematically prepared to coexist as meaningful alternatives to one another.

Publications must be empirical studies with an independent variable (IV), a SP-based manipulation, at least one experimental condition, and a quantitatively measured dependent variable (DV). Qualitative data are also permitted, but quantitative data must be the primary focus in the publication. Empirical work is defined as either experimental or quasi-experimental design. Multiple studies in a single publication are also permitted. A desired standard in quantitative experimental methodology is the inclusion of a comparative control sample. However, it is not always possible to conduct a control sample against an experimental sample. Due to the novelty of experimental studies in the SP field, this criterion is left open. Publications, therefore, are not limited to only those that include a comparative control sample.

The experimental method must measure a behaviour which reveals a potential cognitive bias, based on the psychological definition of cognitive biases. According to the collective efforts of research around the world (from novice to expert), there are currently just over 100 recognised cognitive biases. However, the scholarship of cognitive biases is ever-changing, and no existing list is considered to be exhaustive. Therefore, the criteria are operationally defined in the broadest terms. Cognitive biases are systematic patterns of deviation from the norm, rationality in judgment, or any alteration from a person's previous decision-making state. Additionally, the experimental literature is also treated in a manner that allows further expansion of the present definition.

Measurements of effect include both implicit and explicit measures. Implicit measures include observational data and changes in estimates. Explicit measures are surveys, confidence scores, likelihood ratings, and interviews. Experiments include laboratory and field settings.

4.2.2.2 Search Criteria

To begin, a review is conducted to determine what specific biases the field recognises and whether empirical work has been conducted linking SP and cognitive biases. Primary terms “scenario planning”, “scenario thinking”, “futures thinking” are individually searched. Added to each of these phrases are the secondary terms/phrases “experiment”, “biases”, “heuristics”, and “mental short cuts”. The literature search was conducted between the dates of August 2014 to May 2018 and original publication dates were not restricted when searching through the repositories.

The repositories Scopus, WOS, EBSCO, ProQuest, Emerald, JSTOR, JURN, Open Science Framework (OSF), Google Scholar, Research Gate (R^G), HSTalks, Google books, and the University of Strathclyde Andersonian Library are searched. Several leading repositories and search engines are sourced to help ensure the broadest search possible, in order to capture the widest selection of publications.

All available languages are searched within each repository. SP is a strategy practiced in organisations across the globe. It is taught in higher education in multiple countries, though mostly in the west at this time. One of the leading schools of SP was birthed in France, *La Prospective.* Therefore, it is conceivable that published studies could be in multiple languages. Google translate and native speakers are used to help facilitate the multi-lingual searches.

The following is an example of the search strategies used. In Scopus, the phrase “scenario planning” (with quotation marks) was entered into the *Documents* field. The search options were set to search *All fields* within all the available documents. The *Published* date range minimum was set to *All years* and maximum to *Present*. The *Document type* was set to *ALL*. The *Access type* was set to *ALL*. The initial search resulted in 6,122 documents. The term “bias” was searched for within the documents, to narrow search results. This reduced the number of results to 452. Next, the term “experiment” was searched for within the narrower list of documents. This reduced the number of results to 109. This list was downloaded to a spreadsheet that compiled the master list of documents. The master list was built from the search results of all databases using all primary and secondary search terms/phrases.

4.2.2.3 Data Criteria

The lack of a common conceptual framework used in the SP literature and missing data from some of the documents that comprise the empirical studies prevented the researcher from conducting a meta-analysis. Instead, quantitative information is provided where possible, supported by a qualitative review.

4.2.2.4 Risk of Bias

There are some risks for biased information at the individual study level and across the studies. These potential biases will be reviewed in the next section and discussed in the last section of this chapter.

At the individual level, some studies may have only experimental samples without a control sample. Studies without a comparative control sample will be sub-categorised and discussed in the next section. There may be differences at the participant level due to sampling bias. The sample demographics will be compared and discussed to help reveal any potential biases.

Across studies, there may be a sampling bias (expert vs novice), measurement bias (standard or novel), or a testing bias (process or tool). Possibly, the most important bias to measure across the studies is a publication bias, which is the effort of publishing only significant, positive results. This can skew the understood efficacy of SP and therefore will be measured in this review.

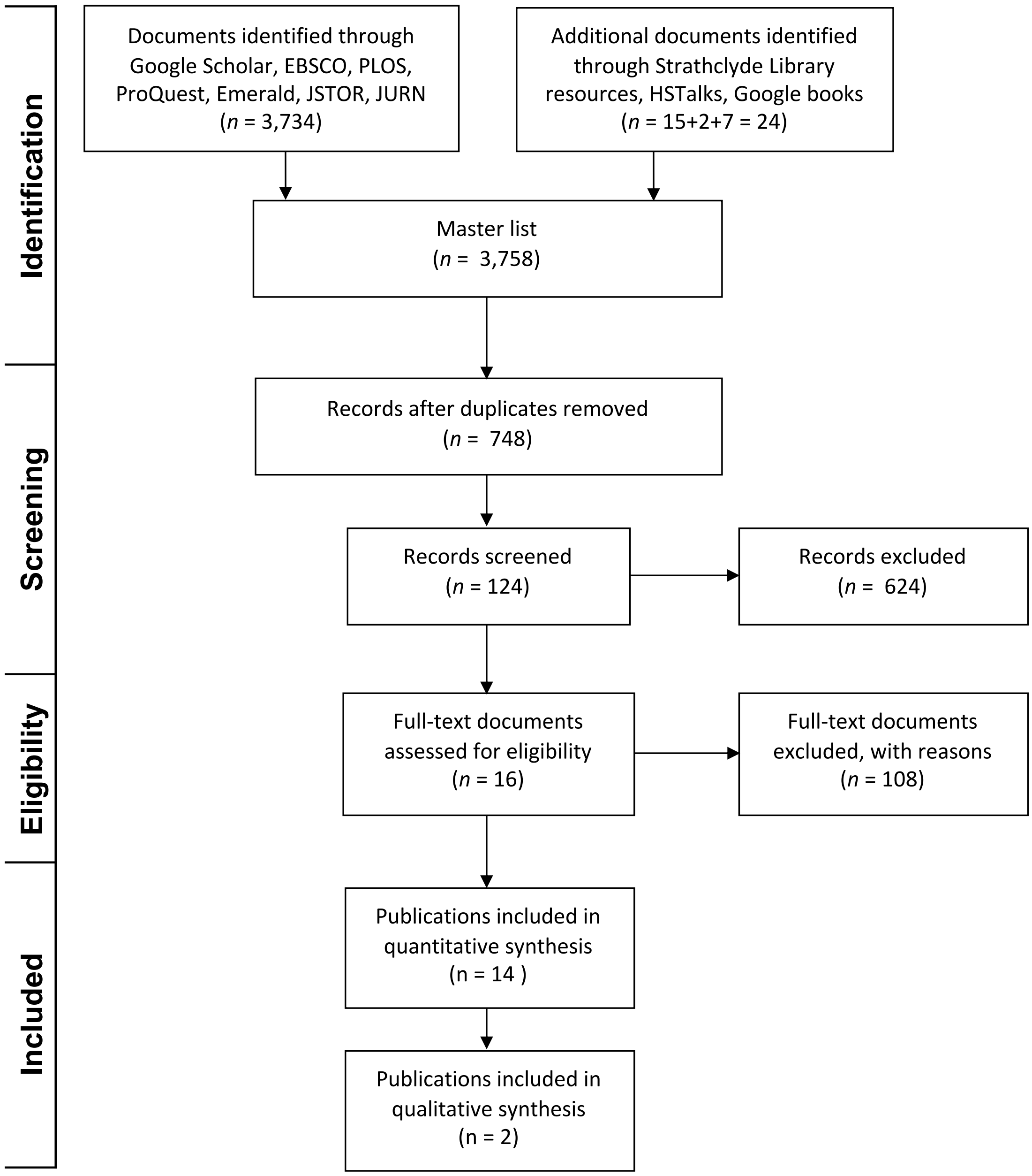
4.2.3 *Results*

The areas of comparison and measure across the resulting documents is informed by the publication standards of empirical research and the American Psychological Association (APA). The main categories are participant sample, method, experimental design, IV, DV, and assessments. Additional categories are also included which are important to this specific review: scenario qualities and focal biases and/or heuristics. Missing categories will also be reviewed.

4.2.3.1 Study Selection

Primary phrase searches and combinations with secondary phrases were performed in Scopus, WOS, EBSCO, ProQuest, Emerald, JSTOR, JURN, Google Scholar, and R^G which returned 3,734 documents in total. The Andersonian Library, Google books, and HSTalks returned 24 documents. The searches produced a master list of 3,758 documents. Using R software, the documents were compared by "Title", "Year", and "Document Type", to locate duplicates. After several iterations of different comparison protocols (e.g. "Title" only, "Affiliation publication" only, "Title" and "Year" together), the grouped selection criteria of "Title", "Year", and "Document Type" proved to be the best protocol. No duplicates were left in the master list and no documents were accidentally removed due to similar selection criteria. After duplicates were removed, the master list shortened to 748 unique documents. All documents that were not articles, books, book chapters, and doctoral theses were removed. This left 124 unique documents. Abstracts were first reviewed for matching eligibility criteria. It is understood that abstracts do not give all the information and therefore cannot be considered a final effective search effort. Abstracts were first read to help quickly sort through the master list and categorise the publications by "maybe" and "definitely not". Those that met some or all of the criteria were triaged to a sub-list for second review. Those that did not meet the criteria were reviewed in full. If the full text revealed that the publication still did not meet all eligibility criteria, it was eliminated from the master list. A full review of the body of the remaining master list resulted in 16 documents: seven experimental articles, six quasi-experimental articles, two case studies, and one book with a collection of four experimental studies. The case studies are pertinent to this search, but are primarily qualitative studies, and therefore do not meet one of the eligibility criteria. However, due to their investigations and the underlying motivations of this review, they are included in the review. Their inclusion is highlighted throughout, to ensure the quantitative and qualitative study differences are understood. Overall, this is a small number of empirical studies. Figure 4.1. presents the flow-chart of search and selection.

Figure 4.1. Flow-chart



4.2.3.2 Publication Characteristics

Publication dates of original work range between 1987-2016, with 52% published by T. J. Chermack and P. E. Tetlock, serving either first or subsequent author, though never published together. Chermack is authored on all six quasi-experimental articles. The distribution of publication years is skewed to the left (earlier years), meaning the majority (81%) of the empirical research was carried out and published in the latter half of the time span, specifically after 2001. This helps reveal potentially an increase in the interest of empirical evidence, specifically in the area of ST, by way of cognitive biases, as the years progress. Table 4.3. presents the journals in which each study is published, the years of publication, impact factor in 2018, and year of first published volume. A journal's impact factor is a measure of the frequency with which the average article in a journal has been cited in a particular year. It is a popular metric for reputation and quality of research accepted into the journal. There are a number of other performance indicators that could be used to assess the reputation of a journal. Each indicator comes with its own strengths and weaknesses in relaying a journal's performance. However, the impact factor is chosen because 1) it is used as an industry standard in assessing publication ranking, 2) it is the most widely accessible measure, and 3) the purpose of including this measure is to give a brief snapshot of the history of published research in this area, since the main purpose of this review is the content of the published research. Included in the table are the two qualitative studies, which are italicized.

The journals cover the topics of futures/forecasting (4), management (2), human resources (2), decision making (1), and accounting (1). Only *Strategic Management Journal* and *Technological Forecasting & Social Change* published further SP cognitive studies after a different journal published the next chronological article. We cannot know exactly why this is the case, but some assumptions are tested. It could be coincidence, a desire by authors to diversify publications, reluctance on editors and/or referees to accept this area of research in their respective fields (therefore artificially lowering the number of available studies), professional affiliations between journals and authors, or possibly a reflection of emerging journals and popularity.

Table 4.3. Scenario planning publishers

Journal	Year of article	Impact factor	1st volume
International Journal of Forecasting	1987	3.387	1985
Strategic Management Journal	1993 2014	5.572	1980
Journal of Behavioral Decision Making	1996	1.791	1988
The Accounting Review	2002	2.319	1926
Princeton University Press*	2005	-	-
Advances in Developing Human Resources	2008	0.575	1999
Human Resource Development Quarterly	2008 2012	3.000	1990
European Journal of Training and Development (fka Journal of European Industrial Training)	2012	1.370	2012 (1977)
Technological Forecasting & Social Change	2007 2013 2013 2013	3.815	1970
Journal of Futures Studies	2014 2016	0.780	2004

* Book publisher

On the issue of potential editorial reluctance, volume 80, issue 4 of *Technological Forecasting & Social Change* (2013) was a special issue dedicated to "Scenario Method: Current developments in theory and practice". The editors' efforts resulted in acceptance of three empirical articles that focused specifically on the cognitive effects as related to SP. Unfortunately, after this special issue, only two other articles were published, and both within the same junior journal.

Understandably, there is an inverse relationship between the maturity of the journal and its impact factor ($r_s(10) = -.612, p = .03$). The older the journal, the higher its impact factor. Since impact factors cannot be accessed for the years of each article's publication, this inverse relationship is used as a baseline measure to help determine whether publications of articles followed with the emergence of new journals. If publications follow the trend of emerging journals, then there should be a positive correlation, but if publication is favoured by more established journals, then there should be a negative correlation. Neither relationship exists with either impact factor or maturity of journal. Given that two authors dominate this arena, it seems unlikely that coincidence is the answer, though this cannot be known with the available data. This leaves the assumptions that possibly authors desired to diversify their submissions, editors were reluctant to publish this path of empirical investigations in their respective fields, and/or the authors have affiliations with the publishers. Whatever the motivations may have been, the result is a spread of published empirical work in five different disciplines across 9 different journals.

4.2.3.3 Article Characteristics

Articles report single (n = 12) or multiple (n = 4) studies. Within the 16 publications, there are 16 experimental studies, six quasi-experimental studies, and four case studies. The result is 26 separate studies with 33 different conditions that make up the body of this review.

4.2.3.4 Participant Demographics

The studies use both novices and experts as their participants. Novices are categorised as university students – undergraduates, MBAs, post-graduates and doctoral candidates. Experts are categorised as employees, business owners, and CEOs within their respective fields of experience, outside the academic arena. This review divides the remaining data along participant lines. Comparing the studies across all publications, the proportion of studies that use either novices (.52) or experts (.48) are almost equal. Possibly due to convenience of accessing novice participants, all but one article that offer multiple comparative experiments use novice samples. All publications compare similar participant samples across experimental conditions, except Sedor (2002), who uses experts in her first study, but MBA students in her follow-up study.

Not all studies report their sample sizes, one study reports ranges for their samples and another study reports a proxy sample measure. Bradfield (2008) reports four workshop groups (i.e. syndicates) that consist of five or six students. This gives a total sample range of n = 20-24 with an average group size range of M = 5-6. Ram & Montibeller (2013), in their three case studies, report only that they work with three different companies. They do not specify the size of each SP group, nor how many participants are involved in the different stages of the study. Gaps in information make it difficult to accurately report sample demographics of all studies. To account for these gaps, the average is taken from the Bradfield study and the Ram & Montibeller is not included. Table 4.4 reports the descriptive statistics for participant samples. When the individual-based studies are divided by novice and expert samples, a histogram reveals that both sample demographics result in the most studies (n = 6) using a near identical sample size range of n = 17-25 and 16-24, respectively. Ranges and standard deviations (SD) for both by novice and expert samples, however, signal that there is a wide difference in participant sample sizes across studies.

Table 4.4. Sample size demographics for individual- and group-based studies

		Novice (*n*)	**Expert* (*n*)**
Individual			
	Median	34	28
	M	41	42
	SD	25.50	43.12
	Range	100	160
	Total conditions	*28*	*23*
Group			
	Median	4	12
	M	5	12
	SD	3.40	2.52
	Range	7	5
	Reported studies	*4 of 4*	*3 of 11*

Note: The Total conditions row reports the total number of conditions for each sample demographic. The Reported studies row reports the number of studies that report sample group sizes out of the total number of studies that report using a group method. *Missing sample data from eight studies

When *SD* is larger than the mean (*M*), this indicates that the data have strong positive skew. The majority of studies have sample sizes less than their *M*; However, *M* is especially sensitive to skew, therefore median scores, which are less sensitive, are offered as well, to give a more accurate impression of sample size popularity. The reported group sizes tell a slightly different story. With studies using novice samples, the groups are nearly 2/3 smaller than those using expert samples.

Also revealed in this initial stage is that articles using novice samples give more complete reporting efforts. Cross-referencing the methods with samples reveals a potentially strong explanation for this standard. All of the studies that report laboratory settings use novice samples (e.g. classrooms and campus spaces), and *potentially* all of the studies that use expert samples, carry out their experiments in the field (e.g. office spaces within an organisation). Tetlock's (2005, p. 269) third study only reports that participants are recruited from "the membership lists of Divisions 18 and 19 (International Conflict; International Security and Arms Control) of the APSA and the Society for Historieans of American Foreign Relations," but not where they are tested. Schnaars & Topol (1987, p. 411) report that participants are randomly assigned to one of four possible conditions and that a" cursory analysis revealed no difference in forecasting accuracy among these three groups. Thus, the results have been combined," but do not specify if the sample of corporate planners were subject to a different experimental setting than the student samples. Therefore both of these studies' locations are inconclusive. One explanation for this difference may be that even though all peer-reviewed, published empirical articles must obtain some form of ethical approval from their associated higher education institution(s), researchers working specifically within these institutions, with populations also from within the same granting institution are more likely to offer increased transparency in academic reporting. However, this division in location presents a potential bias in results. Further to this, there may be an interaction with sample type (novice vs expert) and location (lab vs field) that may further bias the results.

4.2.3.5 Experimental Conditions

A major difference within the SP process is the presence or absence of a facilitator. The fact that most (arguably all) SP workshops conducted within organisations are facilitated – sometimes by more than one facilitator – it is a vital part of this review to compare the discoveries between facilitated and unfacilitated empirical studies. Closely related to facilitation are the experimental methods requiring participants to either generate their own scenarios or review already developed scenarios. This is an important design feature to explore because it resembles the two main interactions executives and management have with SP. There are those who attend the workshop(s) and actively develop the scenarios, and those – often in more senior positions within the organisation – who are presented with the fully developed scenarios as aids for reaching a decision. It is also not unreasonable to assume that if a study's method tasks participants with generating scenarios during their session, that the participants would require some form of facilitated guidance through the multi-stage decision process. However, this assumption proves to be wrong.

The empirical studies reveal that 73% (n = 24) of the conditions require participants to generate their own scenarios (including both qualitative articles, see Table 4.5.). Only two of these conditions provide no active facilitation for the participants as they generate their scenarios. Both conditions are in Schoemaker's (1993) highly cited article, who provides his novice samples with a single page of instructions. Meissner & Wulf's (2013) study is the only other method that also asked novice samples to generate their own scenarios, however, the authors provide a facilitator for the experimental sessions. The majority of the conditions, which span 14 publications, use a scenario 'generated' method. Along with this, the majority of these conditions (n = 20, 83%) use expert samples. This reveals that the majority of studies that use a scenario 'generated' method are also largely facilitated, tested against an expert sample, and conducted in the field. The remaining conditions that use a scenario 'reviewed' method also mostly use novice samples (n = 19, 73%), all but three are not facilitated, and all test within a laboratory setting. Studies that require generating scenarios treat SP as a process whereas those who use a review method treat scenarios as a tool. This is an important distinction to understand because as a discipline, practitioners and facilitators need to be aware whether there are measurable differences in the efficacy of SP that is dependent on active participation.

Table 4.5 also reveals another important feature of the study methods. The last column reports the length of time participants were involved with a single study. Not all studies report this data. Sedor (2002) only mentions mailing or emailing materials to participants, but not how long they spend on the assigned task. Sedor admits that her study sacrificed some experimental control by using this method. In similar fashion, Tetlock (2005) also discusses delivery of the materials, but not how long participants are occupied. Haeffner, et al. (2012) offer a little more information. The first phase gathers data from participants about 2 weeks before their SP workshops. The second phase occurs anywhere between the last workshop and up to two weeks after the last workshop. However, workshop lengths and frequency are not given. Of the session times that are reported, the studies that test within a laboratory – which are almost exclusively novice samples (Schnaars & Topol,1987, being the only potential exception) are 83% shorter, on average, than those tested in the field (Mode_{lab} = 1 day, M_{lab} = 10.4 days, SD = 16.8, Range = 41; $\text{Mode}_{\text{field}}$ = 90 days M_{field} = 59.2 days, SD = 35.9, Range = 84).[13] The timing differences reveal a potential bias in the resulting data.

Novice participants in artificial laboratory settings, given shorter amounts of time to deliberate may reach different conclusions or rely on different cognitive biases than experts within their own organisation's setting, given 5.7 times longer periods of time to deliberate. The modal response gives a more accurate picture of the differences.

---

[13] Tetlock's (2005) five-year session time was removed from the averages due to being an outlier and skewing the results to give an inaccurate picture of field studies. The mode reports all sessions that were ≤ 1 day (1 day, 2 hours, and 45 minutes). However, even if the sessions that were shorter than 1 day were not rounded up, the mode would still be 1 day.

All empirical studies follow some form of the IL model of SP. Two studies, one being a case study, also include methods from the PMT model (Schoemaker, 1995; Ram & Montibeller, 2013).

Table 4.5. Experimental conditions for all studies

Document	Sample	Facilitated	Interaction	Location	Work	Conditions/Control	Session time
Experiments							
(Schoemaker, 1995)	Novice	No	Generated	Lab	Group	1/0	6 weeks
	Novice	No	Reviewed	Lab	Group	1/0	6 weeks
	Novice	No	Generated	Lab	Individual	1/0	1 week
	Novice	No	Reviewed	Lab	Individual	1/0	1 day
	Novice	No	Generated	Lab	Individual	1/0	1 week
(Kuhn & Sniezek, 1996)	Novice	No	Reviewed	Lab	Individual	4/1	45 mins
(Önkal, Sayim, & Gönül, 2013)	Novice	No	Reviewed	Lab	Individual	3/1	2 hours
(Meissner & Wulf, 2013)	Novice	Yes	Generated	Lab	Group	3/1	1 day
(Bradfield R. M., 2008)	Novice	Yes	Generated	Lab	Group	1/0	1 day
(Schnaars & Topol, 1987)	Novice & Expert	No	Reviewed	-	Individual	2/2	1 day
(Phadnis, Caplice, Sheffi, & Singh, 2014)	Expert	Yes	Reviewed	Field	Group	2/0	4-9 days
	Expert	Yes	Reviewed	Field	Group	1/0	4-9 days
(Sedor, 2002)	Expert	No	Reviewed	Field	Individual	2/2	-
	Novice	No	Reviewed	Lab	Individual	2/1	-
(Tetlock, 2005)	Expert	Yes	Generated	Field	Group	2/0	5 weeks
	Expert	Yes	Generated	Field	Group	2/0	5 weeks
	Expert	Yes	Generated	Field	Individual	2/0	5 years
	Expert	Yes	Generated	-	Individual	2/0	-
Quasi-experiments							
(Haeffner, Leone, Coons, & Chermack, 2012)	Expert	Yes	Generated	Field	Group	1/0	-
(Bodin, Chermack, & Coons, 2016)	Expert	Yes	Generated	Field	Individual	1/1	3 workshops
(Chermack & Nimon, 2008)	Expert	Yes	Generated	Field	Group	1/1	3 months
(Hawkins & Chermack, 2014)	Expert	Yes	Generated	Field	Group	1/1	3 months
(Chermack, Van der Merwe, & Lynham, 2007)	Expert	Yes	Generated	Field	Group	1/0	3 months
(Glick, Chermack, Luckel, & Gauck, 2012)	Expert	Yes	Generated	Field	Individual	1/0	12-14 weeks
Case study							
(Ram & Montibeller, 2013)	Expert	Yes	Generated	Field	Group	3/0	2 months

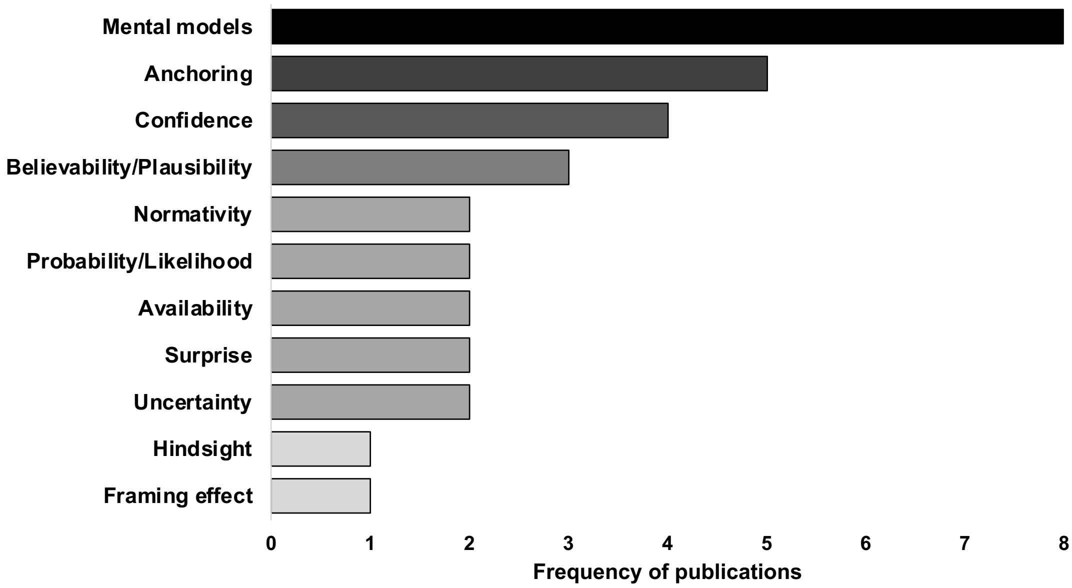
Note: "-" means not specified. Qualitative studies are italicized. Tetlock's (2005) samples are divided between "experts" and "dilettantes". Both samples, however, are professionals, not university students, and therefore both meet the requirements of this chapter's definition of "expert".

4.2.3.6 Cognitive Measures

The publications measure, either directly or indirectly, effects on 11 separate cognitive features. Interestingly, even though all studies test and measure various features of ST, only three publications explicitly name “scenario thinking” as a feature of their experimental studies. Sedor (2002, p. 731) defines ST as “envisioning a sequence of events in which proposed actions lead to future outcomes.” Bodin, et al. (2016, p. 35) defines more of the outcomes of ST, “working through multiple complex and divergent future possibilities”. Kuhn & Sniezek (1996, p. 234) offer an interpretation of Schnaars & Topol’s (1987) conclusions of ST as a bias, itself, “that promotes overconfidence”. However, Schnaars & Topol’s article never directly discusses ST, nor thinking at all. Their article focuses largely on the experience of surprise post-SP. Most of the authors conflate ST with strategic or future-focused thinking, and in one instance “thinking with scenarios” (Önkal, Sayim, & Gönül, 2013, p. 774).

Standardised and novel scales are used to gather the data. Where novel scales are used, content and construct validity reports are also provided (Bodin, Chermack, & Coons, 2016; Chermack & Nimon, 2008; Chermack, Van der Merwe, & Lynham, 2006; Glick, et al., 2012; Haeffner, et al., 2012). Figure 4.2. shows the popularity of each cognitive feature across the publications, ranked from highest to lowest. A wide-reaching attempt is dedicated to measuring changes in mental models of the participants. Almost all corresponding data come from quasi-experimental studies. The most popular cognitive bias that is measured is the anchoring bias and this shows to be potentially correlated with the availability bias and possibly overlapped with the framing effect. Confidence, surprise, and uncertainty all show close relation to one another across the various studies, and are measured in a total of 23 studies, the most common data provided across all publications. As well, believability, plausibility and normativity all show to be correlated, at least superficially.

Figure 4.2. Popularity of cognitive features



Note: Fourteen publications measure more than one bias.

Eight publications, covering 10 studies from one experiment, six quasi-experiments, and three case studies, present a collection of measures that reflect various qualities of mental models (Bodin, Chermack, & Coons, 2016; Chermack & Nimon, 2008; Chermack, Van der Merwe, & Lynham, 2006; Glick, Chermack, Luckel, & Gauck, 2012; Haeffner, Leone, Coons, & Chermack, 2012; Hawkins & Chermack, 2014; Meissner & Wulf, 2013; Ram & Montibeller, 2013). All but one study uses experts, and all use a facilitated, scenario 'generated' method. Eight self-rated surveys assess mental models. The studies are divided across three main themes:

(1) how participants perceive their own decision-making abilities
(2) how they perceive their organisation
(3) how they perceive the SP method

The first theme is assessed through three scales. The General Decision-Making Style Survey (GDMS) is used in two studies and measures five factors (Bodin, Chermack, & Coons, 2016; Chermack & Nimon, 2008). Both samples report an increase in intuitive thinking and a decrease in rational, avoidant, and spontaneous thinking after a SP workshop. The GDMS samples differ on dependent (need for other people's aid) decision-making. Chermack & Nimon (2008) report no change in dependent decision-making, while Bodin, et al. (2016) later report an increase in this same factor. The opposing responses could be due to a number differences between the studies. The two different expert samples are tested almost a decade apart, within different organisations, running different SP workshops (though both following the same basic outline from Chermack's method), with different facilitators, assessing different goals. Any or all of these variables could be influential factors in the difference in perceptions of dependent decision-making. However, closely related to the dependent measure of the GDMS are the two factors measured in the Conversation Quality and Engagement Checklist (CQEC; Chermack, Van der Merwe, & Lynham, 2006). The CQEC measures individual and interaction conversation skills. Both measures increased post-workshop. If interpersonal conversation, communication, and engagement all increase as a product of SP, then it would align that dependent decision-making (need for the aid of other people in the decision situation) would increase as well. Meissner & Wulf (2013) use a three-part decision quality questionnaire designed from Amason's (1996) previous work, to measure factors along the first theme. Novices in the full SP condition report increases in overall decision quality, post-workshop (i.e. accuracy and confidence proxy measures). Interestingly, Schnaars & Topol's (1987) experimental study shows that accuracy is not affected after reviewing multiple scenarios. However, it is important to note, that amongst the differences between both studies, Schnaars & Topol's method required both experts and novices to work individually, whereas Meissner & Wulf's method included group work.

The second theme is also assessed through three different scales. The Mental Model Style Survey (MMSS) measures five factors regarding experts' views of how their organisation generally operates (Glick, Chermack, Luckel, & Gauck, 2012). Post-workshop, experts perceive their organisation as more efficient (top-down, management by objective, and mechanistic), social (a collectivity to which employees belong), and as a system (a series of inputs, processes, and outputs), while less political (single power relations which determine decisions). The drop in political measure may be related, even if superficially, to the increases in perceptions of dependent decision-making, interpersonal conversation, communication, and engagement skills seen in the previous studies. SP appears to have no effect, however, on experts' financial perceptions of their organisation, where the considerations for locus of control remain almost perfectly in the mid-range. The Dimensions of Learning Organization Questionnaire (DLOQ) measure seven factors, with a focus more on learning (Haeffner, Leone, Coons, & Chermack, 2012). Experts report an increase in all factors except continuous learning, which reflects no effect. Again, there is an increase in perceptions of dialogue, collaboration, and systems thinking regarding their organisation. The Gallup Workplace Audit (Q12) measures elements of the work situation and engagement conditions (Hawkins & Chermack, 2014). This scale shows no change in either direction to experts' perceptions of organisational engagement. The Q12 responses appear to contradict other scales by revealing not change in mental models along its factors.

The third theme breaks from the general path of this review, however the insights gained from Ram & Montibeller's (2013) three case studies give pertinent, supportive information in line with the other profiled studies. Experts are tasked with comparing their organisation's standard strategies and to reach predictions against an IL/PMT method of SP. All samples considered SP more challenging by way of engagement and conversation strategies, and more stimulating for idea generation. Most (2/3) of the samples perceived their SP methods to facilitate more intuitive and adaptive thinking and to be a more transparent process.

The mental model surveys reveal some common threads across the themes. Active scenario 'generated' methods promote shifts towards intuitive thinking, engagement with others, increased communication, and viewing the organisation from a systems perspective.

Five publications, covering 20 conditions, measure forecasting behaviours across various scenario conditions (Bradfield, 2008; Kuhn & Sniezek, 1996; Önkal, Sayim, & Gönül, 2013; Phadnis, Caplice, Sheffi, & Singh, 2014; Sedor, 2002). Four studies use a scenario ‘reviewed’ method and Bradfield’s study uses a scenario ‘generated’ method. Most studies sample novices over experts. Sedor (2002) asks experts and novices to give short-term forecasts. Önkal et al. (2013) require novices to generate point, best-case, and worst-case, short-term forecasts. Bradfield (2008) tasks groups of novices to explore the business environment on a medium-term scale to identify and causally evaluate the DF that will become the foundation for their scenarios. Kuhn & Sniezek (1996) ask novices to give five forecasts spanning medium- and long-term horizons. Phadnis, et al., (2014) task experts to evaluate multiple, long-term investment projections (including potentially their own). The quantitative studies report the same forecasting bias across conditions. Participants forecast directions that reflect the direction of the scenario content (best vs worst, increase vs decrease). Sedor reports that when scenarios declare plans to increase future earnings, especially if they also show prior losses, participants give even higher forecasts. In fact, participants appear to reflect the decision-making predictions of prospect theory. When losses are felt more than gains, people are more likely to take risks on future options. Önkal et al. show that best-case scenarios are followed by higher best-case forecasts and worst-case scenarios are followed by lower worst-case forecasts. Kuhn & Sniezek show that this pattern also increases in magnitude as a factor of time. Scenarios that present increasing patterns lead to higher initial forecasts that increase over time, and decreasing patterns in scenarios lead to lower intital forecasts that continue to decrease over time.

Interestingly enough, Önkal et al. (p. 783) suggest that one way to better explore potential anchoring biases would be to ask “participants to give their initial expectations prior to giving any forecast advice” (i.e. pre- and post-scenarios). This is precisely what the quantitative and qualitatives studies from Phadnis, et al. (2014) and Bradfield (2008) reveal, respectively. Phadnis, et al. participants produce a magnitude effect that potentially reflects discoveries from Bradfield’s earlier study. Experts become more supportive of their own extreme forecasts (favourable and unfavourable) after reviewing scenarios. In these pre\post, longitudinal studies, prior forecasts, and decisions are preserved and strengthened through the SP process. Bradfield explores the phenomena of belief perseverance, confirmation bias, experience bias, and overconfidence that may facilitate anchoring biases. Sedor's (2002, p. 738) design actually attempts to ameliorate potential anchoring biases by constructing the high and low time-series of earnings equivalently so "an anchoring-and-adjustment heuristic would affect [annual earnings per share] forecasts equally in all experimental conditions," but an anchoring bias appears within the optimistic forecasts in spite of the balanced design. A closer look, however, at the shared statements between Sedor's (p. 58) profit and loss scenario structures reveal that most of the statements reflect a positive, supportive, and growing business environment: "The strong economy and the publicity given to the health benefits of moderate wine consumption continue to fuel consumer demand for wine. In fact, industry-wide sales of red wines have more than doubled in the past few years." Therefore, it appears that Sedor's participants may be reflecting a similar anchoring bias as the other studies, and not, as first assumed, an optimism bias.

The quantitative measures that show the presence of potential anchoring biases may find support for the cognitive underpinnings through two scenario 'generated' studies, one qualitative, one quantitative (Bradfield, 2008; Haeffner, Leone, Coons, & Chermack, 2012). Bradfield's novice groups conduct exploratory searches of the business environment that result largely in causally linked DF that were founded on recent events "highly publicized in the media" at the time of the workshops (p. 208). Haeffner, et al. (2012) use the Dimensions of the Learning Organization Questionnaire (DLOQ) to measure seven related dimensions of organisational learning perceptions. Their study shows that the dimensions explicitly promoted within the SP workshops have strong effects on perceptions of those same dimensions. Though neither of these studies explicitely test for the availability bias, their results and concluding remarks discuss this strong potential around participant performances. What is not clear, however, in these studies, is whether availability is mediated by a primacy or recency effect.

The two biases of this section are not explicitly tested in any of the studies; However, their potential influences are noted after patterns in results emerge, and therefore warrant inclusion in the larger discussion. Three studies include explorative discussions on the influence of scenario believability and on the decision-making quality of their participants (Önkal, Sayim, & Gönül, 2013; Schoemaker, 1995). To begin, a basic tenet of SP is that scenarios should (or must) be plausible. This feature, however, may introduce its own biases into the decision-making process. It is inconclusive at this time whether plausibility and inflated believability lead to lower decision quality. Önkal et al. (2013) show that single reviewed scenarios are more believable and plausible than multiple differently focused scenarios, and specifically positive scenarios are perceived as more believable (e.g. clear to understand and realistic) than mixed or negative scenarios. Based on the biased behaviour of his novice sample, Schoemaker (1995) concludes that believability in scenarios increases certainty, but that believability may also counter the availability bias in some workshop environments. Specifically, requiring participants to generate opposing scenarios reduces the believability in either, and by proxy, reduces the easy cognitive effort of relying on plausibility via believability.

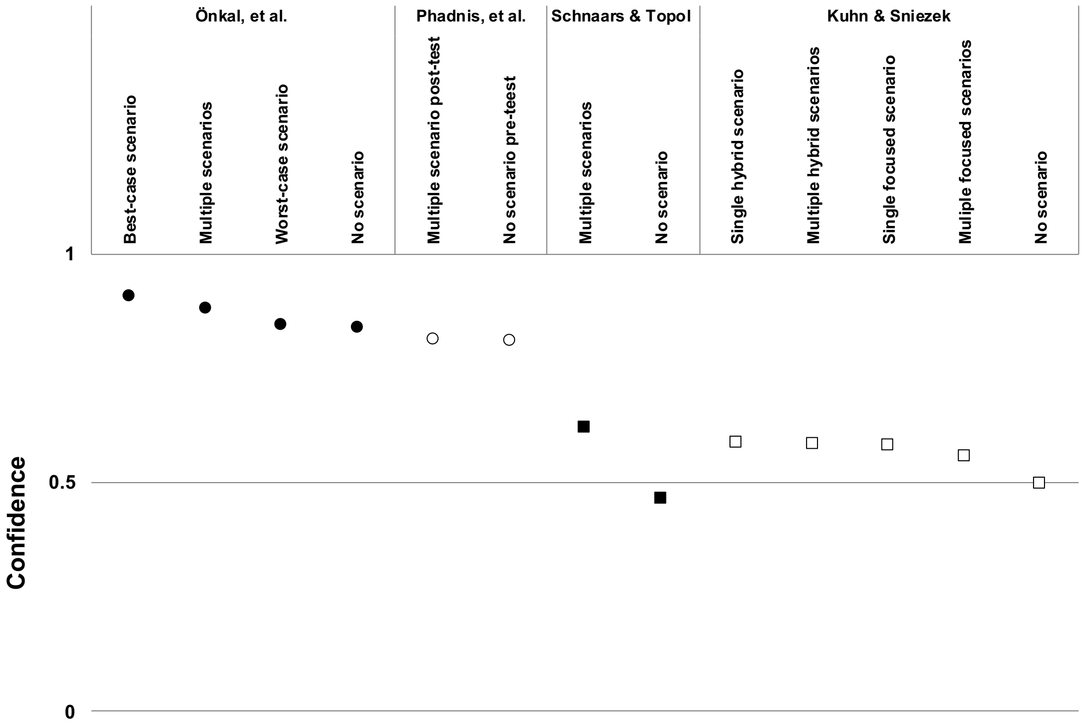
Two publications that cover three studies test the effect of SP on probability and likelihood judgments (Schoemaker, 1995; Tetlock, 2005). In his third study, Schoemaker (1995) finds regular use of the conjunction fallacy, where novices give greater probabilities to multiple causes of a single event in the near future (i.e. short-term) compared to a single cause for a single event, but to varying degrees of severity. Participants in the fourth study committ correlational violations by assuming more positive correlations between events than should be mathematically possible. Tetlock (2005) also finds decision-making violations in three workshops that use a short-term time scale. Experts in the first and second studies give higher likelihoods to plausible future extreme outcomes (improvement or deterioration of a country) after participating in SP workshops. Experts in the fourth study use a hindsight perspective to unpack counterfactual scenarios regarding alternative outcomes. Through the SP effort, participants offer greater probabilities for individually unpacked outcomes than the probability of the whole set of alternative outcomes. Tetlock attributes all three outcomes to the increased imaginability that the process of scenario generation provides. Both authors leave the question open whether these mathematical violations are necessarily a detriment to SP and ST or serve a benefit that has yet to be fully understood.

Four publications measure confidence across 16 conditions (Kuhn & Sniezek, 1996; Önkal, Sayim, & Gönül, 2013; Phadnis, et al., 2014; Schnaars & Topol, 1987). All conditions use a scenario 'reviewed' method, and the three publications that include a control condition, also all use novice samples. The Kuhn & Sniezek and Önkal, Sayim, & Gönül studies include a single 'review scenario' condition, which both report increased confidence scores. Kuhn & Sniezek and Schnaars & Topol show confidence as a function of time, where confidence decreases as distance to a future point increases. In the three studies with a control condition, confidence in forecasting accuracy is shown to increase after reviewing either single or multiple scenarios. The Phadnis, et al. studies stand apart from this trend, which show no change in confidence after reviewing multiple scenarios. However, change in confidence is only part of the story. When the scores are standardized and compared across studies, a shared trend becomes clear.

All confidence scores are standardised and transformed to fall between 0-1. Kuhn & Sniezek (1996) use a 9-point Likert scale to gather confidence scores for five separate decadal predicitons into the future, up to 50 years, from five different conditions. The five forecast confidence scores are averaged across each condition, then transformed. Schnaars & Topol (1987) use a similar simple scale (-3 to +3) to collect six confidence ratings up to 11 years, from four conditions, but they only report the mean confidence-rating for the aggregated experimental conditions and for the aggregated control conditions. Önkal et al. (2013) consider their Surprise Index (SI), which is the average value across all surprise probabilities (scale 0-100%), as a proxy measure of confidence (100-SI), and offer an SI score for each of their four conditions. Phadnis, et al., (2014) calculate confidence as the sum of the products of proportion of votes and average value of the confidence interval, which fall between 0.5-1. The standardized and transformed confidence values are presented in Figure 4.3. Publications are ranked, left to right, from highest to lowest confidence scores. Figure 4.3. shows that all control conditions ("No scenario") report lower average confidence scores compared to their paired experimental conditions. This includes the Phadnis, et al. study, where the pre-test condition can be considered a control condition since participants make predictions without the experience or aid of scenarios. Pretest scores are lower, on average, than the post-test scores. Control conditions are also the only conditions to average a confidence score lower than the midway point (0.5). Even though the Phadnis, et al. study does now allow for values lower than 0.5, the pre-test high scores may explain why their IV (multiple scenarios) results in a lower rise in confidence. It is a phenomenon mentioned in both Schoemaker's (1995) and Sedor's (2002) studies, which is that participants may have experienced a ceiling effect.

As stated earlier, the differences in sample demographics, testing location, and test timing may bias results from these studies. The single study that uses expert samples, also tests them in field settings and offers up to 3 days prior to the workshop to review and prepare information (Phadnis, et al., 2014). The workshop lasts for one day, which isn't as dramatically different from the other three laboratory studies (45 minutes, 2 hours, < 1 day). The experts who are given longer deliberation time and work within their own organisation give confidence scores that fall between the highest and lowest scoring novice studies. What Figure 4.3. helps reveal is that i) working with scenarios has a positive effect on perceptions of confidence, ii) reviewing scenarios appears to lead to increases in forecasting confidence and iii) methodological differences appear to affect confidence across scenario conditions more than the conditions within (e.g. single vs multiple vs none).

Figure 4.3. Confidence scores from empirical studies



Closely related to confidence is the measure of surprise. So close, in fact, that Önkal et al. (2013) consider confidence the inverse of the score from their SI. The authors measure surprise as an anticipatory factor. Schnaars, et al. (1987), on the other hand, measure surprise independently and as a phenomenon that occurs after declaring confidence and receiving feedback. Schnaars, et al., report that the level of surprise felt after discovering errors in their short- to medium-term forecasts is no lower for experts and novices who use multiple scenarios to help them determine their forecasts from those who do not. The main driver to increases in surprise is the level of stability in the market. The more unstable the market reveals to be, the more surprise participants are at the inaccuracy of their forecasts. Önkal et al. show that anticipated surprise significantly decreases with short-term forecasts whether novices use single or multiple scenarios. These studies help reveal an important distinction with measuring the emotional experience of surprise, which is the act of anticipation vs reaction. People may hold poor mental scripts of their future selves, which would cause them to offer inaccurate predictions of their roll in a given future outcome, and that prediction may be biased, in part, from overconfidence in their own faculties.

Two publications, totalling four conditions, measure the effects of SP on uncertainty estimates (Kuhn & Sniezek, 1996; Schoemaker, 1995). Kuhn & Sneizek (1996, p. 233) describe uncertainty as, "beliefs about the range of possible values for that [unknown] quantity, or its variability" and measure it as credible intervals. Schoemaker (1995) measures uncertainty as the inverse of confidence ranges. Both studies have two different conditions, where some participants work with a single scenario and some work with two scenarios. Schoemaker's study suggests that whether participants generate or review single scenarios, their perceptions of uncertainty increase, but Kuhn & Sneizek's study suggests that when participants review a single scenario, their perceptions of uncertainty decrease. Though these initial findings appear to point in opposite directions, it may be the case that the participants who reviewed a single scenario in Schoemaker's study have lower uncertainty intervals, compared to the scenario 'generated' condition. If this is the case, then there would be some support for suggesting the two studies discovered somewhat similar behaviours. However, this is not the case. Schoemaker's scenario 'reviewed' condition produced greater uncertainty intervals compared to the scenario 'generated' condition. Further to this, Schoemaker's study suggests generating high and low scenarios decreases uncertainty, while Kuhn & Sneizek's study suggests that reviewing high and low scenarios increases uncertainty. Schoemaker attributes the decrease in uncertainty to the loss in believability in extreme scenarios, which he suggests overpowers the availability bias. The participants who generated a single scenario in Schoemaker's (p. 202) study were directed to choose "strategically important issues from work or home", and those that generated multiple scenarios created one page scripts that explored conditions that would bring about certain extreme events. Kuhn & Sneizek's participants, on the other hand, developed one paragraph scenarios regarding broadly discussed global issues with few details. Schoemaker's participants had one week to develop their scenarios, whereas Kuhn & Sneizek's participants were given less than 45 minutes to read and evaluate their scenarios. Several studies discuss the biasing effects of elaborative thinking (a factor of time) which can lead to salience of information and believability (*see* Gertner, et al., 2011; Humphreys & Garry, 2000; Petty & Cacioppo, 1986). Further complicating this comparison is that Schoemaker's participants generated scenarios with short-term vantage points (1-5 years), while Kuhn and Sneizek's participants reviewed scenarios with medium- and long-term vantage points (10-50 years). One effect that may come from generating scenarios is that the process leads practitioners to perceive uncertainties differently from merely reviewing them.

Meissner & Wulf (2013) is the only publication to test SP against a traditional form of the framing effect. To recall from Table 4.5, the authors used a facilitated, multi-scenario 'generated' method with novice participants. Their study is one of the few to compare a control group against three experimental groups (full SP, partial SP, and strategic planning). Results reveal that a full SP process has more of a de-biasing effect than a partial SP process, and that a partial SP process may have no more de-biasing effect than no strategic efforts.

Bradfield's (2008) and Tetlock's (2005) studies offer an interesting comparison. All SP workshops focus on short-term time scales, use a scenario 'generated' method, but Bradfield tests novices whereas Tetlock tests experts. With an open-ended SP method, the novices in Bradfield's study generate normative scenarios and are resistant to facilitated prompts to entertain more extreme outcomes. The experts, however, in Tetlock's study are prompted to generate dichotomously extreme outcomes as well as a normative scenario. When asked to reassess their probability judgments for the three scenarios, even though most of the sub-samples assign, overall (pre- and post-judgments), higher probabilities to the normative outcomes (matching Bradfield's novices), they make greater adjustments in the extreme scenario probabilities. In fact, this effect is stronger for the experts vs their dilettante counterparts.

Tetlock's (2005) third study specifically tests SP on hindsight bias. Experts make short-term (5 year) forecasts on real-world events. After the occurrence of the events, they are asked to remember their forecasts before and after a SP workshop designed to address hindsight, using multiple scenario generation methods with counterfactual thinking. Tetlock's study shows that counterfactual SP may help reduce hindsight bias.

4.2.4 *Risk of Bias Across Studies*

Without the original data, it is not possible to statistically determine whether testing, sampling, design, or analysis biases are present across the studies. However, some insights can be discussed from this review.

At the participant level, there are two main demographics, novices and experts. Some studies suggest that, under certain decision contexts, novices (i.e. younger students) and experts (i.e. older corporate executives and managers) have been found to perform similarly (Bateman & Zeithaml, 1989; Meissner & Wulf, 2013). Though Tetlock uses exclusively expert samples, he identifies that even within professional participants, there are cognitive developments related to levels of expertise (i.e. expert vs dilettante, hedgehog vs fox) that can lead to statistically different perceptions and choices. The expert samples make up a large majority of the mental model studies (88.9%). The novice samples make up the majority of the experimental conditions (62.5%). Would more experienced professionals be influenced by SP to the same magnitude or in the same way as novices?

Coupled with global measures of experience are specific SP experiences. Not all of the studies gathered this demographic information. As Chermack & Nimon (2008, p. 367) postulate, "It is feasible that prior exposure to SP may predispose participants to a particular decision-making style." This is potentially one of the biggest gaps in the empirical literature that needs addressing. It is assumed that SP facilitates cognitive changes within practitioners, and the data across all the reviewed studies largely support this assumption. Therefore, one of the next logical questions in the SP efficacy research is whether these cognitive changes are short-term or long-term. Does a single experience with SP carry the same magnitude of effect as several experiences? If there are lasting effects, are they generalised, compartmentalized, or do they have a compounding effect? Experience, globally, as well as within SP practices, may produce a bias in the empirical data that has yet to be detected.

Novices show either increasing or decreasing confidence after SP, while the one expert sample shows no change. In fact, all studies that result in no changes use expert samples (Chermack & Nimon, 2008; Glick, et al., 2012; Haeffner, et al., 2012; Hawkins & Chermack, 2014; Phadnis, et al., 2014), while one study uses a mixture of novice and experts (Schnaars & Topol, 1987). However, the anchoring bias is equally represented across both sample demographics. Sedor (2002) even directly compares performances between novice and expert samples and finds no significant differences in anchoring biases. Tetlock (2005), on the other hand, reports that his experts are consistently more affected by cognitive biases (esp. probability and likelihood violations) than his dilettante samples (professionals without specialised expertise).

Another division within the participants is that all novice samples are randomly assigned to conditions. Expert samples are determined through random (n = 3), stratified (n = 4), convenience (n = 5), or purposeful (n = 3) sampling. This presents potential beneficial as well as misleading biases in the data. Random sampling strengthens the generalisability of data and credibility of methods because performance and individual biases are less likely to be more represented in one condition over the others. However, stratified random sampling can ensure a heterogeneous mix of specific demographics within a group setting, which can help reduce bias across conditions. Bradfield (2008) stratifies across age, qualifications, work experience, and nationality demographics. Phadnis, et al. (2014) stratifies across organisational type. Purposeful sampling is used to gain insights about specific populations (e.g. experts in the transportation industry). Purposeful sampling can also be used to help determine the difference between expertise (novice vs expert, SP knowledge vs no SP knowledge).

At the method level, studies are divided between scenario 'generated' and scenario 'reviewed' methods. There does not appear to be a bias across changes in uncertainty and confidence between the two methods. That is to say, both methods produce increasing and decreasing reports of confidence and uncertainty, even when controlling for number of generated scenarios. Both methods reveal varying magnitudes of the anchoring bias, in the same direction, though whether one leads to greater changes than the other cannot be known with the available data.

Closely linked to scenario interaction is location, field vs laboratory settings. Even though the majority of field settings are coupled with expert samples, and majority of laboratory settings are coupled with novice samples, the division between scenario interaction isn't as extreme. One-third of the scenario 'generated' studies are carried out in a laboratory. The laboratory studies do not share DV across publications, but together they report changes in some quality of decision-making (e.g. mental model and uncertainty), violations in reasoning (e.g. anchoring, normativity, and probability) or in the case of one, reduction in a bias (e.g. framing). The other two-thirds of the scenario 'generated' studies are conducted within organisations, also report changes in qualities of decision-making (e.g. mental model), violations in reasoning (e.g. availability and likelihood estimates), and in the case of one, reduction in a bias (e.g. hindsight). Though the field studies are the only ones to also report no effects, which may reflect a bias in the kind of crystalized thinking that develops as people gain in experience and expertise. These qualitative comparisons show that even though the scenario 'generated' studies do not have large overlaps in DV to make straight comparisons, they do show a largely shared pattern in performance by participants.

Seventy percent of the scenario 'reviewed' studies are conducted in a laboratory setting. Confidence and anchoring biases can be compared by experimental location, and there does appear to be a potential bias. The laboratory study (with novices) reports increasing confidence after SP, whereas the field study (with experts) reports no change in confidence after SP. Again, this may be reflective of a cognitive crystallization that develops as a person gains in experience and specialised knowledge. Both location settings report an anchoring bias in participant forecasts.

The timing differences between studies may also produce potential biases in the data. Table 4.5 shows that the timing of each study varies from as brief as 45 minutes to as long as 5 years. Of those that report timings, there is a sharp divide between studies. Forty-six percent of the reported session times fall between 1-7 days, Mode = 1 day (all but two are experimental, novice samples in the lab), and the remaining 54% fall between 35 days and 5 years, Mode = 90 days (all but two are expert samples in the field). The shorter timed studies report most of the cognitive biases, including all the framing biases and most of the anchoring biases. The shorter studies also report all the confidence scores. The longer timed studies report almost all the mental model results. Of the DV shared across the two timing categories, the long-term studies report increasing uncertainty while the short-term report mostly decreasing uncertainty; both report influences from the availability heuristic; both report violations in probability assessments. Timing can be used as a proxy for deliberation, communication, and engagement, amongst other personal and group qualities. The division of DV across the studies shows that timing is a methodological feature that should be accounted for more clearly in future studies.

The chosen method of analyses also reveals a potential for biased results. With the exceptions of Bradfield (2008) and Önkal, et al. (2013), most of the data are gathered and analyses at the individual level. This level of analysis reveals a bias in our available understanding of SP efficacy. Ninety-four percent of the studies report decision-making effects at the individual level. Half of these studies, however, use a group method to administer the IV, which are primarily the scenario 'generated' methods. The other half, an individual participation design, which is evenly divided between the 'reviewed' and 'generated' methods. What needs to be realised here is that by reporting only the individual experience, the literature is supplying only part the story. The individual is a constant in the process; Therefore researchers should ensure – as they have been – that the individual experiences, cognitions, and behaviours are understood. However, the group, as a separate quale, is also a necessity in SP. To negate the group perspective so completely has left a gaping hole in the literature, and by extension, our knowledge of SP efficacy.

At the publication level, it is important to look at what quality of outcomes are reported. Even though seven articles report no effect of SP on specific features of proposed mental models, surprise, and accuracy, *all* publications report some form of decision-making effects which are attributed to SP. Schnaars & Topol (1987) offer the only study that presents fewer effects than no-effects. Such a pattern across publications reveals a potentially strong bias towards publishing positive results in the field of SP. The lack of control conditions in most of the studies leaves many conclusions opaque. This revelation also highlights issues from earlier in the chapter. The limited published work at this time could indicate a lack of interest in the field to conduct empirical investigations into SP, or it could be an outcome of a publishing bias towards accepting only positive effect studies. Of course, both could at work, as well. We cannot know at this time, but parity in publication of discoveries is by far the most important effort that will strengthen the scientific foundations of SP.

4.2.5 ***Empirical Review Discussion***

Numerous choice behaviours, attributed to eleven cognitive features and biases, are explored – mental models, anchoring, availability, believability/plausibility, probability/likelihood, confidence, surprise, uncertainty, framing effect, normativity, and hindsight – with another seven biases included in the discussion – belief perseverance, confirmation bias, experience bias, overconfidence, optimism bias, primacy effect, and recency effect. Several of the cognitive features are considered barriers to reasoning and decision-making. Table 4.6 summarises the findings from the empirical literature review.

Table 4.6 brings to light a contrasting feature between the theoretically based and the empirically-based dialogues surrounding ST. The theoretical publications explored in Chapter 3 acknowledge that ST is a label used to the describe complex cognitive processes that facilitate SP and determines scenario content; Just as SP is a label used to describe a wholistic process of thinking about, discussing, developing, and

Table 4.6. Summary of cognitive features, definitions, and study measures

Cognitive feature	**Definition**	**Measure**
Mental Models	A collection of representations of external reality, hypothesized to play a major role in cognition, reasoning and decision-making	GDMS, CQEC, MMSS, DLOQ, Q12. MCDA
Anchoring	Relying too heavily on one piece of information to aid decision-making	Forecasts, Identification, Evaluations
Availability	Relying on easily accessible mental information to influence value, probability, likelihood, and other weighted judgments	Exploration, DLOQ
Believability/ Plausibility	Evaluating the logical strength of an argument biased on the believability of the conclusion	Confidence, Estimates, Asymmetry Ratio
Probability/ Likelihood	Estimating the occurrence of future events	Correlations, Point estimates
Confidence	Trust or belief in the accuracy of one's own judgments	Likert scale (1–4; 1–7; 1–9, $^{-}3$–$^{+}3$), 100-SI, Confidence intervals
Surprise	Expectation that the given forecast interval would capture the true value	SI, Likert scale (1–7)
Uncertainty	Perceptions of unknown information	Confidence range, Credible ranges
Framing	Drawing different conclusions from the same information, depending on how that information is presented	Strategy framework, Decision quality
Normativity	Evaluations and reasoning based on perceptions of social norms	Probability adjustment, Punctuated equilibrium induced coagulation
Hindsight	Mistakenly thinking past events were more predictable, before they occurred	Memory recall

planning for the future through a systematic and iterative process. ST is described as a cognitive system (Lintern, 2007), a complex learning and creative process that involves subjectivity, intuition, emotions, rationality and objectivity (Cairns & Wright, 2018a), a cognitive process of imagination, analysis, and judgment (Mackay & McKiernan, 2018). The investigation borrows from these definitions and defines ST as a process of cognitions that includes memory encoding, creative divergent thinking, causal associative thinking and reasoning, parallel thinking, evaluative thinking, systems thinking, and synthesis.

Yet the reviewed empirical work overwhelmingly uses generalised language to present ST as a more singular cognitive function. The thinking process involved in SP is described as "thinking differently" (Bodin, Chermack, & Coons, 2016, p. 22; Ram & Montibeller, 2013, p. 663), "new thinking" (Chermack & Nimon, 2008, p. 359), "changing thinking" (Chermack, Van der Merwe, & Lynham, 2007, p. 381; Haeffner, et al., 2012, p. 524; Hawkins & Chermack, 2014, p. 79), "decision making patterns" (Glick, et al., 2012, p. 496), "flexible thinking about the future" (Kuhn & Sniezek, 1996, p. 232), "envisioning a sequence of events" Sedor (2002, p. 731), "stretching their thinking process" (Meissner & Wulf, 2013, p. 804), "future-focused thinking", "managerial thinking", and "thinking with scenarios" (Önkal, Sayim, & Gönül, 2013, pp. 773-774). Bradfield's (2008) article is one of the few empirical studies that discusses individual features of ST and their barriers: mental models, thinking processes, inductive versus deductive thinking, uphill thinking, cognitive barriers, causal thinking, creative thinking, and freewheel thinking. And yet, Chapter 3's review in conjunction with Table 4.6 reveals there are various features to ST that can be inhibited or exploited by any number of barriers, producing different decision-making outcomes. Discussions regarding ST should make this point clear, and future empirical research into ST should be explicit in the cognitive features that are being tested and measured. The next and final chapter in Section I brings together the IL/ST framework with relevant issues discovered in the field and the cognitive realities of the empirical review to round out the overview of the chosen methodology.

Chapter 5. **Introduction to Empirical Approach**

"It is good to have an end to journey toward;
but it is the journey that matters, in the end."
Ursula K. Le Guin, *The Left Hand of Darkness*

The IL/ST framework developed in Chapters 2 and 3 provides a roadmap for the investigations. The interviews in the first half of Chapter 4 help identify a focal point to begin designing around the framework – high-threat vs low-threat organisational environments. The empirical review in the second half of Chapter 4 presents a selection of cognitive features and biases (i.e. barriers) that help contextualise the methodology. The present chapter brings together these discussions in a synthesis of the methodological position which then informs the design of empirical studies presented in Section II. Four guiding research questions are developed which will drive the empirical studies. Each chapter in Section II presents a collection of empirical studies, which open with a focused literature review specific to that group of studies. In this way, the work is presented in a lock-step effort that allows for the greatest clarity in each section and chapter.

At this time, the field is just breaking ground on empirical support for SP and ST. These early discoveries are promising, but still too few, too weakly supported, and too splintered in methodology to develop a unifying theory. Many authors act as if the field must resign itself to remaining outside the realm of science. A 36th Chamber by virtue of pragmatism.[14] SP planning is a pragmatic practice and as such is heavily context-dependent. Though the literature speaks of SP efficacy as if human

[14] The 36th Chamber of Shaolin (少林三十六房) is the classic re-telling of San Te's Shaolin tutelage (Kar-leung, 1978). Traditionally, training required mastery of the order's 35 kung fu fighting arts, contained within 35 chambers of the Shaolin temple. When faced with his final task to dedicate himself to a single chamber, San Te leaves the temple to create a new (36th) chamber, where he can train laypeople in the basic forms – those who would otherwise never have the opportunity learn the art within the temple, due to lack of discipline, knowledge, and acceptance. Over the centuries, the story has become an allegory.

performance were standardized, psychology says otherwise. The early years of psychology used to hold many features of cognition as almost monoliths (e.g. Behavioural Psychology or the Müller-Lyer illusion) where the same stimulus *should* elicit the same response from any given human. However, as the science of psychology matures, the farther the field moves from this perspective, and so too, should the discipline of SP. The complexity that may be inherent in both SP and ST, as it is in psychology, is no reason to throw in the towel.

5.1 Selection of the IL/ST Framework

The IL/ST framework, presented in Table 3.1, maps the dominant cognitive processes within ST against the existing IL model. The general IL model will inform the stages of SP to be investigated. The IL model is chosen over the more quantitative models for two main reasons. The IL model is the prevailing model used in both SP practice and empirical research (Amer, Daim, & Jetter, A review of scenario planning, 2013; Bradfield, Derbyshire, & Wright, The critical role of history in scenario thinking: Augmenting causal analysis within the intuitive logics scenario development methodology, 2016; Ramírez & Wilkinson, 2014; Ringland, 1998; Tapinos, 2013). The methods developed from the IL model lend themselves the least to a quantitatively based investigation, therefore could benefit the most from the inclusion of quantitative insights, through a mixed-methods approach.

5.2 Identification of a Priming Manipulation

The interviews in the first part of Chapter 4 reveal a grey area in the logic that appears to apply equally across the interviewees (and by extension, their organisations). The grey area is the issue of *perception*. All SP workshops were *perceived* to be of value, and their values were in relation to the *perception* of knowledge sharing and external threats (high vs low). High-threat environments discussed in the interviews and presented in the empirical studies reflect such factors as decreasing profit margins and sales, threats from competitors, and employee satisfaction. Low-threat environments, on the other hand, reflect more business-as-usual factors and opportunities for expansion.

Van der Heijden, et al. (2002, pp. pp. 61-62, Figure 2.1) recognise this same high/low distinction in organisational threat levels, and discuss how perceptions of either elicit standard responses that eventually result in the same equifinality of strategic inertia. However, the empirical work in the second section of Chapter 4 does not appear to support their claim. If practitioners' mental states are influenced by information related to and developed from their SP efforts (as evidenced across section 4.2.3.6 and Table 4.6), then it may be the case that practitioners are susceptible to priming effects from SP information (Oppenheimer, LeBoeuf, & Brewer, 2008; Valdez, Ziefle, & Sedlmair, 2017). Priming is the cognitive phenomenon whereby mere exposure to a stimulus (internally or externally) can unconsciously or unintentionally influence future responses (Weingarten, et al., 2016). Extensive research suggests that people's evaluations of target stimuli can be systematically affected by the presentation of primes, even when people are explicitly instructed to ignore them (*see e.g.* Brownstein, Madva, & Gawronski, 2019).

If external information can reflect different threat-levels, then it may be the case that ST can be primed by differing types of threat messages. Since SP is a process, and ST is an amalgamation of different cognitions, then as newly generated information emerges throughout the process, this process may, in turn, further exacerbate priming effects on the different qualities of ST. Not only could differently-primed ST efforts potentially lead to different SP content, but could potentially influence practitioner's perceived value of the process itself.

5.3 **Cognitive Context in Experimentation**

The empirical review in the second section of Chapter 4 provides a selection of cognitive features and biases that can help contextualise a high-threat vs low-threat priming effort. The following discussion will help develop three research questions which will drive the empirical studies in Section II.

In the category of "cognitive features", mental models are the most popularly tested features. They are treated as reactionary human capacities in the studies, mental capacities to be manipulated – "*How much were mental models changed?*" These features do not serve as priming mechanisms, but rather features to be primed.

Believability, plausibility, and normativity are types of judgments measured in the empirical studies where the scenarios, themselves, serve as priming mechanisms. They are decision-making aids that have the potential to confound priming manipulations within an empirical design and invalidate any potential data. Therefore, the three judgments will need to be controlled for in the empirical designs.

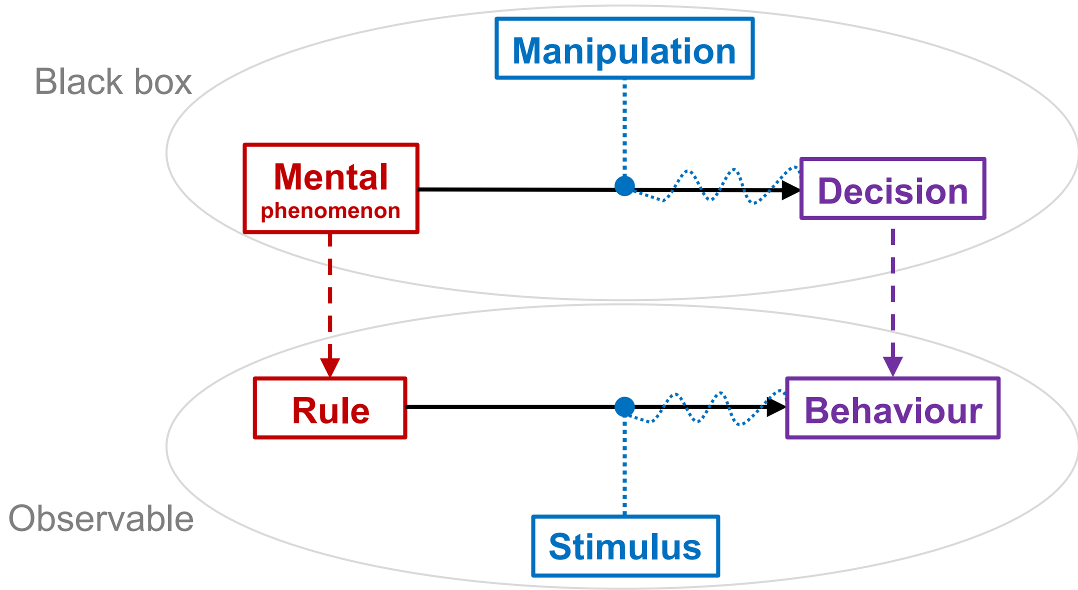
Similar to believability and plausibility are probability and likelihood. They are evaluative judgments based on the priming mechanism of the scenario context. Not priming mechanisms themselves. As well, evaluating probabilities and likelihoods are required techniques in Stage 5 of the IL model. Therefore, rather than be treated as mechanisms of priming, they can serve as potential indicators of ST effects.

Confidence, uncertainty, and surprise are all qualitatively different perceptions, but treated as inverses of the other in the empirical literature, reflecting the cognitive overlap of their judgmental qualities. They do not easily lend themselves easily to serving as priming mechanism for differing threat-level messages; However, confidence and/or uncertainty scores remain reliable standards for measuring perceptions of performance.

In the category of "cognitive biases", hindsight, anchoring, availability, and framing are all tested and measured. As discussed in a number of the empirical papers, such biases appear to relate to and exacerbate other biases. The selection of biases reviewed in the literature serve as a great example of precisely this kind of inter-relationship. For example, information that is more available is shown to often influence (i.e. prime) people's attention and judgments more than less available information (Pollard, 1982; Mamede, et al., 2010). When people use more available information to make judgments and decisions, they are primed by and anchoring to that information (Campbell & Sharpe, 2007; Hess & Orbe, 2013). One method for priming an anchoring bias is to frame dichotomous information with the same outcomes, but in contextually opposite or different directions (Cho & Gower, 2006; Malenka, et al., 1993). If historical perspectives were desired, hindsight would be a key cognitive resource in the thinking process. Detecting a hindsight bias is most effective when using a longitudinal study design, but this method is not proposed for this investigation. Therefore, a method that incorporates availability of information through framing techniques has the potential to serve as an effective priming mechanism to test whether practitioners anchor to perceptions of external threats.

Since cognitive efforts and effects cannot be directly measured, the standard method is to use a proxy. A proxy is an item, event, or behaviour that can serve as an indirect measure of the desired outcome (e.g. a decision) which is itself strongly correlated to that outcome (e.g. making a choice; Sherman & Rivers, 2020). Proxies are necessary when direct measures of the desired outcome are unobservable or unavailable (Grohs, et al., 2018). Figure 5.1 illustrates the logic behind standard empirical detection and measure of cognitions by proxy. We assume there to be particular mental phenomena that directly inform decision-making (top cloud). Along the path of influence (top solid black arrow), manipulations can enter that serve to moderate the level or magnitude of influence (top wiggly dotted line), causing decisions to change to varying degrees or become biased. An example of a mental manipulation that could bias (beneficially or detrimentally) decision-making is a cognitive heuristic. The mind, though, remains a black box, therefore we are unable to directly measure thoughts and decisions. We turn, then, to physical properties to serve as proxies for our manipulations and measurements. Based in established theory (bottom cloud), we rely on specific sets of rules that represent the known (or assumed) cognitive mechanisms of mental phenomena (dashed red arrow). We use these rules to anticipate specific behaviours (bottom solid black arrow) and infer these behaviours to be reliable representations of specific decision-making (dashed purple arrow), within margins of error. To better understand moderating effects on decision-making, we introduce external stimuli to approximate cognitive manipulations (bottom wiggly dotted line) and measure subsequent behavioural changes.

Figure 5.1. Logic for Measuring Cognition



There remains a healthy debate around the science of cognitive measures and designated proxies. The theory of implicit bias that backs the empirical work in Section II is particularly up for debate. A larger survey of the scholarship on implicit bias reveals a host of competing definitions: conscious propositional representations (De Houwer, 2014), unconscious beliefs (Mandelbaum, 2016), attitudes that affect behavior under certain conditions (Olson & Fazio, 2008; Petty, Briñol, & DeMarree, 2007), previously learned attitudes that coexist with newly formed attitudes (Wilson, Lindsey, & Schooler, 2000), and irrational mental states (Brownstein, 2018; Brownstein & Madva, 2012; Gendler, 2008; Madva & Brownstein, 2018). Each operational definition comes with its own range of designated tools and proxies. All take a structural view of cognition and presume that implicit measures (i.e. proxies) capture implicit biases, as much as explicit measures capture explicit awareness. The core of the debate is whether such mental constructs exist, whether they can be adequately measured, and whether they have real-world significance (Brownstein, Madva, & Gawronski, 2019). Greenwald and Banaji (2017) suggest the best way to try to understand what "implicit" can even entail, is to first approach it in an empirical sense, rather than conceptually. This means evaluating types of indirect measures in empirical studies, as opposed to discussing reductive positions on types of (unconscious) mental constructs. Data show reliable and robust differences between implicit and explicit measures (i.e. unconscious and conscious performance), across numerous domains, suggesting a dissociation between implicit and explicit biases (Cameron, Brown-Iannuzzi, & Payne, 2012; Hofmann, et al., 2005). The nature of our dissociation appears to stem, in part, from our general lack of knowledge or awareness about the sources to our implicit biases (Gawronski, Hofmann, & Wilbur, 2006; Hahn, et al., 2014; Hofmann, et al., 2005). Categorising implicit biases as learned dispositions highlights the dual-process argument that implicit measures may not actually be capturing unconscious representations, but rather *associations* between two concepts, attributes, or mental states (Kelly & Roedder, 2008; Mandelbaum, 2016). The nature of how a piece of information is learned (memory encoding), how it is conceptually stored with existing information (memory storage), how it is related to other information (memory retrieval), and how it influences behaviour can all have associative qualities. Therefore, one way to develop implicit measures is to operationally define "bias" as a difference in associative strength (Brownstein, Madva, & Gawronski, 2019; Van Dessel, 2019). Building backwards from this position, the type of implicit bias being measured (e.g. prejudices or learning) will determine the optimal techniques for measure (e.g. reaction time or accuracy). Having stated the logic, it is still necessary to recognise that the mind remains a black box, in particular a black box to itself (for only the mind investigates the mind) and as long as proxy measures built from our physical space remain our main tools for detection, we must continue to identify and acknowledge our limitations in scientific practice.

5.3.1 ***Anchoring Bias***

SP was developed and continues to be utilised for the express purpose of "steering the organization away from the excesses of group think on the one hand and fragmentation on the other" for the purpose of increased agility and future success (Cairns & Wright, 2018, p. 9). To understand whether this level of effectiveness is possible, it is necessary to investigate how the process affects ST, and how these effects, in turn, influence the content of SP.

The review conducted in Chapter 4 reveals that anchoring to information leads to a well-recognised bias that lends easily to developing scenario-related information that is contextually different and could be used as an experimental manipulation (i.e. prime). Decision-making is considered to be, at least partially, affected by an anchoring bias when adjustments – such as forecasts – are generally insufficient and biased toward some initial value (Tversky & Kahneman, 1974). Some authors postulate about the priming qualities of scenarios, where information, either from within the scenario or from scenario development, prepares the participant to think in a specifically guided way. Becker (1983, p. 96) warns against the power of anchoring from scenarios, without necessarily employing the same vernacular.

> Those reading the scenario author's creation frequently believe they are obtaining a view of what the author believes will be. In other terms, the reader often perceives that the description of future conditions prepared by the author was intended to be a forecast of what the future will hold. As used in policy analysis, a scenario is by no means intended to depict a certain future.

This is the case with the Phadnis, et al. (2014, p. 9) study, where experts become more supportive of their own extreme forecasts (favourable and unfavourable) after reviewing scenarios.

> …scenario-based evaluation may allow experts to use their prior knowledge of an asset to think of new ways it may become a strength or a weakness in the environment envisioned after scenario use, and find new reasons to favor (dis-favor) the investment.

Bradfield (2008, pp. 209, emphasis added) also finds support for the notion of belief perseverance.

> …it was apparent that developments envisaged by all syndicates essentially epitomized variations around a common, already well-

articulated midpoint of events that were expected to occur and the order in which they were expected to occur, representing what might be called an ***embedded cognitive script***.

Haeffner, et al. (2012) appear to find a potential explanation for the anchoring bias and belief perseverance. The authors find that the active process of SP had little to no effect on perceptions of continuous learning. Continuous learning is defined by the DLOQ used in their study as, “Learning is designed into work so that people can learn on the job; opportunities are provided for ongoing education and growth” (Marsick & Watkins, 2003, p. 139). The authors attribute the potential loss in continuous learning perceptions to the fact that the SP experience is often a “one-time” effort. However, if a single workshop can bias participants towards their own established scripts, then more workshops may not necessarily be the kryptonite needed for this cognitive phenomenon.

Authors also treat the anchoring bias in their participants as an amalgamated product of other biases. Bradfield (2008) explores the phenomena of confirmation bias, experience bias, and overconfidence that may facilitate anchoring biases.

One element of anchoring that is agreed upon is that its presence is facilitated by availability of information. Information is made available by either memory recall (working, short-term, or long-term) or construction (the process of imaging). These efforts are affected by a number of factors, some known, many potentially unknown. Bradfield (2008) discusses three main influencing factors: salience, recency, familiarity.

Sedor (2002) suggests that SP and ST can inflate practitioners’ beliefs about the plausibility of future events and the likelihood that anticipated outcomes will occur as described in the goal setting stage. They do this by thinking like a human, so to speak. Humans have a proclivity to find causality in events in order to make sense of them (Hume, 1748; Kant, 1783/1994). SP requires humans to think causally in order to make sense of ‘plausible’ future events. The normative scenario development of some of the reviewed studies shows that more easily imagined events are developed first. Information that is more plausible may become more salient (Önkal, Sayim, & Gönül, 2013). Information that is more salient than peripheral information, more recently learned than the rest of the information, and more familiar than the remaining information will be encoded and recalled more easily – and quickly – than the rest. The phenomenal qualities of ease and immediacy are treated as implicit cues that the information is also more important, truthful, and that no other information is needed (Kahneman, Slovic & Tversky, 1982; Shafir, Simonson & Tversky, 1993; Tversky & Kahneman, 1973).

However, the story may be more complex than either Bradfield's (2008) or Sedor's (2002) explanations. Recency of pervasive media information may have played a dominating roll in the decision quality of Bradfield's participants, but so too may have been a primacy effect. Bradfield reports that the information participants generated early in the exploratory thinking stage dominated the remainder of the information they entertained for scenario development (matching Phadnis, et al. (2014) priming discussion). In psychological terms, remembering (or being more influenced by) the first information encountered, which suppresses subsequent information, is known as the primacy effect.

5.3.1.1 By Way of Framing Mechanism

Anchoring biases can be measured using a number of different methods. Northcraft & Neale (1987) provided a selection of different listing prices for local real-estate and measured whether the listing prices affected participants' valuations of a specific property. Wilson, et al. (1996) and Brewer & Chapman (2002) used arbitrary test ID numbers to determine if test responses were biased by the ID numbers. Adame (2016) exploited the priming effects of high and low anchors specifically to bias judgments. Within the context of SP, scenarios served as anchors, whether purposely or accidentally, in the empirical studies of Chapter 4's review (Bradfield R. M., 2008; Kuhn & Sniezek, 1996; Önkal, Sayim, & Gönül, 2013; Phadnis, et al., 2014; Sedor, 2002). In all forecasting measures, the average responses reflected the scenario content. The effect is equally present in both novice and expert samples.

The effect also appears regardless of whether participants generate their own scenarios or review someone else's scenario. With such short workshops, and brief scenarios to review – that include between one and five threat-level statements – the detected anchoring biases appear to be based on almost shallow processing of the organisation's environment, and yet potentially robust. An anchoring bias was measured by comparing decision-making and choices from differently framed scenario exercises (e.g. high vs low, best vs worst, increase vs decrease), often only two. This investigation is interested in understanding the relationship between ST and SP at different stages in the process. Unfortunately, to mimic the methods of past SP studies, where the final artefacts (i.e. scenarios) are used as priming mechanisms for participants to anchor against, would have the consequence of ignoring the process. One of the gaps in our understanding of the process is how much the information affects ST and in return, affects content. Therefore, to help fill in this gap, different priming mechanisms should be created at the *beginning* of the process. Stage 1 could serve as an effective apparatus for attempting to prime an anchor. Borrowing from the methods of previous SP studies, two differently framed organisational environments could be presented: high-threat vs low-threat. Following basic guidelines for framing mechanisms would allow Stage 1's context-rich information to be formalised and built around a theoretically robust framework. In this way, any anchoring bias would be delivered by way of framing effects and help to reduce cognitive noise from confounding variables.

When a decision is based on how the expected outcome of a problem is framed, all things being equal, this is known as a framing effect (Tversky & Kahneman, 1981). Also referred to as the *valence framing effect*, wherein the frame casts the same critical information in either a positive or negative light. Tversky & Kahneman's prospect theory originally detected a framing effect on numerically bounded responses and tested for risky choice reversal between two differently framed problems. Since these early days, framing of information has been shown to affect a wide variety of decision-making efforts, including bargaining behaviours (Neale & Bazerman, 1985), programs for teens (Fagley, Miller, & Jones, 1999), medical decisions (Almashat, et al., 2008), and how the media discuss immigration (Igartua

& Cheng, 2009). From the prolific empirical literature, Levin, Schneider & Gaeth (1998) identified three distinct types of framing effects: attribute-framing effects, goal-framing effects, risky choice-framing effects. Stage 1 will borrow from attribute-framing effect methods because only a single attribute within the given context (i.e. organisational environment) will be the subject of the framing manipulation. This method will also help bring greater control for confounding influences by ensuring all information is standardised except for a single attribute, which will frame the business environment as either high-threat or low-threat, therefore increasing the reliability of results. Taking into consideration the central question, the first empirical research question is posed.

> **Question 1**: Are there measurable anchoring biases in scenario planning content?

5.3.1.2 In Relation to Scenario Thinking

As previously argued, ST is not just a single cognitive effort, but several discrete, interrelating cognitions. The empirical review revealed that anchored decision-making was measured at a global level of ST. That is, practitioners either participated in a SP process, or read scenarios and made final decisions and then completed various decision tasks. With potentially the exception of Bradfield's (2008) qualitative study, the other studies commit an implied error by treating the thought process of their practitioners as encompassing a singular cognitive decision-making feature. Therefore, the present investigation proposes to explore potential anchoring biases on the discrete fundamental features of ST to bring a more comprehensive discussion to the field and add to the theoretical dialogue, by determining whether biasing effects are comparable across different cognitions.

The IL/ST framework maps eight separate, yet inter-related dominant cognitive processes within ST against the existing IL model. This study views cognition from a constructivist view. "Memory retrieval, problem solving, and creative thinking can all involve a constructive search: rather than being simply retrieved as previously stored units, memories and ideas can be constructed from retrieved elements in the course of systematic searches of memory” (Smith, 1997, p. 135). Using the IL/ST framework helps narrow the range of possible proxies by aligning each ST dominant cognitive process against a discrete SP stage. Though the information in successive stages is informed by the sum of previous stages, experimentally speaking, each stage is treated as independent by virtue of i) the different primary forms of thinking and reasoning required to navigate each stage, and ii) the qualitatively different outcomes required from each stage. The ST proxies to be used in this investigation will be the required outcomes of each stage in a real-world SP workshop.

Experimental measures include qualitative (content analysis, theme analysis, and text mining), quantitative (frequency, established scales, and ANOVA) or mixed-method approaches. For the purposes of this investigation, a mixed-methods approach will be adopted across all experimental studies.

Subsequent stages are, by necessity, informed by prior stages. Therefore, if thinking, reasoning, and decision-making can be influenced by differently framed information in Stage 1, then it may be the case that the different qualities of ST would then be influenced by the same initial prime, as practitioners work through the process; However, the magnitude of influence may be qualitatively different. The second empirical research question is posed.

> **Question 2**: Are anchoring biases comparable across different scenario thinking cognitions and scenario planning content?

5.3.1.3 Concerning Confidence Measures

Along with the potential for anchoring biases within ST, confidence has shown to be a major motivator in both decision-making and action (Bénabou & Tirole, 2002; Feltz, 2007). Confidence is an inescapable belief that informs everything we do. The power that confidence wields warrants its inclusion in any investigation into anthropogenic strategic foresight efforts. Along with this, confidence has proven to be a major point of interest in past SP studies. Four publications covering 13 empirical studies, measure confidence in participants who carried out some form of

SP exercise. Their collective results reveal that reviewing scenarios, in general, leads to increases in forecasting confidence, and all experimental scenario groups have higher than midpoint confidence.

Yet, confidence remains a cue for decision-making. Our forward progressions, hesitations, and refusals are based on the strength of our confidence within context. Confidence is not only an internal cue, but an external one as well. We follow confident leaders. We prefer confident politicians. There exists a multi-billion dollar industry dedicated to increasing the impression of confidence in clients' CVs and resumes. C. P. Bowen, in his essay for the *Harvard Business Review*, goes so far as to advocate for advancement spaces within organisations "managers of the future can make decisions and so that they gain confidence to make bigger decisions," concluding that leadership development is about making decisions with confidence; Not with *accuracy*, nor *as a team*, just with individualistic confidence (Moretto, 2012, p. 7). Given that research shows a weak or total lack of relationship between confidence and almost any other scenario-based decision quality, this cognitive feature should be investigated more fully within the realms of ST. To address the gap, confidence is included in the decision-making studies reported in Section II, and develops the third empirical research question.

> **Question 3**: Is confidence correlated with scenario planning performance?

5.4 Researcher's Role

Empirical investigations must take into consideration both the participants – from which data is gathered – and the investigators. In the case of action research, data is gathered through in-depth, participatory activities which require participants to engage in ST (Balogun, Huff, & Johnson, 2003). The most popular methods for data gathering in action research are interviews, observations, and documentation. Though these passive methods serve important purposes, the amount of time and resources required by the investigator(s) can be considerable. To help mitigate such resource-intensive methods, while standardising the delivery of independent variables, a more active role is taken by the investigator. In line with recommendations by Balogun, et al. (2003), the role of the investigator will assume the role of facilitator. Adopting such a role brings the investigator closer to the data and eliminates the need to interpret participant behaviours through a mediator. Closing the gap between investigator and data development/collection lowers the probability of data misinterpretation. As well, closer proximity with the participants can increase clarity and insight for the investigator, both within experimental sessions and data analysis and interpretation. However, efforts must be taken to anticipate and avoid biasing the process by embodying the role of facilitator. In short, the methodological practices of social science laboratory experiments (e.g. A/B testing) are adopted and applied to novel strategy research. Developing the investigator's role highlights a fourth empirical research question that has the potential to aid both future practitioners as much as grounding theory.

> **Question 4**: Can new methods for measuring and understanding scenario thinking be developed for future research?

What is becoming clearer, as the empirical design builds, is the susceptibility of ST to succumb to biasing information and techniques within the process that can affect scenario content. In line with Balarezo & Nielsen (2017), the literature generally does not reflect such an influencing potentiality. Section I serves to expose the complexities that may be at work within ST, and as such, require far more, and far more rigorously controlled experimental studies. Section II will present the empirical studies that attempt to define the types of associative, implicit biases that may motivate decision-making within a SP context and measure the potential magnitudes of identified biases.

Section II – Empirical Work

"What I cannot create, I do not understand."
Richard Feynman

Chapter 6. **Participants and Stimulus Design (Stage 1)**

Section II presents empirical work using mixed-method approaches to help answer the four research questions. Data is primarily quantitative, with supporting qualitative data. One of the proposed strengths of SP is the *process* factor, in that mental models are challenged and changed through learning as they mentally "move" through the process. However, before any series of cognitive *processes* can be understood, dominant cognitions employed within the process should first be investigated. An interesting perspective implied in SP literature is that people are not fully rational, informed, or even terribly exceptional at thinking. SP was developed specifically to counter these assumed shortcomings in cognition and improve individual- and group-level thinking, reasoning, strategy, and decision-making. The series of studies in Section II explore various features of ST, with the express assumption that people are "limited thinkers" and, as behavioural economics promotes, reason with "bounded rationality" (Simon, 1972).

Each chapter focuses on a single stage within the chosen IL model, and each stage emphasises a single dominant cognitive process as an element of ST. The first half of Chapter 6 reports participant sample demographics for the eight empirical studies (stimulus, creative, causal, and evaluative). The second half reports the development and validation of the priming stimulus used as a proxy for Stage 1 of the chosen IL model. Chapter 7 develops a focused methodology for two studies that test Stage 2 creative thinking and presents the work, in full. Building from discoveries of the creativity studies, Chapter 8 completes the methodology for Stage 3 casual thinking and presents two empirical studies. The final chapter in this section builds on the collective discoveries of the previous four studies, develops a focused methodology for Stage 5 evaluative thinking, and presents both an exploratory and hypothesis-testing study.

The decision to test Stages 1-3 and 5 is based, first, on the variable nature of integrating Stage 4 task into the IL process. At the time the empirical studies were in development, the primary IL model taught within the SBS MBA programme integrated a parallel thinking task to create dyadic outcomes for clustered DF at the end of IL/ST framework's identified Stage 6. Therefore, when developing the studies, the dominating IL model within the researcher's programme of study was chosen. Second, given that the more streamlined method proved to be popular in practice due to time-poor practitioners, their practice was chosen to model in an effort to increase the practical value of any empirical discoveries. Third, the decision was also based on resource availability. Though it is preferred to conduct a more comprehensive investigation on all stages of the designated IL model, limitations on time, space, equipment, and participants largely determined the boundaries of available resources for empirical work.

6.1 **Experimental Participants**

All studies were carried out within the Strathclyde Business School (SBS), both in the UK and UAE, and the Cognitive Psychology lab on the campus of Texas A&M University-Corpus Christi (TAMUCC: US). Data were gathered either through the online survey platform Qualtrics (2019) or by paper-and-pen.

Most of the participants (89%) were recruited from the MBA programme within the SBS (UK and UAE). The remaining participants (11%) were recruited from the US. The SBS population was chosen for several reasons. First, they were an easily accessible population by the fact that they must work in the same building throughout their academic studies. Second, they were a fairly homogeneous cohort. The shared demographics were that they were academics, attending the same business programme, in the same university, within the same three years. Third, all participated in at least one "Exploring the International Business Environment" (EIBE) workshop as part of the SBS MBA programme. The EIBE workshops teach an IL model of SP. The US population was sampled for participants due to convenience in availability, as well as their largely overlapping demographics to the SBS samples: academics, enrolled in a business programme, with IL model training.

Each study within Section II sampled a new group of participants. In an effort to diminish as much confounding variables as possible, an unprimed perspective was required at the start of each study. As discussed in Chapter 5, priming theory holds that people's associations and evaluations can be systematically affected by the presentation of primes. If a participant enters the experimental space cognitively affected by a previous experimental prime (i.e. independent variable), then it would not be possible to determine the magnitude of the same priming stimulus within the boundaries of the subsequent study, designed to measure a different cognitive feature. Another reason is the nature of sampling itself. To ensure the same participants received the same prime in multiple, successive studies, it would be necessary to follow the same sample over the course of 3 and half years. This was not possible due to the short-term availability of MBA students enrolled in the EIBE programme, creating full attrition rates after each workshop. A final reason for choosing a new sample of participants for each study stems from assumptions within SP scholarship, as well as the field of cognitive psychology. Leading SP scholars largely assume that participation in the *process* of SP can lead to cognitive changes. Measuring process change is important to the larger scientific investigation of SP, but equally, it is important to understand process change can only be fully understood when investigated against discrete experiences. In this way, the field gains a more comprehensive understanding of ST and potential effects of SP on cognitive landscapes. The main benefit to this form of sampling method is that outcomes have greater statistical power and generalisability, by comparison, due to the higher controls of the experimental boundaries. A more comprehensive discussion on the limitations of the chosen sampling method is presented in Chapter 10.

Another reason the UK/UAE/US academic populations were chosen for sampling was to help preserve ecological validity. The most important factor to preserving ecological validity can be found in the professions of the MBA students. This population closely resembles the target population that uses SP in professional settings. The MBA students were mostly working professionals before enrolling in their respective MBA programmes and many remained employed in their companies while earning their MBAs. Table 6.1 presents the basic demographics recorded from the experimental studies across Section II. Participant locations were automatically recorded during the testing phases of each experimental study (UK, UAE, or US). At the end of all experimental sessions, participants were given the option to record their age (free response), gender (multiple choice plus free response), and the industry they either worked in or affiliated with at the time (free response). The third column gives the total percentage of recruited participants. The fourth column gives the adjusted percentages after data were cleaned. Cleaned data is absent of outliers and non-compliant submissions from participants. The results section of each experimental study in Section II reports the specifics of each sample and method of data cleaning.

Closely tied to preserving ecological validity were attempts to preserve participant motivation. Willingness to commitment, from research participants, is a key, (Balogun, Huff, & Johnson, 2003). There are several discussions regarding the generalisability of results from lab-based experimental investigations due to qualitatively different motivations felt between real-world settings and lab-based settings (Jun, Hsieh, & Reinecke, 2017). There are several ways to address the differences. One way is to "bring the lab to the real world" (Drew, 2018). Another way is to mimic the real world where possible. In psychology, deception is a common practice to preserve intrinsic motivations without inadvertently biasing the participants, of course, free from exploitation and abuse, and ensuring that strict guidelines are met (Boynton, Portnoy, & Johnson, 2013).

Table 6.1 Participant demographics

Demographic	Categories	Percentage (%)	Outliers & Removed Responses (%)
Location	UK	71.6	90.4
	UAE	17.4	7.8
	US	11.0	1.8
Age	18-24	9.2	7.0
	25-29	19.7	16.6
	30-34	14.7	22.9
	35-39	13.4	12.1
	40-44	6.1	6.4
	45-49	2.9	3.2
	50+	2.9	3.1
	No answer	31.1	28.8
Gender	Male	42.1	41.1
	Female	26.8	30.7
	No answer	31.1	28.2
Industry	Oil-Gas	52.4	57.8
	Engineering	1.6	1.6
	Finance	1.3	1.3
	Education	0.8	0.8
	IT	0.8	0.8
	Government	0.5	0.5
	Marketing	0.5	0.5
	Business	0.3	0.3
	Human Resources	0.3	0.3
	Maritime	0.3	0.3
	Medical	0.5	0.5
	Retail	0.3	0.3
	Transportation	0.3	0.3
	No answer	40.1	34.7

Note: SBS located in Glasgow, UK (UK), SBS located in Abu Dhabi & Dubai, UAE (UAE), and TAMUCC located in Corpus Christi, Texas (US). To comply with university ethical standards of human-based research, demographic information was optional for participants to offer. As a result, not all participants answered all the demographic questions. Most of the outliers removed in the studies across Section II were participants who chose not to answer one or more of the demographic questions.

To preserve as much "real world" motivation as possible from the participants, the real world was mimicked (i.e. business vignette content) with a modified form of deception (i.e. priming stimulus). Ethical approval was received, from all participating institutions, each step of the way, before recruiting participants. All participants received the same introductory instructions designed to increase intrinsic motivation.

1. Thank you for your participation. This is a special effort with the University of Strathclyde and SBS.
2. There is no time limit to this exercise.
3. On the next page is a summary of the [the organisation's] goals for the next 5 years. Please read through this now.
4. As a member of the SBS [Texas A&M University-Corpus Christi], your group has been recruited to help the [the organisation] determine how to reach their goals set over the next 5 years. Your cooperation in this exercise holds the potential to bring real recognition to both the SBS and your efforts. As well, the work you do today could have a serious impact within the SBS [Texas A&M University-Corpus Christi].

6.2 Business Vignette Development

In practice, Stage 1 is defined by practitioners conducting background research on the target organisation in order to define the scope of the SP intervention. The purpose is to develop a rich library of information regarding past behaviours, present issues, future goals, and timeline. Information from which practitioners draw as they move through their SP intervention. The function of Stage 1 lends itself to serve as an effective apparatus for developing a priming mechanism to test for anchoring effects in each stage of the IL/ST framework, since, by necessity, all participants will be required to conduct some form of introductory background research on a target organisation profiled for the SP studies. In the field, background preparation can be highly involved and take months of preparation. Such resources, time in particular, are not available in a laboratory setting with volunteer participants. Therefore, organisational information presented to participants must be developed in a way that mimics the most pertinent elements of Stage 1 background research, while remaining usable in a highly time-restricted laboratory setting, and meeting the standards of a sound experimental manipulation. To achieve the aims of brevity without sacrificing comprehension, a shortened business vignette is developed as a proxy for Stage 1. The business vignettes are standardised and empirically validated in a comprehensive effort to preserve soundness and validity.

For experimental purposes, the prime (i.e. business vignette content) is the independent variable, or stimulus. The quality of the prime is informed by the interviews and summary review in Chapter 4 – framing manipulation. The independent variable will present one of two forms of background information: framed as a business in a high-threat environment or a business in a low-treat environment. The independent variable for the experimental studies needs to meet the following criteria:

(1) promote ecological validity
(2) preserve participant motivation
(3) test for anchoring effects

To promote ecological validity, business vignettes were developed as a proxy for Stage 1 and modelled a sample of information commonly sourced in SP workshops. The vignettes include a brief history of the organisation, their present relationship with the external environment, goals, and timeline. Three different organisations are profiled in the vignettes.

There is an ongoing debate in the fields of psychology and economics about participant motivations and the generalisability of data gathered through laboratory-based studies. One of the biggest criticisms concerns the implicit and explicit motivations of laboratory participants, and whether an artificial environment can preserve or inspire the same quality of implicit motivations as a natural environment. Implicit motivations are the internal desires to reach a goal, but occur outside one's conscious awareness (Schulthesis & Brunstein, 2010). Motivations, both implicit and explicit, have a reflexive relationship with attitudes, where attitudes can increase or decrease motivations, and motivations inform changes in or strengthening of attitudinal states (Bansah, 2016; Fodor, 2010). Attitudes and motivations are moderated by credibility/believability (Cheung, Sia, & Kuan, 2012; Metzger & Flanagin, 2015) and content of messages (Mayor & Coleman, 2012; Skinner, Strecher, & Hospers, 1994). This is particularly centred around the consistency and truthfulness of self-report values on surveys. Such criticisms are taken into consideration. To put this into context, practitioners engaging in a SP workshop within their own organisation may be motivated differently to achieve their task, compared to being tasked with the same SP protocol, but for an organisation in which they have no personal investment, and in a location outside that organisation's walls. The choices of the organisations and the information within their vignettes were developed with the explicit purpose of preserving as much implicit motivation in the participants as possible.

To help mitigate potential losses in implicit motivation, vignettes focus on organisations familiar to the participants. Background information is sourced from the organisations' websites and public statements. Level of familiarity (i.e. relatedness), however, cannot be ensured to be consistent across all participants in any given study. Studies show that level of relatedness can affect cognitive processes such as imaginability and information rehearsal, which can lead to different motivational outcomes (Deconinck, Boers, & Eyckmans, 2010; Pighin, Byrne, Ferrante, Gonzalez, & Girotto, 2011; Schacter, Benoit, De Brigard, & Szpunar, 2015). For this investigation, participants could not be pre-screened for their knowledge of specific organisations. To address the uncertainty of familiarity, three different organisations were profiled which offer three different levels of familiarity.

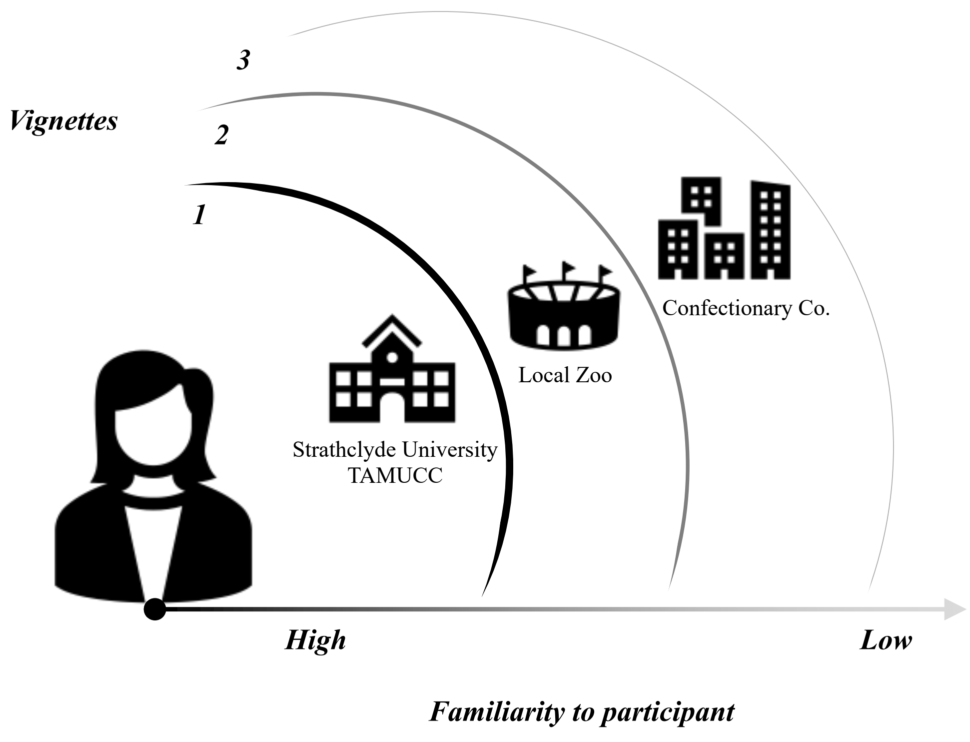
Vignette 1 profiles a real organisation (the participant's university) that all participants within a single experimental session are personally acquainted with (as attending students) at the time of the study. The participants' own university was chosen to increase the believability and saliency of their efforts as much as possible. A vignette about the University of Strathclyde, Glasgow, is prepared for the UK sample, a vignette about the University of Strathclyde, Abu Dhabi and Dubai, is prepared for the UAE sample, and a vignette about TAMUCC is prepared for the US sample.

Vignette 2 profiles a real organisation (the local city zoo). The local zoo is chosen because it represents an organisation that all participants are highly familiar with, but not necessarily personally invested in, to the same level as the university they are enrolled in and attend daily. A vignette about the *Edinburgh Zoo* is prepared for the UK sample, a vignette about the *Emirates Park Zoo* is prepared for the UAE sample, and the *San Antonio Zoo* is prepared for the US sample.

Vignette 3 profiles a fictional organisation (a confectionary company) that borrows features from a real confectionary company. A private organisation with an anonymized name helps ensure no participant has any ability to be more than generally familiar with the industry. Uniform, ambiguous phrasing is used to preserve the same locally based business profile represented in vignettes 1 and 2 (i.e. "The Tammuz Confectionary Co. goals are to expand into the local market…"). Participants are informed the name is fictional to hide the identity of the organisation. By presenting an unfamiliar organisation with a known product, in a known region, greater experimental control can be exercised on information priming because all participants will enter the experimental space with similar levels of knowledge and personal investment, compared to known organisations.

Three organisations are profiled in three separate vignettes ranked by their explicit familiarity to the participant samples. The ranking creates a familiarity scale, illustrated in Figure 6.1, from high (vignette 1), moderate (vignette 2), to low (vignette 3). To increase the uniformity of memory encoding tasks in Stage 1 (i.e. reading a business vignette to gain pertinent knowledge for the subsequent SP tasks), all vignettes featured fictional information about an organisation. The amount of fictional information serves as part of the scale of familiarity. Vignette 1 presents a real organisation that participants are connected with (the local university they are attending), but descriptive information is altered to reflect fictionalised goals and threats. Vignette 2 presents a real organisation that participants are generally aware of (the local zoo in the same city as the university they are attending), but descriptive information is altered to reflect fictionalised public bodies, goals, and threats. Vignette 3 presents a fake organisation that no participants are familiar with (a confectionary producing company), therefore all information is fictionalised. Using a plurality of narratives in an apparatus to deliver the same or similar priming mechanism acknowledges the possible variability that may come from a population that reflexively constructs their social and psychological realities based on understandings of their own individual experiences. Differences in participant output based on vignette may be a reflection of these independent constructs. In that same vein, similarities may reflect shared cultural or social commonalities. Testing the independent variable against a scale of familiarity helps increase generalisability of the results, since data are not gathered from a single organisation's profile.

Figure 6.1 Scale of business vignette familiarity to participants



To test for anchoring effects, statements within each organisation's vignettes (1, 2, 3) present equivalent organisational environments, but with either high-threat or low-threat framed elements. Length (number of words, sentences, paragraphs), complexity (number of headings and sections), number of emotive statements (framing manipulation), and statement structure (order of sections and information) are all controlled for and remain constant between the two matched vignettes. The two differently framed background statements for each vignette group is constructed in a manner that preserves equivalent plausibility between the two extreme high-threat and low-threat environments (Schnaars & Topol, 1987). The vignettes are constructed as follows: Identical organisational background information is presented first, followed by framing manipulation statements, and concluded with the same two goals and restatement of timeline (five years). Vignette statement framing will serve as the independent variable to measure participant responses (dependent variable) for potential anchoring effects.

6.2.1 *Stimuli Priming Validation – First Round*

Whenever a prime is developed, it must first be validated and tested for impact. The priming structure that results in the largest statistically significant difference and at least a medium effect size is considered optimal.

The first round of validation tests for optimal length and complexity for a priming effect. The vignettes need to be long enough to contain the necessary background information of the organisation to give participants a sufficient level of knowledge from which to develop their ST. The length needs to allow for framing manipulations (i.e. differing information) to facilitate any priming effect, if existed. The length, however, needs to be short enough to be manageable within a single experimental session (30-60 minutes). To determine optimal length of vignettes with maximal content information, four vignette structures are developed and tested for priming effect. The vignette structure that produces the largest priming effect will be used as the template for all experimental vignettes.

Structure I is developed to present the most amount of information that could be reasonably read within the expected session time. To reduce difficulty in navigating the information, the ten STIRDEEPER topics are used to develop ten sections of the vignette. Each section begins with a STIRDEEPER heading (e.g. “Society”, “Technology”), then followed by some basic information that coincides with the heading. Some sections include sub-headings, some include bullet-pointed lists, some include both.

Structure II is drafted as a formatting alternative to structure I. It is equally verbose (i.e. same word count and content), but without headings, sub-headings, and bullet-points. The second structure tests whether less formatting and visual cues change the impact of the priming stimulus.

Structure III is developed as an extreme alternative to both structures I and II. This iteration presents the least amount of superfluous information about the organisation and presents only the necessary information in two sentences. The first sentence presents the framing manipulation, and the second sentence presents the goals and timeline.

Structure IV is a modified middle-ground between the formatting of structure I, the verbosity of structure II, and the extreme brevity of structure III. Structure IV presents a single paragraph of background information, with revenue report, size of the organisation, location, and product/service, a second paragraph with three framing manipulations, and a final paragraph that presents the organisation’s goals and timeline.

The first priming validation study uses only vignette 2. Using the vignette that represents a midway point between the two extreme vignettes, on the scale of familiarity, can serve as a heuristic to generalise priming validation results across the vignettes. Furthermore, initially validating the mid-point vignette (2) allows for resources (i.e. time, laboratory space, and participants) to be maximised by avoiding repetitive testing of near-identical business vignettes.

Table 6.2 presents the descriptive statistics across the four structures, comparing high-threat and low-threat priming iterations. Confounding variables (i.e. features) are controlled for, where possible, between the two threat iterations (high vs low) within each structure. This means word count was matched as closely as possible, number of segments (i.e. paragraphs and bulleted lists), sections (i.e. beginning and ending), headings, and framing manipulation (i.e. statements of threat) were matched between the two threat iterations within a structure.

Table 6.2. Frequency table of business vignette 2, structures (I-IV), high-threat vs low-threat

Features	**Structure**							
	I		**II**		**III**		**IV**	
	H	***L***	***H***	***L***	***H***	***L***	***H***	***L***
Words	731	737	714	720	43	49	178	180
Segments	20	20	15	15	1	1	2	2
Sections	2	2	2	2	1	1	1	1
Headings	18	18	2	2	1	1	1	1
Framing manipulation	3	3	3	3	3	3	5	5

Note: Hight-threat (H), Low-threat (L).

Proposition 6.1 If a priming effect is successful, there will be a significant difference between the average output of the two experimental conditions – high-threat vs low-threat.

Appendix A presents the four structures, as they were developed for validation (version 2). Their framing manipulations are identified in **bold and underlined** within each vignette. Once a valid and robust priming structure is verified, the structure will serve as a template for the other two vignettes.

6.2.1.1 First Round Method

Participants ($N = 194$) were recruited from the UK and UAE. They were given a link to an online survey, where they were presented with information regarding the research study. Participants were then randomly assigned a single structure and threat-message for business vignette 2. They were asked to read through the information, then record as many ideas, environmental factors, issues, and events (i.e. responses) as they could think of that could possibly affect the organisation in their path towards reaching their five-year goals. The survey was protected to prevent multiple attempts by the same participant. Table 6.3 presents a breakdown of the organisational profiles presented to reflect the sample demographics.

Table 6.3. Business vignette profiles presented to different participant samples

Stage	Experiment	Vignette	Profile		
			UK	**UAE**	**US**
1	**Priming validation**	2	✓	✓	

6.2.1.2 First Round Results

Structure I (n = 27) results in fewer responses produced from the high-threat framed vignette compared to the low-threat framed vignette, but these differences are not significant, revealing no anchoring effect. However, discussions within the department made it clear that the researcher's own learning style may be too reflected in the structure of the vignettes. The heavily segmented and formatted text – headings, sub-headings, and bullet points – could be affecting the participants' responses. Therefore, a new structure is drafted and tested.

Structure II (n = 25) results reveal the same trend of responses between the two conditions as structure I, but with significance ($F(1,23) = 4.28$, $p = .05$, $\eta^2 = .16$). At a superficial level, this reveals an anchoring effect – high-treat condition produced fewer responses than the low-threat condition. The large variance effect size ($\eta^2 >$ .14) means 40% of the difference is due to differences in variance between the conditions. Further analysis reveals that the low-threat condition contains two participants (n = 2) whose responses are outliers to the data. Once these participants are removed from the dataset, the modified data reveal no significant difference between the framed messages.

Content analysis of responses produced by participants who used either structure I or II shows that almost all participants copied information directly from the long vignettes and entered these as part of their response output. Some participants entered exclusively information from the business vignettes. The goal of the stimulus is to measure the personal knowledge and cognitive abilities of the participants. Since we cannot know if the participants who engaged in the copy/paste efforts would have thought of these responses without the vignette "prompt", a third structure of the vignettes is tested.

Structure III (n = 41) presents only the short framing manipulation, goals, and timeline in the form of two sentences. Participants produce non-significantly differently amounts of DF between the two conditions, showing no anchoring effect, even though the same trend in response division is preserved between conditions. Feedback from some of the participants reported that they were "lost on what to do." This can be due, in part, to the total lack of background information on the organisation. Feedback shows that too little information left the participants unclear how to conceptualise the organisational environment. Therefore, a fourth structure is tested.

Structure IV (n = 101) presents a brief background paragraph and an anchoring effect paragraph, with goals and timeline. This structure reveals a significant difference between conditions ($F(1,99) = 3.99$, $p = .04$, $\eta^2 = .04$, $r^2 = .41$), showing an anchoring effect by condition, with a small effect size due to variance between conditions, and a medium-small effect size due to differences in the two means. No vignette information is copy/pasted, and feedback includes language of believability in the vignette information and even enjoyment by the participants (e.g. "this was fun"), implying potentially stronger levels of internal motivation.

The results of the four structures are presented in Table 6.4 Structure II descriptive statistics are presented with adjusted data, outliers removed. Once outliers are controlled for, it is revealed that structure IV results in more responses, on average, compared to the other three structures. Proposition 6.1 is supported by structure IV results. Furthermore, a negative mean difference across all four validation tests reveals a robust response that participants produced fewer responses within the high-threat condition, compared to the low-threat condition. Therefore, all priming validation tests provide an expectation of quality in anchoring effect between high-threat and low-threat framing manipulations.

Table 6.4. Comparison of business vignette 2 structures (I-IV)

Framing	**Structure**			
	I	**II**	**III**	**IV**
High-threat	9.33 (5.93, 35.15)	6.83 (5.98, 35.79)	10.17 (5.68, 32.27)	11.32 (7.32, 53.63)
Low-threat	12.93 (8.49, 72.07)	11.55 (9.36, 87.63)	12.30 (7.42, 55.04)	14.56 (8.22, 67.64)
Mean Difference	*-3.60*	*-4.67*	*-2.13*	*-3.24*

Note: Means are given first, with standard deviation and variance presented below, inside the parentheses. Structure II is presented with outliers removed (n = 2).

6.2.2 *Stimulus Framing Validation – Second Round*

A second validation test is carried out to determine strength of framing manipulation. To validate the framing impact, an *effect* questionnaire was developed that covered the most common scripts revealed by former and present SP practitioners during the prior interviews and from a year's worth of SP workshops. Table 6.5 presents the effect questionnaire. Each question was assigned a 5-point Likert scale, where the middle value was either "Neutral" or "Neither".

Appendix A presents the remaining two business vignettes 1 and 3 developed into structure IV. Their framing manipulations are identified in **bold and underlined** within each vignette. Structure IV of the three vignettes is used to test for a framing manipulation.

Proposition 6.2 If a framing manipulation exists, the high-threat condition will be perceived as more difficult and stressful, and less likely, realistic, and predictable compared to the low-threat condition.

Table 6.5. Effect questionnaire to test framing manipulation

	Question	Scale
Q1	How likely is it that the organisation will achieve their goals within the time frame?	1 = Extremely unlikely; 5 = Extremely likely
Q2	How easy/difficult will it be for the organisation to achieve their goals within the time frame?	1 = Extremely difficult; 5 = Extremely easy
Q3	How realistic is the time frame for achieving their stated goals?	1 = Extremely unrealistic; 5 = Extremely realistic
Q4	How stressful do you think this business environment is for the employees?	1 = Extremely stressful; 5 = Not stressful
Q5	How stressful do you think this business environment is for their senior members of management?	1 = Extremely stressful; 5 = Not stressful
Q6	If you were one of their senior members of management, how stressed would you be working towards achieving the organisation's goals?	1 = Extremely stressed; 5 = Not stressed
Q7	How predictable is the organisation's future?	1 = Completely unpredictable; 5 = Completely predictable

6.2.2.1 Second Round Method

Participants ($N = 71$) were recruited from the UK and US. Qualtrics was used to deliver the survey and set to prevent multiple attempts by the same participant. This is a between-subjects design. Participants were randomly given one vignette (1-3), with either a high-threat ($n = 35$) or low-threat ($n = 36$) message, answered seven effect questions, and three demographic questions. Only structure IV was tested for second priming validation, using all three business vignettes (. Each question presented a 5-point Likert scale. Question presentation was randomized to avoid response bias by order. Table 6.6 presents a breakdown of the organisational profiles presented to reflect the sample demographics.

Table 6.6. Business vignette profiles presented to different participant samples

Stage	Experiment	Vignette	Profile		
			UK	UAE	US
1	**Framing validation**	**1**	✓		✓
		2	✓		✓
		3	✓		✓

6.2.2.2 Second Round Results

There is a significant difference for all questions, except Q7, between the two framing conditions (high-threat vs low-threat), with large effect sizes due to output differences across the sample ($r^2 \geq .80$). The results are not significantly different across the three different business vignettes.

Exploring the data reveals a clear age/experience division in the sample. Twenty-four percent of the participants have an average age of 21 years old and no long-term professional experience, the remaining participants are professionals (e.g. management, medical doctors, and engineers) with either a post-graduate degree or are presently enrolled in an MBA programme and have an average age of 32 years old. Reflecting on the purpose of the larger investigations, one of the aims is to develop accurate tools for measuring various elements of ST, and to use a representative sample of the types of individuals who use SP in real-world settings, in order to bring a higher level of ecological validity into the studies. With this in mind, the scoring data are divided along clustered age/experience divisions, where the younger, novice sample and older, professional sample are analysed separately. Controlling for age-employment levels reveals a dramatic difference in response patterns.

The younger, novice sample shows no framing impact on any of the questions and no consistency in response behaviours, revealing that their behaviour is no greater than chance. Furthermore difference in means (*MD* = .07) and difference in mean confidence intervals (*CIMD* = -1.23) are smaller than the older, professional sample (*MD* = -.88; *CIMD* = -1.40). A lower difference in mean translates to responses being valued about the same across all questions, and low confidence interval difference indicates that these similar behaviours were consistent across the sample.

In contrast, the older, professional sample reveals an framing impact on all questions except Q7. The difference in responses to Q7, compared to the other six, may be due to the qualitative difference in the topic and the question. It is the only question that asks for predictable/uncertainty thinking. This sample scores consistently across the first six questions, where the high-threat condition has significantly lower ratings than the low-threat condition. Results are reported in Table 6.7. The results support Proposition 6.2. It would be expected to see higher reported stress from the employees and more struggles from the organisation in a high-threat environment while trying to meet the same goals within the same timeline as their counterparts in a low-threat environment.

Participants who read a high-threat business vignette, overall, felt that the organisation would be “extremely unlikely” to achieve their goals, have an “extremely difficult” time achieving their goals within 5 years, which they felt was an “extremely unrealistic” timeframe. They also felt that the employees at all levels, including their own projection of themselves as a manager, would be “extremely stressed” trying to meet the organisation’s goals.

In contrast, participants who read a low-threat business vignette, felt neutral about the reality of the organisation’s ability to reach their goals. They were slightly less favourable with their other assessments, but still gave higher ratings compared to the high-threat condition. On average, they saw the task of reaching their goals only “somewhat difficult”, where the work environment would be “very stressful” for employees and themselves as managers, but closer to “extremely stressful” for imagined managers.

Table 6.7. Comparison of business vignettes (1-3) by condition (high-threat vs low-threat).

Question	Condition M (SD, V)		Effect size (r^2)
	High-threat	**Low-threat**	
****Q1**	1.86 (1.13, 1.29)	3.03 (1.06, 1.12)	-1.07
***Q2**	.93 (.58, .33)	2.26 (.1.05, 1.11)	-1.51
****Q3**	1.87 (1.09, 1.19)	3.08 (1.43, 2.04)	-.93
****Q4**	1.22 (.69, .48)	2.26 (1.16, 1.36)	-1.05
****Q5**	.91 (.83, .69)	1.77 (.95, .91)	-.95
*****Q6**	.88 (.75, .56)	2.72 (1.02, 1.04)	-.85
Q7	2.72 (1.02, 1.04)	2.52 (.89, .79)	.21

Note: Effect sizes comparing means differences (r^2), range across absolute values of small ≥ .20, medium ≥ .50, and large ≥ .80, where $r^2 = |1|$ represents a mean difference equal to one standard deviation.
* $p < .000$, ** $p < .01$, *** $p < .02$

6.2.3 *Business Vignette Discussion*

Three different organisations are profiled in their own vignettes (1-3). Each organisation establishes a different relationship of familiarity (1=high, 2=moderate, or 3=low) with the participants. Vignette 2 was used to validate an initial structure of information (I-IV) that could create the strongest priming effect. Structure IV resulted in the largest significant difference in responses. The three vignettes were then developed into structure IV and tested for framing manipulation validation using an effect questionnaire. The results show that older, professional participants, in particular, are more susceptible to the framing manipulation which reveals an effect across different perceptions of an organisation's efforts and abilities. Table 6.8 summarises the business vignette versions (1-3) and their structures (I-IV) as they were developed, tested, and validated in this chapter.

Table 6.8. Structures (I-IV) developed for each business vignette version (1-3)

Version	**Structure**			
	I	**II**	**III**	**IV**
1				✓
2	✓	✓	✓	✓
3				✓
Development	*tested*	*tested*	*tested*	*validated*

Structure IV provides the best performing content, bringing believability into the task, while resulting in the strongest priming effect. A second round of validation testing shows that participants' perceptions of the inherent threat levels strongly align with intuitive understandings of organisational threat, stress, and difficulty. These differences in perceptions are further validated by participants' self-assessment of their own role within the organisation (i.e. Q6). This, more than anything, is key to understanding the underlying cognitive functioning of an anchoring bias on ST from active SP. Role playing, internal scripts, and emotional intelligence are all methods for internalising information at an implicit level, where biases are proposed to function the strongest. With the methods employed in this chapter, a validated set of primes, developed around a dichotomous framing manipulation (high-threat vs low-threat) will serve as a proxy for Stage 1 information gathering, and be used as the independent variable for all experimental work presented in Chapters 7-9.

Research question 4 of this investigation asks whether novel methods for measuring ST can be developed for future research. From a methodological perspective, providing brief background information is a standard experimental design, in one form or another, in SP and forecasting empirical studies. Participants are given a set of highly controlled information then asked to perform some form of decision making or choice task (e.g. Bradfield, 2008; Kuhn & Sniezek, 1996; Önkal, Sayim, & Gönül, 2013; Phadnis, et al., 2014; Schoemaker, 1995; Sedor, 2002). However, few studies take any pre-development steps to validate the content of the information given to participants, whether for believability or in the case of experiments, strength of priming. For example, in the Önkal et al. (2013) experiment, the authors developed background information about a hypothetical mobile telecommunications company, 18 time-series plots (growth or decline), model forecasts (best- or worst-case), and corresponding scenarios (i.e. best- or worst-case). Participants were provided with only one version of the content-rich information package (e.g. growth + base-case or decline + worst case) before giving their own forecast responses and completing a questionnaire. Analyses compared average performance outcomes between the different best/worst-case conditions in order to draw conclusions on the value of using scenarios. What is missing from their design, however, are validation measures to show whether their profiled company background information was believable, and time-series, forecasts, and scenario contents were perceived as truly different from the others. Önkal et al. attempted to develop experimental primes, but without employing the standards and rigors of psychological experimentation. This investigation, in part, aim to address common design issues, like this, in futures-thinking and management science research. By employing standard experimental guidelines from the science of psychological research, this investigation aims to bridge a gap between the harder sciences of experimental research and the experiential knowledge of business and management.

The methods presented in the second half of this chapter are only two ways to validate psychological tools for effect. For future empirical work in ST, it would be interesting to build a more fully developed framing test, that could be used on a number of SP and foresight studies. To begin, scripts revealed through interviews and workshops would be qualitatively analysed for topical differences. These topics could then be cross-referenced with the stages of the different methods developed from the preferred SP model, to find the most relevant themes within which to develop questions. For example, previous interviews revealed seven topical questions (Table 6.7). Using the IL model as an apparatus, these questions could be divided across the following themes, presented in Table 6.9.

Table 6.9. Possible thematic division of questions 1-7.

Question	Potential theme
Q1, Q3	Timeline
Q2	Organisational Labour
Q4, Q5, Q6	Emotional Labour
Q7	Predictability

Using thematic analysis as a starting point, further questions could then be gathered from existing surveys and related resources. When a question does not fit the existing structure, a new theme emerges and is added. The idea would be to develop more questions than are ultimately needed for a single validation check in order to find those with the strongest relationships with one another (using standardised methods such as structural equation modelling, confirmatory factor analysis, comparative fit index, and incremental fit index) and eliminate the weaker related questions. The final list of questions would then be available for repeated testing to increase validity and robustness. This type of framing validation would be beneficial for any number of projects. Facilitators could use this kid of tool to help determine, prior to SP development, whether the practitioners have dichotomously different views of the same organisation. Of course, researchers could use such a tool for further explorative testing of ST mechanisms, along with revealing strengths and weaknesses within SP.

Chapter 7. **Creative Thinking (Stage 2)**

“You were not there for the beginning. You will not be there for the end.
Your knowledge of what is going on can only be superficial and relative”
William S. Burroughs, *Naked Lunch*

In this chapter, the business vignettes developed as a proxy for IL/ST framework Stage 1 are used to deliver the priming stimuli (independent variable) in two studies designed to measure various qualities of creativity (dependent variable) promoted in Stage 2. The chapter begins with a focused literature review that completes the methodological background for developing two experimental studies to help better understand the creative elements of ST. Five hypotheses are presented. Together, hypotheses 1 and 2 aim to help answer research questions 1 and 2. Hypothesis 3 aims to provide exploratory, novel insights for research question 4. Finally, hypotheses 4 and 5 aim to help answer research question 3.

7.1 **Individual Creativity**

Two studies measure the anchoring bias on the particular style of creative thinking employed early in the scenario process (Stage 2). As discussed in Table 3.1, Stage 2 tasks practitioners with determining as many DF as possible that could have an impact on the organisation and their ability to reach their goals within the designated timeline. The deliberate act of exploring one’s memories to find relevant information in a novel and open-ended manner (i.e. no limits on quantity or quality) is an effort of creativity.

Creative thinking is not a single cognitive act, however, but reflects an interplay of separate, dissociable cognitive mechanisms (Chermahini & Hommel, 2010; Dietrich, 2004; Eysenck, 1993; Heilman, 2005). Though the psychological literature still debates over what precisely “creative thinking” is comprised of and how it is identified, there are some main cognitive mechanisms which dominate the literature. Many propose that creative thinking involves memory retrieval (Smith, 1997), deliberate and spontaneous thinking (Arden, Chavez, Grazioplene, & Jung, 2010; Csikszentmihalyi, 1999; Eysenck, 1995; Martindale, 1997), explicit and implicit thinking (Proctor & Capaldi, 2012), divergent and convergent thinking (Guilford, 1950, 1967) and associative fluency (Lee & Therriault, 2013). Creativity can be deliberate or through unconscious processes. In Stage 2, practitioners must employ their "creative centres" to access their working memory, compare the lexicon of information they possess against a script and rules of the task, judge each piece of information against the rule and determine its appropriateness. The output (i.e. DF) reflects the various qualities of creative thinking. First, there is the act of discovery. ***Discovery*** is achieved through active, deliberate searching for DF. Creativity is also closely linked with the concept of imagination (Singer, 1999). The capacity for imagination is key later in the SP process, as practitioners must combine their discoveries into coherent scenario narratives about plausible futures, indicators, and outcomes. Imagination is particularly beneficial for developing black swan (Taleb, 2010), outlier (Inayatullah, 2008), and peripheral (Ducot & Lubben, 1980) scenarios that break from business-as-usual thinking. The quality of creativity employed in the earlier stages of SP, however, is a different than the later stages.

Of the varied functions that make up creative thinking, the two studies in this chapter focus on divergent thinking. ***Divergent thinking*** is an inductive process that requires the practitioner to generate a broad range of ideas from a particular starting point towards a specified solution through open-ended prompts (Guilford, 1967; Lee & Therriault, 2013). This task is often referred to as "brainstorming" (Bradfield, Derbyshire, & Wright, 2016; Bradfield, Wright, Burt, & Van der Heijden, 2005; Bradfield, Cairns, & Wright, 2015; Davis, Bankes, & Egner, 2007) which is a function of creative thinking (Baas, De Drue, & Nijstad, 2008; Lamm & Trommsdorff, 1973). Joy Paul Guilford was an early pioneer in defining, measuring, and testing creativity. An empiricist at heart, Guilford believed the black box cognitive effort of "creativity" could be objectively studied. He is credited with being the first to propose our present-day understanding of divergent and convergent thinking. Guilford (1950) identified four major qualities of divergent thinking that lend themselves to quantitative analysis: fluency, flexibility, elaboration, and originality. ***Fluency*** is defined in terms of productive output. A creatively fluent person is profiled as producing a large number of ideas. ***Flexibility*** is defined as having the capacity for discovering diverse ideas that relate to a variety of categories.

Elaboration occurs when a person follows an associate path from one idea to the next through several related links. ***Originality*** is related to the statistical infrequency of the idea (Runco & Acar, 2012). The last quality, originality, makes for an interesting feature to measure against the previous three. When speaking of a "creative" idea, the idea is considered more than just original, but it must be in some way original, because inherent to the concept is that all creative things are original in some manner (Runco, 2014b). That originality may take the form of novelty, uniqueness, unusualness, or unconventionality. Therefore, it is important to take this moment to clarify Guilford's use of the term "originality" in order to dispel any misunderstanding due to the ubiquity of the term. These measures of creative, divergent thinking were further explored by Ellis Paul Torrance (1966), who created the still popular *Torrance Test of Creative Thinking*, and Micheal Wallach and Kogan (1965) with their *Wallach-Kogan creativity test*. There are many others as well, but most (arguably all) are expansions of Guilford's and Torrance's original work, and vary in test-specific features that do not apply to this SP exercise. Fluency, being a more global measure, is also the most popular measure to use with creative, divergent thinking tests (Runco & Acar, 2012). However, the other three measures are believed to be more closely related to creativity and offer reliable and unique variance (Runco & Albert, 1985). By offering a more nuanced view of divergent thinking, it is recommended to measure all four qualities, and interpret their results together.

Framing business vignettes with different threat levels can potentially anchor creativity through mood priming. Mood is arguably the most agreed upon predictor of creativity (Mumford, 2003). In their meta-analysis of the relationship between mood and creativity, Baas, et al. (2008) show that positive moods – in the form of activating, approach motivation, and promotion focus – produce more creativity than negative moods. One way to summarize this is that happiness bolsters creativity compared to anxiety or fear. The high-threat vignette presents an organisational environment that is seen as more difficult to achieve their goals and more stressful for management, compared to the the low-threat vignette. Differences in response output that follow Baas, et al.'s findings are also reflected in the first-round priming validation tests for the business vignettes from Chapter 6, where participants given high-threat framed vignettes consistently entered fewer ideas, environmental factors, issues, and events in their responses. Increases in perceived difficulties and stress are often met with increases in anxiety (Melchior, et al., 2007), fear (Lazarus, 1999) and avoidant motivations (Drach-Zahavy & Erez, 2002). Any emotional responses by the participants to the vignettes, however, are most likely to be low given the lack of personal investment in the profiled organisation. Such low-level emotional effects are, indeed, reflected in the effect sizes (r^2) of the first-round stimuli validation analyses. However, a strong framing effect was still detected in the second-round validation test. the task mimics in many ways a typical SP workshop (i.e. preserving ecological validity), where practitioners are projecting images of their selves and others (often faceless stakeholders) into unknown future territories, and therefore are, by virtue of active SP, removed a few steps from an immediate personal investment or familiarity. Differing moods can lead practitioners to conceptually partition or rank the lexicon of their memories differently. Working memory plays a significant role in creative thinking processes (Lee & Therriault, 2013). It is by way of accessing working memory that people discover DF and judge them as relevant or not. Differently framed messages may have the effect of implicitly guiding participants' attentions towards different environmental factors (i.e. anchoring bias). If we think of the environment as posessing 100% of all possible relevant DF, then it may be the case that differing levels or qualities of organisational threat prime different levels of mood in the pracitioners which anchor their awareness to different relevant DF, resulting in high-threat and low-threat framed participants identifying different subsets of the total possible relevant DF. The first hypothesis reflects literature review reported within this section and the validation testing of Chapter 6. The first hypothesis is aligned with helping to answer research questions 1 and 2.

H1: Participants who read a high-threat business vignette will generate lower creative output (fluency, flexibility, elaboration, and originality), than participants who read a low-threat business vignette.

7.2 Group Creativity

There is the view that practitioners who generate ideas on their own could experience unavoidable creative limitations. Generating ideas from an open-ended question can lead to what some refer to as “functional fixity” (Gute & Gute, 2015, p. 19). This refers to people’s tendency to rely on cognitive habits and heuristics during the information generation process. Such routine-like thinking limits the scope of memory and knowledge combinations made during creative thinking. However, some research suggests that making one’s thoughts explicit through a group effort can create an anchoring effect on the verbalized, explicit ideas, therefore limiting the scope of further divergent thinking (Gute & Gute, 2015). A similar anchoring effect during the creative process within a group effort was reported by Bradfield (2008, p. 209), which he called “punctuated equilibrium induced coagulation”.

Creative thinking as an interplay between the exterior and interior landscapes, as well as between the friction and flow of the process (Gute & Gute, 2015). Mackay & McKiernan (2018) argue that thinking thrives on both individual and group creativity, where an interplay of intuition and emotions develop scenarios into socially constructed realities. From the act of discovery comes the potential for ***serendipitous*** insight (Guilford, 1967; Lee & Therriault, 2013). In SP, serendipity is facilitated by the group-level dialogue that practitioners engage in while working through the seven IL stages. Serendipity is the spontaneous discovery of new DF by a single member inspired by the surprise insights shared by other group members. As Burke (1995, p. 289) states, “one thing [discovered] leads to the discovery of another,” and another, and another, etc. Working alone helps one avoid the conformity that could come from group efforts, while working in a group can offer the benefit of unexpected external influences.

Perhaps, then, the method to increase creative thinking is a combination of both individual and group efforts. Adding a group discussion to the individual portion of creative discovery may help achieve such interplay more successfully. The process of sharing each member’s ideas can have the effect of a creative conversation. A benefit to SP is the ability to revisit previous stages at any point during the process.

Therefore, the benefits of individual creative thinking bolstered by group discussions, can continue to be effective throughout the process.

In the group-based portion of the study, each participant will take their own generated DF from their individual creative work and present them to their group. Group SP work requires participants to justify their DF while engaging in dialogue and remaining open to generating potentially new DF, as a group effort. Whereas the first half of the study looks at individual-level cognition, the second half looks at moving from individual-level to the group-level cognitions, and how verbalising justifications to the group, as is standard in SP, may effect creative divergent thinking. The second hypothesis (H2) reflects efforts to move from individual creative work to group-based creative work. Given the potential benefits and limitations of group-based SP, H2 reflects a position similar to individual-level expectations on creativity and anchoring. The second hypothesis is aligned with helping to answer research questions 1 and 2.

H2: Groups who read a high-threat business vignette will create fewer new DF and eliminate fewer existing DF during the group-portion of SP, compared to groups who read a low-threat business vignette.

7.2.1 *Group Composition*

Not all groups are the same. Runco (2014b) suggests small, heterogeneous groups can maximise creative insights. Expert input brings a wealth of knowledge to the table, but can bring traits of inflexibility, as well. Inflexibility can come from a lack of imagination, motivations to preserve one's invested interests, or fear of devaluation of one's knowledge. However, a heterogeneous mix of expertise and knowledge may stimulate healthy group discussions, particularly during brainstorming sessions, such as those experienced in Stage 2. Heterogeneity can also mean including novices in a SP group. Novice group members may bring greater flexibility and open-mindedness as an outcome of less crystalised historical or professional knowledge.

A "small group" is defined as five to six members in this study. Too few group members and the SP workshop risks falling into a rut with less challenging dialogue, lower idea generation, and quick agreement on shared ideas, or worse, groupthink (Janis, 1982). Too many and the group risks losing valuable input from less-forthcoming members at the sake of more outgoing, confident members, or lack of time & energy to allow full input by all members. The cost of sharing an idea – as opposed to thinking it to oneself – is higher in group dialogue, and therefore should be acknowledge when developing SP groups (Runco, 2014b). The personal risks are related to the reasons for inflexibility mentioned in the previous paragraph. Five to six group members allows for at least two experts (to challenge well-developed conventional thinking) and two novices (to bring disparate open-mindedness and flexibility), with the hopes of optimizing the "friction and flow" of the strategic dialogue.

7.2.2 ***Strategic Dialogue***

According to Runco (2014b, p. 159) divergent thinking (i.e. brainstorming) as a group has three basic requirements:

(1) avoid judgment
(2) focus on the quantity (not quality) of ideas
(3) use other group member's ideas as a springboard for one's own

Avoiding judgment requires group members to value each input equally, at least initially. Focusing on quantity over quality maximises fluency. Spring boarding from other group members' ideas becomes an act of serendipity, which itself, can maximise the four major qualities of divergent thinking: fluency, flexibility, elaboration, and originality.

Specific dialogic methods from Schweiger, Snadberg, & Ragan (1986) were adapted in order to control for potential confounding variables that could arise from unstructured group discussions. Ideally, SP facilitators would prefer to minimise the issues mentioned earlier (functional fixity, anchoring, groupthink, lack of input) and maximise Runco's three requirements, with the aim of increasing creative output from the group as a whole. There are, of course, other factors to consider with strategic group dialogue, such as prematurely smoothing over conflict (Brodwin & Bourgeouis, 1984) and social loafing (Karau & Williams, 1993). The two chosen dialogic methods aim to address all these points, whether primarily or implicitly. Schweiger, Snadberg, & Ragan (1986) provide two strategic conversation methods, Dialectic Inquiry (DI) and Consensus (CS), empirically tested, which present dialogic variables for testing against group-level creativity in a SP setting. DI and CS resulted in the highest degree of differences in group decision-making, therefore making them ideal to compare against each other in a group creativity exercise to inform future SP methods. If a heterogeneous group is seen as optimal for group creativity, then by necessity, opposing opinions, ideas and strategies will be brought into the dialogue. The DI method allows for a formal intragroup conflict technique that capitalises on the heterogeneity of a group in order to reach final agreement. The CS method addresses the necessity of reaching group consensus. Regardless of dialogue, however, agreement/consensus in some form must be achieved at each SP stage before the group, as a whole, can move forward.

DI derives from the Hegelian dialectic (Schweiger, Snadberg, & Ragan, 1986). The method uses the reflexive practice of presenting an abstractness and countering with a negative, with the purpose of leading to a concrete, absoluteness of synthesis. DI follows this method by pitting diametrically opposed ideas against each other (i.e. abstractions countered by negatives). Debate is then encouraged in a prescribed manner to maximise conflict and critical evaluation. Mitroff (1982, p. 222) describes the debate as "active, heated, and intense," which can lead to serendipitous insight amongst the engaged group members. Critical evaluations allow each group member to make explicit the strengths and weaknesses of their ideas (Mason & Mitroff, 1981). The intended result is a synthesis that retains the ideas and logic that survive the inquiry. With every positive, however, the negatives should be acknowledged. Unchecked conflict can increase distance between group members, unintended insults, a feeling of devaluation of one's ideas, and overbearing and silenced voices.

Any of these negative outcomes from increased conflict can have inhibitory effects on group creativity, not to mention leaving participants in a foul mood!

CS takes a less "heated" approach to reaching group agreement. Similar to the DI method, group members are encouraged to present their ideas and the logic behind them. However, instead of then engaging in intense, critical evaluation of each idea and its logic, group members freely discuss each idea in a supportive manner, until they reach a group agreement on all the topics (Hall, 1971). Schweiger, et al. (1986) suggest that CS creates a less combative environment compared to other debating methods. Comparison of the DI and CS methods allows this study to extend the Schweiger, et al., study into qualitatively different group settings – of SP– and measure group creativity with valid and robust psychometrics. The third hypothesis reflects the comparative element of dialogic methods, DI and CS, and is aligned with helping to answer research question 4.

H3: Groups who engage in DI will create fewer new DF and eliminate fewer existing DF during the group-portion of SP, compared to groups who engage in CS.

7.3 **Confidence and Creativity**

Confidence measures are taken at the end of each experimental session to help determine whether a relationship exists between message anchoring and confidence. Confidence is shown to be a strong motivator in decision-making (Lichtenstein & Fischoff, 1977; Nickerson, 1998). Fischoff & MacGregor (1981, p. 11) state that, "Forecasts have little value to decision makers unless it is known how much confidence to place in them," but follow-up that confidence has "little value unless forecasters are able to assess the limits of their own knowledge accurately." As knowledge increases, so too can confidence, but at a much faster rate, creating an unreliable bias in overconfidence (Tversky, 1992). As information confirms a person's position, confidence can artificially increase (Kuhn, Weinstock, & Flaton, 1994; Nickerson, 1998; Tetlock & Kim, 1987). Some research shows strong relationships between level of creative output and confidence in the quality of the output (Poon, et al., 2014; Rauth, et al 2010). Lox (1992) shows that perceived threats and uncertainty can have a negative impact on confidence, resulting in lower self-rated scores. While other studies suggest there may be no relationship, under certain conditions (Fisher & Statman, 2003; Koriat, Lichtenstein, & Fischhoff, 1980; Wagenaar & Keren, 1986). Group studies, on the other hand, show a reliable increase in confidence as a product of group-based work (Boje & Murnighan, 1982; Lee, Tinsley, & Bobko, 2002; Sniezek, 1992; Sniezek & Henry, 1989; Zarnoth & Sniezek, 1997). Confidence, as a cognitive motivator, both within and towards others, is considered so ubiquitous in our decision making that it has been identified as an entire category of collected biases, in and of itself (Kahneman & Klein, 2009; Kahneman & Tversky, 1979; Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1974). From these findings, two final hypotheses are developed which align with research question 3.

H4: Average confidence scores will be lower for participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.

H5: Participants will increase their confidence, overall, after participating in group sessions.

7.4 **Ethical Approval for Human-Based Study**

The ethical review for the individual study was submitted through the Department of Strategy and Organisation within the SBS to the University Ethics Committee (UEC) and receive approval on 8 May 2015. Approval for the group study was received from the UEC 21 November 2016.

7.5 **Apparatus & Stimuli**

Two vignettes (2 and 3) are used for the framing stimulus (high-threat vs. low-threat). Vignette 2 is presented to participants who took part in the individual-only study. Vignette 3 is presented to participants who took part in the individual+group study. The strategic dialogue stimulus (DI vs. CS) is developed from the Schweiger, et al. (1986) study which developed their guidelines from Mason (1969) and Mason & Mitroff (1981). The DI and CS instructions are presented side-by-side in Table 7.1. Participants received only one set, either just the DI instructions or the CS instructions. To control for confounding information, the instructions were adjusted to ensure each group received the same number of steps (8). To increase the influence of the dialogic method on each group member, key components were *italicized*, <u>underlined</u>, and **emboldened**. Language was included to help avoid social loafing and increase equity of input from each group member. Copies of the strategic dialogue instructions were printed out, along with business vignette 3, and presented to groups to review at their own pace.

Table 7.1. Strategic dialogue instruction for Dialectic Inquiry and Consensus methods.

Step	Dialectic inquiry	Consensus
1	With your group, let each person discuss her/his driving forces from stage 1, which are printed out and in your group rooms. All group members will take turns presenting their driving forces while the other members listen.	With your group, let each person discuss her/his driving forces from stage 1, which are printed out and in your group rooms. All group members will take turns presenting their driving forces while the other members listen.
2	Presenting one person's driving forces at a time will allow each person sufficient time to explain.	Presenting one person's driving forces at a time will allow each person sufficient time to explain.
3	Group members should listen to each person's reasoning and try to think of plausible assumptions which ***negate*** the reasoning behind each person's driving forces. ***That is, while the presenter is giving their reasoning for their list of driving forces, you, the listener, should try to think of why the reasoning is <u>not fully correct</u> , <u>could be improved</u>, <u>the validity behind their reasoning</u>, and if applicable, <u>give recommendations for improvement</u>.***	Group members should listen to each person's reasoning. Everyone should openly discuss the presenter's reasoning and ask for clarification and elaboration when needed to help everyone in the group understand. The goal is to reach a consensus and understanding.
4	All listeners should be ***active*** in asking clarifying questions and presenting any negating plausible	It is not necessary that everyone agree on all driving forces and their reasoning to reach a consensus. Only that everyone can understand

	assumptions to the presenting person.	and accept each presenter's driving forces on the basis of logic and the willingness to consider them plausible.
5	Both presenters and listeners should discuss openly with the goal of this discussion to reach a final list of driving forces for each person, and the group, that everyone agrees on.	Everyone should discuss openly with the goal of this discussion to reach a final list of driving forces for each person, and the group, that everyone agrees on.
6	Cross out - X - any driving forces from stage 1 that the group agree are no longer needed. However, it may be the case that all original driving forces remain. This will depend on your group discussion.	Cross out - X - any driving forces from stage 1 that the group agree are no longer needed. However, it may be the case that all original driving forces remain. This will depend on your group discussion.
7	If you think of more driving forces, enter them into the survey from the link provided in your email. However, it may be the case that you have no additional driving forces to enter after the group discussion. This will depend on your group discussion.	If you think of more driving forces, enter them into the survey from the link provided in your email. However, it may be the case that you have no additional driving forces to enter after the group discussion. This will depend on your group discussion.
8	When you are finished, whether you entered more driving forces or not, please submit your survey.	When you are finished, whether you entered more driving forces or not, please submit your survey.

Note: The use of "stage 1" in this context refers to the individual session of their participation.

7.6 **Methods**

Participants (N = 218) were recruited from UK and UAE. Some participants took part in the individual-only session of the study (n = 117), while the rest took part in the individual+group study (n = 101). Participants were randomly assigned to a condition, high-threat (n = 106) or low-threat (n = 89) for the individual portion of both studies. For the individual+group study, after participants completed their individual tasks, clustered random sampling was used to assign members to a group. This means all participants who individually read a high-threat framed vignette were randomly assigned to a group with other participants who also read the same high-threat framed vignette. The same sampling method applied to participants in the low-threat condition. Every group had at least two confirmed "experts" in their respective fields, and all members qualified as "novices" with regard to the target organisation in the study. Table 7.2 presents a breakdown of vignette profiles that were used to reflect sample demographics.

Table 7.2. Business vignette profiles presented to different participant samples

Stage	Experiment	Vignette	Profile		
			UK	UAE	US
2	Individual-only	2	✓	✓	
	Individual+group	3	✓		

This is a between and within subjects design. Groups were comprised of either five or six participants. Qualtrics (2019) online survey platform and pen & paper methods were used to administer the stimuli. For the individual-only session, participants were given a link to the online survey which presented them with one of the business vignette conditions, asked to read everything completely, and enter as many DF as they could think of into the online survey. The individual-only participants completed their participation at the end of the individual session.

The individual+group participants performed the same individual session online, but were also given the instructions, *"All DF will be discussed as a group on Friday at 13:00,"* which was the following afternoon. To facilitate the group discussions that were to commence the following day, each participant's DF were printed out onto their own post-it size paper. This would allow all group members to see each other's DF contributions as they discussed them and easily remove the unwanted DF. A stack of blank post-its and markers were provided to each group to allow them to add more DF if they felt this was necessary.

For the group session, groups were semi-randomly given one of two possible instructions for strategic group dialogue. Randomness was controlled for to ensure each business vignette condition (high-threat vs low-threat) received the same approximate assignment to each of the dialogue conditions (DI vs CS). Five participants did not return the following afternoon for the group session of the study. Groups were assigned either the DI method (n = 43) or the CS method (n = 53) for listening, questioning, and developing their unified group list of DF. Participants were allowed to keep, delete, and add as many original DF as they liked for their final group list. Participants were asked to give their confidence scores (Likert scale: 1(not at all) - 5(completely)) twice, once after the individual session and again after the group session. Confidence scores were collected through Qualtrics, immediately after each session. Participants were asked, “How confident are you that you have captured all the relevant driving forces concerning the [company’s] goals for the next 5 years?”

7.7 **Results**

To begin, 23 surveys were removed from the individual-only session for being incomplete. This study uses a product-based approach to creativity. That is, the production of ideas is considered a reflection of creative thinking. Creative thinking is operationally defined as production of DF.

7.7.1 ***Individual***

Hypothesis 1 states that participants who read a high-threat business vignette will be less creative than those in the other condition. To test H1, four measures of creative thinking are analysed using chi-square test of independence and goodness of fit, as well as qualitative content analysis of text. Creative thinking is measured on four qualities: fluency, flexibility, elaboration and originality. These qualities are operationaly defined to reflect the creative output of ST.

- Fluency refers to the number of ideas generated (e.g. how many DF are generated).
- Flexibility is the variety of categories within the generated set of ideas (e.g. how many STIRDEEPER categories apply to the DF).

- Elaboration addresses the level of detail offered in the generated idea (e.g. how many DF in each of the 10 STIRDEEPER categories).
- Originality is based on the statistical infrequency or uniqueness of the ideas (e.g. are there any DF that don't fit into the 10 STIRDEEPER categories).

Quantitatively speaking, the higher the score within any of these qualities, the more creative the thinking. Qualitative analyses are also used to help reveal any differences in creativity by message framing or dialogue conditions, particularly with originality.

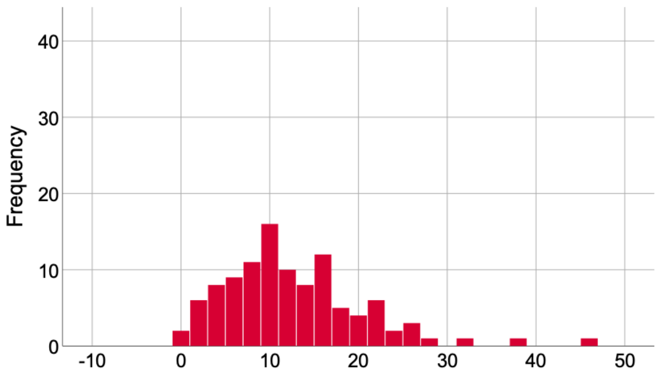
Before proceeding with the expected analyses, the data are explored for outliers. Table 7.3 presents the descriptive statistics. Descriptive statistics for the two conditions reveal similar averages and large ranges in DF. Skewness and kurtosis are calculated to help determine how extreme the distribution of DF may be. Skewness measures the direction of the distribution of DF. A normal distribution will reveal a skewness of zero, however the data have a lower bound because the lowest value for DF is zero, therefore a right-tailed distribution is expected, but an adjusted Fisher-Pearson coefficient > 1.02 will reveal an extreme skewness. Kurtosis measures the severity of the distribution of DF. The formula is adjusted to zero so positive values reveal a heavy-tailed (i.e. outliers) distribution and negative values reveal a light-tailed distribution. Figure 7.1 illustrates the distribution of both conditions. Table 7.3 & Figure 7.1 reveal that there are potentially outliers within the data. Z-scores were calculated, by condition, to determine upper and lower boundaries. Figure 7.2 presents boxplots which reveal the outliers for each condition. Three outliers were removed from the high-threat condition and four from the low-threat condition. All data from these seven participants were removed from the remaining analyses.

Table 7.3. Descriptive statistics by condition for raw data.

	Condition	
	Hight-threat	**Low-threat**
n	106	89
M	12.08	12.69
SD	7.88	9.46
Min	0	3
Max	45	53
Range	45	50
Skewness	1.19	1.86
Kurtosis	2.55	4.18

Figure 7.1. Distribution of driving forces by condition

7.1a. High-threat



7.1b. Low-threat

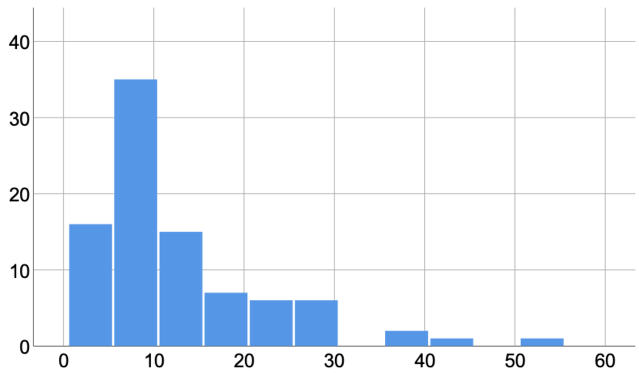
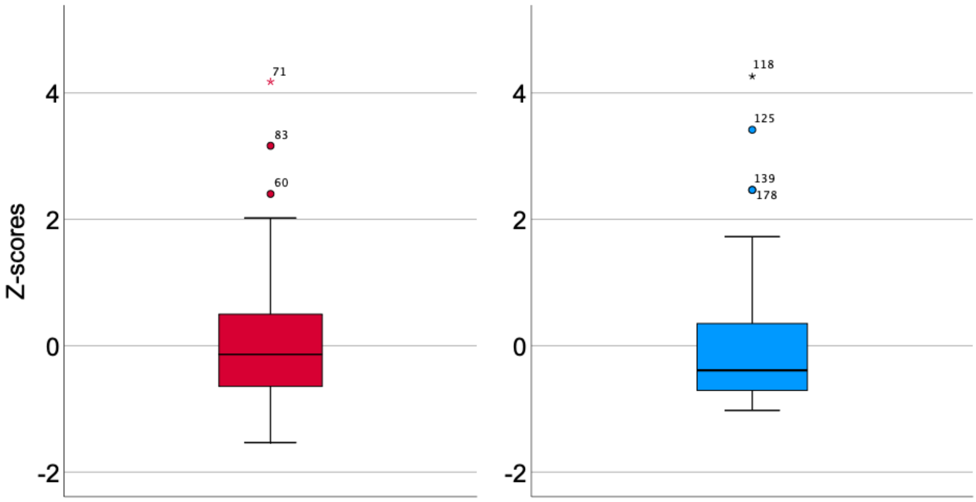


Figure 7.2. Outliers by condition

7.2a. High-threat 7.2b. Low-threat



Note: The numbers within the figure represent the numerical position of the participant in the list as assigned by SPSS. For example "71" is the 71st participant in the list.

Fluency is measured by calculating the average number of DF by condition (high-threat vs low-threat). When the data from all participants (individual-only session and individual+group sessions) are analysed together, there is no significant difference in fluency between conditions. However, if the data are divided by those who participated in the individual-only session from those who participated in the individual+group sessions, an interesting story emerges. Participants who were asked to only enter their DF, with no expectation of group work to follow, produced significantly fewer DF in the high-threat condition than those in the low-threat condition ($M_{high\text{-}threat}$ = 8.44, SD = 5.71, V = 32.59; $M_{low\text{-}threat}$ = 11.77, SD = 7.38, V = 54.53; $F(1,89) = 5.69$, $p = .02$, $\eta^2 = .60$). However, the behaviour reverses when participants expected to discuss their DF with their future group members. The high-threat condition results in significantly *more* DF being produced compared to the low-threat condition ($M_{high\text{-}threat}$ = 13.40, SD = 6.39, V = 40.89; $M_{low\text{-}threat}$ = 10.65, SD = 6.14, V = 37.73; $F(1,95) = 4.37$, $p = .04$, $\eta^2 = .44$).

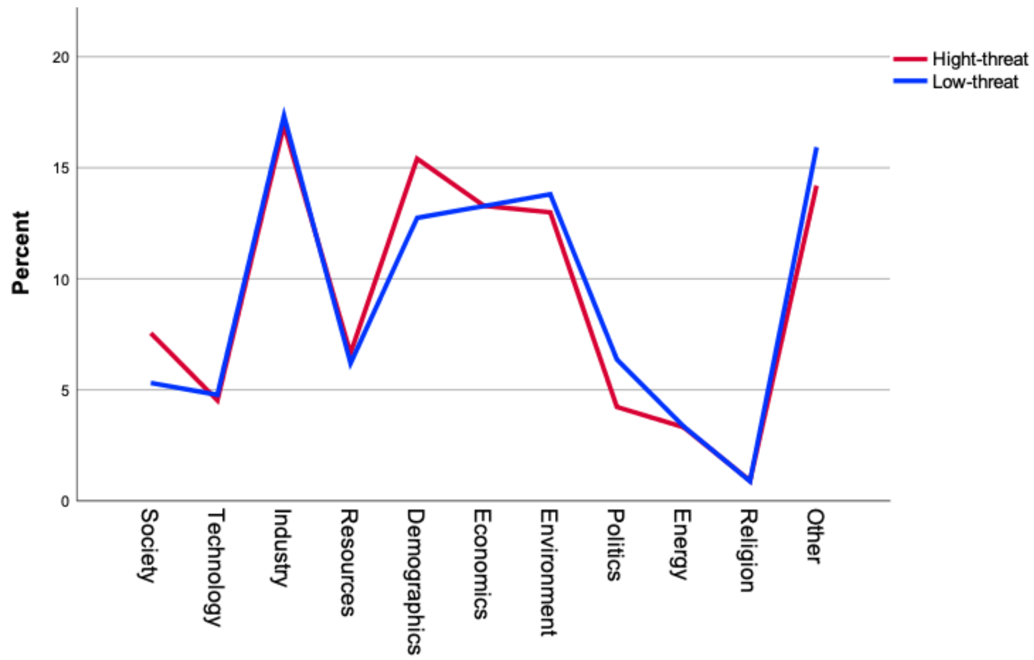
The remaining analyses divide the data between the participants who expected to work alone verses those who expected to work as a group for two reasons. First, prior analyses have now shown that there appears to be a difference between these two states of identity. Second, the qualitative analyses will investigate the specific word choice of the participants. The word choice will be context dependent, that is, specific to the organisation featured in the vignette. Participants in the individual-only session received vignette 2, and participants of the individual+group sessions received vignette 3.

To analyse flexibility of participants' creative thinking, DF are assigned to their corresponding STIRDEEPER category with an eleventh category *Other* for any DF that do not fit into the first ten. For example, "use of zoo for educational purposes" is categorised under *Society*, and "existing asset base (land)" is categorised under *Resources.* To increase accuracy in categorising the DF, three raters were sourced to categorise the data. All raters were familiar with STIRDEEPER use in business models and with SP. The guidelines for raters started with rating on word choice. If a categorical word was in the DF, then it was to be initially assigned to the corresponding category. For example, "changing political landscape" has the word "political" in the description, therefore, it would initially be assigned to the *Politics* category. The next guideline rated on content. Regardless of word choice, if the content of the driving force focused on a particular topic, then the rater was to assign the corresponding category of the topic. For example, "lack of zoo employees who speak Arabic" has the word "employees" in the description, and *Employment* is a subcategory of *Industry*. Therefore, the first rating would be by the first guideline, word choice. However, when considering the content it may be determined that employment is not the main focus of the driving force, but rather language barriers, and a rater may have categorised the driving force by a linguistic focus, which could result in a different category being chosen for the score. Due to time restraints, the three raters received overlapping portions of the data. The proportional division of data ensured a counter-balancing of raters and at least two ratings for each DF. An interclass correlation (ICC_1) was conducted on all raters' scores. DF were scored, discussed, and re-scored until ICC was > 0.90. A high degree of reliability was found between the raters. The average ICC was .952 (95% CI from .815 to .993, $p < .001$). The remaining DF were then individually discussed until an agreement was reached.

The individual-only participants, in both conditions, generated DF in all ten STIRDEEPER categories, as well as DF that did not fit into the ten categories. The participants' creative flexibility in divergent thinking is, at a superficial level, the same regardless of perceived threat in a business environment. The individual+group participants showed only a slight difference in elaborative expression. The high-threat condition considered all but the *Religion* category, whereas the low-threat condition considered all categories. Both conditioners also generated DF that did not fit into the ten STIRDEEPER categories.

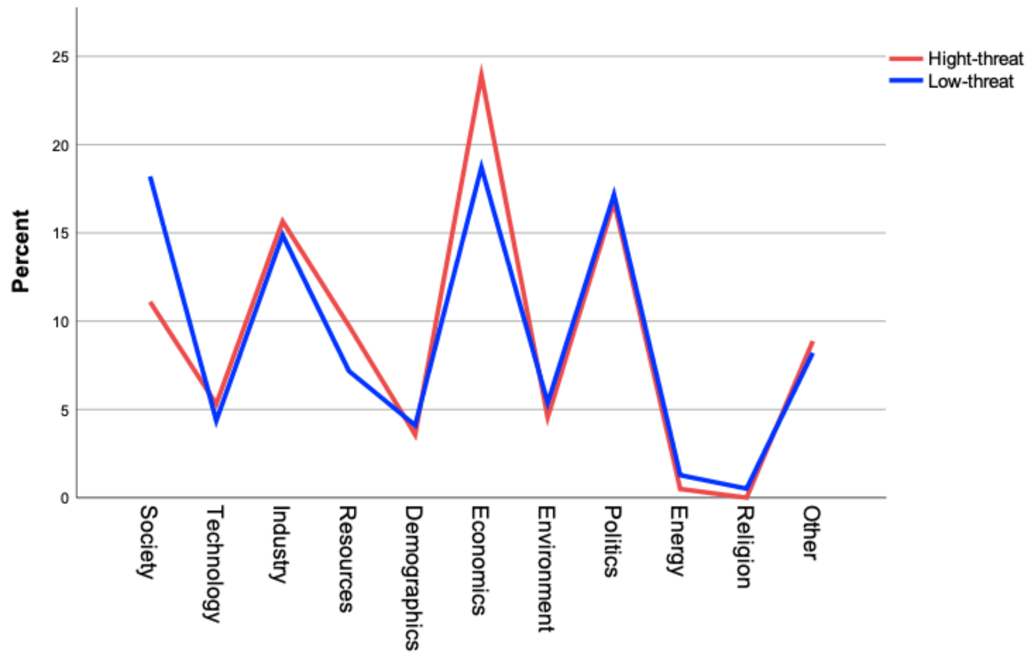
At the individual-only level, elaboration is reported using the proportion of DF in each category by condition. Figure 7.3 shows the proportion of DF from the high-threat and low-threat conditions. A Pearson Chi-Square reveals no significant relationship in elaboration between conditions ($\chi^2(10, 896) = 5.14$, $p = .88$). Both conditions share the same mode response, *Industry.* There is, however, significant differences in elaboration within both conditions ($\chi^2_{high\text{-}threat}\,(10, 331) = 114.02$, $p < .000$; $\chi^2_{low\text{-}threat}\,(10, 565) = 196.88$, $p < .000$). Participants in both conditions elaborated significantly more (> 10%) in *Industry, Demographics, Economics, Environment,* and *Other.*

Figure 7.3. Individual-only elaboration of driving forces divided into STIRDEEPER categories, plus Other



Unlike the individual-only sample, different behaviours are revealed between the conditions when participants anticipate being a group member in the near future (i.e. individual+group). To begin, there is a relationship between the conditions (χ^2 (10, 1201) = 22.72, p = .01). The relationship is illustrated in Figure 7.4. Both conditions also share the same mode response, *Economics*. There are, however, significant differences in elaboration within both conditions ($\chi^2_{\textit{high-threat}}$ (9, 811) = 371.01, $p <$ .000; $\chi^2_{\textit{low-threat}}$ (10, 390) = 203.61, $p <$ .000). Much like the individual-only participants, those consigned to groups elaborated significantly more (> 10%) in the same categories between the two conditions: *Society, Industry, Economics,* and *Politics.*

Figure 7.4. Individual+group elaboration of driving forces divided into STIRDEEPER categories, plus Other



Note: No participants in the high-threat condition considered DF within the category of *Religion*. ICC = .94 (95% CI from .86 to .991, $p < .001$). The remaining DF were individually discussed until an agreement was reached between the raters.

Originality in creative thinking is based on the statistical infrequency or uniqueness of the ideas. First, the proportion of DF by category is calculated (other/(∑fluency). However, this is a global measure, which is the least powerful measure of originality. As stated earlier in the chapter, the anchoring bias could potentially be revealed through different DF being considered more important than others, between conditions. A closer inspection of the DF was conducted using the following packages and libraries in R: tm, tokenizers, dplyr, tidytext, wordcloud2, stringr, ggplot2, tidyr, corpus (R Core Team, 2019).

The data are first cleaned using R programming. The DF are unnested to divide all words from each other. Next, all standard stop words (e.g. “of”, “and”, “for”, “with”, “to”, “a”, “if”, etc.) are removed from the list of DF, along with “"change", "changes", "impact", "impacts", "degree", "effect", "effects", "due", "i.e", "e.g", and "level". The reason for removing versions of "impact", "change", "degree", and "level" is because participants were specifically prompted to begin each DF statement with these terms. This format of creating a DF statement (e.g. "level of investment") designates the beginning words as highly shared between the two conditions, and therefore not of value in determining the key concepts generated by participants within the two conditions. Single numbers (e.g. "1", "2", etc.) are removed with the exception of years and attached numbers (e.g. "3d" and "4d") since these are meaningful uses of numbers. Finally, data are cleaned to standardise the language. For example, "habitats" was changed to "habitat", "govt" and "governments" were changed to "government", and misspelled words such as "vetinary" to "veterinary". This allows for an accurate frequency measure to determine the popularity of the DF concepts.

At the individual-only level, originality in creative divergent thinking is not statistically different. The proportion of DF in the high-threat condition (.14) and low-threat condition (.16) are similar. Interestingly, there appears, at first, to be no anchoring bias on originality of DF. Combining both conditions, there are 136 DF within the *Other* category. Given that participants considered a significantly higher number of "original" DF, which were grouped together into a single category, and the variety of this category has the potential to be broader than the other 10 more strictly defined categories, the *Other* data require further exploration.

The results of the cleaned data reveal that participants in the individual-only session produced 1130 words in the high-threat condition and 2208 words in the low-threat condition. A minority of topics were shared between the two conditions (38%). The highest shared 15 topics are presented in Figure 7.5.

Figure 7.5. Top 15 most popular topics shared between conditions within the individual-only session.

Figure 7.5.a

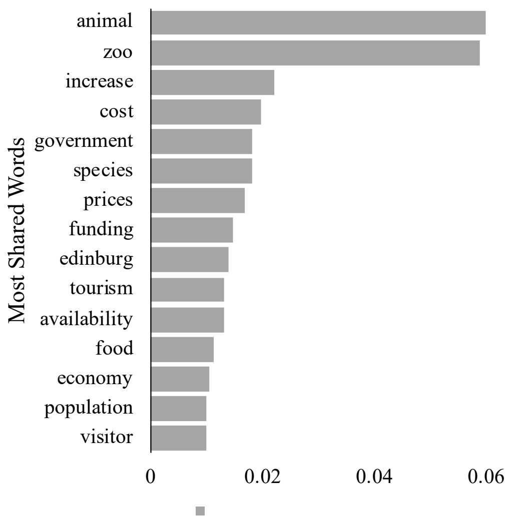


Figure 7.5.b

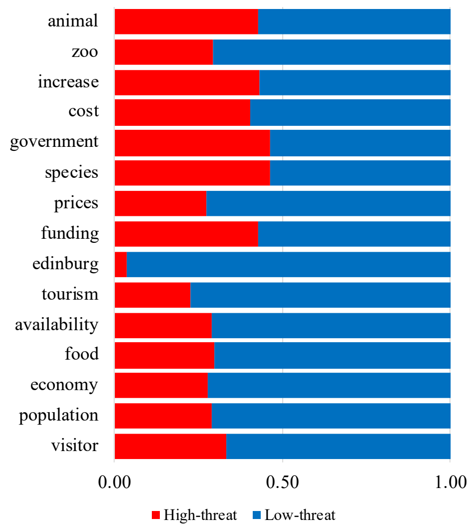


Figure 7.5.a shows the overall distribution. The top two most popular shared concepts are as expected. The profiled organisation was the local zoo. It should be expected that SP participants would mention the *Industry* related topics "zoo" and the main focus of a zoo (i.e. "animals" and "species") more than any others. The remaining topics span the categories of *Economics* ("cost", "prices", "funding", and "economy"), *Society & Demographics* ("population", "visitor", "tourism", "edinburgh"), *Resources* ("food", "availability"), *Politics* ("government"), and apparently a shared focus on "increasing" factors. To understand if participants from one condition were more focused a topic more than the other condition, a comparison of the proportional distribution of mentions by condition is presented in Figure 7.5.b. In every instance, the high-threat condition participants gave less attention to the topic compared to the low-threat condition. The proportions reverse with topics lower down in the ranking, but out of 282 shared topics, only 17.7% of the topics are given greater attention within the high-threat condition.

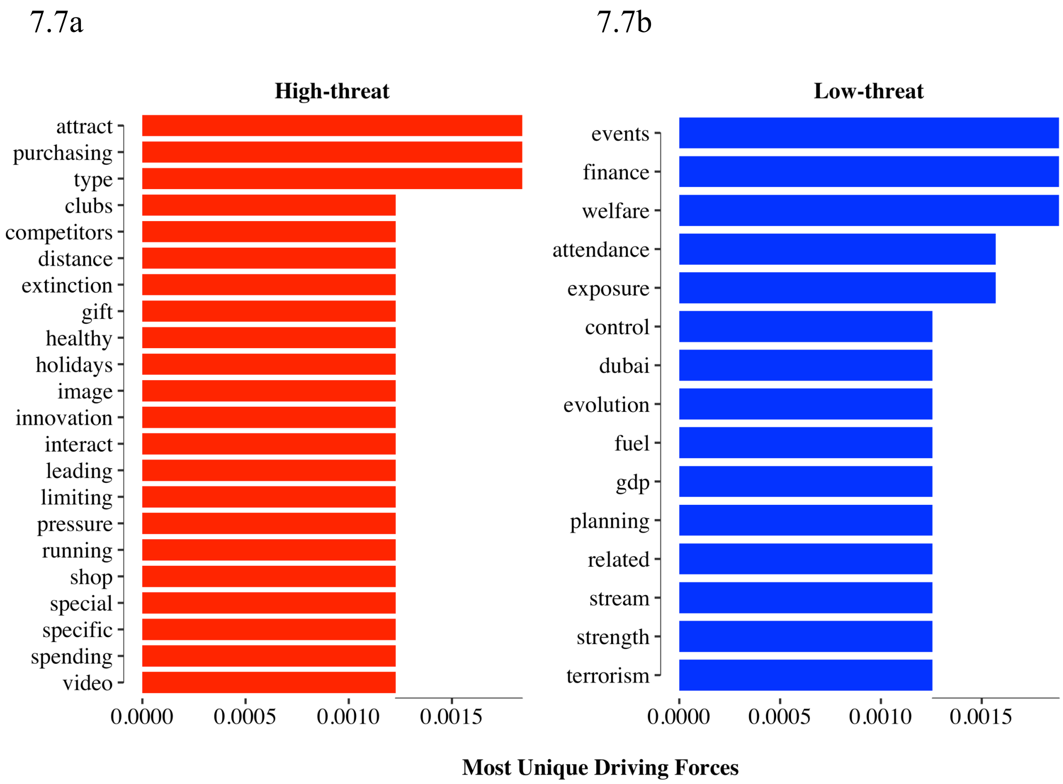
All shared topics are presented in Figure 7.6, where size of topic is related to frequency of mention and color is related to relationship between terms. Even though a minority of topics were shared between the two conditions, they were mentioned

2.5 times more often than the unique, unshared topics, resulting in more overall mentioned topics and topics being shared between the two conditions.

Figure 7.6. Total representation of most popular shared topics within the individual-only session.

Unique topics are those that only appeared within a single condition. Participants in the individual-only session produced 56% fewer unique topics from the high-threat condition compared to the low-threat condition. Topic frequencies were weighted within each condition and ranked by weighting factor. Figure 7.7 shows the top ranked 15 most distinct topics for each condition. Distinct, in this analysis, is a measure of the unique topics that were mentioned the most within a single condition. These topics show where message framing anchored attentions. The first thing to notice is that the high-threat figure (7.7a) presents more topics than the low-threat figure (7.7b). R packages (tidytext and ggplot2) were used to calculate, rank, and then plot topics presents the top topics based on their *cumulative* weighted value. This means that more topics with fewer mentions will be equal to fewer topics with more mentions. As Figure 7.7 shows, the high-threat condition produced not only fewer unique topics overall, but participants mentioned them less frequently as well, compared to the low-threat condition. In all possible measures, the high-threat condition produced lower frequencies of creativity compared to the low-threat condition.

Figure 7.7. Most unique topics by condition within the individual-only session



Figures 7.5b & 7.7 show a strong difference between the two conditions, when we focus on the most salient unique topics within each condition. Participants primed with a high-threat message were more focused on resources and technology, whereas the participants in the low-threat condition were more focused on environmental, politics, and energy factors.

The originality proportions between the two conditions in the individual+group session are not significantly different, with .09 of the DF in the high-threat condition and .08 of the DF in the low-threat condition designated to the *Other* category. Just like the previous individual-only sample, there appears, at first, to be no anchoring bias on originality of DF. Combining both conditions, there are 104 DF within the *Other* category, a large enough sample to require exploration.

Participants in the individual+group portion of the study produced 3390 terms in the high-threat condition and 1109 terms in the low-threat condition, confirming the reversal of behaviour by condition from the individual-only sample. Same as the other sample, a minority of topics were shared between the conditions (41%), but mentioned 2.5 times more often than the unique, unshared topics and topics. The highest shared 15 topics are presented in Figure 7.8.

Figure 7.8. Top 15 most popular topics shared between conditions within the individual+group session.

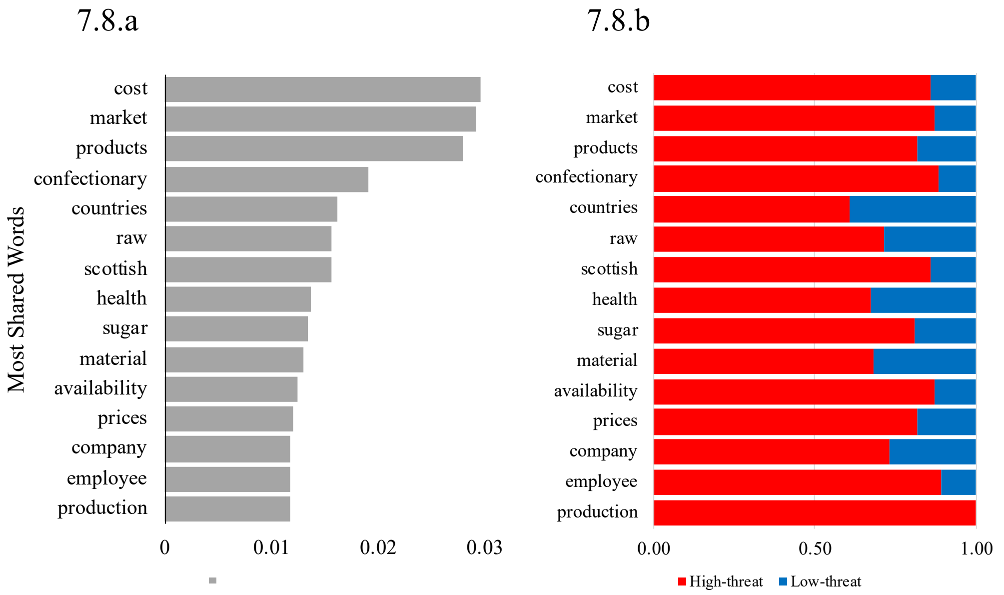


Figure 7.8.a shows the distribution of the most frequently shared topics between the two conditions. Interestingly, when the topic changes from a familiar organisation, like the local zoo, to an unfamiliar organisation, like the unknown confectionary company, participants seem to jointly focus more on the economics of the situation (i.e. “cost” and “market”), and secondarily on the industry-related topics (i.e. “products” and “confectionary”). The remaining topics span the categories *Industry, Resources, Economics*, and *Politics*. Figure 7.8.b illustrates further reflections of the reversal in behaviour between the two individual-sessions, revealing how extreme the high-threat condition participants focused more on each topic compared to the low-threat condition.

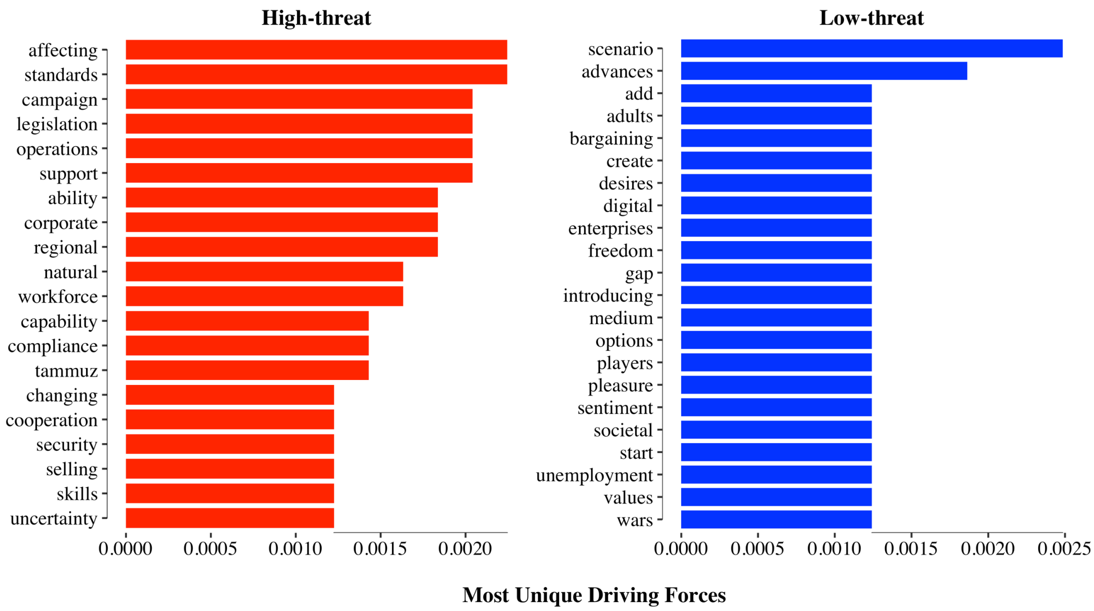
All shared topics are presented in Figure 7.9, where size of word is related to frequency of mention and color is related to relationship between topics. Same as the individual-only sample, even though a minority of words and topics were shared between the two conditions, they were mentioned 2.5 times more often than the unique, unshared words and topics, resulting in more overall mentioned words and topics being shared between the two conditions.

Figure 7.9. Total representation of most popular shared topics within the individual+group session.

Participants in the individual+group session produced 79% of the unique topics from the high-threat condition compared to the low-threat condition. Figure 7.10 shows the top ranked 15 most distinct topics for each condition. These topics show where message framing anchored attentions. As expected, the high-threat figure (7.10.a) presents fewer topics than the low-threat figure (7.10.b), though not as dramatically different as the induvial-only sample, meaning that the high-threat condition produced more unique topics overall and participants mentioned them more frequently, compared to the low-threat cond.

Figure 7.10. Most unique topics by condition within the individual+group session

7.10.a 7.10.b



Note: "tammuz" is the fictional name given to the confectionary company profiled in the business vignette.

As with the individual-only group, Figures 7.8b & 7.10, taken together, reveal a deeper story of differences between the two conditions. Participants primed with a high-threat message were more focused on topics concerned with industry and politics, while participants in the low-threat condition were more focused on resources and societal factors.

7.7.2 *Group*

Hypothesis 2 states that groups who read a high-threat business vignette will create fewer new DF and eliminate fewer existing DF, compared to groups who read a low-threat business vignette. To test H2 states that groups who read a high-threat business vignette will create, ANOVA tests are run to compare means of added and deleted DF.

When participants engaged in the group dialogue sessions, they were given the option to retain as many DF as they found important from the collective of individual group participants' lists, delete all they considered no longer important, and add as many new DF as they agreed upon that may have emerged. Whereas adding DF is a measure of serendipitous insight, which can be seen as a feature of group fluency, deleting DF can be seen as a feature of flexibility in thinking. Table 7.4 presents these results. Participants deleted significantly fewer DF from their individual lists when creating a unified group list in the high-threat condition compared to their low-threat counterparts ($F(1,90) = 5.88, p = .02$, $\eta^2 = .61$). The same participants added significantly fewer DF to their newly unified group list in the high-threat condition compared to the low-threat condition ($F(1,91) = 6.49, p = .01$, $\eta^2 = .67$). Neither framing condition showed a significance within the number of deleted and added DF.

Table 7.4. Driving forces from group sessions by condition

Condition	DF	*M*	*SD*	*V*
High-threat	**Deleted**	1.74	1.95	3.81
	Added	1.05	1.84	3.38
Low-threat	**Deleted**	2.94	2.82	7.94
	Added	2.37	3.16	10.01

One possible explanation for the differences in deleted DF frequencies is that these were due to higher rates of repeated topics. To investigate this possibility, duplication rates were matched by condition. Figures 7.8 & 7.10 help to illustrate these rates. The low-threat condition not only produced fewer topics overall, but these topics were repeated less frequently than those in the high-threat condition. Therefore, it is less likely that participants within the low-threat condition independently discovered the same DF at a higher rate during the individual session compared to the high-threat condition participants.

Hypothesis 3 states that groups who engage in DI will add fewer new DF and delete fewer existing DF during the group-portion of SP, compared to groups who engage in CS. To test H3, ANOVA tests are run to compare means of added and deleted DF. For further support, a 2x2 (condition x dialogue) ANOVA is run on the data to help determine whether an interaction exists.

Participants deleted significantly fewer DF from their individual lists when creating a unified group list using the CS method compared to the DI groups ($F(1,90) = 6.24$, $p = .01$, $\eta^2 = .07$). The CS method groups also added fewer new DF to their unified lists compared to the DI ($F(1,91) = 17.72$, $p < .000$, $\eta^2 = .16$). Neither dialogic method showed a significance within the number of deleted and added DF. The results are presented in Table 7.5.

Table 7.5. Driving forces from group sessions by dialogue

Dialogue	**DF**	***M***	***SD***	***V***
CS	**Deleted**	1.68	1.84	3.38
	Added	.68	1.47	2.15
DI	**Deleted**	2.90	2.84	8.04
	Added	2.70	3.07	9.40

There is an interaction between message framing and strategic dialogue, as well. Table 7.6 presents the interaction results. Participants in the high-threat condition who employed the DI method added significantly more DF as a group, compared to those in the high-threat condition who employed the CS method ($F(1,56) = 30.87$, $p < .000$, $\eta^2 = .36$).

Table 7.6. Interaction of condition and dialogue effects

Condition	**Dialogue**	***n***	**Deleted**	**Added**
High-threat	**CS**	40	1.33 (1.49; 2.23)	.26 (.72; .52)
	DI	23	2.00 (2.49; 6.18)	2.55 (2.35, 5.52)
M* *(SD; V)				
Low-threat	**CS**	17	2.53 (2.45; 6.02)	1.76 (2.11; 4.44)
	DI	21	3.52 (3.11; 9.66)	2.71 (3.66; 13.41)

The large effect size in Table 7.6 is partially due to the average added DF within the CS condition equalling less than one. The small variance score (< 1) helps reveal that not all high-threat/CS groups added DF. The pattern of behaviour is similar in their number of deleted DF, but this is not significant. Though the pattern is also similar for the low-threat condition, in both added and deleted DF, these similarities are not significant.

7.7.3 *Confidence in Creativity*

Hypothesis 4 states that individual confidence scores will be, on average, lower for participants who read a high-threat business vignette compared to participants who read a low-threat business vignette. To test H4, a test for independence is run on confidence scores between conditions, before and after group sessions, and dialogic method at the group-level. Hypothesis 5 states that individual confidence scores will increase overall, after participating in group sessions. To test H5, ANOVA is run (condition x dialogue) to determine whether main effects and/or interactions exist.

Confidence scores were only collected from the individual+group session (before and after the group sessions). Scores were accidentally not recorded for the individual-only participants. Further, not all participants in the individual+group session offered their confidence scores (n = 92). Participants were asked:

> Q1) Given your knowledge after today's group discussion, how confident are you that at the end of stage 2[15], you captured all the relevant driving forces concerning [the organisation's] goals for the next 5 years?

Participants generally held higher confidence (>3) before and after their group session, in their abilities to have discovered all the relevant DF. There is no significant difference in confidence scores between conditions. This holds for both individual-level and group-level effects. However, all confidence scores, by

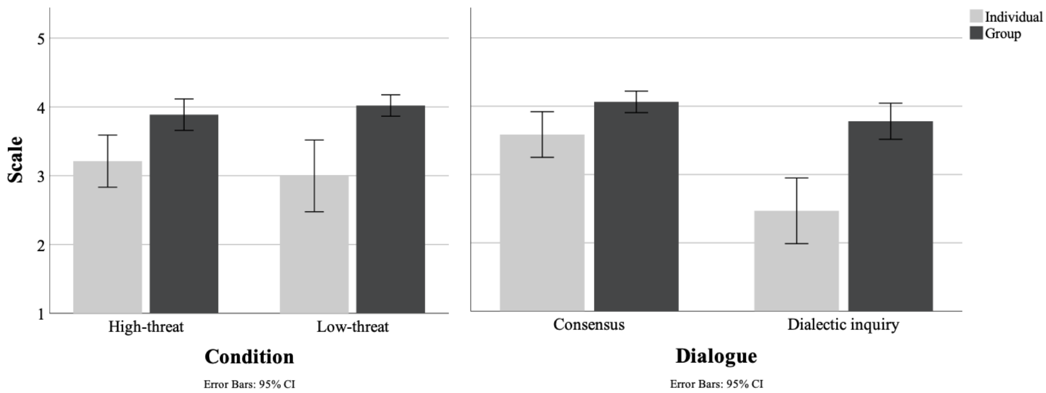
---

[15] The use of "stage 2" in this question is in reference to the two stages of the causal thinking study: "stage 1" was the individual session and "stage 2" was the group session.

condition and by dialogue, increased after the group sessions. The increase is significant after the group sessions ($t(66) = -5.84$, $p < .000$, 95% CI [-.76, -.38]), overall, as well as within both threat conditions, ($t_{high\text{-}threat}(37) = -4.90$, $p < .000$, 95% CI [-.87, -.36]; $t_{low\text{-}threat}(28) = -3.77$, $p = .001$, 95% CI [-.82, -.21]). There is also an effect of dialogue on confidence, with a significant increase in confidence after the DI sessions ($F(1,65) = 3.95$, $p = .05$, $\eta^2 = .06$). These changes -are illustrated in Figure 7.11.

Figure 7.11. Change in confidence by condition

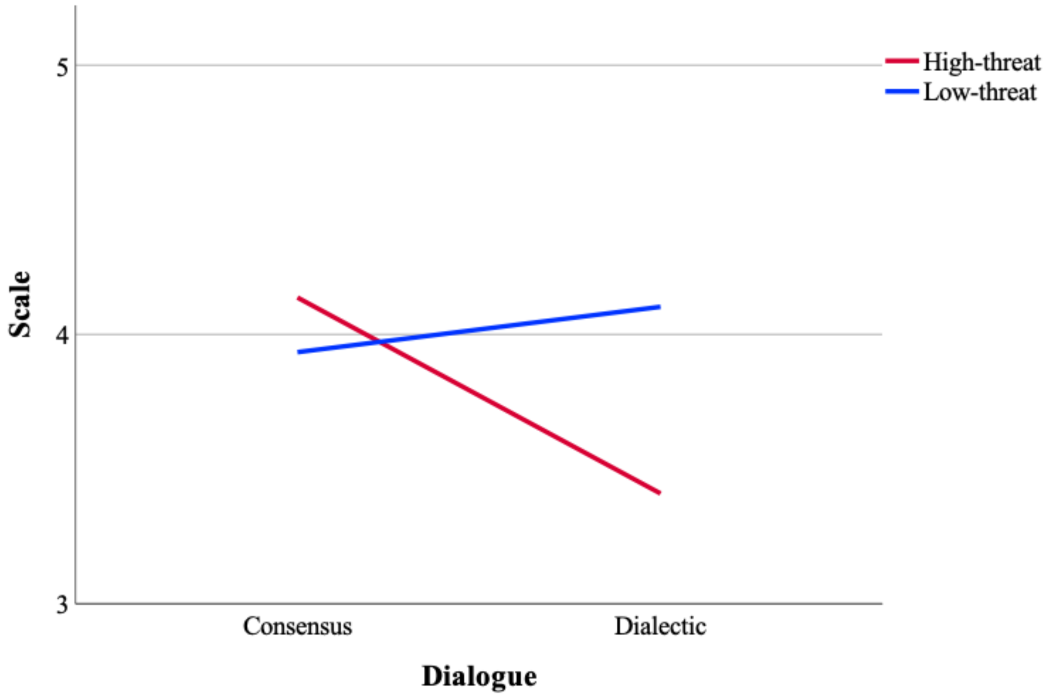
7.11.a 7.11.b



Note. Error bars show a decrease in confidence variability after group sessions.

There is an interaction between the threat condition and dialogue ($F(1,26) = 9.47$, $p = .005$, $\eta^2 = .27$). Figure 7.12 illustrates the interaction.

Figure 7.12. Interaction of condition and dialogue on confidence scores



Along with their confidence, participants were asked eight follow-up questions after the group sessions. [Likert scale: 1(strongly disagree) – 5(strongly agree)]

Q2) Based on the discussions with my group, we created a collection of plausible driving forces with valid reasoning.

Q3) My group's final set of driving forces came from the group's combined effort, but not necessarily the averaged outcome of all members' individual contributions.

Q4) I would be willing to use this group method again on future scenario planning projects.

Q5) Working with this method in my group was an enjoyable experience.

Q6) The group discussion and decision process made me critically reevaluate the validity & reasoning I had for my driving forces from stage 1.

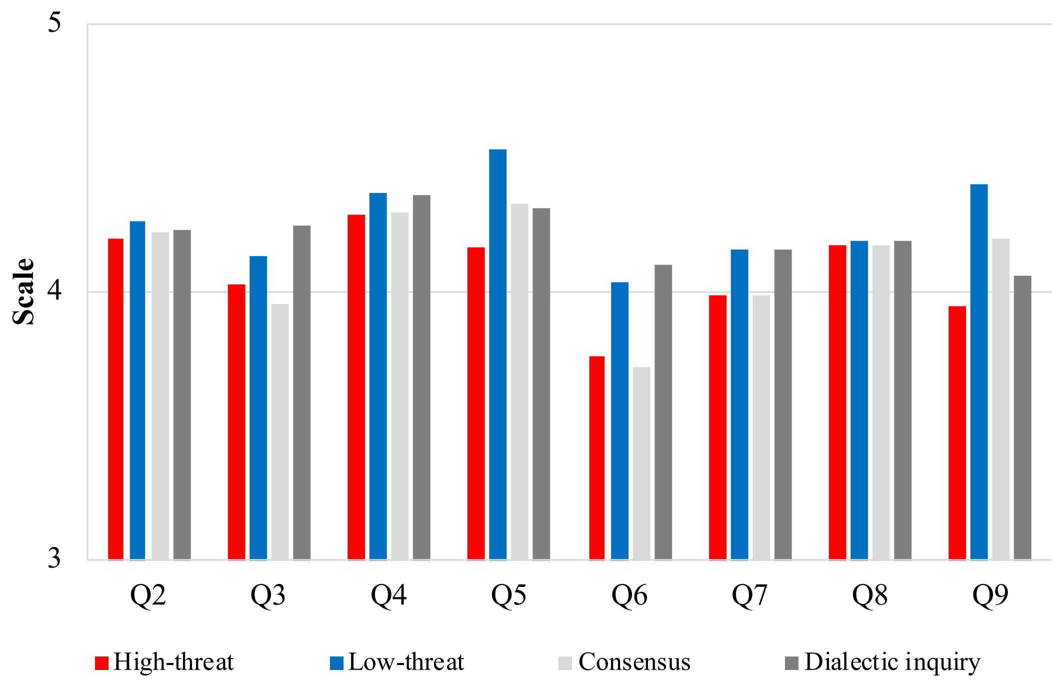
Q7) The group discussion and decision process revealed valid recommendations & reasoning I had not considered before, during stage 1.

Q8) My group's recommendations and reasoning for my final list of driving forces were correct.

Q9) I am satisfied with my final list of suggested driving forces.

Averages for each question by condition and dialogue are presented in Figure 7.13. The range of possible responses is from 1(strongly disagree) to 3 (neither agree nor disagree) to 5 (strongly agree). All averages are greater than the mid-point (3), showing overall agreement with all questions. The y-axis begins at the mid-point (3) and the x-axis presents questions (Q) 2-9.

Figure 7.13. Mean responses to group questions by condition and dialogue



As illustrated in Figure 7.13, in response to questions 2-9, participants in the high-threat condition answered with lower agreement to all questions, compared to those in the low-threat condition. The high-threat condition participants gave significantly lower "agreeable" scores for Q5 ($F(1,66) = 4.59$, $p = .04$, $\eta^2 = .07$) and Q9 ($F(1,66) = 6.79$, $p = .01$, $\eta^2 = .09$) compared to the low-threat condition. Given participants' overall higher confidence after the group session, it would be expected that question 9 would also reveal significant differences. Though there was no significant difference between the two dialogic methods, participants who engaged in CS scored lower agreeableness to all the questions, *except* 5 and 9. All scores, however, are greater than "undecided" (>3), showing a more positive attitude towards both dialogic methods overall.

Some selected feedback from the participants may shed light on their performance and scale values. Several participants who used the dialectic method said they "really enjoyed the study," and expressed appreciation for introducing them to a new, formal method for group discussions. One group who used the CS method expressed frustration with determining their group list. They had trouble determining where their individual boundaries should be within the group discussion, and how much of their opinion they should express (too much vs too little).

7.8 Creative Thinking Discussion

Two experimental studies were developed to test qualities of creative thinking employed in Stage 2 of the IL/ST framework. The first study measured creativity from participants working alone, as individuals. The second study measured creativity from participants who first worked alone, then as a member of a group. Business vignettes 2 and 3 were used to deliver the priming stimulus in order to measure for anchoring effects. Analyses reveal similarities in creative thinking between individual-only and group-level output. Surprisingly, though, the data reveal an apparent difference in creative output at the individual-level, depending on whether a person has expectations of justifying their ideas to a group or not. Table 7.7 presents a summary of the key findings from this chapter.

To help answer research questions 1 and 2, H1 tested for anchoring biases between high-threat and low-threat framed messages at the individual level. *Fluency* and *originality* revealed to be the most sensitive features of creative thinking to differently framed messages. Participants anchored to the differently framed messages and were, on average, more prolific in their creativity when working with a low-threat business environment, compared to a high-threat environment (even though vignettes focused on the same organisation, aiming for the same goals, in the same financial positions, with the same employee profiles). H1 is supported when participants have no expectations of group work, however, the specific quality of their anchoring biases reversed when they anticipated group engagement afterwards.

Furthermore, when participants expected to discuss and justify their ideas to a group, creative *flexibility* also proved to be sensitive to message framing and revealed an anchoring bias that partially supports the first hypothesis.

Arden, et al. (2010) state that using simple, single-scale creativity tests to draw conclusions about a person's overall creative potential can be problematic. Runco & Acar (2012) support this sentiment when they suggest that fluency, alone, should not be considered an efficient measure of the complex and variable nature of creativity. The process of and capacity for creativity play gatekeepers to a whole host of cognitive efforts. This is why creativity in the realm of ST is measured across four major qualities of divergent thinking: fluency, flexibility, elaboration, and originality. Each quality is considered an interdependent process within creativity. By borrowing the methods of Guilford, Torrance, and others in the field of psychology, this study hopes to present a deeper understanding of the creative elements in ST to help improve the efforts of both facilitators and practitioners, alike.

The most interesting insight from the creativity study is the unexpected difference in creative output that appears to be motivated by the individual's potential group-level identity. Behaviours completely reversed or globally altered on all measures of creativity, depending on whether participants believed their ideas would or would not be scrutinised in the near future by a group of peers.

There is a surprisingly large proportion of DF that did not fit within the ten distinct, and common, categories. These DF included topics such as awareness, legalities, war, and specifics to the participant's environment, such as "Expo Dubai" and "Abu Dhabi Vision 2030". Overall proportions between .08 to .15 of the total number of DF fit outside the STIRDEEPER categories. This made for one of the largest groups of DF. Originality encompasses not only a statistical improbability, but can also include unconventional and even strange concepts, which can make them difficult for others to adopt or understand, due to their break from the norm (Runco, 2014b). In SP research, this quality of decision-making must be brought to the forefront and discussed.

Table 7.7. Summary table of samples and conditions by hypothesis

Hypothesis	Dependent variable	Measure	Individual-only		Individual+group		Group			
			High-threat	**Low-threat**	**High-threat**	**Low-threat**	**High-threat**	**Low-threat**	**Dialectic Inquiry**	**Consensus**
1	DF	Fluency	Anchoring bias, fewer average DF	Anchoring bias, more average DF	Anchoring bias reverses, more average DF	Anchoring bias reverses, fewer average DF	-	-	-	-
		Flexibility	No anchoring bias, all categories covered	No anchoring bias, all categories covered	Partially supported anchoring bias, no Religion DF	Partially supported anchoring bias, all categories covered	-	-	-	-
		Elaboration	No anchoring bias, same categorical spread	No anchoring bias, same categorical spread	No anchoring bias, same categorical spread	No anchoring bias, same categorical spread	-	-	-	-
		Originality	Partially supported anchoring bias, lower word popularity	Partially supported anchoring bias, higher word popularity	Partially supported anchoring bias, higher word popularity	Partially supported anchoring bias, lower word popularity	-	-	-	-

Table 7.7. *(continued)*

Hypothesis	Dependent variable	Measure	Individual-only		Individual+group		Group			
			High-threat	**Low-threat**	**High-threat**	**Low-threat**	**High-threat**	**Low-threat**	**Dialectic Inquiry**	**Consensus**
2	DF	Added	-	-	-	-	Anchoring bias, fewer DF	Anchoring bias, more DF	-	-
		Deleted	-	-	-	-	Anchoring bias, fewer DF	Anchoring bias, more DF	-	-
3	DF	Added	-	-	-	-	-	-	Anchoring bias, more DF	Anchoring bias, fewer DF
		Deleted	-	-	-	-	-	-	Anchoring bias, more DF	Anchoring bias, fewer DF
4	Confidence	Likert	-	-	No anchoring bias, but trend shows higher average score	No anchoring bias, but trend shows lower average score	No anchoring bias, but trend shows lower average score	No anchoring bias, but trend shows higher average score	-	-
5	Confidence	Likert	-	-	Anchoring bias, lower score compared to any group composition		Anchoring bias, higher score compared to individual		Anchoring bias, higher score compared to individual	

One of the tenets of SP is that the process helps break the narrowed decision-making style of business-as-usual thinking, "use of scenario methods takes you beyond bounded thinking and enables you to think about the future from multiple perspectives at the same time" (Cairns & Wright, 2018b, p. 6). Therefore, the presence, prolific nature, and quality of creative originality should carry extra importance in ST studies.

Focusing on the specific topics generated during the studies, it is interesting to note that from both business vignettes (2 and 3), participants primed with a low-threat environment were the only samples who explored the potentials of conflict, which can be seen in their mentions of "terrorism" and "war" (Figures 7.7b & 7.10b). Intuitively, it would seem that perceiving an environment with higher threat levels would prime participants to associate their memory searches towards similarly related topics to "threat", such as war, conflict, strife, hazard, risk, and intimidation. However, just the opposite appears to have happened, which may reveal a subtle difference in priming and anchoring effects. Even though the differently framed messages appear to anchor creative thinking on various interdependent creativity levels, it may be the case that semantic priming is affected differently, and may be less salient than the conscious effort to "think more broadly". Unexpected commonalities bring about questions of extraneous influences. The samples were primed with the same threat-level, the content regarded different businesses, participants between each low-threat condition (individual-only and individual+group) were located in different countries, and the different experimental study sessions were overseen by different experimental facilitators. Any extraneous co-varying commonalities remain within the samples' demographics and any unregulated external information (e.g. media, literature, conversations). However, the likelihood of confounding influential information being shared amongst the two independent low-threat level samples has a lower likelihood of confounding information being shared amongst the high-threat samples that tested at the same time, in the same spaces. Therefore, the likelihood of the two low-threat samples discovering the same DF is greater than two between session samples (low-threat vs high-threat). Taking these controls into consideration, it is not unreasonable to attribute the similarities in behaviour to the priming mechanism of the business vignette framing.

Also addressing research questions 1 and 2, H2 tested for anchoring biases at the group-level and data support the hypothesis. Participants who read low-threat framed messages about an organisation were more inclined to eliminate existing DF, supplied by each member, and create more group-level DF, revealing stronger preferences for information adjustment, compared to those working with a high-threat business environment. More activity in either removing or adding DF changes the group's list from merely a collection of individual efforts to a product that reflects a third entity, that of the *cooperative group mindset*. Serendipitous insights were bolstered in groups that perceived fewer threats within the organisational environment. Just as interestingly, individual-only behaviours were mimicked in the group-level behaviours, which remained resistant to any group dialogue effects. Altogether, anchoring biases appear at both the individual and group levels.

To address research question 3, H4 tested for an anchoring bias in participants' confidence in their abilities to capture all the relevant DF for an organisation, given their goals and timeline (5 years). At both the individual- and group-level, there were no significant differences in self-evaluated confidence scores, by condition, even though trends mostly followed the hypothesised direction. H4 is not statistically supported, however, the increase in confidence after the group sessions was greater and higher for participants primed with a low-threat business vignette, which lends credibility to the hypothesis. Expanding to the macro level, H5 tested for differences between self-reported confidence scores after working alone, then after working as a member of a group. Overall, participants, reported they felt more confident about the scope of their creative output after participating in their group session. This held true for both conditions, regardless of output. The lower variability in confidence scores after group sessions, compared to individual work, also reflects a homogeneity in group effects to confidence. The strength of increased confidence priming after the group sessions is further supported by in the same behaviours exhibited within both dialogic methods. Though analysing confidence data against dialogic method was not discussed in the hypotheses, the analyses help reveal a robust effect group dialogue can have on each member's confidence in their collective work. Including self-reported confidence in this study affords us the opportunity to compare and contrast implicit motivations with explicit self-reflection.

To address research question 4, H3 tested for anchoring biases between dialogic methods (DI vs CS). Groups who employed the DI method of debate, which encouraged a reflexive manner of dialogue to maximise critical evaluation, were more inclined to create more group-level DF as well as eliminate existing DF, revealing similar group-level creative output to perceptions of a low-threat business environment. Adding and retaining more DF in final group lists, much like the low-threat framing effect, reveals not only serendipitous insights, but group dynamics that facilitate adoption of individual member's input. Though H3 is not supported, the differences between methods are statistically significant and highlight the importance of gaining greater understanding of how influential the mechanisms for group-based dialogue can become. One follow-up question is whether it was beneficial for the DI group members to be more willing to eliminate existing DF, compared to the CS group members? The SP method is an iterative process where each stage builds from the information of the previous stages. If information, such as DF, is eliminated at some point from the process, then it is more likely to remain eliminated from the rest of the SP. With regard to the participants, we cannot be certain whether this was an act of removing unnecessary, superfluous information or eliminating vital information. This study offers budding evidence to support the claim that strategic dialogic methods formalised 30 years ago, in a wholly different and possibly antiquated business culture, can continue to exhibit similar effects on managers today. Such a discovery echoes Godet's (2000b, p. 8) discussion on the longevity of business management tools, "Yesterday's tools are still useful today. Indeed, the kind of problems encountered, even if the world changes, often remain similar." Though the research within this thesis takes a process view, Godet's thoughts can still find perch in this chapter.

In real-world settings, SP practitioner (e.g. management and executives) know that they will be working as a group towards a common goal, therefore we can be assured that most, if not all, the practitioners will see themselves as potential group members, and as such may be more likely to mimic the individual creative thinking efforts of the individual+group sample. The results of this study suggest that if a facilitator wants to maximise the creative output from the practitioners, they will want to ensure that management's research into the organisation, as part of their preparation for the SP intervention, includes information that frames the organisation in both a low- and high-threat environment, to bring a heterogeneous mix of perspectives and information to the session. Whereas those practitioners who perceive more threats to their organisation may think more fluently when preparing their research for the group sessions, the group members who perceive fewer external threats will bolster the chances of serendipitous insights during the group discussions. The next step would be to remain aware of the how ST is affected by both information anchoring *and* process throughout the intervention. Heated, challenging dialogues may increase spontaneous, serendipitous insights, but may also lead towards elimination of valuable knowledge. One suggestion of modification is for the facilitator to coach the practitioners in the DI method, but abstain from instructing practitioners to eliminate DF. Instead choosing to focus on any additional insights that may arise through the challenging dialogue and suggesting only repetitious DF be combined. This is one method to capitalise on their increased flexibility in thinking while helping to avoid damaging priming in the wrong direction. Chapter 8 builds from the data and knowledge discovered from the individual-level and group-level, Stage 2 studies. Two empirical studies will attempt to measure the next stage in the IL/ST framework (Stage 3) against the prevailing cognitive effort (associative causal thinking).

Chapter 8. **Causal Thinking (Stage 3)**

> "One way to explain the complexity and unpredictability of historical systems, despite their ultimate determinacy, is to note that long chains of causation may separate final effects from ultimate causes lying outside the domain of that field of science."
> Jared Diamond, *Guns, Germs, and Steel*

The present chapter builds on information gained from the Stage 2 creative thinking empirical studies in Chapter 7 and expands into testing qualities of associative causal thinking (dependent variable) promoted in Stage 3 of the IL/ST framework. Whereas the creativity studies compared individual-level output to group-level output, the causal thinking studies focus primarily on the individual level at this time. As stated in Section I, in order to gain the greatest understanding of ST mechanisms, function, and influences, we must look at individual-level, as well as group-level efforts. Insights from Stage 2 studies reveal that there may be a qualitative difference between working with DF supplied through a consistent cognitive effort (i.e. same practitioners through the process, primed with the same framing message) and working with DF supplied through more varied cognitive efforts (i.e. different practitioners join the process, primed by differently framed messages). This is an important distinction to recognise, given the variability of real-world SP interventions. No two interventions are alike, and one of the common variables is whether the same team of practitioners work together developing the same homogeneous outputs, or new practitioners enter and leave the space, whereby default, introduce novel outputs, creating a heterogeneous mix of information. In an effort to increase generalisability of results by preserving as much ecological validity as possible (i.e. mimicking real-world settings), this chapter explores the potential homogeneity/heterogeneity of SP interventions. Therefore, starting at the individual level, two studies are developed in this chapter which explore potential anchoring effects on causal thinking, compared between homogeneous stimuli and heterogeneous stimuli.

The chapter begins with a focused literature review that completes the methodological background for developing two experimental studies to help better understand how we translate our assumptions of causality within SP boundaries.

Seven hypotheses are presented. Hypotheses 6-10 are aimed at helping to answer research questions 1 and 2. Hypothesis 11 aims to help answer research question 3. Finally, hypothesis 12 aims to addresses research question 4, by introducing novel methods for understanding SP spaces and measuring ST output.

8.1 **Causal Thinking**

Practitioners must switch from an idea generation mode of creative, divergent thinking to a categorically different cognitive effort of causal thinking. Causal thinking is a necessary component to ST and includes, in part, associative processing, causal reasoning, and causal mapping. During Stage 2 creative process, several mental networks are activated that access a combination of working memory, long-term memory, and explicit and implicit searches, containing related concepts or ideas. The cognitive and SP move into Stage 3 activates initial associative processing efforts, where the purposeful, yet implicit, spread of activation retrieves and recombines these related concepts (i.e. DF) into novel products (i.e. clusters; Lee & Therriault, 2013; Mednick, 1962).

According to Hastie (2015) causal thinking begins with the recognition of the regularity of events (*where there is x, y follows, all x-similar objects are followed by y-similar objects*). Causal thinking is then supported by the understanding of intervening efforts (*if I do x, y will follow*). Then belief in counterfactuals can developed (*if z instead of x occurs, maybe y does not follow*). We assume to live in a causal universe with an order to causal relationships (Hastie, 2015; Hume, 1748; Kant, 1783/1994). We think of these relationships in either a predictive order (cause-effect, $x \rightarrow y$) or a diagnostic order (effect-cause, $y \rightarrow x$). People can also reason in both directions (transitively, $x \leftrightarrow y$) along the causal chain. When deciding how to causally connect the DFs from Stage 2 into separate, unique clusters in Stage 3, each DF can be assessed (i) predictively as a cause ($x \rightarrow y$), (ii) diagnostically as an effect ($y \rightarrow x$), or (iii) transitively ($x \leftrightarrow y$) within a causal chain.

SP refers to this effort as “clustering” and the products as “clusters”. Clustering is an effort similar to a type of free-association known as chain association, where a person is given a stimulus (e.g. a single DF within a group) and finds the first associated item (e.g. another DF), related on any level, in any manner, and subsequently finds additional associations, each of which relate to the previous associated item(s) (Marron & Faust, 2018). Traditionally, chain association is unrestricted, however, within SP, boundaries exist only as far as the amount of available information (i.e. DF) and focus on the organisation's goals. The task is designed to be pragmatic and organic.

Instructions for discovering causality are open-ended, where participants are tasked with finding "any" connection, avoid focusing on positive or negative associations, and any value of strength across a connection. Further, participants are instructed not to focus on outcomes. This is to help avoid projecting norms and desires onto the effort which may bias the type and frequency of causal associations practitioners perceive between the as-yet associated DFs and limit the final formation of plausible scenario futures. Studies suggest that causal reasoning is altered by projected beliefs in outcomes, and that differently imagined outcomes will produce differently reasoned causal sources (Hastie, 1984; Monson, Keel, Stephens, & Genung, 1982). Just as with the previous stage, Stage 3 moves from intuitive to deliberative efforts.

The purpose of creating clusters is multi-fold. There is good consensus that human cognition has limitations to information processing (Gigerenzer & Gaissmaier, 2011; Kahneman & Tversky, 1979; Miller, 1956; Simon, 1972). By clustering various DFs into related groups, practitioners are able to create a more manageable workspace. Clusters allow practitioners to systematically focus their thinking on smaller groups of information (Cairns & Wright, 2018a). To an extent, clustering can be seen as a method to accommodate natural limitations with information processing by maximising the utility of cognitive heuristics (Flach & Hoffman, 2003; Hertwig & Todd, 2003; Payne & Bettman, 2004; Sowa, 1987; Gigerenzer & Brighton, 2009). The result is better processing of information and effective ST.

Kahn and Weiner (1967, p. 6) make it clear that a clustering stage is an integral part to SP when they say, "Scenarios are hypothetical sequences of events constructed for the purpose of focusing attention on *causal processes* and decision-points" (emphasis added). The act of clustering prompts practitioners to include more nuanced understandings to their information. When causally clustered together, DF cease to be disparate, stand-alone factors that offer little information, and transform into more context-rich events. This, in turn, helps practitioners build scenarios from their understanding of real-world mechanisms, which increases believability. One consensus of cognitive science is that people construct working models based on their understandings of the world around them, and use these models to develop strategic decisions, which eventually lead to some form of action – including inaction (Hodgkinson, Bown, Maule, Glaister, & Pearman, 1999).

Causal knowledge about an organisation's environment can help management make more adaptive decisions as the future emerges. Futures thinking is supported by causal knowledge by allowing better understanding of consequences (effects) given any number of plausible circumstances (causes), and by helping an organisation navigate effective interventions (Hagmayer & Fernbach, unpublished paper). Without the explicit inclusion of a causal thinking stage in SP, the driving events, motivations, and progression of relationships that could affect the organisation remain underdeveloped for the practitioner, rendering SP less informative and less supportive.

Through causal reasoning, inferences of the mechanisms that connect causes to their effects is modulated by knowledge of generalised causality (Fugelsang & Dunbar, 2005). We generalize experiences and create general causal schemas (Kelly, 1972; Tversky & Kahneman, 1980). Causal schemas are expressed through causal mapping (Axelrod, 1976). Causal maps visually represent the pattern of causal inter-relationships between factors (Hodgkinson, Bown, Maule, Glaister, & Pearman, 1999; Kaplan & Norton, 2000). By creating causal maps, practitioners can focus on the temporal order of events (Huff, 1990; Montibeller & Belton, 2006). Clustering is the product of causal mapping in SP. Practitioners use clusters to represent the pattern of causal relationships between related DF. Clusters, in turn, can be interpreted through any number of causal models (Blalock Jr., 1985).

The open-ended method promoted in the IL model is designed specifically to help counter limiting biases internalised from any number of experiences and dispositions. Studies show that certain biases in causality exist where reasoning in one direction, particularly predictive order thinking, is normatively stronger than transitive thinking (Bes, Sloman, Lucas, & Raufaster, 2012). As well, we have a tendency to focus on a single causal source ($y_1 \leftarrow x \rightarrow y_2$), rather than multiple ($x_1 \rightarrow y \leftarrow x_2$) (Fernbach, Darlow, & Sloman, 2011). One explanation for this heuristic is imaginability. When alternative causes are difficult to imagine, the focal cause will tend to dominate judgments in a non-normative manner. The easier imagined cause is more salient, and therefore more believable. We often support this qualitative effort *post hoc* with stronger probabilistic mental calculations (or justifications). Another explanation for this asymmetry can be found in the concepts of sufficient and necessary reasoning. If one cause is sufficient to explain the relationship, then all other causes have a tendency to be discounted. If more than one cause is necessary, then more than one cause will more likely be entertained in the explanation.

However, some studies show that people violate this asymmetric attribution and judge multiple causes for an action as more likely than a single cause. This type of attribution bias is known as the conjunction error (Leddo, Abelson, & Gross, 1984; Tversky & Kahneman, 1983). One possibility is that this bias may be due to levels of familiarity with pre-constructed causal schemas. If a topic is familiar, a person may access causal schemas in long-term memory and apply those for better understanding. As well, the more similar the causes are to one another, the more likely they will be considered together (Fugelsang & Dunbar, 2005). Other conjunction errors may be due to judgments of probability. If a person is judging the likelihood of occurrence, then the more causal factors that are present, the more likely an effect will occur (Zuckerman, Eghrari, & Lambrecht, 1996). These explanations are pertinent to ST. Schoemaker, for example, found that participants committed conjunction fallacies when assessing multiple causes (Schoemaker, 1993). However, given the presence of other biases (overconfidence, availability, believability, and simulation), it may be a benefit to SP that practitioners identify multiple causes for an effect. One benefit is that it could make unlikely scenarios more believable. Another benefit is that this form of causal reasoning may be a product of broader thinking, which the field of SP promotes as one of the practice's biggest benefits (Chermack, 2011; McKiernan, 2017).

8.1.1 *Anchoring via Causality*

Hastie (1980, 1984, 2015) has stated in several publications that causal attribution processes begin with information seeking both externally (the environment) and internally (memory). External and internal information are updated together using causal reasoning skills. The marriage or divorce of new and old of information is what informs cluster development. An anchoring bias has proven to be not only fairly robust across studies, but also potentially ubiquitous in effect. Hodgkinson, et al. (1999) suggest that such biases are not restricted to merely probability choices made within laboratory settings. In fact, the anchoring bias may very well be a far more influential cognitive mediator with the type of complex strategic decisions that make up the full profile of ST, and employed in SP. If this is the case, then there may be an anchoring effect on the early stages of causal thinking. The creativity measures from the previous studies reveal partial anchoring biases that more often result in less prolific output from participants after reading a high-threat framed message, compared to a low-threat version. Furthermore, these biases are strongest with participants who do not expect to justify their work to a group of peers. Building from these discoveries, participants in Stage 3, causal thinking studies will work at individuals with no expectation of future group tasks. They will be tasked with identifying basic causal relationships between DF. Their efforts will link DF together into *ad hoc* clusters. A cluster, therefore, is comprised of two main elements, DF and causal links. The first two hypotheses test these assumptions and help to answer research questions 1 and 2.

H6: Fewer causal associations between DF will be generated from participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.

H7: Fewer DF will be integrated into clusters, overall, from participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.

A potential outcome from the number of DF integrated into a cluster, when working with an existing list of unassociated DF, is the number of clusters that results from participant associative efforts. Generally speaking, in absence of new DF being generated during the clustering process, the fewer DF that are clustered together, the more clusters, overall, that can potentially be produced. If high-threat framed messages anchor participants towards integrating fewer DF with fewer causal associations between, then overall, there should be a difference in the number of clusters between conditions. The next two hypotheses test this assumption and help to answer research questions 1 and 2.

H8: Clusters will, overall, be less complex from participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.

H9: More independent clusters will be generated from participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.

In line with the complimenting qualitative analyses used to help understand the relationship between Stage 2 creativity and framing effects, qualitative analysis can help understand whether message framing affects the kinds of DF participants may focus on, as they explore their causal relationships. In particular, there may be an effect on which DF become a focal point for cluster developments. To help determine focal DF, their STIRDEEPER categories will be used to help with analysis. The next hypothesis tests this assumption and helps to answer research questions 1 and 2.

H10: There will be differences in focal concepts, by STIRDEEPER category, between threat conditions, which will reveal an anchoring bias.

8.1.2 *Confidence and Causality*

Same as in Stage 2 studies, confidence measures are taken at the end of each experimental session to help determine whether a relationship exists between message anchoring and confidence after efforts of causal thinking. Differences between conditions were not significant after the creativity studies and revealed a slight reversal when moving from individual work to group work. Even though no confidence scores were collected after then individual-only creativity sessions, group creative behaviours reveal to mimic more closely the measured individual-only behaviours, therefore, the next hypothesis builds from the logic, and helps to answer research question 3.

H11: Confidence will be lower for participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.

8.1.3 *Creativity from Causality*

In the IL model, even as associations are developed between existing DF, practitioners are encouraged to introduce more DF as they see fit. This method is supported by the larger body of studies on free-association tasks, which have shown to enhance creative thinking in several domains (*see* Runco, 2014a; Mednick, 1962; Freedman, 1965; Roth, 1975; McFadzean, 1998). So it could be expected that the more effective the free-association in Stage 3, the more likely practitioners are to identify and incorporate novel DF. Even though creative thinking is dominant in Stage 2, it is a cognitive effort that remains employed throughout the full process. For example, in a more organic SP setting, like a workshop, practitioners are encouraged to develop more DF as they cluster their existing DF generated during the previous stage. The effort serves two main purposes. First, it is assumed that no matter the group dynamics, all pertinent DF will not be discovered only within the single, early stage of the workshop. Second, it is also assumed that the more information that can be included in scenario development, the greater the benefit of SP. It is not uncommon to experience fresh insights during the causal, associative phase of thinking. This is the act of serendipity, discussed in Chapter 7, that stems from new modes of thinking (i.e. associative thinking, causal reasoning). As observed in SP workshops, practitioners will initially, intuitively cluster a group of DF. As they begin creating explicit causal association between them, they also begin to develop a story to explain the strength of the association (i.e. causal reasoning). In this exercise, sometimes two DF are believed to be causally related, but the association is too weak to create a credible story. As has been pointed out in previous chapters, credibility and believability are key components to any scenario work. Therefore, they identify new DF that moderate the association, and develop a causal chain between the two original DF, thus increasing such factors as believability. At the end of each Stage 3 study, practitioners will be asked whether they believe any DF are missing from their presented list of DF that they feel are important. Space will be provided for entries. Stage 2 studies present the corresponding hypothesis (H1) for the creativity follow-up testing.

8.1.4 ***Heterogeneous vs Homogeneous Information***

Yaniv (2011) conducted framing effect research on individuals and groups, where group composition reflected a mix of members who either received the same framing message (homogeneous) or a mixture of framing messages (heterogeneous). His study reveals that participants reflect framing effects differently, depending on whether they are working with similar or dissimilar information. Specifically, framing effects are attenuated when participants work with dissimilar information (heterogeneous). Yaniv's studies measure differences at the group level. Though the causal thinking studies in this chapter aim to measure decision-making at the individual level, the stimuli will include DF developed by participants in the Stage 2 creativity studies. A DF list can serve as the outcome of a group of Stage 2 participants. The list can either contain DF from participants who were given the same framing message, or a mix of the two different framing messages. Therefore, borrowing Yaniv's language, Stage 3 causal thinking studies develop two different methods (heterogeneous vs homogeneous) that reflect the framing-primed DF (high-threat vs low-threat) from Stage 2 studies.

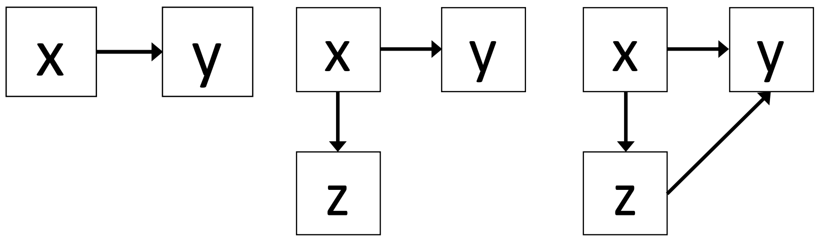
The first study tests for an anchoring bias, working with a heterogeneous mix of information (i.e. DF). The heterogeneous design aims to reflect real-world SP interventions where practitioners come from disparate perspectives and, therefore, may provide a broader scope of DF. In this study, heterogeneous is operationally defined as an equal representation of DF from Stage 2 output that were developed by participants in either threat condition (high vs low). However, the heterogeneous design only allows for a partial understanding of the causal thinking story, with respect to the larger picture offered by ST. Discussions with scenario facilitators brought to light that in SP workshops, clustering efforts would only be done on a homogeneous list of DF. That is to say, DF would have been created by a group of practitioners who received the same information in Stage 1 and created their lists together in Stage 2. Using a homogeneous list of DF would also increase the ecological validity of Stage 3 studies. Would there be a difference in performance with a homogeneous mix of DF? Would an anchoring bias be more pronounced? A second design is tested. The second study tests for anchoring biases, working with a homogeneous mix of information. Homogeneity is operationally defined as a list of DF from Stage 2 that were developed by participants assigned to the same threat condition as the participant reading the list in the Stage 3 study. Therefore, DF developed after reading a low-threat vignette in a Stage 2 study will be supplied to participants in a Stage 3 study who also read the same low-treat vignette. The methods are identical for the high-threat condition. The homogeneous design adds to the ecological validity by reflecting other forms of common, real-world SP where a group of practitioners is maintained throughout the workshop, and therefore are more likely to work with the same information, where their output (e.g. DF) may be more closely related. The last hypothesis tests this assumption and helps to answer research question 4 by introducing novel methods for understanding a SP space and measuring ST output.

H12: Observed anchoring bias trends in the heterogeneous study will be intensified in the homogeneous study.

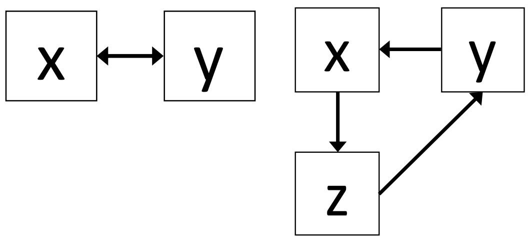
Causal associative thinking is operationally defined across several outputs. Frequency and complexity (DF and causal connections) of unique clusters, unassociated DF, concentration of causal DF vs effected DF, loops, and focal DF. There are many orientations for cluster construction. The least complex includes two DF and one causal connection (CC) between them (i.e. “1 level complexity”, Figure 8.1.a). The next complex orientation is to increase the DF, which by default will increase the number of CC (Figure 8.1.b). The next level in complexity is to increase CC (Figure 8.1.c). Within clusters, CC can go both ways, creating a transitive relationship (Figure 8.1.d). Finally, transitive complexity is increased by creating loops, where no beginning or end can be discerned (Figure 8.1.e). Complexity increases and changes in quality by the number of DF within a cluster, number of CC between DF, and orientation of CC.

Figure 8.1. Cluster construction and orientations.

8.1.a. Least complex 8.1.b. More complex 8.1.c. Most complex



8.1.d. Transitive 8.1.e Loop



Note: The nodes (x, y, and z) are DF and the arrows represent the associated CC and the direction of causal relationship between the connected DF.

8.2 **Ethical Approval for Human-Based Study**

The ethical review for this study was submitted through the Department of Strategy and Organisation, within the SBS to the UEC and received approval on 21 November 2016.

8.3 **Apparatus & Stimuli**

All business vignettes (1-3) are used as primes in the heterogeneous study. DF for vignettes 2 and 3 are sourced from the individual portion of Stage 2 studies and categorised by the 10 STIRDEEPER categories. Vignette 2 is supplied with 116 DF and vignette 3 with 66 DF from Stage 2 studies. Vignette 1 was not tested in any of the Stage 2 studies, therefore a pilot study (n = 13) is carried out to source DF on the topic (the university). The pilot survey was delivered online and randomly presented either the high-threat or low-threat version of vignette 1 to each participant. In order to maximise the number of DF per STIRDEEPER category, pilot participants were presented with only four randomly selected STIRDEEPER categories and asked to fill in as many DF as they could think of within each category. No participant was made aware that there were six other categories. This method reduced the need for extensive post-hoc categorising of DF by different raters, in order to reach a consensus, while attempting to maximise participant output. The pilot study yielded 75 DF for vignette 1.

Within each vignette group of DF, repeats within the same category were combined to reduce the occurrence of that specific topic to one. For example, under the Resource category, three different participants (P) entered the following:

P1: “Change in land availability”
P2: “Degree in land availability”
P3: “land availability”

The three DF were combined to create a single DF “Change in land availability”. The purpose of eliminating near-identical DF is to create a repository of *unique* DF. DF were associated with their category and the threat-level (high vs low) of the business vignette read by the participant who entered the DF. DF were then randomized within each category to ensure a heterogeneous mix of threat-levels. Four DF were chosen from each category to be included in the Stage 3 study. This 4x category selection resulted in 40 unique DF for Stage 3 study (20 sourced from high-threat condition, 20 sourced from low-threat condition). The 40 DF were randomised and presented on a single PowerPoint slide, creating six templates (Appendix B reports all templates used in Stage 3 studies). The study was offered in .pptx format to accommodate as many participants as possible with available software, while meeting the unrestricted requirements of causally connecting DF with each other. The final distribution (*n*) for each vignette sub-sample in the heterogeneous study are reported in Table 8.1.

Table 8.1. Heterogeneous study business vignette sub-samples.

Condition	**Vignette**	***n***
High-threat	1	8
	2	12
	3	4
Total		*24*
Low-threat	1	6
	2	27
	3	4
Total		*37*

Vignette 2 was tested in the homogeneous study. Due to resource restrictions (time and participants), vignettes 1 and 3 were not able to be test with homogeneous samples. The original list of 116 DF produced from Stage 2 studies was used. The list of DF was divided by threat level (high vs low) according to the original sample, creating two sub-lists, one of high-threat DF and another of low-threat DF. DF were categorised by the 10 STIRDEEPER categories, randomized, and four DF were chosen from each category to be included in the Stage 3 homogeneous study, resulting in 40 unique DF for the high-threat message and 40 DF for the low-threat message.

8.4 Methods

Due to the global presence of the different samples, and the nature of the EIBE workshops at the time, a team of two additional researchers (SBS faculty) helped facilitate the experimental sessions in the UAE. In this chapter, we are collectively referred to as "the team". The addition of a facilitator team allowed the UAE testing method to use a double-blind method. The UAE-based team did not know which condition their participants were randomly assigned. The double-blind method is a stronger method for testing because it helps to reduce or eliminate certain experimental biases. Participants (N = 98) were recruited from UK and UAE. The heterogeneous study sample (n = 61) was tested in both UK and UAE, while the homogeneous study sample (n = 37) was tested only in the UK. Table 8.2 presents a breakdown of vignette profiles that were used to reflect sample demographics.

Table 8.2. Business vignette profiles presented to different participant samples

Stage	**Experiment**	**Vignette**	**Profile**		
			UK	**UAE**	**US**
3	**Heterogeneous**	**1**		✓	
		2		✓	
		3	✓	✓	
	Homogeneous	**2**	✓	✓	

This is a between subjects design, with randomly assigned conditions (high-threat vs low-threat), measuring decision-making at the individual level. PowerPoint (pptx) was used to administer the stimuli and collect the data. Banxia's (2017) Decision Explorer (DE) software package, SPSS, and R were used to analyse data. DE is used as a supporting tool to store and explore the complexity expressed by the participants, as revealed through their clusters. Participants were randomly assigned to a condition – heterogeneous (high-threat (n = 24) or low-threat (n = 37)) and homogeneous (high-threat (n = 17) or low-threat (n = 20)) – presented with a pptx file that contained a single slide with 40 DF directly associated with the organisation, and 44 red arrows. The order of the 40 DF on the page was randomized. Participants were asked to link the red arrows between the DF they perceived as having a "causal relationship", where the arrow illustrated the order of the relationship. They were also instructed how to quickly make more arrows when they needed them. Once participants felt they were finished, they submitted their ppt file. Their final product, with arrows connecting DF, is referred to as their “model”. After submitting their model, participants rated their confidence in “capturing all the relevant influential links”. A final follow-up question asked whether there were any DF missing from the list (i.e. template) that they felt were important.

8.5 **Results**

The sub-sample distributions by vignette are not uniform across the three vignettes. Further to this, the entire sub-sample collected for the low-threat condition using vignette 1 had to be removed ($n = 6$). Participants randomly assigned to the low-threat, vignette 1 sub–sample either did not complete the study (e.g. submitting blank templates) or produced incomprehensible models (e.g. arrows placed around the screen connecting no DF). It is not clear why this particular group of participants were more egregious in their behaviour compared to the other samples. Conversations with the team revealed some potential links to the causes: participants came from the same EIBE workshop, had the same research assistant leading their group, did not receive accurate information regarding their role in the experimental study, and participants may have been pressured for time. The unbalanced sample distribution created by data removals eliminates the possibility of comparing conditions (high-threat vs low-threat) across all vignettes (1-3). The final analyses, therefore, only include vignettes 2 and 3 from the heterogeneous study because only these two groups include samples from both conditions. However, due to the low sample size in both conditions for vignette 3, inferential statistics are only possible for vignette 2, and the whole sample (N) where necessary.

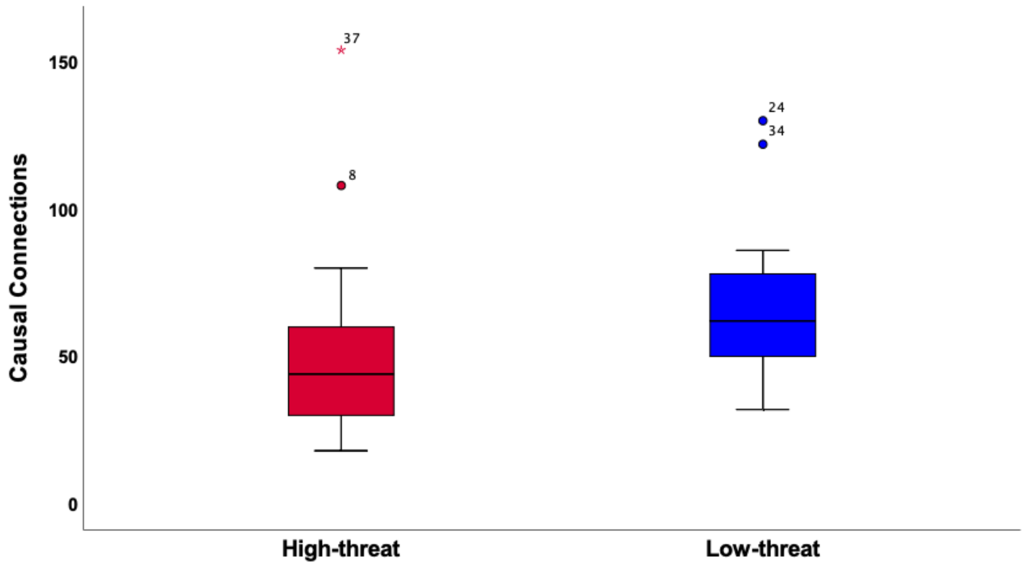
Participants created single CC between a potential of 40 different DF that cover 10 different categories. Participants were not specifically instructed to create clusters of DF, but to link one to another to represent the presence and direction of a causal relationship. Therefore, any clusters existing outside a 1-level complexity can be said to be built largely outside the conscious awareness of the participant. To make the limits of cluster complexity explicit, the least complex cluster would include two DF and one CC (e.g. Figure 8.1.a), the most complex cluster would include 40 DF and 40*39 CC. The latter would be a cluster that includes all available DF and a transitive relationship between every DF. All pptx files were converted to DE files (see Appendix C). The DE software package was used to analyse each participant model and extract unique clusters from within (Banxia, 2017). DE extracts clusters by determining the relative intensity of linkages between concepts (i.e. DF, nodes). The analysis uses a scoring system for calculation. Each DF is first scored for having a link (+) or no (-). Those with links are followed to the next DF to determine the presence of a continuing link (+) or no (-).

> The analysis takes no account of the concept text and meaning, as the underlying assumption is that the meaning of a concept is gained both from its content and from the concepts to which it is linked, and it is the relationships which are important in this analysis… Essentially a cluster analysis tries to determine relatively isolated "islands" of concepts where there are a minimum of "bridges" between the islands. Therefore the contents of the clusters produced are mutually exclusive" (p. 68).

Well separated, small clusters indicate a small amount of bridging between the different parts of the model. The parameters are set at a minimum size of "5" (the smallest allowed) and a maximum target size of "40" (the highest number of DF). Internal search for clusters is designated to begin with the first DF (1) and end with the last DF (40). By expanding the internal search to all available DF, the minimum size of "5" is adjusted to account for 1-level complexity. The seed set is designated to "none" which ensures the search criteria includes the whole template. These parameters ensure no DF and CC are eliminated from the analyses.

Initial analyses of clusters showed anomalies and extremely large variances within the homogeneous study. Therefore percentiles were calculated for the total number of CC to determine if outliers existed, by condition. Figure 8.2 shows the designated outliers for each condition, which all lie at the higher end of the count. Two outliers were removed from the high-threat condition (≥ 108 CC), and two from the low-threat condition (≥ 122 CC). Though these outliers tell an interesting story regarding causal thinking, which is discussed later in this chapter, they are treated as a separate sub-groups to the analyses. The justification for this division is because real and powerful behavioural differences in causal thinking are lost in the analyses when the outliers are averaged into the data. However, when divided along the division of ≥108 CC, an almost threshold is discovered which reveals more nuanced behavioural changes that brings greater value to the interpretations of the data.

Figure 8.2. Outliers by condition.



The number of CC developed between the DF is analysed. This is the first measure of cluster complexity. There are three types of CC that could be created within the clusters: directional, transitive, and loop. Directional CC are those illustrated in Figure 8.1.a-c, where the single direction of the arrow shows a direct causal relationship from one DF to the next. A transitive CC exists when two arrows are directed to and from the same two DF, representing a dual relationship. An example of this is given in Figure 8.1.d A loop is developed when multiple DF have a causal relationship in a manner that leaves neither a starting causal DF nor a final effect DF (Figure 8.1.e). These are chains of events that have no beginning or end. Loops also include transitive CC.

The total number of CC (TCC) created by each participant, within condition, and within vignettes is calculated (see Equation 8.1.). The sum of DF (d), occurring i times, multiplied by the number (n) of CC (c) stemming from each DF is calculated. Since 40 DF were supplied in the original template, the sum of d_i cannot exceed 40 for any single participant. However, n can occur any number of times between a and b, where a is the lower limit and b is the upper limit of CC that can stem from a single DF. It would seem intuitive to automatically designate $n = 1$ as the lower bound, however, this assumes that every cluster will be certain to have at least one DF with a single CC. Simple transitive clusters disprove this assumption by creating two CC between two DF. Therefore the lower limit is left undefined along with the upper limit.

Equation 8.1. Total number of causal connections (TCC)

$$TCC = \sum_{n=a}^{b} (\, d_i c_n)$$

For example, Pt81 developed seven DF with one CC, five DF with two CC, three DF with three CC, and nothing more (7*1 + 5*2 + 3*3) showing that Pt81 developed 26 TCC during their clustering task. Table 8.3 compares TCC from both samples (heterogeneous and homogeneous) by condition (high-threat vs low-threat) and vignette (2 and 3). Within the heterogeneous sample, though both vignettes result in fewer TCC in the high-threat condition than the low-threat condition, showing an hypothesised trend, these differences are not significant. Within the heterogeneous sample, hypothesis 6 is not supported.

However, differences by condition are more extreme within the homogeneous sample. The high-threat condition developed significantly fewer CC than the low-threat condition ($F(1,31) = 8.51$, $p = .007$, $\eta^2 = .22$). The difference also has a moderate effect size. Within the homogeneous sample, hypothesis 6 is supported.

Table 8.3. Causal connections by condition and vignette.

Condition	Vignette	*M*	*SD*	*V*
Heterogeneous				
High-threat	**2**	58.33	12.96	167.88
	3	43.50	12.79	163.67
Total		*56.58*	*16.70*	*278.78*
Low-threat	**2**	66.37	19.99	399.70
	3	58.00	8.17	66.67
Total		*65.29*	*19.01*	*361.21*
Homogeneous				
High-threat	**2**	43.33	18.37	337.52
Low-threat	**2**	60.61	15.67	245.55

Note: No significant difference at any level. The large variance in the low-threat condition of vignette 2 is explained by the higher proportion of DF with > 6-CC, compared to the single DF in the high-threat condition with > 6-CC.

The number of DF within each cluster is averaged by condition, by vignette. Size is determined by the number of DF within each cluster. Table 8.4 reports the findings of both samples. Within the heterogeneous sample, there is a wide difference in variance between conditions, which is largely the effect of vignette 3. Vignette 2 results in fewer DF, on average, per cluster within the high-threat condition, compared to the low-threat condition, however, this difference is not significant. Closely linked to this behaviour is the lack of significance between the total number of unused DF by condition, by vignette. Unused DF are those DF that were not causally associated with any other DF within a participant's model. At every level, the differences are not significant. These results show that there is no anchoring bias on the frequency of DF when organically creating clusters with a heterogeneous mix of information. Within the heterogeneous sample, hypothesis 7 is not supported.

The homogeneous sample reveals a slightly more extreme anchoring bias, in comparison. Even though the average number of DF within a cluster follows the hypothesised trend, the differences are only approaching significance and with a large variance within the low-threat condition ($F(1,31) = 3.32$, $p = .08$). It would then seem expected that the remaining unused DF should also show a similar outcome.

Surprisingly, though, there is a significant difference in the number of unused DF by participant, between conditions, ($M_{\text{high-threat}} = 15.07$, $SD = 9.35$, $V = 87.50$; $M_{\text{low-threat}} = 7.94$, $SD = 5.72$, $V = 32.76$; $F(1,31) = 7.22$, $p = .01$, $\eta^2 = .19$). What this indicates is that clusters are roughly focused on the same average number of DF (~6), but as a group, the high-threat condition focused on a smaller sub-selection of the 40 DF provided in the template, whereas the low-threat condition broadened their focus and incorporated more DF. Within the homogeneous sample, hypothesis 7 is partially supported.

Table 8.4. Driving forces per cluster by condition and vignettes.

Condition	Vignette	*M*	*SD*	*V*
Heterogeneous				
High-threat	**2**	4.84	2.36	5.55
	3	7.11	7.26	52.76
Total		*5.41*	*3.96*	*15.66*
Low-threat	**2**	6.58	2.95	8.75
	3	4.66	.66	.43
Total		*6.34*	*2.84*	*8.05*
Homogeneous				
High-threat	**2**	4.81	2.14	4.59
Low-threat	**2**	7.22	4.72	22.29

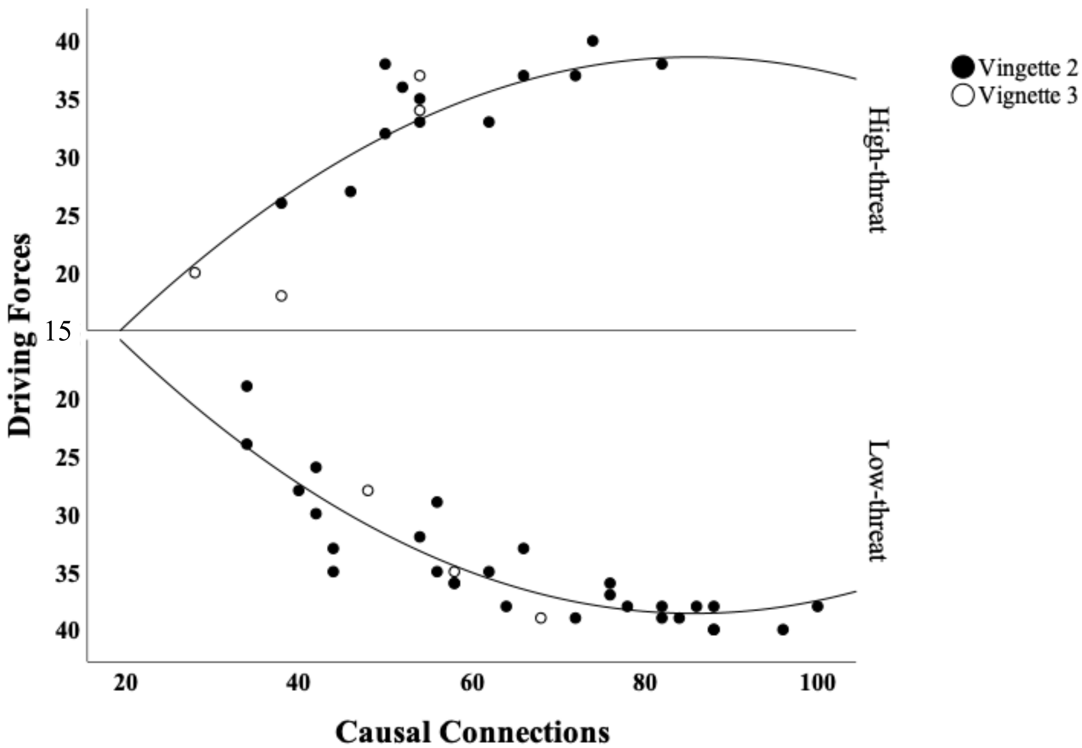
Note: No significant difference at any level

To understand cluster complexity, it is necessary to look at the relationships between the two main cluster factors (CC and DF). From a research perspective, this chapter introduces novel methods for understanding causal thinking within a SP context. As such, no single analysis can give a comprehensive picture of cluster complexity, therefore a series of analyses are used to help understand the qualities of cluster complexity and how these relate to potential anchoring biases.

The first analysis measures the frequency of CC to *used* DF to help determine whether CC concentration (number of CC per DF) is different between conditions, vignettes, and studies. Understanding levels of CC concentrated around DF within each model provides insight into a quality of cluster complexity.

Figure 8.3 plots the relationship between factors (CC and DF) from both vignettes 2 (black dots) and 3 (white dots), by condition (high-threat vs low-threat), within the heterogeneous sample. Each dot represents the total number of DF and CC in a single participant's model. Data are divided by condition, with high-threat samples plotted in the top half and low-threat samples plotted in the bottom half. The figure illustrates that as the number of CC increase, so too does the number of DF within a model, for both conditions. A Pearson test reports significant strong positive correlations between the two factors for vignette 2, within both conditions ($r_{\text{high-threat}}(12) = .76$, $p = .005$; $r_{\text{low-threat}}(27) = .85$, $p < .000$), but not for vignette 3. The small sample size from vignette 3 may explain the lack of significant relationships at this time.

Figure 8.3. Heterogeneous sample relationships between causal connections and driving forces.



Note: The y-axis zeros at the centre line.

A regression curve for vignette 2 data is plotted against each scatterplot (high-threat and low-threat) in Figure 8.3, to help understand the relationship between conditions (Equation 8.2).

Equation 8.2. Quadratic regression curve for vignette 2.

$$y = -.8 + .92 * x - 5.37E - 3 * x^2$$

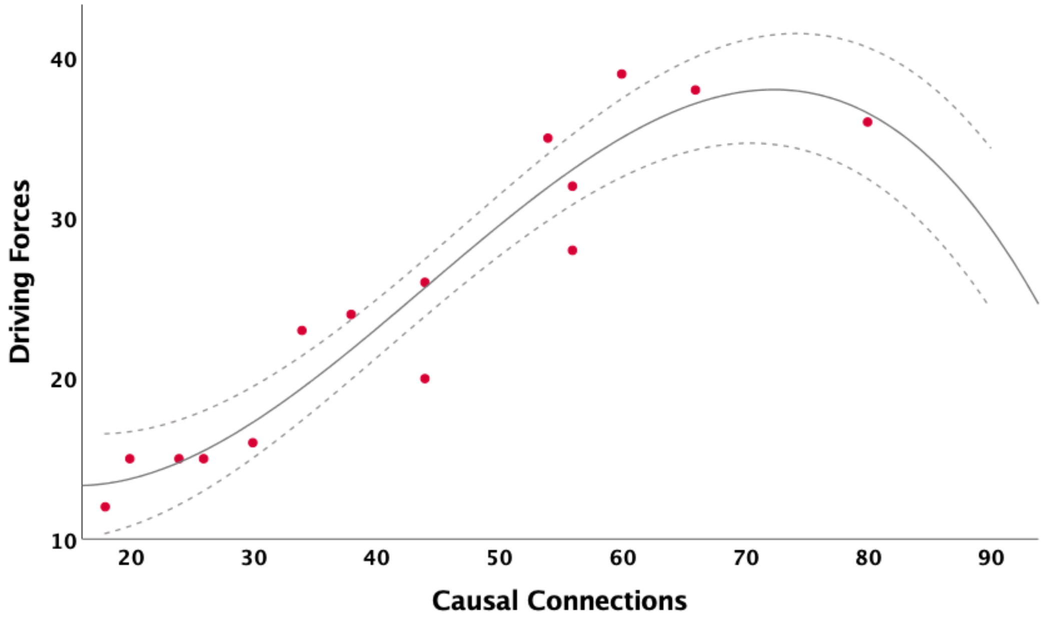
The results reveal a stronger correlation, better fitting curve, and smaller confidence interval within the low-threat condition of vignette 2 ($R^2_{\text{high-threat}} = .66$, $S = 3.00$, CI [10.27, 28.79]; $R^2_{\text{low-threat}} = .81$, $S = 2.94$, CI [15.22, 23.43]). The sample size is too small within vignette 3 to compare. The results indicate that participants within the low-threat condition may have built more 1-to-1 relationships within their models, and more consistently, where CC concentrations are lower (≤2 CC per DF), compared to the high-threat condition. Within the heterogeneous sample, CC concentration complexity results indicate no initial support for hypothesis 8.

The homogeneous sample tells a slightly different story with vignette 2. Figure 8.4 plots the same relationships between factors (CC by *used* DF) by condition (high-threat vs low-threat) within the homogeneous sample. As the number of CC are created between DF, so too does the number of DF increase within a model, both show a strong positive correlation that is significant ($r_{\text{high-threat}}(14) = .93$, $p < .000$; $r_{\text{low-threat}}(17) = .61$, $p = .007$).

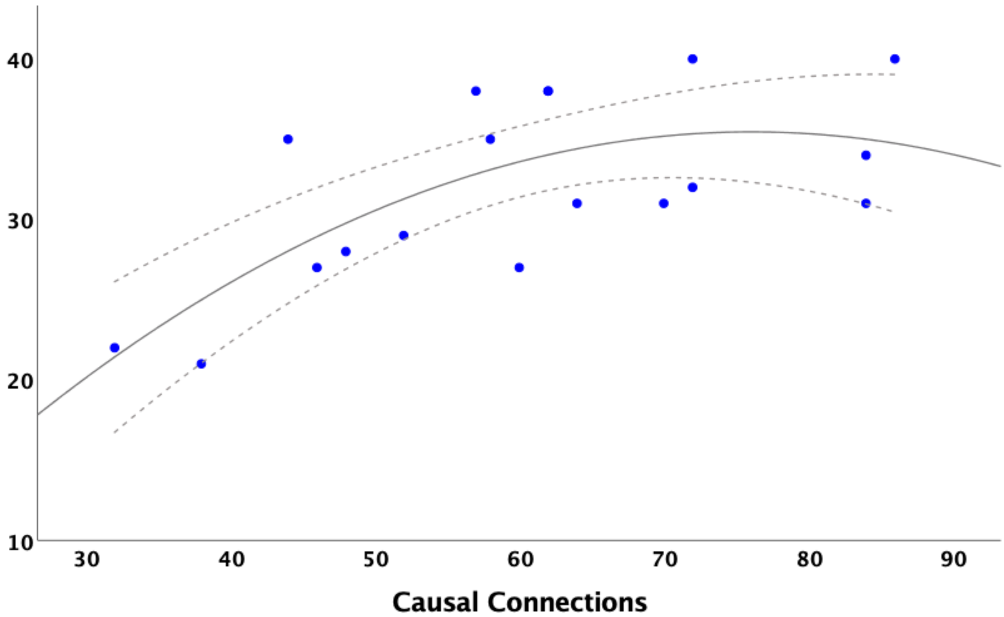
The scatterplots for each condition reveal asynchronous relationships between conditions. Both conditions develop a concave action, but reveal divisions in clustering behaviours. To quantify this difference, a regression curve is plotted against each scatterplot with 95% CI bordering each regression line. Confirming earlier asynchronous assumptions, the best fitting regression curve for the low-threat condition follows Equation 8.2, but for the high-threat condition, a different curve is required (Equation 8.4).

Figure 8.4. Homogeneous sample relationships between causal connections and driving forces.

8.4.a. High-threat



8.4.b. Low-threat



Note: The midline represents the best fitting regression curve and the peripheral lines are the 95% CI from the mean.

Equation 8.3. Cubic regression curve for high-threat condition.

$$y = .71 + .56 * x + -.01 * x^2 + 9.6E - 5 * x^3$$

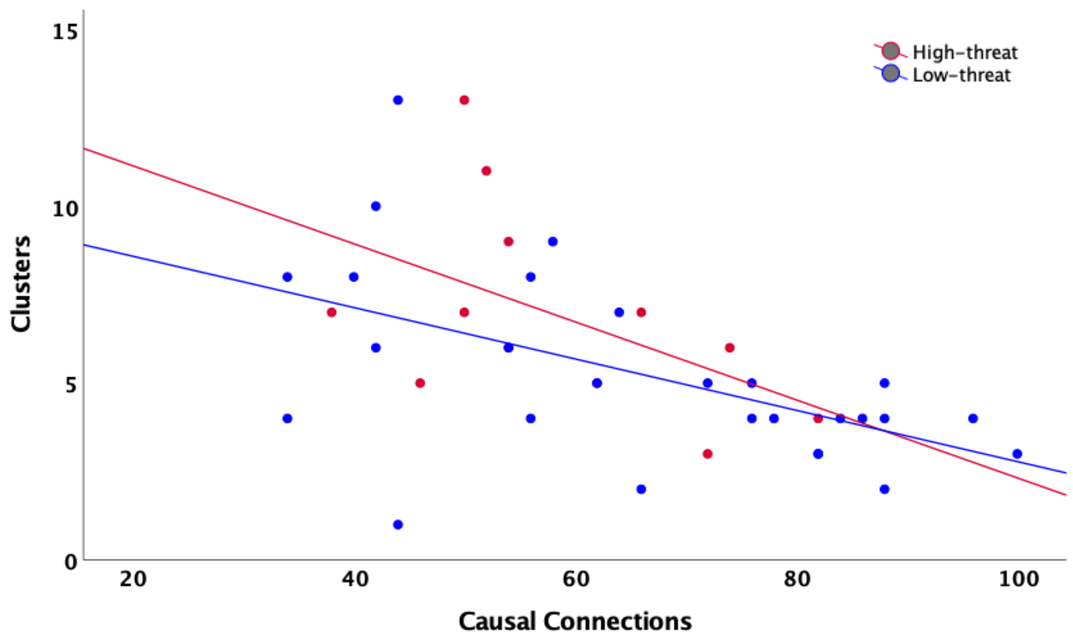
Results reveal not only cubic regression behaviours for the high-threat condition ($R^2_{\text{high-threat}} = .79$, $S = 1.37$, CI [8.28, 20.24]; $R^2_{\text{low-threat}} = .35$, $S = 4.14$, CI [16.68, 29.54]), but a stronger correlation and lower confidence interval (though not considerably smaller) where curve fit is significant ($p < .000$) for the high-threat condition. The results help show how participants within the high-threat condition were more inclined to build more 1-to-1 relationships (similar to the heterogeneous sample), where CC are simplistic (≤2 CC per DF), compared to the low-threat condition. Within the homogeneous sample, CC concentration complexity results indicate partial initial support for hypothesis 8.

However, a minority of participants created high levels of CC concentration, where a single DF became a focal point for a model with multiple CC linked to/from it. The outcomes are causally complex relationships within a fairly simplistic cluster. Though a more dramatic behaviour is exhibited in the high-threat condition (explaining the more complex cubic regression curve), this is also revealed in the data from the low-threat condition, *when outliers are included*. The four outlier data points that were removed from analyses are > 90 CC and < 20 DF. If we preserve the full low-threat dataset (i.e. outliers included) the results reveal a threshold of CC concentration complexity for both conditions, but larger for the low-threat condition. Taking into view the full dataset of both conditions further confirm an initial partial support for hypothesis 8 within the homogeneous sample.

Stated at the start of this section, the two Stage 3 studies introduce novel methods for understanding causal thinking within a SP context, and as such require further explanatory analyses, to better understand qualities of cluster complexity output. The next complexity analyses explore the relationship between CC and clusters, by condition, for vignette 2. Due to the low sample size of vignette 3 within the heterogeneous sample, only vignette 2 is explored in this section, for all samples.

Though participants were not instructed to create clusters of DF, they were instructed to “identify causal connections” between DF. By default, clusters were automatically generated. Within the heterogeneous study, Figure 8.5 illustrates the relationship between production of CC and the organic emergence of unique clusters (i.e. groups of causally linked DF that are not shared with other groups). Each dot represents the total number of CC and clusters within each participant’s model. A trend line is provided for each condition in Figure 8.5. Within vignette 2, both conditions exhibit the same behaviour. A positive trend line illustrates that as participants added CC to their model, they consequently made more clusters, implying that participants created consistently smaller clusters. Conversely, a negative trend line illustrates that as CC increased within a participant’s model, they consequently made fewer clusters, implying that participants implicitly saw connected DF as additional factors to ever more complex clusters. The relationship within each condition is generally weak for both conditions, but significant within the low-threat condition ($R^2_{\text{high-threat}}$= .25, S = 2.62; $R^2_{\text{low-threat}}$ = .29, S = 2.33; $r_{\text{low-threat}}(25)$ = -.54, p = .004).

Figure 8.5. Heterogeneous sample relationships between causal connections and cluster frequency for vignette 2.



The results help illustrate that the low-threat condition produced more consistently complex clusters than the high-threat condition. The second round of complexity analyses support hypothesis 8 within the heterogeneous study.

Figure 8.6 illustrates the conditional CC by cluster relationship, within the homogeneous study. The figure reveals that within the high-threat condition, as participants added ever increasing numbers of CC, they consequently made smaller clusters (i.e. positive trend line). However, in the low-threat condition, their behaviour inverses. As participants added more CC to their models, they implicitly saw them as additional factors to ever more complex clusters, rather than more unique clusters. The relationship is weak for both conditions, but significant within the high-threat condition ($R^2_{\text{high-threat}}$= .27; $R^2_{\text{low-threat}}$ = .16; $r(14)$ = .52, p = .05). The results help illustrate that the high-threat condition produced more consistently simplistic clusters than the low-threat condition. The second round of complexity analyses support hypothesis 8 within the homogeneous study.

Figure 8.6. Homogeneous sample relationships between causal connections and cluster frequency.

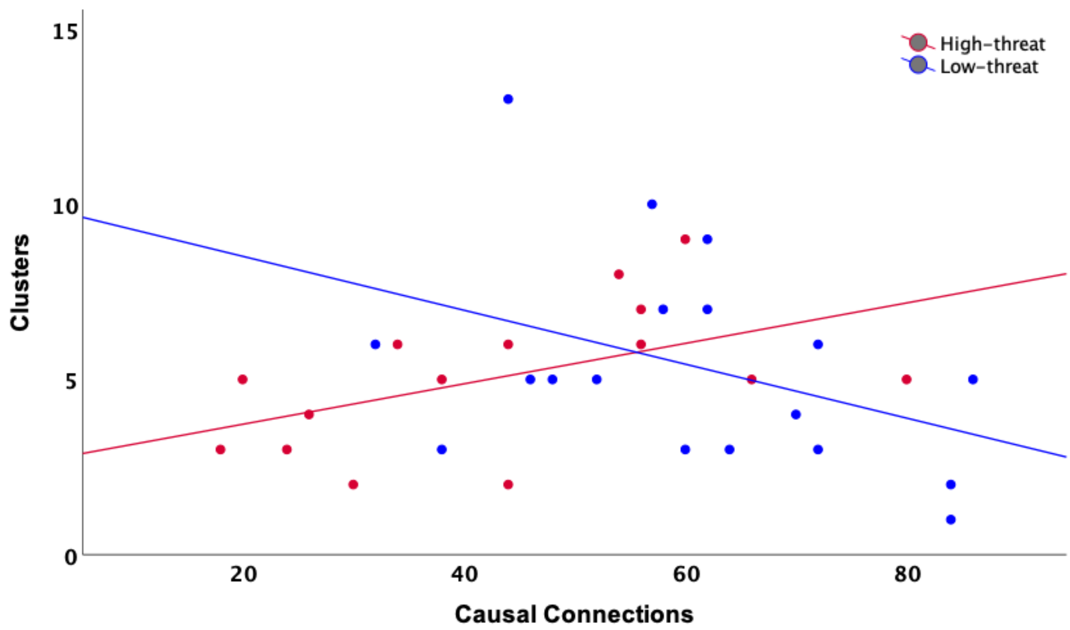


Table 8.5 reports the average number of CC within a cluster, by study, condition, and vignette. The heterogeneous study shows significant differences between conditions within vignette 2 ($F(1,37) = 4.05$, $p = .05$, $\eta^2 = .05$), but not within vignette 3. The homogeneous study also shows significant differences between conditions within vignette 2, with a large effect size ($F(1,32) = 14.84$, $p < .000$, $\eta^2 = .87$). The final complexity analyses support hypothesis 8 within both the heterogeneous and homogeneous studies for vignette 2.

Table 8.5. Average causal connections per cluster by condition and vignettes

Condition	**Vignette**	***M***	***SD***	***V***
Heterogeneous				
High-threat	**2**	10.33	6.24	38.98
	3	13.52	16.33	266.71
Total		*11.33*	*8.91*	*79.37*
Low-threat	**2**	17.54	11.62	135.11
	3	9.11	2.35	5.501
Total		*16.45*	*11.22*	*125.89*
Homogeneous				
High-threat	**2**	9.47	4.91	24.08
Low-threat	**2**	17.69	18.91	357.60

The average number of clusters developed by each participant is calculated. Table 8.6 reports the descriptive statistics of both studies. In the heterogeneous study, the table shows the wide difference in variance between the vignettes (2 and 3) by condition (high-threat vs low-threat). Further a slight reversal in behaviour is shown with average cluster frequency by participants within each condition. Vignette 2 results in more average clusters within the high-threat condition, while vignette 3 results in almost identical cluster averages between both conditions. However, there is no statistical difference at any level. In the homogeneous study, even though the high-threat condition produces fewer clusters with less variance, compared to the low-threat condition, this difference is not significant. Hypothesis 9 is not supported in either study.

Table 8.6. Cluster frequencies by conditions and vignettes.

Condition	Vignette	*M*	*SD*	*V*
Heterogeneous				
High-threat	**2**	6.92	2.88	8.27
	3	6.50	4.04	16.33
Total		*6.81*	*3.06*	*9.36*
Low-threat	**2**	5.22	2.07	7.33
	3	6.75	2.06	4.25
Total		*5.42*	*2.66*	*7.05*
Homogeneous				
High-threat	**2**	5.07	2.05	4.21
Low-threat	**2**	5.39	3.01	9.08

Qualitative analyses give some interesting levels of detail to causal reasoning efforts of ST within each condition. Anchoring biases are not just revealed in the prolific and complex nature of the cluster constructions, but also through the topical focus of the clusters.

Clusters are individually analysed to determine whether any single DF is treated as a focal point, at the beginning (causal source), middle (within the chain), or end (final effect) of the causal chains. A focal DF is one that has multiple CC directed at it (effects), from it (causes), or both (central). Focal DF are determined by first calculating the average number of CC per DF by condition, then isolating all DF that have a total number of CC above the mean, by participant. Next, all DF that are separated from the continuous cumulative count of CC by at least one additional CC are then analysed for content. For example, within vignette 2 the high-threat condition has $M = 1.69$ CC per DF, and low-threat has $M = 1.89$. Both are rounded up to 2, and all DF > 2 CC are isolated, by participant, by condition. Participant Pt140 created 22-DF with *1-CC*, 11-DF with *2-CC*, 4-DF with *3-CC*, then they skip off the cumulative path and created 1-DF with *8-CC*. The focal analysis recognises that Pt140 produced a single focal DF. By contrast, Pt135 created 19-DF with *1-CC*, 4-DF with *2-CC*, and 1-DF with *3-CC*. The focal analysis does not recognise that Pt135 produced any focal DF.

Qualitative analyses reveal a striking difference between conditions within the heterogeneous study. Figure 8.7 reports the results. The outer most circle contains the specific DF, the middle circle aligns the DF with its STIRDEEPER category, and the inner most circle reports the position of the DF within the causal chain. In the high-threat condition, only two participants developed one focal DF in each of their models. Both focal DF were at the end of a cluster chain and considered a factor that was affected by six ("Tourism") and eight ("Level of public interests") different DF, respectively.

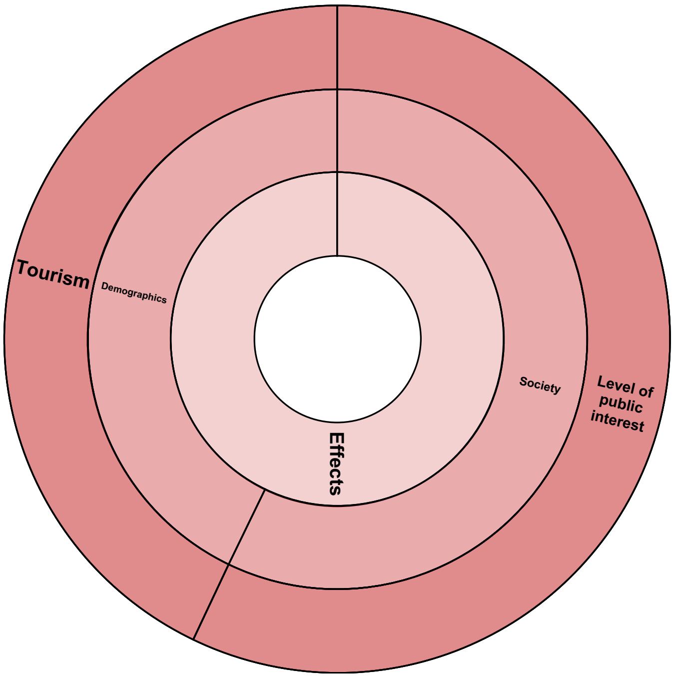
By contrast, 52% of the low-threat condition created at least one cluster with a focal DF. A single demographic DF, "Tourism", proved to be the most popular and treated almost exclusively at the end of causal chains – reflecting a similarity to the high-threat condition. For the Abu Dhabi sample, the society outlier DF "Impact from Abu Dhabi Vision 2030" was treated as the collective causal source for at least seven other DF in the clusters. Differences in focal DF (frequencies and causes vs central vs effects) between conditions, shows support for hypothesis 10 within the heterogeneous study.

The homogeneous study also reveals striking differences between the two conditions, but qualitatively different from the heterogeneous study. Within the high-threat condition, 20% of the clusters are constructed around at least one highly focused DF. The most highly focused DF are Politics (3), Economics (3), Society (3), and Demographics (2) (see Figure 8.8). The number of participants who built focal clusters with the specific DF is within the parentheses that follows the topic. Society and Demographics focal DF are located at the end of the causal chains, Politics DF are almost entirely at the beginning of the causal chains, Economics DF are divided between one acting as a main causal source and the other two acting as final sources.

What is most remarkable about these results is that developments of the same DF by different participants were oriented in the same manner within their respective clusters! For example, only two participants highly focused on "Impact of tourism", and both developed this DF in the same way within their own respective clustering efforts by making it the highly-focused end to their causal chains.

Figure 8.7. Focal driving forces by condition

8.7.a. High-threat



8.7.b. Low-threat

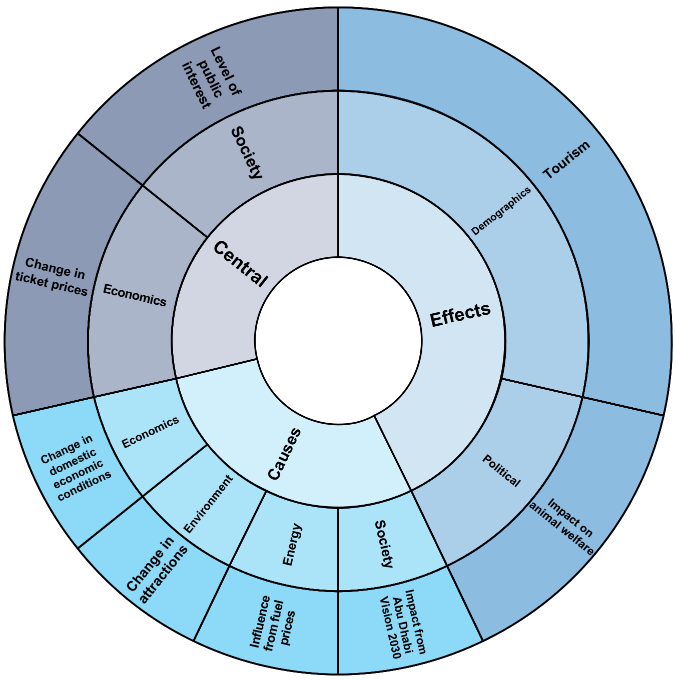
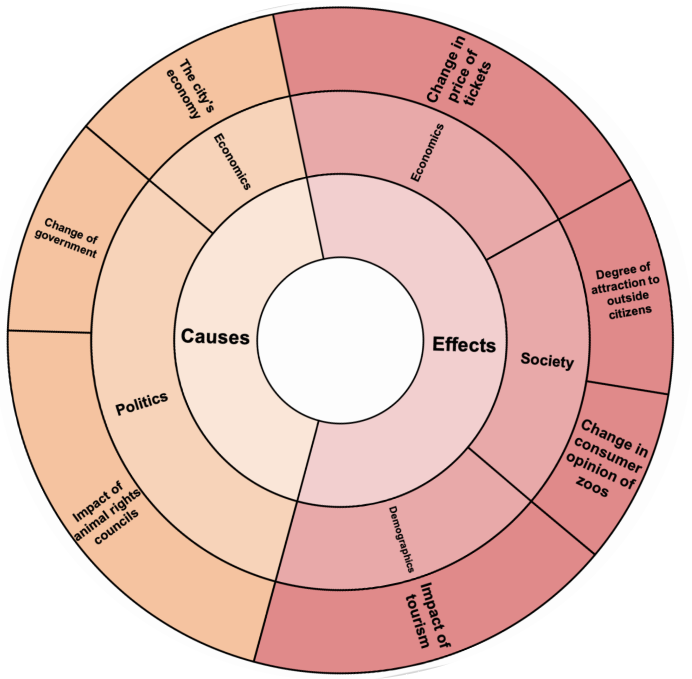
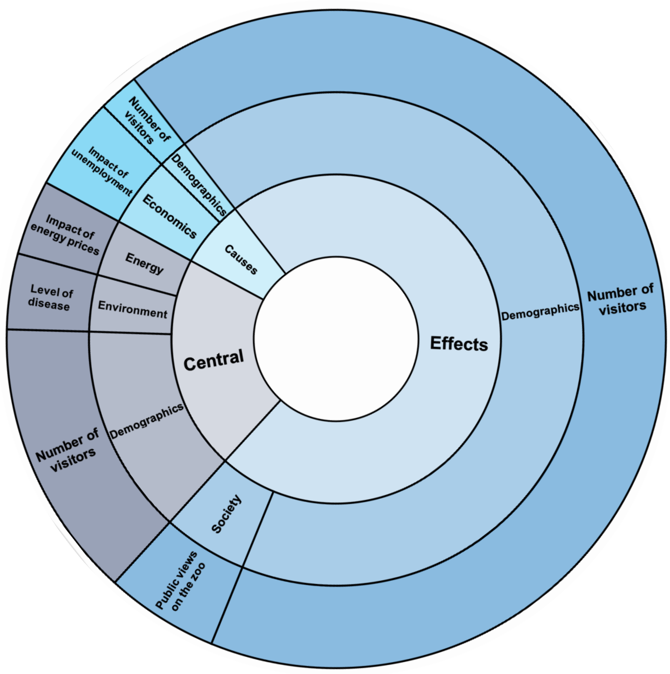


Figure 8.8. Focal driving forces by condition

8.8.a. High-threat



8.8.b. Low-threat



There also exists a strong focal effort within the low-threat condition, but with different outputs. Seventeen percent of the clusters are constructed around a focal DF. The most highly focused DF are Demographics (12), Society (1), Economics (1), Energy (1), and Environment (1). A single demographic DF "Number of visitors" proved to be the most highly focused DF across 12 unique clusters. Most of the highly-focused DF are located at the end of the causal chains, less than a fifth are centrally located, and a minority (<10%) are designated to the beginning of their respective causal chains. Unlike the high-threat condition, different functions are seen for the same DF (see Figure 8.8). Taking both conditions together, a minority in DF alignment shows some level of shared intuitive functioning of DF within the given organisational environment. Differences in focal DF (frequencies, topics, categories, and causes vs central vs effects) between conditions, shows support for hypothesis 10 within the homogeneous study.

Confidence scores were taken after participants completed their clustering task. Table 8.7 reports the descriptive statistics for both studies by condition and vignette. Though scores for vignette 2 follow the hypothesised trend in both studies, the differences are not statistically significant, and therefore do not support hypothesis 11.

Table 8.7. Confidence scores by conditions and vignettes.

Condition	**Vignette**	***M***	***SD***
Heterogeneous			
High-threat	**2**	3.45	.87
	3	3.50	.50
Low-threat	**2**	3.62	.61
	3	3.50	.53
Homogeneous			
High-threat	**2**	3.22	.91
Low-threat	**2**	3.37	.88

Table 8.8 reports the descriptive statistics for both studies by condition and vignette. From vignette 3, two participants from the low-threat condition created further DF, and they submitted only one DF each. Since no DF were created by participants in

the high-threat condition, no inferential analyses can be run. However, the stark presence/absence of DF between conditions shows a trend in the hypothesise direction. Vignette 2 resulted in slightly more creative activity. Half of the participants from each condition (high-threat = 50%, low-threat = 52%.) submitted more DF. The high-threat condition produced a lower average number of novel DF, following the hypothesised trend, but the differences are only approaching significance (p = .06). Within the homogeneous sample, a larger percentage of participants, overall, between conditions (high-threat = 76%, low-threat = 75%) offered further DF. Participants, however offered about the same number of further DF, revealing no significant difference between average additional DF by condition. Hypothesis 1 is not supported in either study for serendipitous creativity.

Table 8.8. Serendipitous post-causal thinking by conditions and vignettes.

Condition	Vignette	*M*	*SD*	*V*
Heterogeneous				
High-threat	**2***	1.67	1.21	1.47
	3	1	-	-
Low-threat	**2***	3.43	2.06	4.26
	3	1	-	-
Homogeneous				
High-threat	**2**	1.47	1.18	1.39
Low-threat	**2**	1.30	1.26	1.59

* p = .07

8.6 Causal Thinking Discussion

Building from the data provided from the Stage 2 studies, two further experimental studies were developed to test qualities of associative causal thinking employed in Stage 3 of the IL/ST framework. Both studies measured individual-level thinking, after receiving DF from earlier group efforts. Though all vignettes were originally delivered in a randomized method to participants, only vignettes 2 and 3 resulted in large enough sample sizes to further analyse. Even within this sub-sample, only vignette 2 samples proved to remain large enough after data cleaning to perform inferential analyses across for hypothesis testing. It would appear, when all measures are brought together, that causal thinking, as a function of ST, is susceptible to output biases after anchoring to differently framed messages. Further to this discovery is support for Yaniv's (2011) claim that groups who comprise a heterogeneous mix of information (whether through preparatory research, members' different fields of expertise, age, gender, education levels, etc) will attenuate potential biasing effects, compared to homogeneous groups. To look at the two samples side by side, the patterns become clearer. Table 8.9 reports a summary of the key findings from this chapter. Only the outcomes of vignette 2 from the heterogeneous sample are reported in the table, alongside the homogeneous sample. Anchoring biases are found almost entirely within the homogeneous sample (twice as many as the heterogeneous study), and only strongly within the qualitative measures of the heterogeneous sample, therefore, showing support for H12.

Due to the novelty of both study methods, a number of other data features could be explored and measured for anchoring biases. The literature review did not explicitly lead to including many of these other features. However, the information they convey is not only empirically interesting, but potentially informative for both future research investigations and practitioners. Therefore, additional exploratory analyses for Stage 3 studies are provided in Appendix D, along with discussions on the novel discoveries.

To help answer research questions 1 and 2, H6-H10 tested for anchoring biases across several output dimensions, between high-threat and low-threat framed messages at the individual level. The homogeneous study reveals that when participants work with DF that were developed by groups of participants who all read the same framed messages about the same business (Stage 2), then message framing has a stronger effect on assuming causality between recognised DF (H6), which DF are incorporated into a person's model of the business environment (H7), how long chains of events may continue (H8), and the specific topics that become most salient for addressing an organisation's goals within a given timeline (H10).

Table 8.9. Summary of samples and conditions by hypothesis

Hypothesis	Dependent variable	Heterogeneous		Homogeneous	
		High-threat	**Low-threat**	**High-threat**	**Low-threat**
1	DF	No anchoring bias, but trend shows *fewer* new DF	No anchoring bias, but trend shows *more* new DF	No anchoring bias	No anchoring bias
6	CC	No anchoring bias, but trend shows fewer average CC per model	No anchoring bias, but trend shows more average CC per model	Anchoring bias, fewer average CC per model	Anchoring bias, more average CC per model
7	DF	No anchoring bias, but trend shows fewer average DF per cluster	No anchoring bias, but trend shows more average DF per cluster	Partially supported anchoring bias, fewer average DF per model, Similar trend per cluster	Partially supported anchoring bias, more average DF per model, Similar trend per cluster
8	Complexity	Partially supported anchoring bias, less complex clusters	Partially supported anchoring bias, more complex clusters	Anchoring bias, less complex clusters	Anchoring bias, more complex clusters
9	Clusters	No anchoring bias, but trend shows more on average	No anchoring bias, but trend shows fewer on average	No anchoring bias	No anchoring bias
10	Focal DF	Anchoring bias, only two categories and both as effects	Anchoring bias, several categories evenly divided between cause, effect, and central	Anchoring bias, even divide between cause and effect with several focal categories	Anchoring bias, highly focused on a single category with mostly effect positioning
11	Confidence	No anchoring bias, but trend shows lower average score	No anchoring bias, but trend shows higher average score	No anchoring bias, but trend shows lower average score	No anchoring bias, but trend shows higher average score
12	Study	Fewer anchoring biases across analyses		More anchoring biases across analyses	

When compared to participants who received DF from groups who read differently framed messages about the same business (Stage 2), the heterogeneous study reveals that framing effects are still *mildly* detected across all measures, but strongest with the specific topics participants find most salient and integral for the organisation's strategic efforts. The difference in topical focus may explain why participants developed shorter chains of events after reading a high-threat framed vignette.

It is important to discuss the covarying risk a heterogeneous design introduces. Any anchoring biases that may be present run the risk of being potentially masked by noise of the heterogeneous design. Or more severely, the detected effects in the analyses are actually measuring a heterogeneous effect. The heterogeneous study brings together potentially disparate information. The DF developed by participants in Stage 2 studies aligned with their world view as defined, in part, by the framed vignette they read. Several studies have already shown that people prefer to attend to and evaluate information of plausible (i.e. consistent) hypotheses, rather than implausible (i.e. inconsistent) hypotheses (Fugelsang & Dunbar, 2005). By mixing DF from Stage 2 studies into a heterogeneous template, it may have been the case that half of the DF were inconsistent with the world views of Stage 3 heterogeneous participants, as developed by the framed vignette they read. Therefore, inconsistencies between the 40 DF and participant expectations and causal reasoning may have led participants to selectively attend to and associate some DF (≤ 20) over others (Fugelsang & Dunbar, 2005). The qualitative results in both this chapter and Appendix D provide important explanatory support for the discussion. The heterogeneous clusters show that even though there was a difference in focal DF between the two conditions, there were also considerable differences within each condition (e.g. transitive and loops). Where we can see some alignment of thought, is within the shared focal categories across the two conditions and the shared DF across the outlier (focal) analyses. Both conditions were heavily focused on demographics and society, while "Tourism" and "Level of public interest" are mutually seen as major future factors that will be affected by a series of different DF.

To address research question 3, H11 tested for an anchoring bias in participants' confidence in their abilities to capture all the relevant causal connections between a list of DF. Similar to Stage 2 studies, there were no significant differences in self-evaluated confidence scores, by study and by condition, even though trends mostly followed the hypothesised direction. The consistent trend, however, of low-threat messages leading to higher average confidence scores is interesting, in and of itself. P-values are only one method for understanding the sensitivity of different cognitive motivations. In this light, confidence should continue to be studied against decision-making efforts, particularly when the assumption is that the two phenomena should have no relationship. Further to this, it is important to note that in both Stage 2 and Stage 3 studies, average confidence scores for all conditions remain higher than the mid-point (>3). Perhaps the studies are beginning to detect a separate bias within participants' self-evaluation of their own confidence. Both the lack of strong framing effects and consistently higher confidence may be revealing how little confidence is related to knowledge and performance.

To address research question 4, H12 tested for anchoring biases between studies. There remain major gaps in the narrative being built around SP and ST. One of them concerns the susceptibility scenarios can reflect from external (i.e. environment) and internal (i.e. the mind) influences on practitioners. Information dis/continuity as a product of group input is absent from scenario planning and futures-thinking scholarship. Stage 3 studies add to not just our knowledge, but as practitioners, our awareness of mental sensitivities within scenario practitioners. Stage 3 studies further support group-based research that show how biases of varying qualities can become more prevalent in groups that share commonalities, compared to other group orientations that do not (Janis, 1971; Tetlock, 2005; Yaniv, 2011).

8.6.1 ***Types of Causality***

What is not tested in Stage 3 study is the type of causal relationship created between DF. Did participants use a normative approach to ST? This would result in identifying expected causal relationships based on past regularity of those relationships. Most agree that it may be impossible to create a scenario free of normative inclusions, if nothing else, at the implicit level. So it is not unreasonable to expect some normative reasoning.

However, SP finds its strength in guiding practitioners towards seeing the dynamic, unexpected, unusual, and black swans. Einhorn and Hogarth (1982, p. 25) emphasize that "as full range of *different* types of causes as possible" are required when dealing with the inescapable issue of various futures' uncertainties. Derbyshire & Wright (2017) address the need for employing different types of causal thinking during scenario planning by way of Aristotle's theory of causality. Aristotle holds that a vital condition for a successful investigation requires understanding all the types of causes of the world (Hocutt, 1974).[16] Of Aristotle's four main causes, "efficient cause" – the primary source of the change or rest – is probably the most widely employed causal perspective used in SP, more commonly known as cause-and-effect (Derbyshire, 2016; Derbyshire & Wright, 2017). Unlike the other causal forces, efficient cause represents a very strong time perspective from the practitioner. Creating clusters from the available DF is an effort in thinking along a linear time line with a chronological perspective of events. However, due to this more salient chronology, many practitioners only reference efficient causes in their causal thinking. They fail to engage in deeper considerations, which leads to little more than a shallow consideration of the emerging futures (Wright & Goodwin, 2009). Part of the reason people focus only on efficient causes is because it is less cognitively taxing. In this way, efficient causal thinking is a heuristic. However, the purpose of ST is to challenge conventional thinking and improve decision-making. Did participants use this narrower reasoning effort to determine causal relationships? We cannot know the answers to these questions with the methods and data of Stage 3 studies, but understanding such features of causal thinking could greatly inform the efficacy of SP.

[16] Aristotle's general account of the four causes is repeated between *Physics* II 3 and *Metaphysic*s V 2.

8.6.2 *Future Methods*

Stage 3 studies only investigate individual-levels of causal thinking, and not group-levels, like Stage 2 studies explored. Given the constructivist view of SP research, it would be valuable to the theory to understand individual-level, group-level, and transitional thinking. All bring strengths and weaknesses to data interpretations.

A transitional study would be a within-subjects design. Within-subjects designs offer greater statistical power over between-subjects designs, and in the case of the SP method, within-subjects designs more closely resemble the reality of ST. A one-day method could task participants to provide CC between DF on their own (individual), then as a group with other participants (individual+group) who read either the same threat-message (homogeneous) or different (heterogeneous). A potential drawback for the one-day design is that it may prove exhausting for participants. Participants took, on average, 30-60 minutes to complete their tasks. An individual+group one-day method has the potential to require participants to remain in a high-level thinking and reasoning exercise for 2 hours or more. Such a design is cognitively, emotionally, and physically draining. Therefore, the validity of results would be suspect.

An alternative transitional, within-subjects design could attempt to limit DF to shorten the overall time for the participants (e.g. thresholds or max/min requirements). Any limitations, however, also limit potential data points from a single group. In turn, such a design weakens validity and generalisability of the study.

A third within-subjects design could mimic the methods of the individual+group study in Stage 2 experiments. The individual and group portions of the study could be divided across two days, where the individual session was held on the first day and the group session was held the next day. However, attrition rates are increase by an order of magnitude when participants are required to return for further testing. This is a well-documented behaviour across medical, psychological, and sports studies.

Alternatively, a between-subjects design could be developed that compliments the present individual-only studies of this chapter with a second group-only study. Even though the results from a between-subject design could be argued as a proxy for ST behaviours, statistically, the power would be much lower, introducing greater issues behind abilities to generalise any knowledge gained.

Following the step-wise process of this investigation, Chapter 9 builds from the data and knowledge discovered from Stage 3 studies. Two final empirical studies will attempt to measure the next stage in the IL/ST framework (Stage 5) against the prevailing cognitive effort (evaluative thinking).

Chapter 9. **Evaluative Thinking (Stage 5)**

The revolution will not be televised…NBC
will not be able to predict the winner at
8:32… The revolution will not be televised.
Gil Scott-Heron c.1970

"The official future will not take place…
most studies that address the futures also
tend to be cheerily (or gloomily) dogmatic:
this *will* take place. But it won't. At least,
not the way you think."
Pamela McCorduck & Nancy Ramsey c.1996

The final empirical chapter builds on information gained from the Stage 3 associative causal thinking studies in Chapter 8 and expands into testing qualities of evaluative thinking. For this investigation, the order of experimental studies moves through the IL/ST framework from Stage 3 to Stage 5. As stated at the beginning of Section II, the overall design moves the testing process from Stages 3 to 5 in part to the variable nature of integrating Stage 4 task into the IL process. At the time the overall experimental design was in development, the primary IL model taught within the SBS MBA programme integrated a parallel thinking task into Stage 6 of the IL/ST framework. Therefore, when developing the studies, the dominating IL model within the researcher's programme of study was followed. Furthermore, a more streamlined method proved to be popular in both practice and praxis, due to time-poor practitioners. Offering empirical evidence that follows a more streamlined method can help increase the practical value of any discoveries. The evaluative studies preserve an individual-level testing focus and a heterogeneous vs homogeneous comparative method. The main purpose is to help preserve continuity between the studies, thereby increase applicability of results, which is largely addressed in research question 2 (*Are anchoring effects comparable across different ST cognitions?*). The chapter begins with a focused literature review that completes the methodological background for developing two experimental studies to help better understand how we evaluate future-oriented perspectives of impact and predictability. SP scholarship is sparse with discussions and empirical research around evaluative judgments, outside forecasting points and predictions (*see for example* Kuhn & Sniezek, 1996; Önkal, Sayim, & Gönül, 2013; Sedor, 2002). Experimentally testing Stage 5 tasks requires methodological techniques that break into new territory more so than the previous empirical chapters. Therefore, Stage 5 studies begin with an exploratory approach, much like a pilot study, to gather initial understandings of potential behaviours and outcomes. Propositions (1-3) are developed from the literature review. Exploratory analyses follow from the propositions which help inform whether the chosen method is informative enough for addressing the four research questions. A brief discussion is provided to summarise some of the knowledge gained from the exploratory study. Discoveries revealed in the exploratory discussion lead to the development of six hypotheses and a second experimental design. Hypotheses 13-15 are developed to help address research questions 1 and 2. Hypothesis 16 aims to help address research question 3. Finally, based on surprise discoveries from the exploratory study, hypotheses 17 and 18 address research question 4.

After practitioners employ their creative centres to think divergently, then their causal centres to think associatively, they next switch to a categorically different cognitive effort, evaluative thinking (often absent of numerical quantifiers). In Stage 5, practitioners evaluate the level of future *impact* each cluster could have on the organisation reaching its stated goals, within the timeline, and the level of *predictability* (i.e. a measure of uncertainty) of each cluster of events occurring. The aim of this stage is to focus efforts on exploring and testing the most pertinent information that emerges from the process, for later scenario development. It is possible, of course, to develop scenarios using all the clusters that exist by the end of Stage 3. However, the outcomes of Stage 3 clustering efforts, more often than not, result in numerous unique clusters of various complexities. For example, a brief survey of clusters developed in EIBE workshops held within the Department of Management Science at SBS shows that the majority of SP groups develop between nine and 12 clusters from Stage 3, with an average of 120 DF developed since Stage 2. Aligning with similar outcomes, Chapter 8 studies revealed the average numbers of clusters organically developed from a list of only 40 DF fell between five and seven. It is not uncommon, and in fact encouraged, for several hundred DF to be identified during a SP workshop. The results lead to a large collection of clusters that can eventually become divided into their own thematic sub-groups. To create scenarios from all the available clusters within a single workshop would not only be more resource taxing, the task risks cognitive overload for the workshop practitioners, and the end result is a collection of numerous scenarios that achieve little more than information overload. Chermack (2011, p. 223) even considers this one of the top 10 pitfalls of SP, "*Too many scenarios.* Do not use more than four scenarios… Having more than four scenarios is overwhelming for decision makers and complicates the project." In short, more information is not always better at this stage of the process. To ameliorate such potential downfalls, the IL model triages clusters by those that may have the highest levels of impact on the organisation, but are the most difficult to predict (i.e. the highest levels of unpredictability).

9.1 **Anchoring Bias and Processing**

The two Stage 5 studies aim to measure different anchoring biases in non-numerical evaluative thinking. Evaluative thinking is "critical thinking applied to contexts of evaluation" (Buckley, Archibald, Hargraves, & Trochim, 2015, p. 376). The levels of critical thinking practitioners apply to their evaluations of clusters varies on a continuum from shallow to deep. Shallow evaluations refer to quick, implicit-level processing, where heuristics are heavily employed, whereas deep processing is more deliberative, elaborative thinking, guided by greater motivation and individual abilities to assess the issues (Chaiken & Trope, 1999; Deutsch & Strack, 2006; Kahneman, 2011; Petty & Bruñol, 2012; Tversky & Kahneman, 1974). Shallow processing can occur when people have incomplete information, and infer, instead, what a conclusion will be. In reality, people arguably always have incomplete information. Deep processing is facilitated when people run into conflict and must analyse the details of the situation, order their thoughts, a explicitly pay attention to something. Evaluative thinking benefits from both modes of thinking. Anchoring biases are believed to affect shallow processes more than the deep processes, as is argued in the previous studies. Shallow processes, however, also inform the slower, more deliberative deep processing of information (Kahneman, 2011). Therefore it could be the case that an anchoring bias affects all or any level of evaluative thinking.

A popular technique in the IL model is to give evaluations of impact and predictability without stated numerical values. Measures are given qualitative values: “low” to “high”. This design is used for several reasons. The main benefits for an experiment to include numerical values in forecasting would be to compare against different samples, as a measure of difference, and/or the true values, as a measure of accuracy. A measure of difference is, in fact, analysed within this study, but exposing participants to numerical values for this comparison is not necessary. There is no standard by which practitioners could know the exact value of impact or predictability for any given cluster, therefore a measure of accuracy is not applicable to this study. Another issue with making numeric values explicit is that these could be translated into anchors for participants’ open-ended evaluations and introduce confounding biases to the study. One way to address this could be to use standardised probability measures, such as those found in the US Intelligence Community Directives (2015), where probability and likelihood (i.e. uncertainties) language is standardised. However, as Rowe (2010, p. 2) discovered, people perform equally deficient at translating uncertainties using either estimates of verbal or numeric form, and that “formalized systems for defining probability terms do not perform well.”

9.2 **Impact**

Evaluations on potential future impacts seek a measure of magnitude. How great will the impact of the events be on the organisation’s operations? Will there be no impact, affecting nothing of the operations, supply chain, profit margins, ROIs, price stabilities, overhead, labour costs, etc.? Will there be moderate or great impacts? From the EIBE manual, “’Impact’ in this case relates to the impact of the driving force on the client’s areas of interest” (Wright & Bradfield, 2015, p. 11). Clusters are more than a single focus of impact. They are constructed from a number of DF with various forms of causal associations between them. It is conceivable to assume that within a single cluster the various DF could be perceived to have different levels of impact on the organisation’s future. This is another reason why numerical values are not made explicit in the valuations, and rather, intuitive valuations are elicited. The discussion leads to development of the first proposition for the exploratory study.

Proposition 9.1 The range of clusters across an impact scale offers a measurable difference for understanding anchoring biases on evaluative thinking.

9.3 **Predictability**

Unlike magnitudes of impact, evaluations on predictability seek a measure of probability. The EIBE manual describes predictability in terms of the DF within the clusters. "'High Predictability' means we are reasonably certain that the driving force will play out in ways that are fairly well understood; 'Low predictability' means that we have no clear idea which of a number of ways it might go" (Wright & Bradfield, 2015, p. 11). Notice here that predictability and confidence are closely associated with the task. Everyone uses probabilistic thinking. This type of thinking does not require special knowledge in probabilistic calculus to be utilised in daily life (Johnson-Laird, 1994). In fact, when presented with a problem that requires a probabilistic answer, people often do not follow the propositional calculus to assess the probability, but rather tend to rely on shallow processing heuristics, such as similarity in proximally close factors, past behaviours and trends (Kahneman, 2011; Rowe, 2010). However, probability evaluations can also be affected by deep processing. Gregory et al. (1982) argued that additional cognitions such as familiarity can increase imaginability, and the ability to imagine can increase saliency. Increasing the salience of an event "may lead a person to believe more strongly that the event will actually happen" (Dougherty, Gettys, & Thomas, 1997, p. 137). Therefore, the more practitioners deliberate and familiarise themselves with the clusters, the more likely they are to increase their probability evaluations. So it may be the case that both shallow and deep processes can affect predictability evaluations. From the discussion, a second proposition is developed.

Proposition 9.2 The range of clusters across a predictability scale offers a measurable difference for understanding anchoring biases on evaluative thinking.

Evaluating both impact and predictability are carried out together in Stage 5. Some IL techniques divide the two tasks, where practitioners evaluate clusters by one measure first (e.g. impact), then on the second measure (e.g. predictability). Other techniques task practitioners to evaluate each measure together for each cluster. Practitioners visually perform both tasks by placing representations of clusters (i.e. post-it notes, colour coded shapes, or headings) on a physical matrix, where impact is represented across one axis and predictability across a perpendicular axis, more commonly known as the Impact/Predictability (I/P) matrix. Regardless of chosen technique, practitioners must perform both evaluations, and due to the dual effort, an interaction of evaluations develops. The discussion helps develop a third proposition.

Proposition 9.3 The interaction of impact and predictability as a matrix is important to SP and therefore should be taken into consideration when understanding anchoring biases.

9.4 Exploratory-Heterogeneous Study

The first Stage 5 study is designed to help explore the types of methods and analyses that could help detect evaluative thinking. The chosen design is of a heterogeneous mix of clusters. This design is chosen, first, because it offers a base-line for measuring potential anchoring bias. It reduces confounding variables (as reported in the Stage 3 studies). All participants, in both framing conditions, are given the same stimulus (i.e. clusters) with which to work. In this way, potential anchoring biases can be better isolated to the threat-message instead of stimulus content.

Several methods were explored to aid the construction of heterogeneous clusters for the first Stage 5 study. The first method proposed to develop new DF and clusters as the stimulus for Stage 5 participants - the "New DF/Clustering" method. New DF could be generated by the previous research team[17] (see Chapter 8), followed by clustering of their DF. This method promises the quickest turn-around in time, and

[17] The research team is faculty within the SBS who have working knowledge of IL models of SP. Though they only directly supported Stage 3 studies, they remain a potential resource for expert knowledge in SP content and practices.

offers the bonus of sourcing information from potentially a well-informed group of people (e.g. expert sample). However, the "New DF/Clustering" method is not used because it risks biasing the stimulus in a manner that could remain undetected. As the Stage 3 studies and others have shown, a generally homogenous group of people are more susceptible to producing biased content than heterogeneous groups, however these similarities/differences are defined. A homogeneous group of experts risks confounding the data. With the concern of time resources, perhaps Frith (2020, p. 1) said it best when she said, "Fast Science is bad for scientists and bad for science." Alternatively, a pilot study could be conducted where a random sample of participants identify DF and construct their own clusters. This method offers more control over creating a heterogeneous group. The "Pilot" method, however, is also not used. Neither method is chosen due to resource constraints in one area and availability in others. New data would require investments in analyses (e.g. code, weight, and select the clusters for a heterogeneous study). Which is to say analyses and time-investments would be similar to a third method that is proposed. The existing repository of DF from Stage 2 studies and clusters from Stage 3 studies could be analysed to help inform cluster construction for the first Stage 5 study. The "Existing Data" method is chosen because it helps eliminate extraneous work required to create a new data set, offers a higher level of ecological validity by using data from participants who read the same business vignettes from the start, the highest level of consistency by maintaining the same repository of data used in existing studies, and does not appreciably increase the amount of researcher's time required to develop the stimulus. A heterogeneous cluster is one with a balanced representation of high-threat and low-threat DF created in Stage 2 studies. The cluster construction also attempts to reflect a balanced representation of threat-message influence by containing the strongest causal associations between DF shared between the two threat-level conditions from Stage 3 studies.

9.4.1 ***Ethical Approval for Human-Based Study***

Ethical review was submitted through the Department of Management Science (fka "Strategy and Organisation"), within the SBS to the UEC and receive approval on 24 January 2018.

9.4.2 *Apparatus & Stimuli*

All business vignettes (1-3) are prepared as primes in the exploratory-heterogeneous study. Cluster features of frequency, complexity (DF and CC), and topic are controlled for. Cluster frequency is determined by the STIRDEEPER categories with an addition of an *Other* category. Each cluster focuses on a single category. The results from Stage 3 studies help determine cluster complexity. The average number of DF per cluster in the heterogeneous study is M = 5.41 and in the homogenous study, M = 6.00. Therefore all clusters in the first Stage 5 study include six DF. The CC between DF in Stage 3 studies show that participants favoured simplistic constructions with mostly ≤2 CC per DF. This model is followed, and after conducting factor and cluster analyses (explained the next paragraph), clusters are designed to have an average of eight CC. By maintaining the same complexity profile for all clusters, we can control for some of the potential extraneous variables.

To create a heterogeneous mix, cluster relationships are analysed from the heterogeneous models in the first Stage 3 study. DF are already divided into a balanced representation of 50% high-threat and 50% low-threat influenced variables. This balance is carried through to the division of DF across the 10 STIRDEEPER categories. Factor analyses are conducted to measure direct relationships between DF and displayed in a matrix of frequencies. The matrix results in 40 DF with a numerical value representing its associative relationship with the other 39 DF. Figure 9.1 illustrates a sample of the matrix built from vignette 2. It shows the first 20 DF presented in the template from the heterogeneous low-threat condition. The first column and first row are numerical representations of the first 20 DF. The matrix then sums the total number of CC between each DF from the low-threat condition within the heterogeneous sample, colour coded by frequency. To control for double representation of connections, DE software (Banxia, 2017) is used to count only one direction of connections. The software labels effected DF as "heads" and causal DF as "tails". Either setting (heads or tails) results in the same frequencies, but it is necessary to pick one, otherwise each connection would be counted twice. For example, if Cluster A was comprised of DF1 ⇒ DF2, then DF1 would be labeled a “tail” and DF2 a “head”. Without causal direction restrictions, DF1 would be given a connection frequency count of “1” with DF2, and vice versa. The frequency total would then show “2” CC between DF1 and DF2, skewing the frequency counts by an order of magnitude. As well, by restricting counts to causal directions – in this case “tails” – transitive connections are also counted accurately. For example, if Cluster B was comprised of DF3 ⇔ DF4, then restrictions allow for DF3 to be given a frequency count of “1” with DF4, and vice versa. The frequency total would then show “2” CC between DF3 and DF4, accurately reporting the number of CC.

Figure 9.1. Frequency matrix of driving forces from Stage 3 cluster outcomes

DF-TAIL	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	0	0	0	0	0	0	1	0	0	4	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	2	0	1	1	0	0	1	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	1	1	3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	2	0	0	0	0	0	2	0	1	0	0	0	0	0	0
5	0	0	0	5	0	10	0	0	0	1	0	4	0	0	0	0	0	0	0	0
6	0	6	2	0	0	0	0	0	0	1	1	5	0	0	0	0	0	0	0	0
7	0	1	0	0	0	0	0	0	0	0	1	0	2	5	0	8	0	0	1	3
8	2	6	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
9	0	1	0	0	0	0	2	0	0	1	1	0	0	4	0	0	1	4	0	0
10	0	4	1	0	0	1	0	1	1	0	3	2	0	0	0	0	0	1	0	0
11	0	2	1	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0
12	0	4	0	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0
13	0	1	1	0	0	0	0	0	0	0	0	0	0	6	1	0	1	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	4	1	0	1	3	0	0	0	1	0	0
16	0	3	0	0	0	0	2	0	0	0	0	0	2	5	2	0	1	0	2	1
17	0	0	0	0	0	0	0	0	0	0	0	0	1	10	1	0	0	0	0	0
18	0	0	0	0	0	0	0	0	4	2	2	0	0	3	1	0	0	0	0	0
19	1	1	0	0	0	0	1	0	0	0	3	0	0	0	0	3	0	0	0	1
20	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	2	0

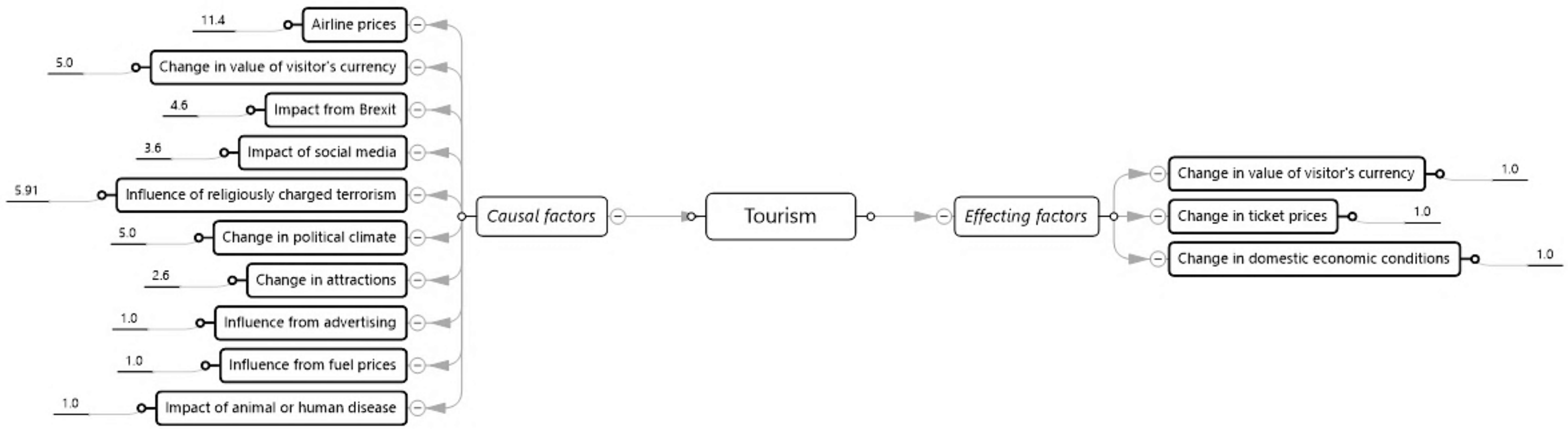
Results are crossed with the threat-level condition (high vs low) from the Stage 3 study that produced the causal associations. The threat-level proportion of every association ≥2 is next determined (e.g. in the matrix, green and red colour coded frequencies). The proportion is multiplied with the frequency, then the sum of the two conditions’ products is given to create a single weighted rank score for each associated DF. Each association is ranked from highest to lowest.

For illustrative purposes, Figure 9.2 shows the proportionally ranked direct associations for “Tourism” from vignette 2, where the association appears in both framing conditions from the first Stage 3 study. The top three DF that affect "Tourism" the most are "Airline prices", "Change in value of visitor's currency" and "Impact from Brexit", where the only three DF identified in both conditions to be affected by "Tourism" are (in ranked order) "Change in value of visitor's currency", "Change in ticket prices" and "Change in domestic economic conditions". Content analyses are carried out to isolate as many independent clusters as possible (those that do not share the same DF). Clusters are then adjusted, where needed, to meet the following criteria: six DF, an average of eight CC, and align with one STIRDEEPER category as its theme.

The categories of Society and Demographics share a number of associated DF. This is somewhat understandable, given the many similarities between the two categories. To work with this major overlap, the two categories are merged together into a single cluster theme *Soc*/*Dem*. In addition, a tenth cluster is created with a selection of associated DF from a mixture of categories, including non-traditional DF generated by participants in the previous studies. This cluster theme is categorised as *Other*. The same measures are taken to ensure there is a balanced representation of previous studies' threat-levels within the cluster.

To help ensure associations between DF in the *Other* are comparable to the previous nine clusters, opinions are sourced from both members of the previous research team and a small focus group (n = 3). The development methods result in 10 unique clusters with the same complexity profile (DF, CC, threat-level). To control for explicit influences of themes, all clusters are given a number (1-10) in lieu of a title (see Appendix E).

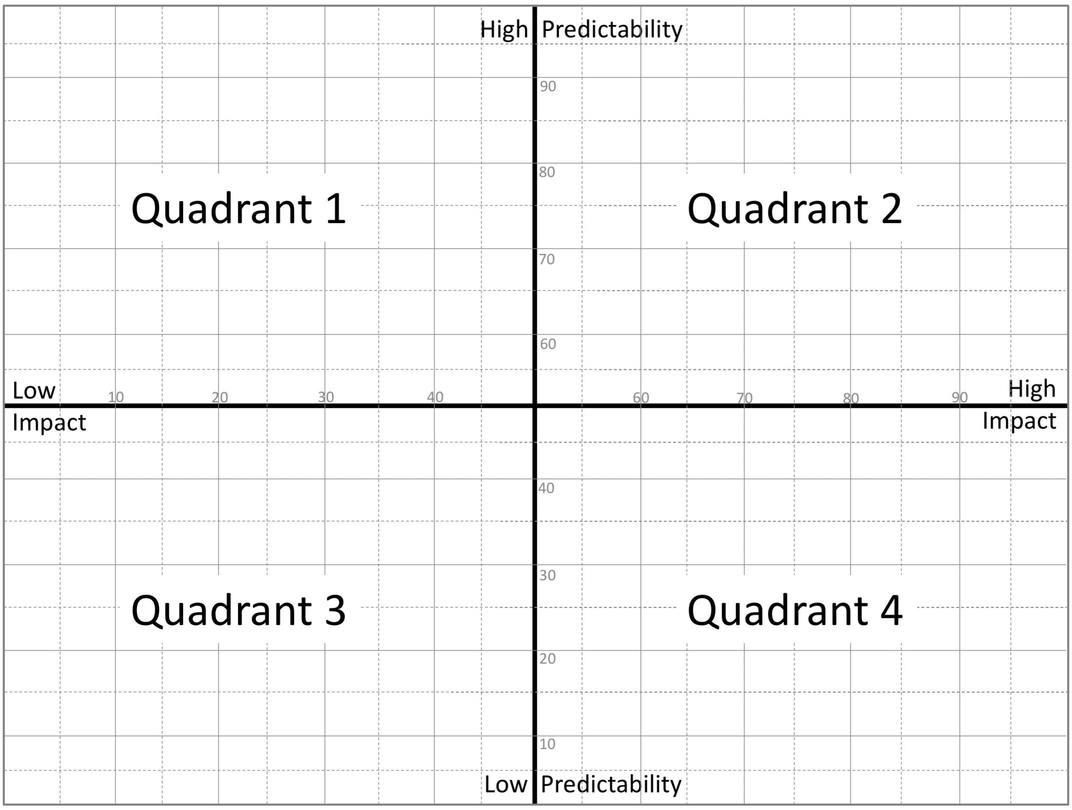
Figure 9.2. Proportionally ranked influencing driving forces to “Tourism”



To illustrate a cluster's projected impact and predictability, a blank I/P matrix is developed to plot the 10 clusters against. The I/P matrix is a common platform in different SP schools. The specific I/P matrix used for this study is from the IL model's methodology, promoted in several major publications and taught within the MBA programme at SBS (Cairns & Wright, 2018a; Chermack, 2011; Van der Heijden K. , 1997; Van der Heijden, Bradfield, Burt, Cairns, & Wright, 2002). There are no numerical values to the axes. As stated in previous chapters, a major aim of this study is to promote ecological validity alongside standard psychological experimental methods. By providing numerically empty matrices in this study, participants must rely entirely on a form of "fuzzy" cognition (Reyna & Brainerd, 1995). This more closely mimics the type of ST promoted in the IL model. Impact runs along the x-axis from "Low" to "High". Predictability runs along the y-axis, also from "Low" to "High". Another way to represent predictability is to value uncertainty along the y-axis. If uncertainty is used, though, the axis values must be reversed, and run from "High" to "Low". This is because predictability and uncertainty are dichotomous opposite, though qualitatively different. The more predictable an event becomes, the less uncertainty associated with the event, and vice versa.

The I/P matrix is a 2x2 space that creates four quadrants (see Figure 9.3). Quadrant 1 designates the upper left corner of the I/P matrix. This space represents where clusters are considered to be highly predictable (HP), but have a lower impact (LI) on the organisation's future. Examples of factors within clusters that may be more likely to end up in Quadrant 1 are demographics, such as population of 10–15-year-olds in the next 5 years. Everyone who will fall within that age range are already born, so we can predict with low uncertainty the approximate population size and from this, the conditions surrounding their future lives, work, health, etc. Quadrant 2 identifies the upper right corner. This space represents the clusters that are considered both HP and highly impactful (HI). Clusters that include events such as certain elections, may be found in this quadrant. Quadrant 3 identifies the lower left region of the I/P matrix. Cluster that are considered to have both low predictability (LP) and impact (LI) are placed here. Clusters in this region have the highest potential of being edited from final scenario considerations due to their perceived low importance (impact) and difficult predictability. An extreme example of a cluster in this region might include events such as the local shop's choice of product expansion. Quadrant 4 represents the lower right I/P matrix space. Clusters here are considered to be difficult to predict (i.e. LP), but have the potential to be HI. Black swan events are often thought to fall within this region, as well as potentially volatile stock behaviours and pre-war power relations. The IL model prioritises clusters from this quadrant for further scenario development. Figure 9.3 is shown with axis values and gridlines. These are only visible in the master analysis template for the researcher. Participants are given a blank I/P matrix template that only contains the two axis lines and axis labels "Low Impact", "High Impact", "High Predictability" and "Low Predictability.

Figure 9.3. Master I/P matrix



9.4.3 *Method*

Participants ($N = 44$) were recruited from the UK. All experimental sessions were held in one of the computer labs within the SBS. When participants arrived at their lab session, they were seated at a computer that preserved as much space between all attendees as possible to limit visibility of other participants' efforts. After reading through the instructions, purpose of the study, and submitting their signed consent to volunteer, the session began. Each participant received one of two possible packets. Both packets included further instructions (p 1), the business vignette (p 2), and 10 colour copies of the clusters (pp 3-12). Half the packets included a high-threat framed vignette and half a low-threat framed vignette. Each packet's framing was unknown to the instructor, which created a double-blind randomization. Participants were randomly assigned to a condition, high-threat ($n = 20$) or low-threat ($n = 24$). Table 9.1 presents a breakdown of vignette profiles that were used to reflect sample demographics.

Table 9.1. Business vignette profiles presented to different participant samples

Stage	**Experiment**	**Vignette**	**Profile**		
			UK	**UAE**	**US**
5	**Exploratory-Heterogeneous**	**1**	✓		
		2	✓		
		3	✓		

Participants were instructed to read through the business vignette on their own (about 2 minutes). Once completed, the researcher briefly explained the practice of SP and the history of where the 10 clusters' information had been sourced (i.e. participants in previous studies; about 10 minutes). To familiarise the participants with the function and content of the clusters, the researcher read through each cluster out loud with the attendees (about 20 minutes). The clusters were projected on two large screens at the front of the room to help participants follow along, as well as being represented on paper within their packets. All participants were encouraged to ask all the questions they had, along the way, and to make notes in their packets to help them better understand the clusters. After the cluster review, participants were instructed to open the .pptx file on their computers.[18] The file presented participants with a blank I/P matrix template one the first slide and 10 small boxes on the second slide. Each box held a number from 1-10 that corresponded with the numbered cluster in their packet. A blank I/P matrix was projected on the two large screens at the front of the room where the instructor explained the axes and how to move the numbered boxes around the space. Participants were first instructed to only focus on the impact levels of the clusters (x-axis). Once all attendees completed this first step, they were instructed to focus on the predictability of the clusters, and to adjust their placement along the y-axis. Before and after each axis' instructions, participants were reminded to reference their vignettes to help them think about the future of the company. Participants worked at their own pace and were encouraged to ask questions at any time. When each participant was finished, they saved their file to the computer[19] and left their packets. They were encouraged to keep a copy of their participant agreement form. Packets were matched with their completed matrix files to label the condition for each (high-threat vs low-threat).

9.4.4 *Results – Exploratory Heterogeneous Study*

Due to recruitment issues, only vignette 1 was completed by both condition samples and tested, high-threat (n = 15) and low-threat (n = 19). Therefore, the following analyses regard the "University of Strathclyde" profile. One participant from the low-threat condition is removed from the analyses due to being unable to complete the clustering task. Five clusters are removed from the analyses due to participants failing to plot them on their personal I/P matrices. The high-threat condition is missing one cluster, each, from *Politics*, *Energy*, and *Religion*; low-threat is missing

[18] One lab session did not include computers. Participants instead completed their cluster placements on a paper print out of the I/P matrix. The printout was from the same .pptx used in the computer sessions. This made it possible to scan the participants' responses back into the .pptx matrix for analysis.

[19] For those in the paper and pen session, they just included the single I/P matrix sheet with their packet.

two clusters from *Religion*. This means the *Religion* category has three fewer clusters, the most missing, calculated into the analyses. The two leading assumptions as to why participants failed to plot all their clusters: i) the participant didn't realise they hadn't completed the task or ii) the participant was unable to determine the high/low values for the cluster, and therefore skipped it. In the case of the first assumption, *Religion* is the 10th cluster, and it may be the case that participants plotted their clusters following numerical order, and therefore, after plotting the majority (90%) of the clusters, failed to realise there was one more. Though this cannot be determined with the data that are collected, some or all of the participants who failed to plot their *Religion* cluster may have felt unnecessary pressure or distraction from other participants completing and leaving the lab before them, thus causing them to rush the end of their task. In the case of the removed participant, they quit the study early, reporting that, even after the training and Q&A time, they didn't understand how to determine the impact and predictability of the clusters, and felt lost. From this participant's admission, the second assumption arises. If one participant can feel lost and confused, then potentially more than one participant may feel the same way, whether about the whole or part of the task.

To conduct the quantitative analyses, each cluster is given an x (impact), y (predictability) coordinate. The coordinates are derived from the values designated across each axis, from 0(low) – 100(high), as illustrated in Figure 9.3. This results in the central point where the two axes cross to have a coordinate of 50,50. The master I/P matrix template is used to plot all x,y coordinates. Participants' matrices are transferred to the master template and the central point of each numbered cluster box designates the x,y coordinate for that cluster. All coordinates <50 are designated "low" and all coordinates >50 are designated "high" for the axes.

Proposition 9.1 is explored through various impact measures. First, clusters are divided by their condition (high-threat vs low-threat), then by their impact value (0-100), to help determine if message framing anchors evaluations of future impact. LI clusters (<50) are those placed in left hemisphere of Quadrants 1 & 3 of the I/P matrix, and HI clusters (>50) are those placed in right hemisphere of Quadrants 2 &

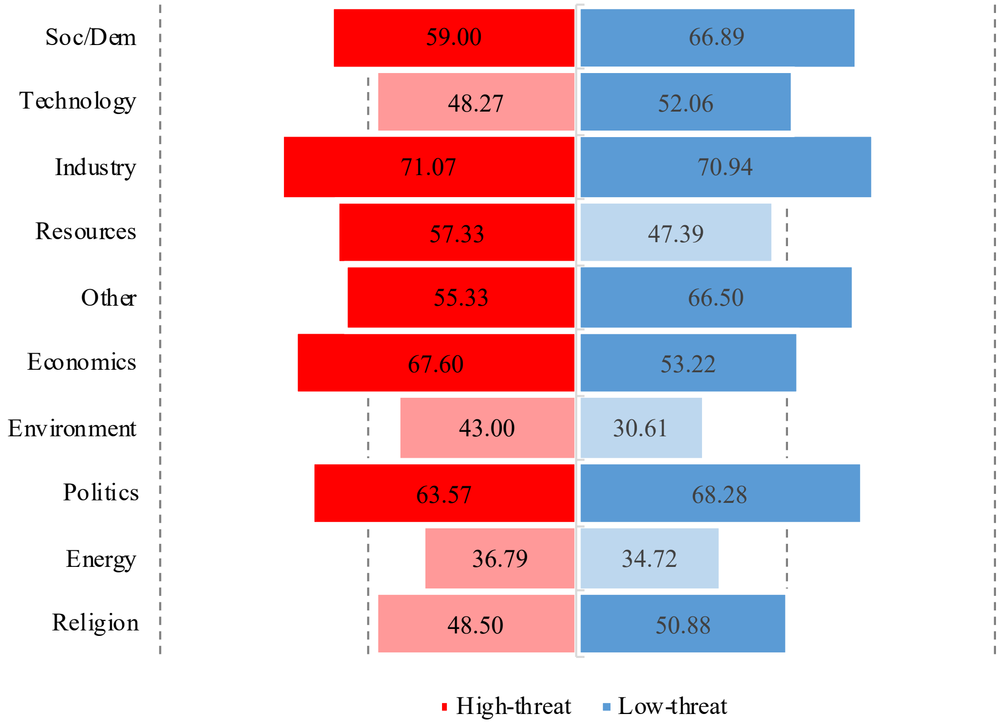
4. Impact proportions of each hemisphere by condition are reported in Table 9.2. A significantly higher proportion of clusters are considered to be HI (i.e. highly impactful), for both threat condition ($\chi^2_{high\text{-}threat}(1) = 10.49$, $p = .001$, 95% C.I. [.44, .96]; $\chi^2_{low\text{-}threat}(1) = 5.49$, $p = .02$, 95% C.I. [.14, .84]). Of note is that the high-threat condition also produced a smaller distribution across their cluster, as seen in the smaller confidence interval, compared to the low-threat condition. There is no significant difference between conditions, revealing no strong anchoring bias on global impact measures by threat message.

Table 9.2. Impact proportions by condition

Condition	**Hemisphere**	
	LI	**HI**
High-threat	0.37	0.63
Low-threat	0.41	0.59
Total	*0.39*	*0.61*

Clusters are next divided by their STIRDEEPER category and the average of x,y coordinates are calculated for each category. Figure 9.4 compares the average distribution of all cluster categories across each axis. The x-axis is along the vertical middle, dividing the two conditions (high-threat vs low-threat). Dotted lines represent the 50-point and 100-point markers of the axis. Every categorical bar that crosses the 50-point threshold illustrates that the average evaluation for that category is considered "high", and conversely, every bar that does not cross the same threshold illustrates that the average evaluation for that category is "low". Impact distributions of the 10 clusters are significantly different within both conditions, with medium-strong effect sizes ($F_{high\text{-}threat}(9,137) = 3.05$, $p = .002$, $\eta^2 = .17$; $F_{low\text{-}threat}(9,168) = 7.32$, $p < .000$, $\eta^2 = .28$), revealing that average values are fairly distanced from either side of the threshold (= 50). The categories evaluated to be the most impactful within the high-threat condition are *Soc/Dem, Industry, Resources, Other, Economics*, and *Politics*; within the low-threat condition are *Soc/Dem, Technology, Industry, Other, Economics*, and *Politics*, with *Religion* on the cusp. There is, however, no significant difference between conditions, revealing no strong anchoring bias on categorical impact evaluations by threat message.

Figure 9.4. Distribution of clusters along impact axis



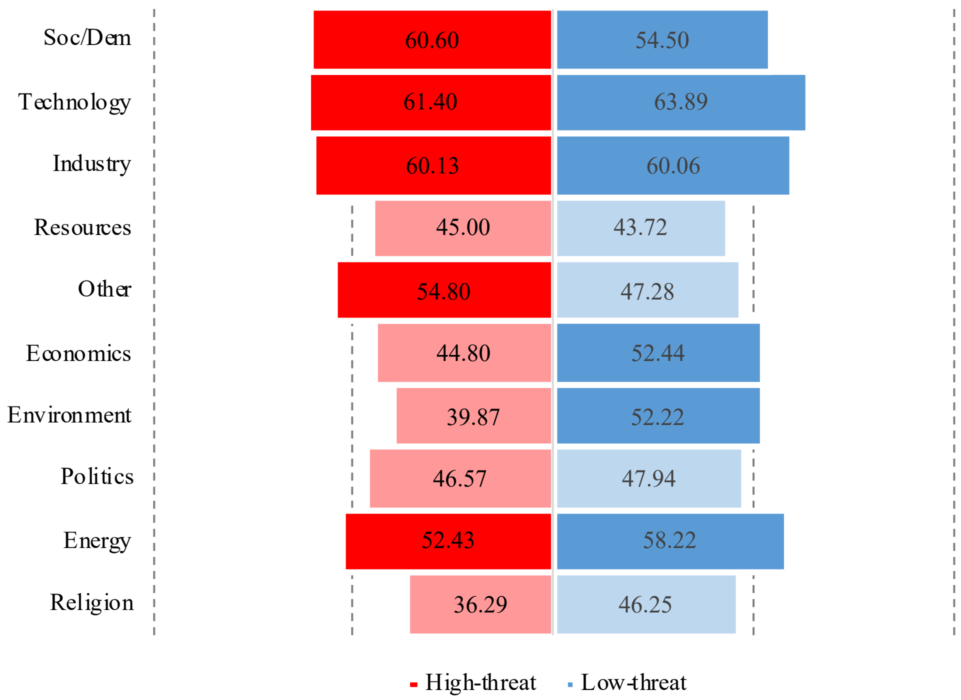
Proposition 9.2 is explored through the same measures as impact, but across the predictability axis. First, clusters are divided by their predictability value (0-100), to help determine if message framing anchors evaluations of predictableness. LP clusters (<50) are those placed in lower hemisphere of Quadrants 3 & 4 of the I/P matrix, and HP clusters (>50) are those placed in upper hemisphere of Quadrants 1 & 2. Predictability proportions of each hemisphere are reported in Table 9.3. Similar trends to impact evaluations are seen in both conditions. A higher proportion of clusters are considered to be HP (i.e. highly predictable), for both conditions; However, differences within both conditions are not significant, revealing that average values closely border the threshold (= 50). Analyses between conditions reveal no significant differences either. At a global level of predictability, data show no anchoring biases.

Table 9.3. Predictability proportions by condition

Condition	**Hemisphere**	
	LP	**HP**
High-threat	0.49	0.51
Low-threat	0.47	0.53
Total	*0.48*	*0.52*

Clusters are next grouped by category, and averaged along their x,y coordinates. Similar to impact results, predictability distributions (Figure 9.5) are also significantly different, but only for the high-threat condition ($F(9,137) = 2.30$, $p = .02$, $\eta^2 = .13$). The categories evaluated to be the most predictable within the high-threat condition are *Soc/Dem, Technology, Industry, Other,* and *Energy*; within the low-threat condition are *Soc/Dem, Technology, Industry, Economics, Environment,* and *Energy*. Furthermore, there is no significant difference between conditions, revealing no strong anchoring bias by threat message.

Figure 9.5. Distribution of cluster categories along predictability axis



Note: Vertical dashed lines represent the midway point (50) and highest (100) values on each axis. The central line marks the lowest axis value (0).

Proposition 9.3 is next explored by looking at the relationship of clusters across impact and predictability evaluations, by condition. Proportions are determined by quadrant (Figure 9.3) and reported in Table 9.4. The data reveal that both conditions result in similar trends – the largest proportion of clusters are placed in Quadrant 2 (HI/HP), the second largest proportion In Quadrant 4 (HI/LP), followed by Quadrant 3 (LI/LP), with the smallest proportion placed in Quadrant 1 (LI/HP). The divisions are significant within the high-threat condition ($\chi^2_{high\text{-}threat}$ (3) = 14.86, p = .002), but only approach significance for the low-threat condition ($\chi^2_{low\text{-}threat}$ (3) = 7.38, p = .06). The results support Table 9.2 hemispheric summaries, in that the majority of clusters are considered highly impactful (i.e. HI). Interestingly, the lowest proportion quadrant is not the extreme opposite of HI/HP, which would be Quadrant 3 (LI/LP), where both impact and predictability are low. Instead, predictability appears to be more evenly distributed across the quadrants, which, again, supports the previous table's results.

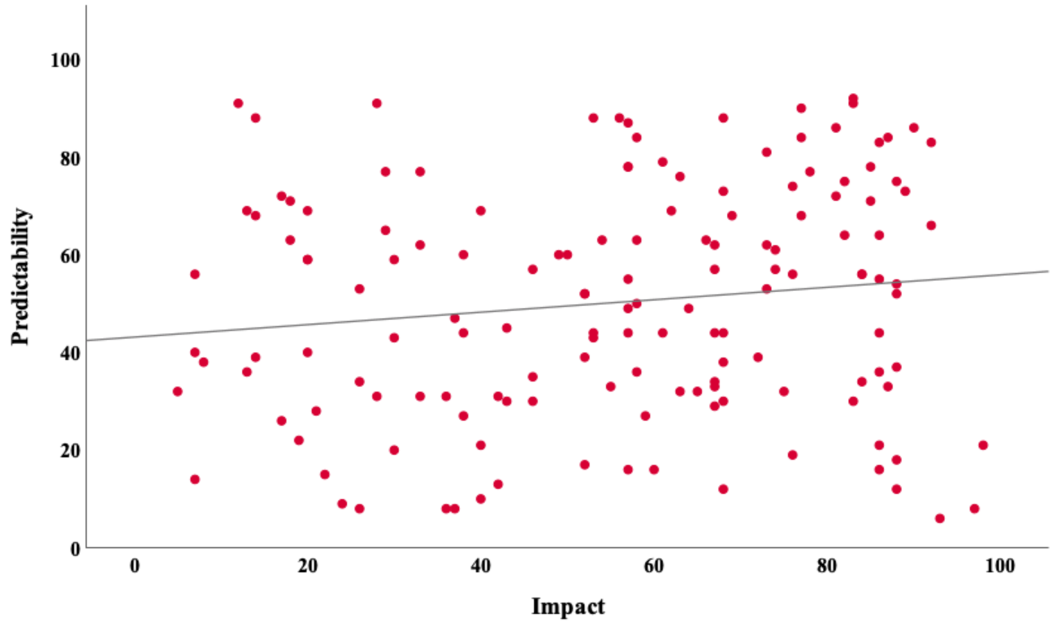
Table 9.4. Proportion of clusters by condition by quadrant

Condition	**Quadrant**			
	1 (LI/HP)	**2 (HI/HP)**	**3 (LI/LP)**	**4 (HI/LP)**
High-threat	0.14	0.37	0.22	0.27
Low-threat	0.20	0.33	0.21	0.26
Total	*0.18*	*0.34*	*0.21*	*0.27*

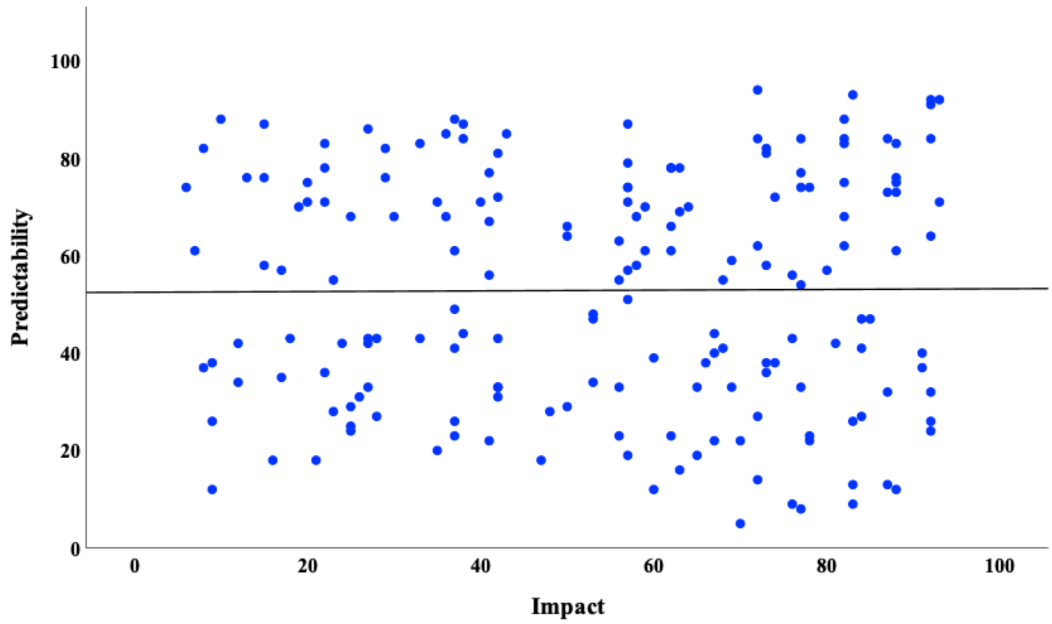
The next analyses is a manipulation check, based on the behaviours of the missing data points from participants. Correlations in distribution between axes is explored to help determine whether numerical titles for clusters had a confounding effect on distribution. This is particularly important to test, given the higher proportion of missing *Religion* clusters from participants' final matrices, which was anonymized with the numerical title "Cluster 10". If clusters are considered ordinal data due to their numerical titles (1-10), then a Spearman's rho correlational test should be calculated. However, no emphasis was placed on the order of the clusters during the instruction portion of the study, and we cannot know *a priori* if the numerical titles influenced participants to plot their clusters ordinally. Therefore the more appropriate test is the Pearson correlation. Both tests (Spearman's and Pearson) reveal no significant correlation between the impact and predictability axes, for either condition, though there is a slightly greater relationship within the high-threat condition. Individual clusters are plotted in Figure 9.6 by their coordinates and their regression equations are given in Equations 9.1 and 9.2, respectively.

Figure 9.6. Scatterplot of clusters by condition across axes

9.6.a. High-threat



9.6.b. Low-threat



Note: z-score confidence intervals are given to adjust for the constant. 95% $CI_{High\text{-}threat}$ [-.03, .30]; 95% $CI_{Low\text{-}threat}$ [-.14, .16].

Equation 9.1. Regression line, high-threat

$y = 43.13+0.13*x$

Equation 9.2 Regression line, low-threat

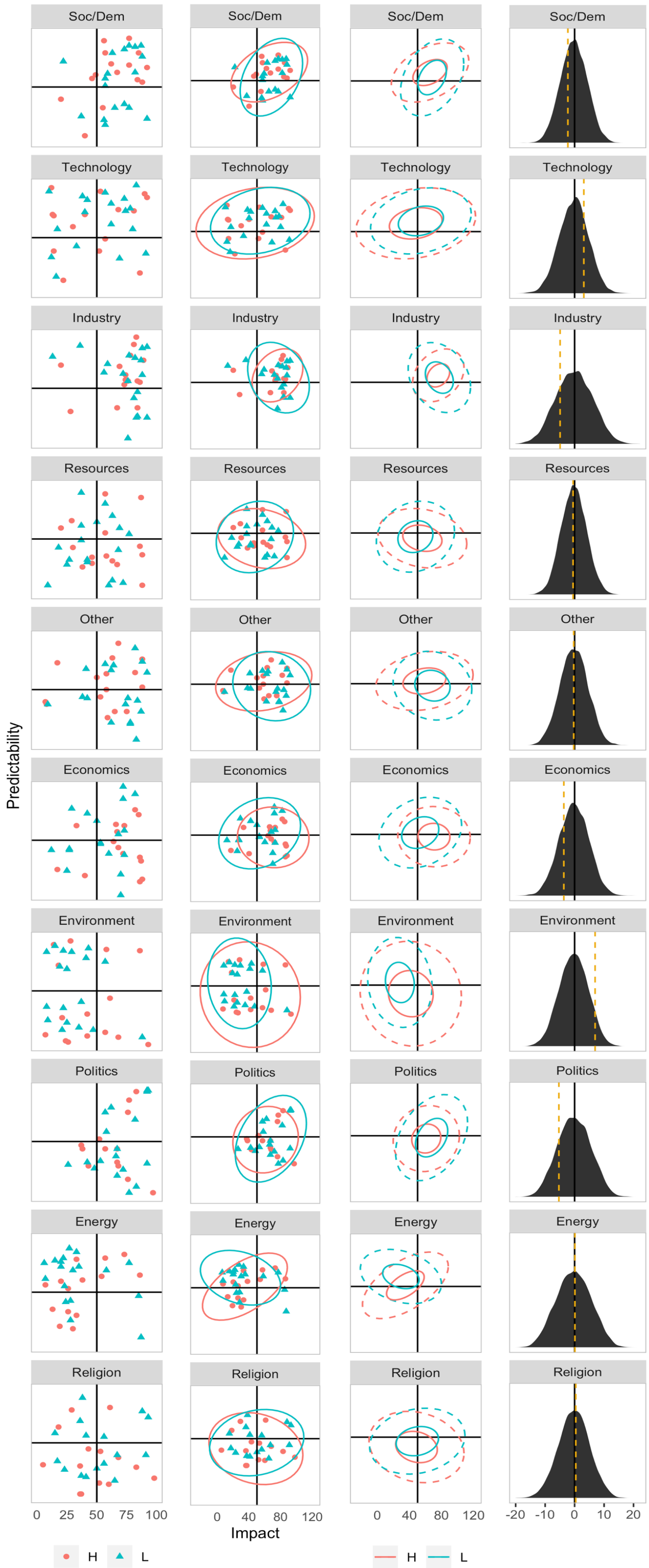
$y = 52.47+7.18E\text{-}3*x$

Results from the Chi-square tests run against both impact and predictability hemispheric distributions revealed a potential difference in spread across the I/P matrix. Spread represents the arrangement of clusters across the I/P matrix in relation to one another, as opposed to in relation to hemisphere (i.e. impact or predictability) or quadrant (i.e. interaction of impact and predictability). Measures of spread can help reveal a more granular level of message framing effects on evaluations by category. For example, if the frame of a message focuses attentions similarly for a category, then we should expect to see a smaller spread across the categorical clusters in one condition compared to the other condition, independent of hemisphere or quadrant.

Figure 9.7 illustrates how spread is analysed by category, by condition. First, clusters are divided by category, then condition, and plotted on their own I/P matrix (Figure 9.7.a). Red circles represent the high-threat condition and blue triangles represent the low-threat condition. Next, bivariate *t* distributions are calculated by category and condition. A *t* distribution is used because the population distribution is not known, and it cannot be determined whether the sample distribution is a normal distribution. Using the stat_ellipses function on R (2019), *t* distributions are calculated by category with a 95% CI ellipse plotted, by condition (Figure 9.7.b). Differences in spread are revealed through three different ellipse orientations, i) the two condition ellipses do not overlap within a category, ii) ellipses are different in size, or iii) ellipses have different orientations on the matrix. Figure 9.7.b shows that all ellipses overlap by category, but do not all have the same orientation. Confidence intervals are adjusted for 50% and 95%, to determine if variability in spread requires slightly more sensitive thresholds for detection. Figure 9.7.c illustrate the two CI, where inner ellipses indicate 50% of the distribution and the outer ellipses indicate 95% of the distribution. So far, results suggest that condition has no noticeable effect on spread within each category, except potentially *Environment*. To quantify average spread within each category by condition for significance testing, the hypotenuse for each cluster is first calculated using its x,y coordinates. Next, the mean for each category by condition is determined. Differences in mean hypotenuse between conditions, by category is finally calculated. Maximum possible difference between categories is 141, with a minimum of zero (i.e. identical spread). The range of mean differences for all clusters is -4.9 to 7.0. Though no differences actually result in zero, the differences show a remarkable similarity in spread. In order to determine if these differences are statistically significant, a bootstrapping approach is used (Figure 9.7.d). Sample sizes are small, and the population distribution is unknown, therefore a nonparametric bootstrap method is chosen. This form of bootstrapping avoids making assumptions about the population prior to analysis, avoiding risks of inaccurately skewing sample distribution, and without deriving the sampling distribution explicitly. Using the R package car::, parameters are set to reassign the condition label 10,000 times, with replacement, within each cluster, and calculate the difference for each reassignment. Figure 9.7.d. shows the bootstrapping output. The true difference-values (orange dashed line) all appear well within the expected distribution (black bell curve) under the assumption that condition has no true effect. Calculating the proportion of bootstrap values less extreme than the observed difference-values results in statistical confirmation that any differences in spread are not significant. Bootstrapping reveals, however, potential differences across categories, between conditions. Potential differences are seen the categories *Technology, Industry, Economics, Environment,* and *Politics.* Measures of spread serve as an indicator for potential anchoring biases when stimuli become more homogeneous.

Figure 9.7. Bootstrapping distribution by category

9.7.a. 9.7.b. 9.7.c. 9.7.d.



Exploratory tests also reveal a second potential bias that may be independent of framing effects. The bias appears to be a preference for evaluating clusters along the higher side of each scale. The preference is labelled a "high-side bias".

Proposition 9.1 explores evaluations of a cluster's future impact. Asymmetric distributions of clusters are revealed across the impact scale, with the majority clusters in total and by category evaluated as having greater future impacts (>50). When cluster values are paired by category between conditions, there is a significant positive correlation between threat conditions across the impact scale ($r(10) = .79$, $p = .007$). Variance and t-test results, alongside Pearson correlations in this section reveal that both conditions result in a high-side bias along the impact axis.

Proposition 9.2 explores evaluations of a cluster's predictability. Asymmetric distributions are also revealed across the predictability scale, which is similar across both conditions, showing the majority of clusters evaluated as more predictable than chance (>50), but not statistically different within conditions. Figure 9.5 reveals asymmetric distributions as well, but only significantly different within the high-threat condition. However, when cluster values are paired by category between conditions, there is a significant positive correlation between threat conditions across the impact scale ($r(10) = .69$, $p = .03$). Variance and t-test results alongside Pearson correlations in this section reveal a weaker, but still present high-side bias along the predictability axis.

Proposition 9.3 explores the relationship between impact and predictability evaluations. Interestingly, within the high-threat condition only, the distribution of clusters by category across the I/P matrix is significantly positively correlated between the impact and predictability axes, revealing an interaction of high-side bias by category ($r(147) = .14$, $p = .05$, $r^2 = .02$).

9.4.5 ***Heterogeneous Study Discussion***

In this exploratory-heterogeneous study, several features of clustering behaviour are highlighted. The features of hemispheric (high vs low) and quadrant (interaction) distribution are compared between conditions and categories which reveal unexpected insights. Though there appears to be no statistically supported anchoring bias, trends are revealed, and other behaviours detected. The propensity to assume the majority of clusters will have a higher than chance impact and predictability appears, at first glance, to be a possible bias in ST, which this study terms “high-side bias”. This effect is stronger within the high-threat condition compared to the low-threat condition. Perhaps is it the case that participants were familiar with a number of DF within the clusters, therefore making them more salient and influencing levels processing. More salient information is shown to be more believable, and levels of believability may have led to clusters appearing more predictable. Or perhaps the cluster constructions were inequivalent with regard to impactful DF, and the majority of clusters included more DF with higher perceived impacts, while a minority of clusters included fewer to no high-impacting DF. The type of data required to understand more nuanced evaluative thought processes of the participants were not gathered in this study. However, the study brings to the forefront further methods of exploration to better understand how practitioners evaluate potential events in the future.

As stated earlier, in the IL model, clusters designated to Quadrant 4 are often isolated from the rest of the clusters to further develop into scenarios. This technique is an effort of pragmatism, as much as it is strategic. Quadrant 4 clustered events are seen to hold qualities that could be highly impactful on the organisation’s abilities and path towards achieving their goals within the stated timeline, while also being the most difficult to predict. These events present the highest potentiality for disruptions on the organisation. Highly impactful events that are easily predictable can be more effectively planned for and/or exploited for benefit. However, impactful, influencing events that cannot be tracked, in turn cannot be effectively planned for or exploited. This means opportunities for growth could be missed, leaving the organisation losing out to their competition, and disruptions become surprises, debilitating the organisation. In short, focusing on highly impactful clusters with low predictability for scenario development is a type of risk-averse technique.

Both conditions appear to agree that the cluster largely focused on a *Politics* theme has is generally considered more highly impactful with lower predictability. The difference in the average I/P matrix values is near zero, as well, between conditions. This may seem not only less surprising, but in fact, an insightful move, when both the conditions' samples and the contents of the political cluster are examined. The participants were UK-based, MBA students, most with established careers within the EU. The cluster includes the DF, "Impact from Brexit", and the timing of this study was in the spring of 2018. This was after the successful Brexit vote, where this period in time was labelled a "transition period"[20]. During the same time period, the UK forfeited decision-making powers within the EU (2019), ambivalence regarding the Irish border increased sharply, and decisions/proposal from the UK Government regarding Brexit were not being ratified. This point was perfectly punctuated when Michel Barnier, European Commission's Head of Task Force for Relations with the UK, stated, "Nothing is agreed until everything is agreed" (Castle, 2018). Brexit was a factor of saliency. Its ubiquitous presence in all matters of British life at the time seem to have inoculated it from such shallow processing effects as the anchoring bias.

Within the high-threat condition, the clusters largely designated to Quadrant 4 are *Resources* and *Economics.* One way to interpret this is that when the present business environment is perceived to have higher threat levels, issues of supply and monetary behaviours take on more uncertainty qualities as their perceived future impact increases. To check for potential information bias, vignette 1 was reviewed for mentions of resources and economic factors. There are no direct mentions of economic factors within the vignette. There are, however, resource scarcity statements such as "departmental closures", "decreasing enrolment rates", "thus decreasing the number of available courses and degree paths". Though both conditions share the same focal statements on departments, enrolment rates, courses,

[20] *Google Trends* show searches for this phrase increased by 73% during the phase of this study, from the UK alone.

and degrees, the quality of scarcity only appears in the high-threat condition. This may very well be a reflection of qualitative anchoring bias.

Within the low-threat condition, only one cluster met the same criteria, *Other.* This cluster contains DF from a mixture of categories, including non-traditional DF (i.e. those designated "original" by creativity standards established in study 1). The *Other* cluster includes personal, familial, and governmental support systems, university features, and confidence in future successes. Some features of this cluster are under the control of the organisation, and others are not. Perhaps, when the present environment is perceived to have lower threat levels, the 'wild card', in so many words, takes on greater uncertainty and impact qualities.

When the mean differences in spread were calculated, even though differences are minimal, it is revealed that the high-threat condition plotted less distance between their clusters compared to the low-threat condition, 60% of the time. As understood from the Stage 3 studies, this trend may be an indicator of an anchoring bias that has been diminished due to the heterogeneous quality of the stimuli. Therefore, one way to help determine if the quality of the stimulus (i.e. clusters) is the case is to conduct a homogeneous study and compare differences.

Unfortunately confidence scores were not recorded for this study. Missed confidence data was due to technical errors that did not allow for the desired software to be used, which caused delivery and submission options for the participants to change at the last minute.

Just as with Stage 3 studies, to better understand ST, it is important to run a homogeneous comparative study. Hypotheses are developed from the discoveries of the exploratory-heterogeneous study and confidence measures are taken.

9.5 Hypothesis Homogeneous Study

Even though a number of behaviours are not significant, trends reveal potential anchoring biases between the two threat-message conditions that warrant further awareness in testing. Overall, the high-threat condition appears to favour a high-side bias more than the low-threat condition. The novel quality of a bias that appears in the data (i.e. high-side bias) offers an indication of the kind of behaviour that may be anchored to differently framed messages. Furthermore, results from Study 3 partially support Yaniv's (2011) work that shows biasing effects increase with homogeneous information vs heterogeneous information. New stimuli will be required for a comparative homogeneous study. The previous study's clusters were constructed with a heterogeneous blend of DF and CC developed from different threat-level primes in earlier studies. Homogeneous clusters will require DF and CC that are held constant within threat-messages across Stage 2 and 3 studies. The effort will create two different stimuli for each condition in the present study. A difference in threat levels between conditions, compounded by a difference in cluster content between conditions may prime participants to evaluate clusters not only differently, but to a greater magnitude than the exploratory-homogeneous sample. This final study will offer further evidence on the relationship between group factors and the anchoring bias. Building from the exploratory-homogeneous study's three propositions, along with discoveries in the data, a series of hypotheses are developed for testing.

The high-threat condition resulted in stronger significant differences in overall distribution (i.e. more clusters evaluated with higher impact scores) and a larger effect size, compared to the low-threat condition. Though statistical strength slightly reversed by condition, at the categorical level, both conditions resulted in statistical differences in distribution, in the exploratory-heterogeneous study. The first hypothesis tests this assumption and helps answer research questions 1 and 2.

H13: There will be an anchoring bias on the distribution of clusters, where participants presented with a high-threat business vignette will show a high-side bias on more clusters across the impact axis compared to participants presented with a low-threat business vignette.

Trends in evaluative differences across the predictability axis between conditions reveal a potential anchoring bias, but possibly not as sensitive to framing effects as impact evaluations. At the categorical level, the high-threat condition resulted in stronger significant differences in distribution, compared to the low-threat condition, though the low-threat condition evaluated more categories overall as highly predictable. The next hypothesis tests this assumption and helps answer research questions 1 and 2.

H14: There will be an anchoring bias on the distribution of clusters, where participants presented with a high-threat business vignette will show a high-side bias on more clusters across the predictability axis compared to participants presented with a low-threat business vignette.

Cluster spreads are not significantly different, but there appeared to be a trend for the high-threat condition to have a smaller spread compared to the low-threat condition. The difference in behaviours were also not consistent across categories. This may be an indicator of attentional susceptibility to framing effects. The next hypothesis tests this assumption and also helps answer research questions 1 and 2.

H15: Participants presented with a high-threat business vignette will group their clusters closer together (smaller spread) than participants presented with a low-threat business vignette.

As with previous studies, confidence measures are collected at the end of the task to help determine whether a relationship exists between message anchoring and confidence. All studies that analysed confidence measures have shown the same hypothesised trend in scores. From these findings, the next hypothesis tests the same assumption to help answer research question 3.

H16: Participants presented with a high-threat business summary will have lower confidence in their efforts compared to participants presented with a low-threat summary.

Finally, the novel discovery of a potential high-side bias required further understanding. If a high-side bias appears across all measures, regardless of threat-message condition, then this may indicate a more robust implicit cognitive bias. Helping reveal an implicit bias on evaluative judgments that is more resistant to other external influences (e.g. framing effects) has the potential to be a critical piece to the ST puzzle. Novel discoveries from the exploratory-heterogeneous study lead to the final two hypotheses, which help answer research question 4.

H17: There will be a high-side bias across the impact hemispheres.

H18: There will be a high-side bias across the predictability hemispheres.

9.5.1 *Ethical approval for human-based study*

Ethical review was submitted to both the Department of Management Science within the SBS and the Department of Psychology & Sociology within TAMUCC. Approval was received from the SBS on 29 November 2018 and from TAMUCC on 18 February 2019.

9.5.2 *Apparatus*

All business vignettes (1-3) are prepared as primes in the hypothesis-homogeneous study. The same complexity profile developed in the exploratory-heterogeneous study is maintained for each cluster in the hypothesis-homogeneous study: six DF, an average of eight CC, and align with one category as its theme. To create 10 clusters for two different threat conditions, first, a cluster analysis is conducted on Stage 3 data, to determine the strongest connections between the high-threat DF and then between low-threat DF. Second, a factor analysis is conducted on each DF repository, separated by condition, to determine the strength of associations and popularity between high-threat DF and then between low-threat DF. All DF used in the exploratory-heterogeneous study are eliminated. All DF shared between the two conditions are eliminated. To ensure the new clusters are believable, similar efforts employed to create the heterogeneous *Other* cluster in the previous study are used in the present study. Opinions are sourced from both members of the research team and a small focus group (n = 4). Twenty clusters are created, 10 for each condition, that aligned with the same themes as the exploratory-heterogeneous study. The same I/P matrix developed for the exploratory-heterogeneous study is used (see Appendix F).

9.5.3 *Method*

The method is the same as the exploratory-heterogeneous study, with a few modifications. Participants (N = 40) were recruited from UK and US. US participants (from TAMUCC) were incentivized to participate through course credit in the Department of Psychology and Sociology. Two sessions were held, one in the SBS computer labs, and the other in the TAMUCC computer labs. In all sessions, participants were randomly assigned to a condition, high-threat (n = 22) or low-threat (n = 18). Table 9.5 presents a breakdown of vignette profiles that were used to reflect sample demographics.

Table 9.5. Business vignette profiles presented to different participant samples

Stage	**Experiment**	**Vignette**	**Profile**		
			UK	**UAE**	**US**
5	**Hypothesis-Homogeneous**	**1**			✓
		2	✓		✓
		3	✓		✓

The exploratory-heterogeneous study revealed an issue with participants failing to plot all 10 of their clusters on their respective I/P matrices. To help reduce the probability of this occurrence, instructions were modified to emphasise the risk of failing to plot all 10 clusters, and explicitly reminded participants to double check that all 10 clusters were present on their I/P matrix. At the end of their sessions, no participants stated they were unable to complete the task and all submitted completed I/P matrices that included all 10 clusters.-After participants submitted their I/P matrix, they were asked to give a confidence score, then fill in three demographic questions.

9.5.4 ***Results – Hypothesis Homogeneous Study***

Seven participants are removed from the final analyses, three from the high-threat condition and four from the low-threat condition. Due to the incongruency of cluster content for vignette 1 from the TAMUCC sample, along with resource constraints that prevented completion of vignette 2 samples, only vignette 3 resulted in enough submissions to analyse. Therefore, the following analyses regard a “confectionary company”. It is important to note that the order of the analyses follow the order of hypotheses. Analyses are similar or identical to those used in the exploratory-heterogeneous study, but not presented in the same order as the previous study.

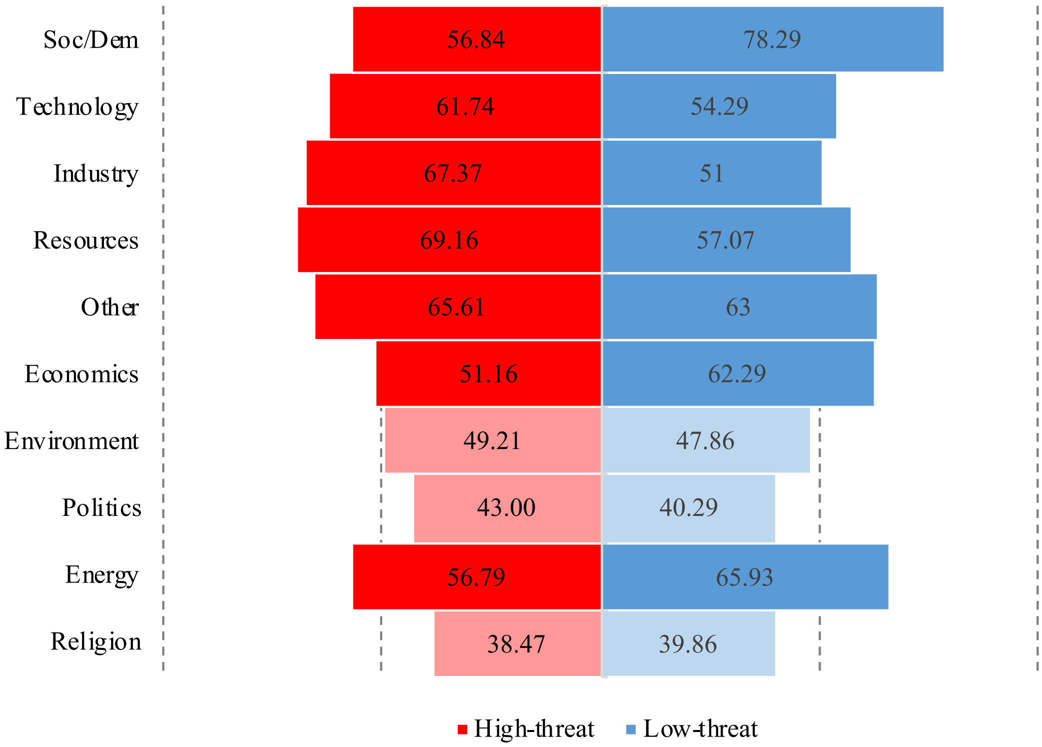
To test hypotheses 13, clusters are each assigned an x,y coordinate along the I/P matrix. This gives each cluster an impact value (x = 0-100) and a predictability value (y = 1-100). Clusters are divided by condition (high-threat vs low-threat), then by their impact value (<50 vs >50). Impact proportions of each hemisphere by condition are reported in Table 9.6. There is a significant difference in impact values within both conditions ($\chi^2_{high\text{-}threat}$ (1) = 8.22, p = .004; $\chi^2_{low\text{-}threat}$ (1) = 4.97, p = .03). There is no significant difference between conditions, however, revealing no strong anchoring bias on global impact evaluations by threat message.

Table 9.6. Impact proportions by condition

Condition	**Hemisphere**	
	LI	**HI**
High-threat	.40	.60
Low-threat	.395	.605
Total	*.40*	*.60*

Cluster are next grouped by category and divided by condition. Averages are taken from their x,y coordinates, by category, and compared within and between conditions. The distribution of the 10 clusters by impact evaluations is illustrated in Figure 9.8. Cluster distributions are significantly different within both conditions ($F_{high\text{-}threat}$ (9,179) = 2.80, p = .004, η^2 = .12; $F_{low\text{-}threat}$ (9,130) = 3.03, p = .003, η^2 = .17). Categories projected to be the most impactful within the high-threat condition are *Soc/Dem, Technology, Industry, Resources, Other,* and *Energy,* with *Economics* on the borderline; within the low-threat condition are *Soc/Dem, Technology, Resources, Other, Environment,* and *Energy,* with *Industry* on the borderline. The distributions, however, have a significant positive correlation between the threat conditions along the impact axis ($r(10) = .49$, $p = .03$, $r^2 = .24$), revealing no strong anchoring bias categorical impact evaluations by threat message. Hypothesis 13 is not supported.

Figure 9.8. Distribution of clusters along impact axis



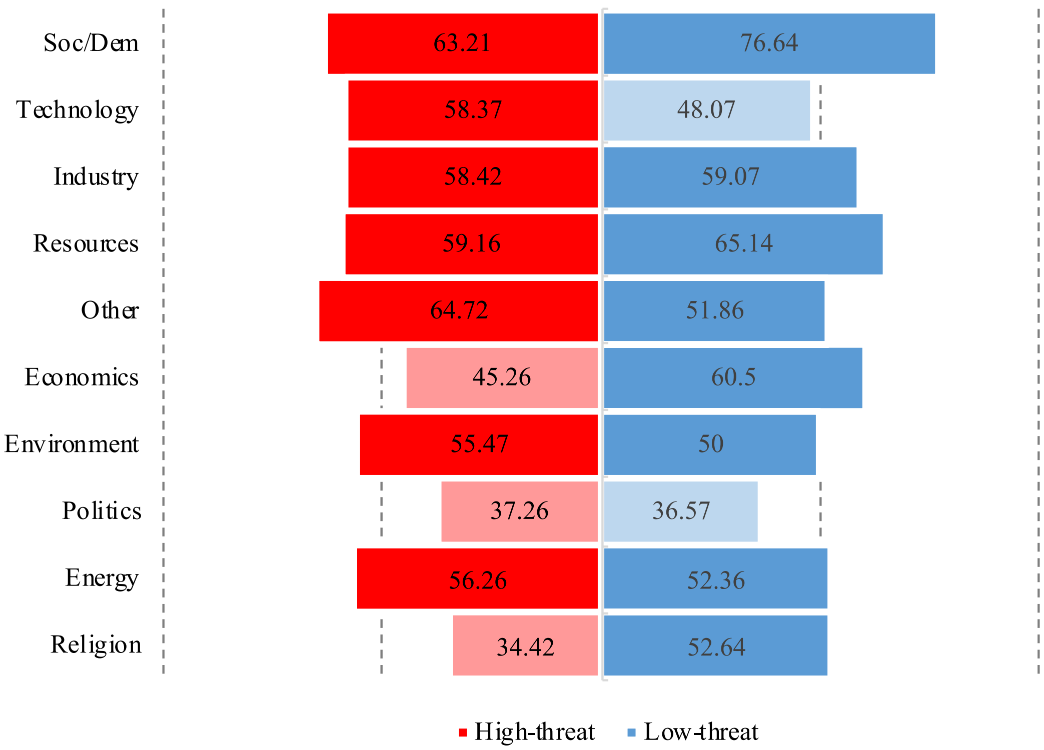
To test hypothesis 14, cluster are divided by condition, then by predictability value. Proportions are calculated along the predictability axis. Proportions are reported in Table 9.7. There are significant differences of predictability values within both conditions, though the high-threat condition is just approaching significance ($\chi^2_{high\text{-}threat}(1) = 3.63$, $p = .057$; $\chi^2_{low\text{-}threat}(1) = 5.32$, $p = .02$). There is no significant difference between conditions, however, revealing no strong anchoring bias on global predictability evaluations by threat message.

Table 9.7. Predictability proportions by condition

Condition	**Hemisphere**	
	LP	**HP**
High-threat	.43	.57
Low-threat	.40	.60
Total	*.42*	*.58*

At a categorical level, predictability evaluations are significantly different within both conditions ($F_{high\text{-}threat}$ (9,179) = 3.42, p = .001, η^2 = .15; $F_{low\text{-}threat}$ (9,130) = 2.81, p = .005, η^2 = .16). Distributions are illustrated in Figure 9.9. Discussion follows on the next page.

Figure 9.9. Distribution of cluster categories along predictability axis



Categories projected to be the most impactful within the high-threat condition are *Soc/Dem, Technology, Industry, Resources, Other, Environment,* and *Energy*; within the low-threat condition are *Soc/Dem, Industry, Resources, Other, Economics, Energy,* and *Religion* with *Environment* on the borderline. Distributions, however, have a significant positive correlation between conditions along the predictability axis ($r(10) = .49$, $p = .03$, $r^2 = .24$).[21] Though there is a difference in predictability values across the 10 categories, their differences are too similar between conditions to be considered effected by an anchoring bias. Hypothesis 14 is not supported.

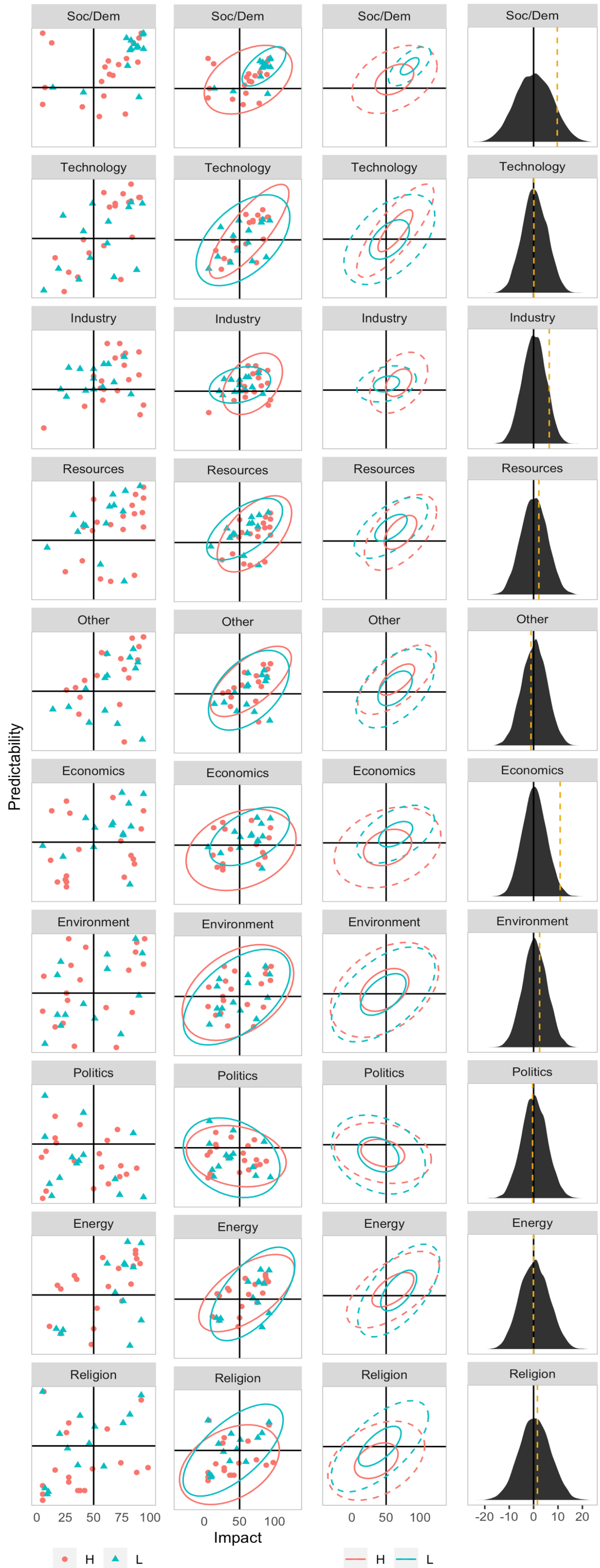
To test hypothesis 15, x,y coordinates are potted by category, divided by condition (Figure 9.10.a), the broadest CI (95%) are developed into ellipses and plotted (Figure 9.10.b), followed by two CI threshold ellipses (50%-95%) as a more sensitive measure of spread (Figure 9.10.c), and finally bootstrapping analyses are plotted to help determine whether spread are significantly different between conditions, by category (Figure 9.10.d). Analyses reveal further migrations outward from chance (as illustrated in (Figure 9.10.d) with a true-difference orange line). Clusters themed around *Soc/Dem, Industry,* and *Economics* show to be the most sensitive to threat-messages, but statistically remain within the curve, therefore not significant. Further to this, the behaviours leading to these three categorical differences are the opposite of what was hypothesised. The high-threat condition resulted in larger spreads of their clusters within these themes, compared to the low-threat condition. Hypothesis 15 is not supported.

---

[21] The reported results are not a typo, Pearson correlations by conditional for impact and predictability measures came out identical when rounded to the second decimal place.

Figure 9.10. Bootstrapping distribution by category

9.10.a. 9.10.b. 9.10.c. 9.10.d.



To test hypothesis 16, participants were asked to indicate "how confident you are that you determined the future of the clusters correctly" at the end of the study. Confidence was scored on a Likert scale, 1(not at all) – 5(completely). Not all participants gave a confidence score ($n_{high\text{-}threat} = 20$, $n_{low\text{-}threat} = 16$). Table 9.8 reports mean, median, and mode for both conditions. Scores between the high-threat and low-threat conditions were not significantly different. Both conditions, however, resulted in higher than midpoint scores. Though scores follow the hypothesised trend between conditions, the differences are not statistically significant, and therefore do not support hypothesis 16.

Table 9.8. Confidence scores by condition

Condition	**Vignette**	***M***	***SD***	**Md**	**Mode**
High-threat	**3**	3.86	.73	3.80	3.00 5.00
Low-threat	**3**	3.93	.76	4.00	3.00 4.00

Finally, to test hypotheses 17 and 18, additional analyses are run as extensions of analyses in Sections 9.5.4.1 and 9.5.4.2. Both sections reveal that participants in both conditions exhibit a bias toward evaluating the impact and predictability of clusters overall, and by category, as higher than the mid-way point (> 50). The proportion of clusters by condition by quadrant is next calculated and is reported in Table 9.9. Similar to the exploratory-heterogeneous study, the largest proportion of clusters are placed in Quadrant 2 (HI/HP) and is significantly different for both conditions ($\chi^2_{high\text{-}threat}(3) = 35.01$, $p < .000$; $\chi^2_{low\text{-}threat}(3) = 31.08$, $p < .000$). These divisions support earlier summaries provided in Tables 9.7 and 9.6, where the majority of clusters are considered both HI and HP.

Table 9.9. Proportion of clusters by condition by quadrant

Condition	**Quadrant**			
	1 (LI/HP)	**2 (HI/HP)**	**3 (LI/LP)**	**4 (HI/LP)**
Low-threat	0.20	0.33	0.21	0.26
High-threat	0.14	0.37	0.22	0.27
Total	*0.18*	*0.34*	*0.21*	*0.27*

Scatterplots reveal the nature of correlations between the axes, by condition (see Figure 9.11). Distribution of clusters across both axes are significantly positively correlated within both conditions, and confidence intervals are given on the standardized z-scores ($r_{high\text{-}threat}(189) = .42$, $p = .000$, $r^2 = .18$, 95% CI = .292, .553; $r_{low\text{-}threat}(140) = .42$, $p = .000$, $r^2 = .18$, 95% CI = .270, .575).[22] Regression equations are given in Equation 9.3. and 9.4. Both conditions show a high-side bias, and a more severe bias than the exploratory-heterogeneous study. Hypotheses 17 and 18 are supported.

Equation 9.3. Regression line high-threat

$y = 31.2+0.39*x$

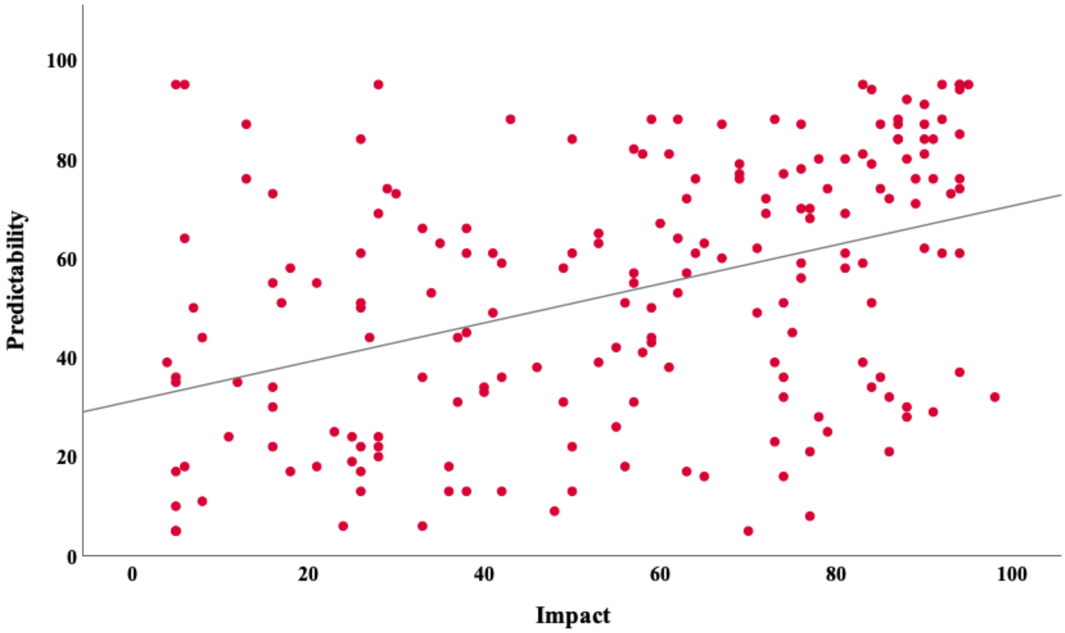
Equation 9.4. Regression line low-threat

$y = 33.21+0.39*x$

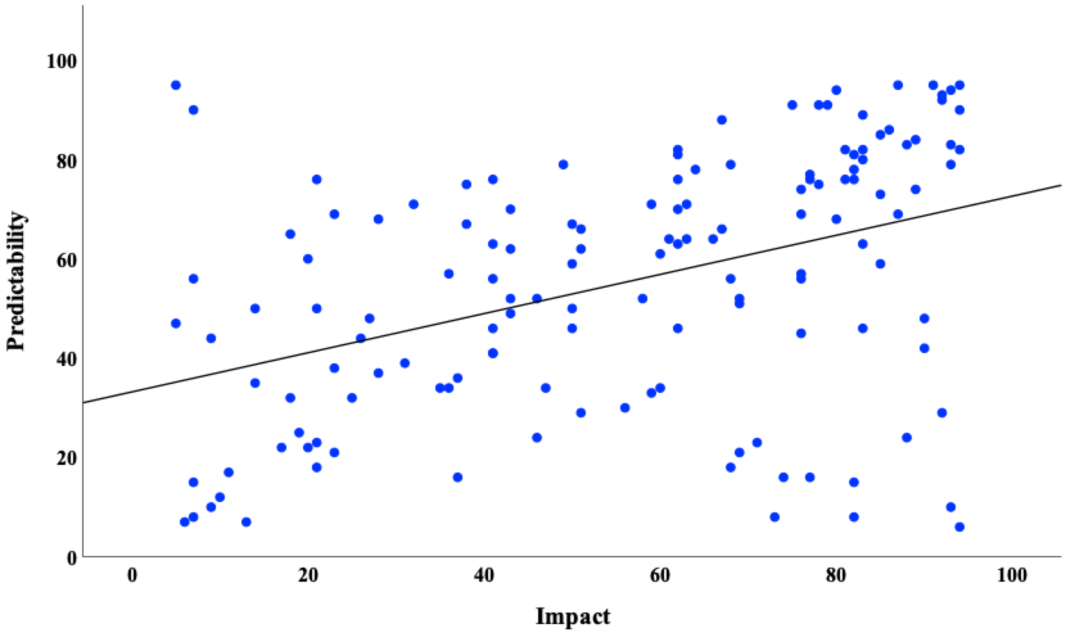
[22] Reported statistical values and regressions are not a typo, their correlations came out identical when rounded to the second decimal place.

Figure 9.11. Scatterplot of clusters by condition across axes

9.11.a. High-threat



9.11.b. Low-threat



9.6 Evaluative Thinking Discussion

Data produced in both Stage 2 and 3 studies provided the stimuli developed (i.e. clusters) for the two Stage 5 studies. Both studies measured individual-level thinking, after receiving clusters built from DF and CC informed by prior studies' outcomes. Though all vignettes were originally delivered in a randomized method to participants, only vignettes 1 and 3 resulted in large enough sample sizes for further analysis. Due to the novelty of experimentally testing ST cognitions, particularly within a SP framework, for biasing effects from threat-messages, the scholarship did not provide clear lines for well-established, robust methodological designs. Therefore an exploratory, pilot study was first developed to gather empirical evidence. The data revealed potential biasing trends by condition, but also surprisingly, a new quality of bias, which was labeled the "high-side bias". Hypotheses were developed, following trends in the exploratory results, that aligned with the four research questions, and a second study was designed and carried out. As well, building on discoveries from Stage 3 studies, the exploratory-heterogeneous study was complimented with a homogeneous paired study, that not only allowed for hypothesis testing, but comparative outcomes across studies to further understand potential impacts of consistency in group and information influences. Table 9.10 presents a summary of the key findings from the second, hypothesis testing, study.

In both studies, regardless of content, participants exhibit a high-side bias when thinking about future impact severity and probability of occurrence. This high-side bias appears to become more severe when the content under review aligns with the framed messages of previous participants, thus exhibiting further complexity in the anchoring effect. Even after controlling for sample demographics, such as age, education, and professional status, the bias persists. Much like the trend towards higher confidence in all the studies, the high-side bias in predictability is potentially worrisome. The potential novel discovery of a high-side bias presents new opportunities for exploring new methods for measuring ST, and directly addresses research question 4.

Table 9.10. Summary of samples and conditions by hypothesis

Hypothesis	Dependent variable	Hypothesis-Homogeneous	
		High-threat	**Low-threat**
13	Clusters	No anchoring bias	No anchoring bias
14	Clusters	No anchoring bias	No anchoring bias
15	Clusters	No anchoring bias	No anchoring bias
16	Confidence	No anchoring bias, but trend shows lower average score	No anchoring bias, but trend shows higher average score
17	Clusters	Higher impact scores overall and categorically	
18	Clusters	Higher predictability scores overall and categorically	

By projecting higher probabilities for the occurrence of events, participants may be lulling themselves into a false sense of security, assuming a control of information where it does not exist and placing blinders on their ability to further explore and test potentially high impacting events. MacKay & McKiernan (2004) discuss a closely related bias, foresight bias, of which the high-side bias may very well be a feature. The foresight bias is considered to emerge from the combination of shallow perceptions of historical experiences and facts (e.g. ideologies and norms), hindsight bias, and creeping determinism (Fischhoff, 1975). This leads to dogmatic (over-confidence or over-pessimism), oversimplified views of the future. “It results in logical (and structural) path-dependencies, faulty reasoning and, ultimately, a poor understanding of the future” (MacKay & McKiernan, 2004, p. 165). Tetlock’s team (2005) recognises a similar bias in their ‘foxes and hedgehogs’ sample. Foxes (seasoned industry experts), show more susceptibility to assign higher likelihoods to more possibilities than their hedgehog counterparts (business professionals). This results in foxes becoming “entangled in self-contradictions” (p. 190). “‘Imaginability’ appears to drive the inflations of subjective probabilities (p. 197). Bradfield may have encountered a similar phenomenon with his study’s participants (called 'syndicates'). He states, "it was apparent that developments envisaged by all syndicates essentially epitomized variations around a common, already well-articulated midpoint of events that were expected to occur." Though Bradfield's discussion is largely concerned with how this behaviour related to cognitive schemas, it is, none-the-less, an eerily similar behaviour. Participants rate a series of future events as more likely to happen, more predictable, and more impactful than the situation seems conducive towards.

One of the common behaviours from students learning in the IL model of SP reflects precisely these biases. Students often make their initial group-based evaluations across the I/P matrix with the majority of clusters designated to the higher probability hemisphere. This reflects a quick, shallow processing effort. Part of the curriculum is addressing this bias by instructing students to think about the realities of the 'known' probabilities of the events. Using strategic inquiry methods, instructors challenge students to make more explicit declarations of their knowledge on the cluster and its individual factors. The purpose of the strategic inquiry is to allow deeper processing of the information. After the strategic inquiry step, many student groups reach new agreements and adjust the placement of clusters on the I/P matrix. Related, in Stage 2, participants also commit a similar bias in the first round of determining "uncertainties" and "predetermines" as they build their list of DF. They commonly assume more DF are predetermines. A similar strategic inquiry step is introduced to the method, to allow for deeper processing of the information. Most of the time, participants realise they cannot project the direction, nor even the occurrence, of the DF, and change it from a "predetermined" to an "uncertainty".

Possibly the most surprising result from Stage 5 study is the similar distribution of cluster categories across the I/P matrix for the hypothesis-homogeneous study. Participants received *differently* framed futures and *different* DF within their clusters, but they still projected similar evaluations of impact and predictability of the clusters at the categorical level for the same business. It is possible that short vignettes – representing Stage 1 research and information gathering – do not relay enough framing difference to anchor the participants' perceptions of the future impact and predictability of clustered events. However, the two conditions' distribution behaviours were also greater than chance. A lack of anchoring bias does not assume identical evaluations from separate groups with different, yet related, information. It may be, then, that the cognitive effort of non-numerical evaluations of the related features (impact and predictability) are more stable in the face of certain information biases, in this case the anchoring bias. By extension, the cluster *categories* implicitly carried a shared understanding across the participants, regarding the same company. Both conditions averaged the same distribution alignment across the I/P matrix on 7 out of the 10 categories. This may also explain why their confidence scores were similar as well (though this is a trend across all studies).

Another surprising discovery is the reversal in spread between the conditions when the stimuli changed from heterogeneous information to that of greater homogeneity. With a mix of information that was constructed from different prior influences, the high-threat message appeared to possibly anchor participants' clusters to a more central value, or perhaps perceive less variability between the different clustered events, compared to the other condition. However, just the opposite occurred when the information was constructed from similar/identical prior influences. Though the effect was smaller. Another explanation may be that this is a reflection of shallow vs deep processing. As stated earlier, shallow processing can occur when people have incomplete information and rely on inferences to reach conclusions. The vignette used in the exploratory-heterogeneous study was of the university that the participants were attending at the time, as students. Even though students are not generally prompted to think of their higher education institutions in terms of "a business", they are, none the less, closely affiliated with the daily workings of the organisation. Whereas the vignette used in the hypothesis-homogeneous study was of an anonymous confectionary company. No participant had personal knowledge of the company and, by necessity, had less knowledge of the organisation. More familiarity entails more knowledge, and more knowledge will provide more opportunities (and potentially motivation) to engage in elaborative thinking and reasoning (i.e. deep processing). Therefore, it may be the case that the heterogeneous sample engaged in deeper processing than the homogeneous sample.

Where we see agreement between the conditions in the heterogeneous sample with the *Politics* cluster, we see no agreement across conditions within the homogeneous sample (though both conditions project LP for *Politics* as well). In fact, the high-threat condition is so severely high-side biased in cluster distribution, that no cluster categories share space in Quadrant 4, where they would be separated from the others for further scenario development. There are potentially two categories, however, within the low-threat condition, that would develop the remainder of the group's scenarios. These are *Technology* and *Environment.* This is where the differences between the vignettes may become more obvious. In the heterogeneous sample, participants were evaluating the future likelihood of events against their university. In the homogeneous sample, participants evaluated a confectionary company's future. To check for any previously unseen information bias, vignette 3 was reviewed for mentions of technology or environmental factors. There are no technological factors mentioned, but there is one environmental statement. The presence of this statement may have increased the immediacy of the information and caused the *Environment* cluster to become more salient, due to recency effects. By creating more salient information, participants could have implicitly dedicated deeper information processing to this cluster. If deeper levels of evaluation occurred, they may have led participants to become more aware of other potentially associated factors and events. Deeper awareness of more factors could have led to greater levels of uncertainty introduced into the valuation process – as each factor understandably brings with it its own inherent levels of uncertainty. This unexpected overlap in potential information-load hints at insights presented in Wright & Goodwin's (2002) paper. Participants were given a short reading that presented one of two possibly framed vignettes. The sample that was given a task that encouraged participants to "think harder and longer about their choice" altered the way they valued the inherent risk of the options (p. 1063).

This chapter completes Section II and the empirical studies. Each chapter focused on a single stage within the chosen IL model, and each stage emphasised a single dominant cognitive process as an element of ST. Chapter 6 reported participant demographics for the eight empirical studies (stimulus, creative, causal, and evaluative), and the development and validation of the priming stimuli that were used across all remaining experimental studies. Chapter 7 reported two studies that tested Stage 2 creative thinking using standardised methods borrowed from cognitive psychology. Chapter 8 built from the outcomes of the creativity studies and developed two novel studies that tested Stage 3 casual thinking, borrowing both creativity measures as well as developing novel measures to detect framing effects. Chapter 9 built, in a step-wise process, on the cumulative outcomes and knowledge of the previous chapters and developed two novel studies that attempted to break new experimental ground in order to advance the discipline's narrative around ST and cognitive biases. The final section will now present a comprehensive discussion on the empirical discoveries, followed by implications for both industry and academia based on discoveries from the body of research in this thesis.

Section III – Discussion and Implications

"The beginning of knowledge is the discovery
of something we do not understand."
Frank Herbert

Chapter 10. **Empirical Discussion**

The central question for this thesis asks how cognitive biases in ST effect the content of SP. The IL model, with discrete stages of progression, is chosen to develop a foundation for empirical study. ST is recognised as a cognitive system that encompasses a number of discrete, interrelating cognitions that are employed during SP. The IL/ST framework is developed that pairs dominant cognitive processes within ST against key stages in the chosen IL model and serves as a roadmap for the empirical studies. The central question is divided into four research questions that guide the empirical portion of this investigation. Data collected through practitioner interviews and a review of existing empirical SP studies reveal potentials for a number of biasing effects both from, as well as on scenario content. Anchoring to information is identified as a well-recognised bias in the literature, yet existing SP research largely fails to address the bias and understand practitioners' levels of sensitivity to such a bias throughout SP. A framing mechanism based on external threat-levels is chosen as the apparatus for all experimental studies. The framing mechanism is used to help reveal whether different ST processes, dominant at different IL stages, are sensitive enough to produce implicit anchoring biases in scenario content. Eight empirical studies are carried out: two to validate the framing mechanism and six to test the mechanism against different IL/ST stages. All studies reveal biases in the data, where some are statistically stronger than others. A discussion on how the outcomes address each of the four research questions is presented in this chapter and implications for both academia and practitioners are discussed in the final chapter.

The main contributions this investigation brings to the SP discipline are i) providing a framework against which we can understand SP and ST in a novel, yet cohesive manner, ii) bringing awareness to how we could and should use interdisciplinary knowledge to better understand practitioner-driven practices, iii) providing evidence of how prolific undetected biases can motivate ST and alter the outcomes of SP, and iv) discovering potentially a new bias, referred to as the 'high-side bias'. Table 10.1 presents a summary of comparisons across the empirical chapters, showing how each study addressed the four research questions.

Table 10.1. Summary table of the research questions (RQ 1-4), studies (Stage 2, 3, and 5), and hypotheses (1-18)

Stage	**Hypothesis**	**Answer**	**Results**
RQ 1) Are there measurable anchoring biases in scenario planning content?			
2	1	Yes	Abilities to think fluently and originally are most sensitive to biasing influences at the individual level.
	2	Yes	Abilities to think fluently are most sensitive to biasing influences at the group level.
3	6	Yes	Homogeneous information leads to greater sensitivities to biasing influences.
	7	Yes	Homogeneous information leads to greater sensitivities to biasing influences.
	8	Yes	Homogeneous information leads to greater sensitivities to biasing influences, but heterogeneous information only partially mitigates biasing effects.
	9	No	Biased information does not influence perceptions of complex causal relationships.
	10	Yes	Biasing influences cause participants to focus on driving forces differently.
5	13	No	Abilities to evaluate impact are not sensitive to biasing information.

Table 10.1. (*continued)*

	14	No	Abilities to evaluate predictability are not sensitive to biasing information.
	15	No	Abilities to evaluate events in relation to others are not sensitive to biasing information.
RQ 2) Are anchoring biases comparable across different scenario thinking cognitions and scenario planning content?			
2	1	Yes	Creativity measures were adopted at the individual-level – fluency, flexibility, elaboration, and originality – with originality adapted for qualitative analysis.
	2	Yes	Creativity measure, fluency, was adopted at the group-level, with originality adapted for qualitative analysis.
3	6	Yes	Creativity measure, fluency, was adapted for causality output.
	7	Yes	Creativity measure, fluency, was adapted for causality output.
	8	Yes	Creativity measure, elaboration, was adapted for causality output.
	9	Yes	Creativity measure, fluency, was adapted for causality output.
	10	Yes	Originality adapted for qualitative analysis.
5	13	Potentially	Simple proportional measures.
	14	Potentially	Simple proportional measures.
	15	Potentially	Geometric relational measures adapted for spacial placement.

Table 10.1. (*continued*)

RQ 3) Is confidence correlated with scenario planning performance?			
2	4	No	Only trends show potential sensitivity to biasing influences.
	5	No	Group work biases self-evaluations of performance more than individual performance.
3	11	No	Only trends show potential sensitivity to biasing influences.
5	16	No	Only trends show potential sensitivity to biasing influences.
RQ 4) Can new methods for measuring and understanding scenario thinking be developed for future research?			
2	3	Yes	Standard creativity scales and dialogic methods can be adapted for ST measures. Abilities to think fluently are most sensitive to group dialogues.
3	12	Yes	Some standard creativity scales can be adapted for causality measures and causal complexity can be quantified. Abilities to think causally are most sensitive to biasing influences when information is homogeneous.
5	17	Yes	Evaluations across a non-numerical matrix can be understood using geometric relational measures. Impact evaluations are higher on average.
	18	Yes	Evaluations across a non-numerical matrix can be understood using geometric relational measures. Predictability evaluations are higher on average.

10.1 The Presence of Anchoring Biases

The first research question asks whether there exists measurable anchoring biases in SP content. In order to answer this question, tools are required that both affect and detect cognitive motivations. Stage 1 is identified as an effective apparatus for developing a framing mechanism to test for anchoring effects. This is because participants are required to be familiar with Stage 1 information (i.e. background research on a target organisation) in order to engage in any of the other IL stages. Therefore, Stage 1 information can be used across all stages in a manner that preserves real-world SP methods, while also standardising an independent variable (i.e. framing mechanism) across all studies. Priming and validation tests help develop the most robust framing stimuli in the form of dichotomously paired business vignettes: one with high-threat framed messages and the other a structurally identical business vignette, differing only with low-threat framed messages. Creating stimuli through repeated development and validation testing helps clarify the boundaries of an experimental space, whether lab or field. Clearly defined limitations allow for greater accuracy in not just data interpretations, but generalisability of discoveries. Validation tests and clearly defined limitations to a study are especially important in fields like SP where the process is practitioner-driven, through intuitive efforts, and knowledge is built in pragmatic, constructed efforts. SP studies reviewed in Chapter 4 reveal, largely through absence of discussion in the respective articles, that validations of stimuli have not been a major focus for SP researchers within empirical efforts. The chosen methodology addresses the validation gaps in SP studies by using psychological theory against the backdrop of business models. As SP researchers, bridging business with psychology should be a norm in our work. Psychology is the scientific study of the mind and its functions, especially those affecting behaviour in a given context. SP is a series of efforts in thinking about the future within a business context. Research into SP is, by virtue of the practice, psychological research.

Three vignette pairs are developed for three different organisations. Each organisation reflects a different level of general familiarity to the participant. The purpose is to provide a plurality of priming stimuli that explore how participants relate to information through their constructed social and psychological realities, and how these relationships, in the face of framing mechanisms, manifest in the data. Absent in SP empirical literature is acknowledgement that priming stimuli are not all the same, just as participant samples, even from within the same population, are not the same. This investigation offers one standard of many, borrowed from psychological disciplines, to develop and validate stimuli within the context of not just SP, but forecasting, futures-thinking, and any strategy related study interested in understanding sensitivities in decision-making.

Table 10.2 shows the vignettes used for testing Stages 2, 3, and 5. Differences based on vignettes could show support for the strength of individual, socially constructed internal narratives, while similarities could show support for shared cultural or social narratives. For a full comparative analysis, it would be necessary to have all three vignettes tested across all sample profiles (UK, UAE, and US). Table 10.2 shows that this is not the case. For a basic comparative analysis, it would be necessary to have all three vignettes tested, at the very least, within a single sample profile (UK, UAE, or US). The table shows that this method was attempted in Stage 3 and Stage 5 studies. Within the Stage 3 heterogeneous study, only vignettes 2 and 3 samples returned usable data, therefore eliminating basic comparative analyses. Within Stage 5 studies, only vignette 1 was completed in the homogeneous study, and vignette 3 was the only sample to result in enough submissions to analyse.

A minimal comparative analysis requires at least two vignettes within a single sample profile. The table reveals that this was achieved only in the Stage 3 heterogeneous study, and as reported at the start of the results section, vignette 3 sample is too small to perform comparative inferential statistics. Therefore, even though priming and validation tests helped develop a statistically robust framing mechanism, experimental results do not allow for comparative results at this time, across vignettes.

Table 10.2. Business vignette profiles presented across studies and samples

Stage	**Experiment**	**Vignette**	**Profile**		
			UK	**UAE**	**US**
2	**Individual-only**	1			
		2	✓	✓	
		3			
	Individual+group	1			
		2			
		3	✓		
3	**Heterogeneous**	1		~~✓~~	
		2		✓	
		3	✓	✓	
	Homogeneous	1			
		2	✓	✓	
		3			
5	**Exploratory-Heterogeneous**	1	✓		
		2	~~✓~~		
		3	~~✓~~		
	Hypothesis-Homogeneous	1			~~✓~~
		2	~~✓~~		~~✓~~
		3	✓		✓

Note: Checks indicate the vignette was delivered to participants. Crossed out checks indicate vignette samples that were too small to analyse or absent due to no submissions.

Another method for minimal comparative analysis would be to compare responses by vignette across sample profiles. The table shows that this method is potential possible in Stages 2, 3, and 5. It is reported in Chapter 7 that behaviours between samples did not statistically differ. The same is true for Chapter 8 results. The UK sample in Chapter 9 is not large enough (n < 10), unfortunately to allow for comparison across samples. Limitations from these mismatched outputs are discussed later in this chapter.

Creative, divergent thinking is associated with Stage 2 efforts in the IL/SP framework. Stage 2 appears to be highly sensitive to anchoring biases. Different threat-levels communicated within organisational backgrounds affects prolific abilities of divergent thinking, as well as the kinds of external factors participants focus on. Anchoring biases also appear to have a compounding effect with future performance expectations. Whether participants anticipate presenting and defending their own ideas to a group of peers or not, alters which external factors they focus their attentions on. Surprisingly, this "role expectation effect" results in similar behaviours whether participants are working alone or in a group. The expectation of discussing their decisions in the future appears to be a potential confounding prime that results in different behaviours. The similarity in individual-only and group-level behaviours may also partly be a reflection of both being homogeneous environments (i.e. same framing message backed output).

Similar to creativity, causal, associative thinking appears to be sensitive to anchoring biases. Searching for causality between existing DF is associated with Stage 3 efforts in the IL/ST framework. The same dichotomously different threat-levels appear to affect how people construct causality between existing factors. Almost all measured biasing behaviours are pronounced when scenario information is presented in a homogeneous manner, where framing message is held constant throughout the stages. Interestingly, all anchoring biases are attenuated when information from the previous stages are presented as a heterogeneous blend of factors. In fact, there are no statistically verified anchoring biases when information from Stage 2 is a heterogeneous mix of creative ideas generated from different groups of participants who received different messages about the same organisation.

Unlike the previous two stages, evaluative thinking does not appear to be sensitive to anchoring biases. Evaluating future impact and predictability of scenario content is associated with Stage 5 efforts of the IL/ST framework. Message framing (high vs low) as well as stimuli mixes (heterogeneity vs homogeneity) are preserved in the method, yet no anchoring biases are detected in the outcomes. Several questions arise from all the studies, but some become more pertinent to the Stage 5 studies. When a behaviour or phenomenon is not detected, researchers ask themselves two main questions: *Is the behaviour truly not there?* and *Am I using the correct tools to detect it?* As discussed in Chapter 5, the mind remains a black box and therefore we cannot directly measure thoughts and decisions. To work within these limitations, we use established theory to help establish rules that represent mental phenomena, and use physical properties to serve as proxies for our manipulations and measurements (Grohs, Kirk, Soledad, & Knight, 2018). Based on the rules and proxies, we anticipate specific behaviours and design tools to measure for those behaviours. Chapter 9 reports very light theoretical support for defining evaluative thinking, in relation to anchoring biases, and even less on understanding evaluations through Stage 5 behavioural measures. The result is that the tools developed in this study to measure a hypothesised phenomenon may not be the correct tools to measure Stage 5 cognitions. It may also be the case that evaluative thinking may not be as sensitive to framing effects as the previous two cognitive tasks. Just as likely, the stimuli developed to serve as a proxy for Stage 1 may be an effective tool for measuring anchoring biases in some IL stages, but not in others. As with all science, new ground is being broken. To discover the answers, more testing with new methods must be conducted. This issue is further discussed in the next section, which addresses research question 2.

10.2 **Comparison of Anchoring Biases**

The second research question asks whether anchoring biases are comparable across different ST cognitions and SP content. The creativity studies use standardized analyses for measuring creative output. Fluency, flexibility, elaboration, and originality measures are borrowed from Guilford (1950), Torrance (1966), and Wallach & Kogan (1965). These measures are complimented with qualitative, content analyses that use text mining. Dialogic measures are also taken, to help understand effects from group-level engagement. The creativity studies are the only ones that offer group-level alongside individual-level testing, therefore there are no group-level comparissons in performance possible across the ST studies. Creativity measures reveal that external threats on an organisation do not necessarily affect all modes of creative, divergent thinking. Divergent thinking, by way of fluency measures, appears to be the most sensitive, at both the individual-only and group-levels (as measured by added and deleted DF). Differences in organisational threat, irrespective of familiarity with the organisation or location of the participants, appears to consistently affect which external factors participants pay the most attention to, when searching for potential future influences. The remaining creativity measures show that within the scope of exploring memory banks for influencing factors, participants remain equally limited; In that categories are elaborated on the same, with as much difference between categories. The ubiquity in flexible thinking – identifying factors in all STIRDEEPER categories, plus more – potentially shows a strength in the IL model. By simply directing participants to first "think of as many external factors" as possible, appears to prime participants to effectively cover a wide range of standard categories (i.e. STIRDEEPER) while also identifying a substantial number of outlier factors. Deeper dives with content analyses reveal striking effects in unique topical focuses, which account for the majority of the outlier DF (i.e. outside the standard STIRDEEPER typology). This is a key finding given the goals of SP, which include the ability to "'stretch' thinking about the future and widen the range of possible alternatives" (Van Notten, Rotmans, Van Asselt, & Rothman, 2003, p. 434), "identify a range of new threats and opportunities" (Volkery & Riberio, 2009, p. 1199), and facilitate a "creative experience that generates a heartfelt 'Aha!' from your managers" (Wack, 1985b, p. 140); Not to mention that the content created in the early stages of SP determines the breadth of information explored in the later stages. The other key finding is the "role expectation effect" that appears between the two individual samples. Implications from these discoveries are discussed in the last chapter.

However, as previously mentioned, one unifying theme does appear to tie the six ST studies together. The individual-only and group-level studies reflect a form of homegeneity in information, at an early stage in SP. The method to develop homogeneous and heterogeneous group information to serve as stimuli for the causal and evalutative studies extends the theme across all studies. This measure in information consistency may provid valuable insight into the efficacy of SP methods. Comparitive measures are discussed along these lines in this section and implications on the technique are discussed in the next chapter.

The associative, causal thinking studies adapt existing methods for analysis, both from causal and creative theories. Causality features are largely defined by Hastie (2015), Hume (1748), and Kant (1783/1994). Causality measures adapt creativity measures of fluency (i.e. frequency of DF, CC, and clusters) and elaboration (cluster complexity). Stimuli are adapted from Yaniv's (2011) work with heterogeneous vs homogeneous group information. The quantitative measures are complimented with qualitative, content analyses. Both studies reveal that messages about external threats on an organisation most strongly affect how prolific participants identify causal relationships between external factors, matching threat-level outcomes with the creativity studies. Increasing external threat-levels that could potentially be acting upon an organisation leads participants to more limited exploration of causal relationships between factors key to the organisation's future. The effects are amplified when information is maintained across Stages 1, 2, and 3, indicating differences in whether SP practitioners work with the same group throughout a workshop (i.e. homogeneous), or new members join midway through (i.e. heterogeneous). Content analyses help clarify *how* participants develop their scenario perspectives, where topical focuses alter from more diverse to more focused cluster constructions, depending on framed message.

There are, however, some consistencies across the conditions. If we reference the exploratory analyses in Appendix D, vantage point is almost identical across conditions, where participants largely use a diagnostic mode of reasoning, which appears to be resistant to any anchoring or information consistency effects. Thinking diagnostically results in participants reasoning with more conjunction fallacies. What is truly fascinating about the discovery is the issue that this mode of inference and reasoning is considered to be more difficult (Bradfield, 2004; Hastie, 1984, 2015). We are conditioned to think in a forward fashion, from here to the future, in our causal reasoning. To reason in the opposite direction is thought to require greater cognitive reasoning and associative skills. However, these previous assumptions may not apply to the quality of causal reasoning utilised when engaged in ST. Spontaneous creativity as a factor of causal reasoning is also similar across framing conditions, showing no anchoring bias. Though, spontaneous creativity appears to be greater from participants who work with homogeneous information, the average trends between the homogeneous and heterogeneous samples match the average trends for individual-only and individual+group samples from Stage 2 studies, revealing continuity in behaviours across both stages.

The evaluative thinking studies are largely exploratory in nature, therefore novel measures are developed for both studies. Frequency measures are taken across axes that could possibly serve as comparitive results to fluency measures of the previous studies. A novel measure is developed alongside standard frequencies, which helps understand how messages of organisational threat can affect average perceptions of cluster content (i.e. spread). One of the most interesting outcomes of the framing studies is the almost complete lack of anchoring bias in evaluative thinking. Whether participants are given the heterogeneous or homogeneous clusters, they valued their impacts and predictability almost identically to their counterparts who read a different threat level to the external environment. This may very well show that the IL model of employing evaluative thinking after a series of creative and causal thinking exercises leads to a more stable perspective of the future. Though it is important to note that 'stability' does not necessarily equate to 'accurate' or 'beneficial' knowledge. Much like the complimenting creativity measures, the resistance of evaluative thinking to any anchoring biases is as likely, at this point, to be a reflection of another bias as it is to be symptom of the robustness of the IL model. And, as the data show, another bias did appear to emerge. In all conditions, participants exhibited a high-side bias, which was intensified in the homogeneous condition.

As revealed across all studies, an anchoring bias is present for some cognitive features of ST and not for others. This is an incredible insight that must be better understood in SP literature, not to mention foresight and forecasting scholarship. Thinking about the future, whether through reasoning, deliberation, or decision-making, is not a single cognitive task. And yet, this is precisely how the empirical studies, in particular, frames ST. The studies reviewed in Chapter 4 treat the process of ST as an all-encompassing effort that can be reduced to a single identity (i.e. ST) and a single measure (e.g. forecasts). However, to strategise for the future requires a system of diverse cognitions, each with their own unique profile (Lintern, 2007), neurological pathways (Runco, 2014a), and language (Gentner, Holyoak, & Kokinov, 2001). ST is a collection of these features expressed through a cognitive structure. Providing evidence that a single manipulation does not, in fact, have the same effect on the different structural components of ST, speaks strongly to the incomparable nature of cognitive functions. As well, the studies support the view of ST as a gestalt, and to understand the whole, we must also understand the parts. Furthermore, we find that ST is more than an internal activity of the individual. ST is also an effort of distributed cognition, where…

> …a joint activity… is distributed across the members of a work or social group and their artifacts. Cognition is distributed spatially so that diverse artifacts shape cognitive processes. It is also distributed temporally so that products of earlier cognitive processes can shape later cognitive processes. Most significantly, cognitive processes of different workers can interact so that cognitive capabilities emerge via the mutual and dynamic interplay resulting from both spatial and temporal coordination among distributed human agents.
> (Lintern, 2007, p. 398)

The altered, even reversed behaviours of participants when they moved from individual to group activities, and between heterogeneous and homogeneous group information, show that ST is even more complex than we have been acknowledging. Future research must take these realities into consideration and account for such complexities.

As discussed in earlier chapters, the extant SP literature generally has not reflected on further biases introduced during scenario development. The present investigation appears to suggest that participants enter with a proclivity for the anchoring bias, and certain stages within SP can exacerbate the bias. The discoveries contribute to understanding greater complexity in the phenomenon of the anchoring bias.

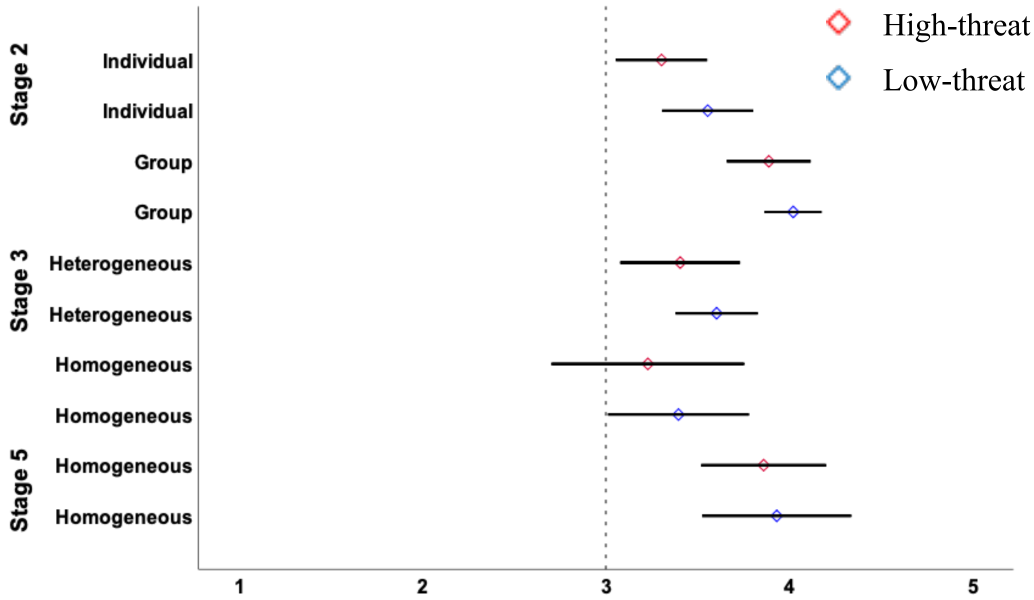
Traditional and early experiments exploring the anchoring bias use single questions that ask for simplistic quantitative responses (i.e. point estimates) to illustrate the strength of anchoring biases. Empirical studies in the SP scholarship, though not all explicit in their efforts to measure an “anchoring” bias, help reveal the beginning complexities that can attribute to practitioners’ proclivity to anchor on information. The quantitative studies mostly use a tradition point estimate method for data collection. The qualitative studies extend the narrative with exploratory work to help illustrate further areas research can be expanded, to better understand potential anchoring biases. The studies in Section II show that anchoring biases can be detected and measured on a number of different behaviours, both quantitative (e.g. frequency and complexity) and qualitative (content analysis).

10.3 **Practitioner Confidence**

The third question asks whether a relationship exists between performance across the studies and participants’ confidence in their performance. This question, possibly more than the others, really brings to the forefront issues of SP efficacy. In order to understand the efficacy of SP, we must understand ST and what this means to practitioners (both as individuals and as group members). Confidence has shown to be a major motivator in both decision-making and action and informs everything we do (Bénabou & Tirole, 2002; Feltz, 2007). It is a cue for decision-making, and as such, help us determine the value of our action. In all studies, there are consistent trends that support an anchoring bias. Participants who read about high-threat conditions, regardless of cognitive effort, are consistently less confident than their low-threat counterparts (see Figure 10.1). No differences, however, are strong enough to statistically support the argument that confidence is affected by differently framed messages of the same business environment. In fact, the only statistically valid change in confidence comes from engaging in homogeneous group discussions that employ creative thinking. The specifics of the dialogue do not appear to strongly affect confidence, and the product (adjusting DF lists) of their group sessions do not appear to affect confidence, even when their feedback reports otherwise. What appears to have the largest effect on confidence is engaging in 30 minutes of active group discussions. This effect, as well, is similar in nature across both dialogic methods. On average, participants feel greater confidence in their efforts after sharing and defending their ideas with others.

It is important to highlight that another, stronger, bias emerges from the data, specifically, a high-side bias. In a series of different cognitive tasks where participants are asked to make judgments and decisive choices about content that has no standard by which to gauge accuracy, all participant conditions report greater confidence (> 3) in their decision making.

Figure 10.1. Mean confidence scores by study



Note: 95% confidence interval is represented with each plot.
(<3 = low confidence, 3 = indifference, >3 = high confidence)

The only condition that results in participants reporting lower confidence is one of the high-threat conditions within the causal thinking study, which brings some support to the hypothesis that an anchoring bias could affect confidence. However, the robustness of this self-directed perception against the larger priming efforts for an anchoring bias helps reveal the untrustworthiness of using confidence as a guide for accuracy. As Kahneman and Klein (2009, p. 521) conclude from a series of past studies, “The answers that come to mind are typically held with substantial confidence, and the victims of anchoring manipulations confidently deny any effect of the anchor." The words of the researcher's former mentor, Miguel Moreno, on the first day of cognitive psychology lecture, come to mind, "Being sure, feels good."

10.4 New Methods for the Future

The fourth research question asks whether new methods for measuring and understanding scenario thinking could be applied to future research. Every experimental study in this investigation borrowed techniques and tools from psychology and applied them to an exemplar of SP methods. Guiding this investigation was the development of the first-of-its-kind IL/ST framework, which maps dominant cognitive processes within ST against the IL model. By linking complimenting, yet qualitatively different features of ST against discrete, interrelated stages of SP, the methodology introduces new language into the research. SP can be viewed through the lens of ST, and vice versa. Actions and content of SP output can, and should, be understood more comprehensively through the mental states of practitioners engaging in ST.

10.4.1 *Intuition and Biases*

Linking ST with SP is particularly useful in an intuitive practice. Intuition is knowing something, without knowing you know it (Kahneman, 2011). The popular attitude with intuitive insights is to accept the spontaneous discovery of the insight 'as is', and no systemic deconstruction of the process will reveal meaningful knowledge on how one produced the insight (Gladwell, 2007). However, the IL model appears to adopt the ideas more closely related to the respective works of Klein and Kahneman, that through a specific process of construction, deliberation, and systematic thinking, insights will not only be revealed, but their sources will be understood. It is a wonder, then, why we still call it the "Intuitive" Logics model.

Anchoring biases are well-documented, decision-making and reasoning phenomenon (Jacowitz & Kahneman, 1995; Kahneman & Klein, 2009; Mussweiler & Strack, 2000). It's presence is further revealed from the stimuli (i.e. business vignettes) used to help prime differently framed impressions of the same environment. Anchoring biases function at an implicit level, and therefore the effects often fall outside a practitioner’s awareness (as evidenced when comparing performance outcomes with confidence ratings).

10.4.2 *High-side Bias*

Along with revealing anchoring biases through several iterations of different stages in the IL/ST framework, this investigation also reveal a potentially new bias, identified as a high-side bias. The first this new bias is mentioned is in the Stage 5 studies, where participants evaluate the potential future impacts and predictability of various clusters, related to two different organisations, higher on both scales. The second time a high-side bias appears is across the 10 confidence measures that follow the completion of each stage in the IL/ST framework. Both the emergence of the bias at a specific stage in the process, as well as the ubiquity of the bias on internal motivations of confidence is striking. The field of behavioural science has led to thousands of publications discussing discoveries of potentially hundreds of cognitive biases, to date. What qualifies as a bias and does not, depends on a number of factors, the least of which is robustness in evidence. As with all scientific pursuits, more testing is required; However, the discussion has begun. Framing and anchoring bias research is largely limited to simple priming mechanisms asking for numerical responses. This investigation has now shown that framing effects and anchoring biases can appear from more than simplistic decision-making, and be measured in a variety of responses that go well beyond a single numerical prediction. The precedence is set to open the field to the awareness that other biases may be in action as well. Something as subtle, yet ubiquitous, as a high-side bias may be motivating practitioner decisions at any point within SP.

The field would benefit from borrowing methods and tools popularised in other disciplines, as this investigation has done, to help reframe how we understand SP and ST. The methods and analyses used across the studies borrow from classic and contemporary knowledge in psychology, economics, and sociology. As Godet (2000b, p. 8) reminds us,

> Yesterday's tools are still useful today. Indeed, the kind of problems encountered, even if the world changes, often remain similar. If we ignore our accumulated heritage, we deprive ourselves of powerful levers, and waste a great deal of time reinventing the wheel. The memory of our methods must be kept alive so as to improve upon them.

Stage 2 studies show that the use of traditional creativity measures (i.e. fluency, flexibility, elaboration, and originality) in conjunction with standard strategy tools (e.g. PESTEL or STIRDEEPER) can promote richer knowledge of practitioners' cognitive processes and attentional focus. The more granular view afforded through such AI efforts as text-mining, has the potential to make novel insights more salient (e.g. the trend of low-threat conditions leading participants to focus on higher-threat factors such as "war" and "terrorism"). Stage 3 studies illustrate how accounting for multiple aspects of causal thinking (i.e. size and composition of clusters, as well as content) can lead facilitators and practitioners, alike, to see important patterns in reasoning that would otherwise go unnoticed (e.g. less vs more complex causal reasoning, and predictive vs diagnostic reasoning). This speaks directly to the level of deliberation employed by practitioners in the moment, as well as abilities to think in systems, and even personal investment. Stage 5 studies help reveal the plurality of expressions (large and small) from biases in ST. Using such measures as distribution and spread of evaluated choices alongside standard strategy tools for orienting thoughts has great potential for maximising discussions amongst practitioners who are regularly short on time. The value of utilising novel research methods that blend qualitative and quantitative analyses can also be found in how the resulting insights help inform facilitators where to focus their attentions as workshops progress.

10.5 **Limitations and Gaps**

As with all investigations, there are a number of limitations in the designs which have left gaps in our knowledge and highlight the methodological challenges of such work. Several points are identified and elaborated in this section.

10.5.1 *Levels of individuality*

A major limitation in the present series of studies, as well as all empirical and case studies, is the lack of comparison between individually generated and group negotiated decisions. A major challenge for all the studies, and any future studies, is the length of time participants can engage in the tasks. University ethics boards often have regulations that limit participant time to less than 60 minutes. However, even when ethical approval is given for longer experimental sessions, participant motivation is a limited resource that diminishes over time, whether due to unappealing experimental material, tasks, or mental/physical fatigue. Issues with time severely limited the designs for the studies, and unfortunately left several group-based methods on the cutting floor.

To advance our comprehension of ST, and the efficacy of SP, individual- and group-level thinking, reasoning, and decision-making must be better understood. Both as individual efforts and in conjunction with the other. There are three main options to empirically explore these cognitive efforts:

- Independently – Measure individual thinking in isolation and group thinking in isolation, then compare the results in a between-subjects design.
- Ordered 1 – Measure the change in outcomes when participants move from an individual thinking task to a group thinking task (e.g. Stage 2 studies). This would be a within-subjects design, which offers greater statistical power.
- Ordered 2 – Measure the change in outcomes when participants move from a group thinking task to an individual thinking task. This would also be a within-subjects design and would afford greater knowledge of effect sizes when compared to the independent outcomes.

Stage 2 studies show that, at least for divergent, creative thinking, the “role expectation” has a different effect on the participants compared to performing as a lone participant. Further, the rate of creativity lessens dramatically when participants move to their group sessions. This may largely be due to their existing lists of DF, before the group sessions. It would be good to run a further study where participants only have a group session, to determine if creativity is facilitated better by group-only participation or individual-to-group participation.

As was discussed in Chapter 3, individual decision-making may utilise heuristics specific to isolated reasoning and thinking and can succumb to different types of biases compared to group decision-making (Bennett, 1990; De Bono, 1987; Eden, 1992; Eden & Ackermann, 1998). Considering this knowledge, studies should be designed which measure for these potential differences. All statistical methods used in the studies to analyse the individual-ST data can be used for group-ST data.

10.5.2 *Distributed Cognition*

Along with group-level ST, a perspective that should be brought into the larger ST/SP research is awareness of how much distributed cognition is part of the ST structure (Cole & Wertsch, 1996; Hutchins, 1991). Distributed cognition is also referred to as the *extended mind*. The theory explores how individuals off-load, share, store, and in many other ways distribute their knowledge to other sources for storage and retrieval. External storage units (that is, external from one's personal mind) can be digital, mechanical, and social, to name a few. Distributed cognition offers a framework for studying cognition rather than a type of cognition. It is important to recognise that workshops create a reflexive space, where practitioners and facilitators are actively engaged in a circular relationship, influencing one another, and responding based on cues each give, as was the case with the dialogic methods in Stage 2 study. Cognitive efforts are shared, loaned, passed, and gathered from the other actors in the workshop space. Such research could include analyses of audio and video data from a strategy workshops, conversation analysis, and assessment of group dynamics.

Extended analyses could be applied to the group portion of Stage 2 to help determine whether relationships exist between the participants or their individual DF lists and the resulting unified group list could lead to identifying predictive factors. For example, features such as length and elaboration of categories could be measured against the final group lists to help determine if editing efforts were guided by popularity of category or the prolific measure of an individual's list. Conversations could have been recorded and analysed using software such as NVivo to help determine if there was equity in members' input, degree of supportive and critical language, and how successful participants engaged in their assigned dialogic method.

10.5.3 *Content Accuracy*

A limitation all studies contain is a lack of clarity in the accuracy or efficacy of the content that participants generated. Stage 2 studies show that high-threat messages result in fewer DF generated by participants, compared to their low-threat counterparts. However, it may be that learning of greater threats to an organisation leads participants to focus on the most important, or highest risk, DF. This group may be saving cognitive resources by paying more attention to those factors that eventually prove to the be the most important in the end. Then again, it may be exactly how it looks, which is that this group thought of fewer DF. Stage 3 studies show that the same high-threat message can lead to differently complex clustering compared to the low-threat related participants. Further investigations and real-world case studies are needed to better understand the benefits of these differing behaviours.

10.5.4 *Ubiquitous Comparisons*

Another limitation to the present investigation is the lack of full comparisons between heterogeneous vs homogeneous conditions. The issue of information consistency is a regular feature in SP workshops. Reported in many case studies and position pieces is the issue that practitioners are time-poor professionals, often unable to participate in the full SP experience (Cairns & Wright, 2018b). Some attend one of a series of workshops, while others participate in just the early interview stages. As well, IL scholars suggest introducing "remarkable people" or external experts, midway through a SP workshop (Bradfield, Cairns, & Wright, 2015; Bradfield, Wright, Burt, & Van der Heijden, 2005; Van der Heijden, 1997). The results are a selection of heterogeneous conditions. Even when anchoring did not appear to have a strong effect on the choices of the participants, the difference in scenario information (i.e. DF, models, and clusters) did have effects. Given the regularity of heterogeneous conditions within real-world SP, it would be good to expand this research to understand the effects of consistent vs inconsistent information more fully.

10.5.5 *Vantage Point*

A surprise from the data that would be helpful to further explore is the domination of diagnostic thinking patterns in Stage 2 studies. Causal thinking discussion in scenario literature almost entirely assumes that predictive, forward-casted thinking is more natural and easier to perform, and therefore should be the expected norm with practitioners. However, this proved not to be the case once ST was dissected into its structural parts and specifically measured for effect. Diagnostic thinking is one of few features measured that is stable across conditions. Future ST research should make a strong effort in testing the natural proclivity of directional thinking (diagnostic vs predictive). Such studies should also compare and contrast the frequency of the conjunction fallacy. Tversky & Kahneman present a picture of negative outcomes when such a fallacy is used in reasoning. Then Schoemaker opens the discussion that such a fallacy might actually serve a beneficial purpose under certain (SP) circumstances. What remains a gap in our knowledge, however, is whether one direction of thinking over the other (predictive vs diagnostic) is more beneficial. Along with this, whether certain associated fallacies serve their own beneficial purposes for scenario development and mental challenges.

10.5.6 *Lack Of Control Conditions*

A major limitation in the power of the analyses comes from the lack of control groups throughout the studies. If a control group is possible, then a true baseline is needed to contrast experimental conditions against. Without a control group, comprehension of impact, effects, and even generalisability is muted. To best address the claim that SP, and even more simpler scenario review, can and should challenge mental frames, a powerful comparison would be a between- and within-subjects design that compares no process to a full scenario process. A version of this method is used in Meissner & Wulf's (2013) study, where the authors test a full, partial, and no scenario process in a between-subjects design. Their conclusion is that a full scenario process taught participants how to think more critically, which is measured by the degree of bias participants exhibited in a simple framing test. These are promising results, but Meissner & Wulf's study presents the same shortcoming as all other SP empirical studies. The authors treat ST as a single cognitive function and generalise the outcomes of a single choice task to a real-world, scalable platform. To build on the existing research, a more powerful design could include two samples of participants (S1 and S2). S1 gives a series of foresight responses regarding organisation 1's future. S1 then participates in a SP workshop with the purpose of giving a series of foresight responses for organisation 2, where measures are taken at each stage. S2 only participates in a SP workshop with the purpose of giving a series of foresight responses for organisation 2. Performance outcomes (e.g. creativity, causality, evaluations) regarding organisation 1 from S1 serves as the control group. The same performance outcomes regarding organisation 2 from S1 serve as a within-subjects comparison, and from S2 as a between-subjects comparison. Differences between each comparison help reveal the effect of SP on cognitive structures.

Sometimes, though, a control group is not possible. In the case of factorial designs, two variables (i.e. factors) are tested together. This is the case Section II studies. It is possible that the designs could have been reduced further, to incorporate a control condition that did not experience a framing manipulation. However, one of the main questions a researcher must ask when designing an experimental study is whether the information gathered from a control condition will be meaningful to the study and answer the research question. At the beginning of this investigative journey, it was determined that the comparative knowledge gained from control conditions would not answer the questions being asked. However, it is important to note at this point, that a comparative control condition may prove necessary and beneficial with other designs aimed at detecting and measuring different cognitive biases.

10.5.7 *Confounding Variables*

Along with a lack of control conditions, there are potential, unknowable gaps in understanding stemming from confounding variables. The primary cause could be due to the chosen factorial design. Unavoidable external changes in the environment and participant populations introduced confounding variables into the designs of all studies. The main confound is the use of different business vignettes across each study. The changes in participant behaviour could be due to the anchoring bias, but could also be due to vignette effects. As discussed in Chapter 6, three different vignettes were designed specifically to help determine whether level of familiarity and feelings of personal investment with an organisation affected ST, either independently of any framing effects, or interactively. This was not possible to determine and is a major limitation on this study. However, the plurality of stimuli does not necessarily weaken the discoveries made within the three study designs. "An operational imperfection is better than some non-existent perfection" (Kapferer, 1990). As the results help illustrate, there are several behaviours that remain unchanged across the different conditions (e.g. confidence averages and high-side biases). To fully answer the hypotheses, however, another round of studies, that control for confounding variables more completely are needed.

Another potential source of confounding variables is the research team that was temporarily utilised for some of the Stage 3 studies. The lead researcher led all experimental sessions in the UK and US. Travel restrictions prevented the lead researcher from personally facilitating two of the UAE sessions that required a physical actor in the computer lab to direct participants to computers, ensure they had access to the online materials (i.e. Qualtrics survey), and answer any questions that may arise. The other team members were briefed on the specifics of the study and given a script to use that told participants how to access the online survey and to email the lead researcher with any questions. The UAE-based team members only served to ensure participants were able to access the online materials so they could complete their part of the study and did not alter then study's content in any way. It cannot be known how much influence they may have had on the session or the participants within the session. The team members could have had a more familiar relationship with some or all of the participants that attend the sessions that afternoon, compared to the lead researcher and the remaining participants. It is possible they could have been less familiar, in comparison. Differences in familiarity could have contributed to attendance preferences within the samples, as well as possibly why the whole UAE sub-sample who received vignette 1 within the heterogeneous group (n = 6) had to be removed due to incomplete or incomprehensible models. However, the addition of other facilitators also provides the only opportunity for any confounding influences brought on by the lead researcher's implicit biases to be potentially removed from the sessions. Awareness of researcher bias is vital to good scientific practice. The additional team facilitators allowed their sessions to utilise a double-blind method. In this way, helping to reduce experimental biases.

10.5.8 *Full Scenario Thinking Exploration*

Due to limitations in various resources (e.g. time, finances, location, participants), this present investigation was unable to develop a full exploration of the anchoring bias against ST. Such an investigation, though, could take any number of expanded designs, depending on the series of questions.

First, individually testing the other stages would allow for fuller understanding of structural components of ST. Section II studies only cover four of the eight IL stages, but from these, effects of an anchoring bias and information consistency are detected. Further research should be carried out on the latter stages of the IL model, to test for potential effects with the other components of ST: memory encoding, parallel, systems, episodic-future, political, counterfactual, and prospective thinking. As well, any number of qualities within each of these can also be explored. For example, Stage 3 studies only measure one form of causal thinking related to clustering. As Aristotle[23], Hume (1748), Hastie (1984, 2015), and many others have discussed, there is a plethora of ways to think and reason causally, and just as many ways to measure these phenomena. An addition to the Stages 3 and 4 of causal mapping would be to include values of relationships between DF. In Stage 3 studies, no value was given to relationships, just direction. However, not all relationships are the same, just as not all effects are the same. One type of valuation that could bring more

[23] *Physics* II 3 and *Metaphysic*s V 2

granular understanding and nuance to a SP exercise is the designation of positive (+), negative (-) and no relation (0) quantifiers (Ülengin, et al., 2010). Hodgkinson, et al. (1999) use a similar method. The authors instruct participants to map causal relationships by drawing arrows between nodes (i.e. variables) to indicate direction of causation, and value the arrows ±3 points to represent the nature and strength of the causation, where +3 represented the strongest possible positive effect and -3 represented the strongest possible negative effect.

Second, testing participants through a full SP workshop, gathering measures along the way, would create a homogeneous condition from which we could understand the interacting efforts of ST. Studying the full process would also help address the question whether homogeneous conditions, where the same practitioner group remains from beginning to end, would exacerbate any anchoring biases that appear at early in the process, and whether introducing new players into the process would attenuate any existing biases. Third, testing different groups of participants with the same variable manipulations (e.g. more samples given vignette 1, 2, or 3) could better show the level of stability of ST, the robustness of SP, and statistically show the effect sizes of manipulations through longitudinal designs. This last point is particularly difficult to justify with empirical evidence, while trying to increase ecological validity. For example, all experimental studies presented in this thesis took place over the span of four years. These four years started at the Scottish independent vote at the end of 2014 and ended with the post-Brexit "negotiations" at the start of 2019. Trends in the data show that salient factors for the participants changed over time. For example, all mentions of "Scottish independence" or "Scottish referendum" appear exclusively within the two experiments of Stage 2 studies, which took place between the years 2014-2016. Mentions of "Brexit" in any capacity as a DF appear with all experiments for Stage 2 and 3 studies. As a consequence of its popularity, a Brexit DF was factored into the clusters for the heterogeneous portion of Stage 3 study.

10.5.9 *Participant Motivation*

An often discussed, but little addressed, aspect in both SP and empirical studies is the issue of participant motivation. Motivation is key to ownership of knowledge and information gained during SP workshop sessions. The more practitioners (i.e. executives, stakeholders, management) feel ownership of the information and process, the more likely perceptual changes will take hold. “We contend that if practitioners and researchers wish to achieve stakeholder ownership within a participatory planning process, they need a clear understanding of its characteristics (or attributes), the philosophy underlying its achievement and the implications for project governance, engagement processes, staffing, time and budget” (Soste, et al., 2015, p. 251). Existing empirically-backed knowledge regarding motivational aspects of participants in SP studies remains the elephant in the room. How powerful are the results of scenario studies that appeal to the most shallow level of information processing? The studies presented in this thesis attempted to overcome this barrier by creating action-oriented experimental designs, where participants created content in a similar manner to SP workshop requirements. However, as with all human-based testing within laboratory conditions, the scalability of these results will require further field testing. Suggestions are offered to guide further empirical work in practitioner motivations, especially in group environments (Druskat & Pescosolido, 2002; Fullan & Miles, 1992; McMillan & Chavis, 1986; Soste, et al., 2015; Spreitzer, 1996; Stirling, 2006).

(1) low role ambiguity (clarity of purpose)
(2) a wide span of control, which ‘opens up’ rather than ‘closes down’ conversations
(3) a sense of socio-political support
(4) access to information and resources
(5) a participative climate
(6) ownership of process
(7) the need for continuous learning
(8) the need for heedful inter-relating

10.5.10 ***Laboratory vs Field Testing***

The final limitation to the series of studies stems from the chosen methodology. There is, understandably, a long and well-argued discussion on the issues with generalising discoveries in the lab to real-world settings. Scaling up from artificial laboratory settings to complex real-world environments often falls short. The entire field of *implementation science* is dedicated to precisely this cross-over. The methodology for this investigation, though, borrows from the hard sciences with respect to empirical inquiry. Necessary to the validity of the results is the ability to control as many extraneous variables as possible, within the testing space, and amongst the variables. To maintain high levels of experimental controls, laboratory settings were integral to the design of the studies. As a result, several outcomes showed a trend in the direction of the hypotheses, but lacked the statistical power to be considered true differences affected by the IV. In future studies, advice of the researcher's former psychology advisor will be taken, and metaphorically "smack the participants with a heavier IV". That is, develop much farther extremes with the framing manipulation. Such an effort may also more closely mimic real-world conditions, where perceived threats to one's organisation are often felt deeply and create an atmosphere of stress and anxiety that is qualitatively different from business-as-usual primed atmospheres.

Still, it can never be fully assumed that performance in the laboratory will mimic performance in the field and vice versa. Every variable in the environment counts and must be counted (Dilip, 2015). Balogun, et al. (2003) offer some compelling methodological suggestions to researching strategic behaviours. Their guidance addresses the issues of extraneous variable control, resource limitations, and data collection. The authors offer a list of criteria for conducting action research in the discipline of strategy.

Methods must meet the following criteria (pp. 200-201):

(1) Provides evidence/data that is both broad and deep because it is
    a. contextual
    b. longitudinal
    c. facilitates comparison across sites
    d. can be collected at multiple organizational levels.
(2) Elicits full and willing commitment from informants because it is
    a. interesting enough to engage organizational commitment
    b. enjoyable enough to sustain commitment over time.
(3) Makes the most effective use of researcher time because it
    a. collects
    b. organizes
    c. analyses, large and varied amounts of evidence.
(4) Anchors the majority of questions being asked in organizational realities because it
    a. is sensitive to multiple definitions of critical issues
    b. addresses problems of interest and relevance
    c. involves organizationally based collaborators.
(5) Goes beyond research-based feedback to
    a. contribute to organizational needs
    b. provide informants with personally useful insights
    c. inform the content of further collaboration.

Balogun, et al. (2003) suggest a blending of case study and laboratory methods to achieve meaningful action research. Addressing each criteria in order; (1) Research should be conducted in the field, within organisations; (2) The experimental method, such as SP workshops, serves to benefit the organisation; (3) The researcher *is* the facilitator; (4) Research findings and organisational interests are equally prioritized; (5) Build longitudinal, micro-level research that offers regular feedback. The authors challenge fellow researchers, "if we are to move beyond archival data and limited questionnaires to gather the kind of in-depth information on strategizing discussed in this paper, we must ask much more of ourselves and our colleagues in organizations" (p. 220).

10.5.11 ***Transparency***

A major focus of the work throughout this investigation has been to maintain the highest levels of transparency in reporting. Critical evaluation of the theories and replicability of the methods are cornerstones to good scientific practice. The researcher acknowledges that even with the greatest care taken, information may be lost in translation, so to speak. Therefore, the last section to discuss potential limitations of the research, provides a review of the efforts taken to report with as much transparency as possible. Chapter 1 introduces the methodological foundation to scenario work and argues from a constructivist position – socially, psychologically, organisationally, dialogically, and reflexively. To approach a constructivist position using empirical methods, a pragmatic methodology is used. Adopting a pragmatic approach allows designs and discoveries within the thesis to be contextually defined, but not deny the potentiality for more generalisable applications. Chapter 5 synthesises discoveries from the multiple scenario reviews and establishes boundaries for the empirical work, by way of proxy testing in a cognitive context. Possibly most important to the collective body of work within this thesis is the systematic development of the priming stimuli (i.e. business vignettes) presented in Chapter 6, which serves as the independent variable across all the experimental studies. There are an uncountable number of ways to develop, validate, and deliver an independent variable. The methods used to create the independent variables were specifically chosen to reflect real-world SP as closely as possible, but also to resonate as much as possible with SP researchers. One reporting feature missing too often in empirical studies within the SP discipline is validation of the independent variable. Without understanding the strength or quality of a manipulation, impact is lost, to an extent, when it comes to interpreting the data. Transparency and higher standards in design are why a better part of a chapter is dedicated to just reporting the development of the independent variable. Furthermore, the appendices provide the developmental roadmap to explicitly show the evolution of the independent variable across the priming and framing validation tests. The work reported in Chapter 6 and the appendices is explicitly aimed at helping to establish higher levels of method and design reporting in academic writing.

10.6 Key Contributions and Further Developments

The main contributions from this investigation are stated at the beginning of this chapter. The first is the IL/ST framework developed to guide the empirical studies. The IL/ST framework is not the only framework possible. Given the variety of prevailing SP models, and the black box the mind continues to remain, there is possibility for any number of pragmatically constructed frameworks that bridge the stages (or steps) of a SP method with their associated unique, dominant cognitions. At the very least, the methodology introduces a new and more structured approach to understanding the practice.

The methodology is interdisciplinary. Bringing the language, techniques, and tools of other empirical systems of inquiry into business models has introduced a new way to understand SP. Specifically, by using psychological methodology, this investigation moves SP research from action and content focused, to practitioner focused (i.e. mental states). SP is a practitioner-driven process, and because of this, SP should be understood through the practitioner as much as it is understood through the scenarios, themselves.

In researching one bias (i.e. anchoring), the methods appear to have revealed another potential bias (i.e. high-side). Understanding practitioner sensitivities to biasing content speaks directly to the efficacy of both the process and outcomes of SP. Researchers and practitioners must stop using language in their research that implies the discipline is immutable and begin to own the knowledge that SP is truly a pragmatic practice, intuitions are susceptible to undetected influences, and, as evidenced in consistently high confidence scores, our awareness of the effects is considerably limited. The final chapter brings together the key contributions of the work and applies them to the field, with recommendations for practical pursuits.

Chapter 11. **Implications for Scenario Planning and Thinking**

"The mind can go either direction under stress—toward positive or toward negative: on or off. Think of it as a spectrum whose extremes are unconsciousness at the negative end and hyperconsciousness at the positive end. The way the mind will lean under stress is strongly influenced by training."
Frank Herbert

"The invention of the ship was also the invention of the shipwreck"
Paul Virilio

This chapter brings together the collection of work from the interviews, reviews, and empirical studies, to present a list of implications for SP interventions. The IL model that guides this thesis follows eight distinct stages for SP. Each stage emphasises a specific cognitive quality of ST, with the assumption that they all work together in a complex web of functionality. A relationship between practice and cognitions is reflected in the IL/ST framework. Though SP methods are pragmatic, with any number of dimensions proving necessary/unnecessary for a given SP intervention, there are a few best-practice suggestions provided by the field, supported by the empirical discoveries, and presented in this thesis. All implications focus on maximising the capacity of ST. The practical assumption is that maximised ST will increase efficacy of SP and improve strategising within organisations. The anchoring bias is a well-documented cognitive mechanism that has now been shown to affect any number of decisions, entirely outside the awareness of the decision-maker, with a chimichanga complexity. Some strategists even use the phrase "short-terminism" to illustrate the effects of the anchoring bias on strategy (Desjardins, 2018). Anchoring on more salient, available information and making strategic decisions from this form of limited knowledge can lead organisations toward decisions that can be rationalised now, in the short-term, but add less long-term value.

One of the strongest messages to come from the present investigation is that participants – and by extension, practitioners – are cognitive followers. To be a cognitive follower is to possess a cognitive style that accepts incoming cues and information with too little critical evaluation of its content, source, or purpose

(Kahneman, Slovic, & Tversky, 1982). Cognitive styles develop the cognitive maps that act as heuristics for managers (Franco, Meadows, & Armstrong, 2013). As a result, heuristics are relied upon more heavily to reach strategic decisions, increasing the probabilities of making biased or sub-optimal decisions (Kahneman, 2011; Tversky & Kahneman, 1973; Tversky & Kahneman, 1974). Throughout most of our daily lives, these same heuristics and associated biases prove beneficial. They streamline our day, lessen unnecessary efforts (physical and mental), and keep us alive. However, within the context of a strategy workshop, these same cognitive maps can be, and have been, detrimental to achieving effective action (Hodgkinson & Wright, 2002).

11.1 **Facilitators**

If practitioners are cognitive followers, then it must be the case that facilitators assume the role of cognitive leaders, if SP is to become (or remain) an effective process for organisational survival. This does not mean facilitators must do the thinking for practitioners. Rather, their scenario expertise must include more than the knowledge of methods and models. Expertise must include technique. Facilitators would serve well if they can have expert knowledge in human thinking and reasoning, as both individuals and members of a group, awareness of biased decision-making cues, and be equipped with a toolbox of elicitation techniques to counter inhibiting biases and exploit beneficial biases (Van der Heijden, 1997). Empirical research is one method of providing facilitators with the necessary knowledge and techniques. With this in mind, scenario research could develop investigations (particularly experimental and quasi-experimental) that aim to answer the questions: *How can facilitators identify a practitioner's mental models? What processes affect practitioners' mental models?* and just as importantly, *What techniques can facilitators employ to promote practitioners' strengths and counter weaknesses that stem from their mental models?* In this finale chapter, implications of the empirical studies to SP as a practice and as a field of research are discussed.

11.2 **Setting the Scene in Stage 1**

Most important to the process is the preparation. It is a standard task in all SP schools to initially identify a problem, goal, or issue at the start (Bradfield et al., 2005). An effective SP intervention requires extensive planning and preparation. This initial point is illustrated well in the follow-up analyses from Moyer's (1996) "British Airways" case study. The author recognised that the SP team "underestimated the amount of work involved in developing plausible but challenging scenario stories... the team would have benefited from having a full-time analyst working on the problem... more contact by the team with external experts would have been useful... facilitators were doubling as presenters and vice versa," and many of the commitment issues with their team "could have been overcome if the presenters had been included more in the planning and running of each workshop" (p. 179).

11.2.1 ***Building a Scenario Team***

Success is greatly increased by ensuring at least one expert facilitator is designated to the task, with multiple co-facilitators (Bradfield, 2008). For large-scale SP interventions, such as with the FEMA, ESPAS projects, and British Airways case study, multiple facilitators with specific designations within the workshops may be necessary, in order to meet the needs of the tasks (FEMA, 2018; Hoorens, et al., 2013; Moyer, 1996). The British Airways case study included five facilitators, each responsible for the workshop phase (Moyer, 1996). Facilitators are able to remain more vigilant than practitioners with respects to regarding unnecessary constraints and reducing them where possible (Duckett, et al., 2017).

Even more so, creating a heterogeneous SP group with at least one external expert (e.g. remarkable person), will greatly increase the chances of avoiding standard pitfalls that stem from bounded rationality (Simon, 1972). Boland & Collopy (2004, p. 11) exemplify this notion when they state, "The more ways of thinking we have available to us, the better our problem-solving outcomes can be." This is a dimension of the second domain of Chermack's (2011) SP theory, "Decision-Making Theory", and a key concern with SP. The final six empirical studies presented in Chapter 7-9 show that heterogeneous groups appear to attenuate certain biases that can come from the information at hand.

Related to group success is ensuring the decision-makers comprise at least part of the workshop practitioners. Volkery and Riberio (2009, p. 1201) explain, "The main impacts of scenarios often result more from the process of developing them rather than from any published product describing the scenarios that were created." It is in the process that such importance lies because this is when mental models are made explicit and challenged (Chermack, 2011). The effects of challenging mental models are what increase adoption of the knowledge gained from SP. Further, "it has implications for the extent to which people trust scenarios and thus use them" (Volkery & Riberio, 2009, p. 1202). As one practitioner reported, after participating in a series of SP workshops, "In painting that picture it helped build relationships, it helped build that sense of common purpose and understanding. I think that's very powerful" (Bowman, et al., 2013).

11.2.2 *Setting the Agenda*

The reported studies within this thesis use brief vignettes as a proxy for Stage 1. In real-world settings, facilitators work with the client to identify and contact internal management (sometimes higher executives such as the CEO) who will participate in the intervention. During Stage 1, facilitators and the management team set the agenda, conduct background research, define goals, and determine the timeline. These practitioners will bring their own cognitive styles, maps, perceptions, and awareness to the process, and it is at this stage that such themes as threat-levels can begin to take form in the background information. Each of the empirical studies show that changes in even simplistic background information can cause participants to anchor in such dramatically different ways, that their input to the intervention becomes skewed. Based on the discoveries throughout this thesis, one of the cues a facilitator could learn to recognise is levels of diversity in the management team's early information searches. Van der Heijden (1997) supplies a well-established technique, by way of Stage 1 individual interviews, for gathering information from practitioners which could make these cues more salient. Van der Heijden's basic interview process is designed to "trigger" a conversation and is summarised below (pp. 145-148).

(1) "Ask the interviewee to briefly relate how (s)he came to be in their present position."
(2) "List their concerns and uncertainties… in the business environment."
(3) "Imagine that your future is a good one, rolling out as you would like it to be," what would it look like?
(4) "This question is followed by a similar one in which the world develops in an undesirable direction."
(5) "What pivotal events can you identify in the past of this organisation, good or bad, that should remain in our memories as important lessons for the future?"
(6) "What major decisions with long term implications is the organisation facing at the moment, decisions that need to be tackled in the next few months?"
(7) "What major constraints are you experiencing inside or outside of your organisation that limit you in what you can achieve in your business situation?"
(8) "Please consider the situation in the future when you will have moved on from your current position, to the next job or retirement, what do you hope to leave behind that people will associate with your period in office. What do you want to be remembered for?"

By triggering (i.e. priming) managers to make explicit what they find most important in their business's environment, the facilitator can determine not only where their compass is pointing (e.g. high or low levels of external threats), but how many compasses are pointing in the same direction (i.e. level of diversity), within the management team. A key to this format's success is in the order of delivery, as much as the content. Patterns, if they exist, are more likely to be identified after the interviews, during the analyses of them.

11.3 Facilitation Through the Stages

Stage 2 (the workshop) typically begins with introductions. This is the first opportunity for facilitators to apply expert techniques for countering any identified biasing cues from the interviews. Research discoveries, like the present investigation, help facilitators build a toolbox of expertise (e.g. cue detection, dialogue techniques, and biasing awareness). For example, if background research interviews reveal that management perspectives are largely concentrated on a single quality of external threats (low or high), then facilitators can expect greater limitations on the management team's ST. They could attempt, instead, to apply early intervention techniques such as explicit awareness reinforcement and countervailing information.

11.3.1 *Reinforcement Technique*

Generalised language can be employed to increase awareness and reinforce the message that certain forms of biased thinking can affect abilities to creatively explore the external (future) environment, identify causal relationships between the various external factors, and value their performances in the future effectively or even accurately (Sukhera & Watling, 2017; Tetlock, 2005). Reinforcing awareness of biased thinking moves practitioners from being reflexive to reflective, and towards a metacognitive form of self-reflection. Verbal reinforcement technique can prove to be particularly helpful at any stage within the process, whenever practitioners begin to engage in stagnating efforts such as group-think (Janis, 1982), defensive avoidance or unconflicted adherence (Janis & Mann, 1977), or political thinking (Kaplowitz, 1990). Stagnating efforts risk severely limiting the scope of the management team's ST by entering them into a shared, reinforcing cognitive space (Wright & Rowe, 2011; Yaniv, 2011). If facilitators detect stagnating cues, one technique they can use is to intervene with guiding questions such as *"What do you think is not possible?"*, *"What have you left out because you thought it was implausible?"* Guiding questions prime practitioners to engage in purposeful divergent and/or counterfactual thinking, increasing the chances they will identify original content (e.g. DF, causal relationships, stakeholders, etc).

Reinforcement techniques, however, are weak in isolation and require implementation of associated counter-measures to ensure ST is maximised in the moment. As Bradfield (2008) discovered with his SP participants, when groups of practitioners reach a state of punctuated equilibrium induced coagulation, basic verbal reinforcement techniques are rendered inert. One major difference between Bradfield's study and Stage 1 study in Chapter 6 is that Bradfield identified ST coagulation within his participants later in the process. This difference in timing is precisely the phenomenon Wright, et al. (2008, p. 229) recognise, "Once the coping patterns have been deployed, recognition of the need for, and the value of, scenario planning may be lost." Such concerns are further supported by the heterogeneous vs homogeneous studies in Chapters 8 & 9. Biased decision-making appeared to be greater when information supplied by previous groups of participants was homogeneous in content, as well as with the present participant's framing manipulation.

11.3.2 *Countervailing Information*

Providing countervailing information shifts a management team's perspective away from a concentration on high-threat factors, towards a more heterogeneous mix of factors. Bradfield (2004, p. 39) states that, "getting individuals to make large, creative leaps in their thinking is exceptionally difficult as they are inevitably constrained by their cognitive anchor." Results from Stage 2 studies suggest that this type of mental shift can increase creative output on several levels. Whereas those practitioners who perceive more threats to their organisation may think more fluently, group members who perceive fewer external threats will bolster the chances of serendipitous insights during the group discussions. One source for credible countervailing information could come from Van der Heijden's (1997) pre-interview process. Question 3 primes managers to think beyond the negatives, and open-ended questions 2, 5, 6, and 8 offer potential for explorative responses. Using information shared in the early interviews also reduces the work-load for facilitators.

11.3.3 *Individual vs Group Behaviours*

The data suggest that another technique to maximise the kind of creativity employed in Stage 2 is for facilitators to begin this stage at the individual level, then move to the group level. To maximize creative output at the start of this stage, facilitators could emphasise in the introductions that practitioners will work on their own and as a team throughout the workshop(s) and will have to negotiate their own contributions alongside others' on its usefulness. After the process has been explained to the management team, the task of identifying DF begins with each practitioner creating their lists on their own. After a pre-set time limit, or when practitioners begin to slow down, the facilitator then groups practitioners together (no more than 5 to a group if possible) and the creative process continues at the group level.

11.3.4 *Dialectic Inquiry*

As we've come to realise, though, not all group interactions are the same. Contrary to expectations, employing more active Hegelian dialectics within groups bolsters creative, serendipitous thinking. If management teams engage in little debate, then facilitators can expect less diversity in thinking – the worst outcome leading to group-thinking stagnation. Low levels of debate, idea development, and quick agreements amongst members of the group can all serve as biasing cues for the facilitator. If any of these cues are detected, the facilitator can step in and offer some dialogic guidance. Such guidance should encourage in a prescribed manner to maximise conflict and critical evaluation. Another technique would be to avoid such possibilities altogether by standardising the dialogic method within SP. To transition from the individual task to the group task, an additional step could be added that introduces *how* to work as a group. Participants in the group-based sessions of Stage 2 studies were given no more than a few minutes of introduction to their respective dialogic methods before engaging in the group task. Even this brief moment of familiarising themselves with the method produced measurable differences in performance. Of course, with a method based on increasing defense of personal ideas and critical evaluation of others' ideas, negative outcomes are possible and should be monitored. Unchecked conflict can increase distance between group members, unintended insults, a feeling of devaluation of one's ideas, and overbearing and silenced voices. Cues for negative group conflict are a single group member speaking more (or less) than others, senior members speaking up the most, and scenario information (e.g. list of DF) decreasing rather than shifting or increasing.

11.3.5 ***Deliberation Time***

As Runco (2014b, pp. ch1, 10) illustrates, when it comes to creative thinking, originality becomes a factor of time, "The notion that original ideas come late in the associative chain implies that we should take our time when faced with a problem, to insure that we get to those remote ideas." In essence, cues, such as the information supplied in the business vignettes (i.e. information gathered during Stage 1), are apt to focus attention on the obvious and known at the expense of creative thinking (Bradfield, 2008). The more novel, unique, unusual, or unconventional ideas will emerge *after* the standard, common, and business-as-usual content is identified and moved to the side. Armed with this knowledge, facilitators should be aware that shorter face-to-face time will limit originality and novelty in thinking. One solution could be to adopt schedules that allow practitioners ample time to deliberate on the exploratory, creative efforts to generate lists of DF, as well as all new information that is discovered throughout the stages. This is, understandably, a difficult requirement to achieve, given the time-poor nature of practitioners. However, the less pressure practitioners are put under to debate and engage in ST, the more likely they are to discover novel insights (Janis & Mann, 1977; Wright, et al., 2008).

11.3.6 ***Serendipity as a Factor of Time***

This phenomenon is further supported in the group data gathered from Stage 2 studies and the secondary creative data in Stage 3 studies. Though participants in the Stage 2 studies expressed an average higher confidence in their own individual abilities to identify all the "necessary driving forces for the organisation to reach their 5-year goals," some groups still managed to identify up to 15 more novel DF that were not included in any of the group members' original lists. Stage 3 studies reveal that even after engaging in half an hour of causal thinking, participants were still inspired towards serendipitous thinking, and identified more novel DF, beyond their casual models. What is incredible about both of these discoveries is that there was no active facilitation after the groups (Stage 2 study) and individuals (Stage 3 study) began their creativity sessions, and both received only a single statement that references further creative, divergent thinking: Stage 2 study's dialogic methods, "If you think of more driving forces, enter them into the survey from the link provided in your email," and Stage 3 study's follow-up instructions, "Can you think of any other influences that weren't mentioned?" Both studies help show that even the lowest level of facilitation can have a measurable impact on ST.

11.3.7 *Clustering*

Taking into consideration the causal evidence from Stage 3 studies, facilitators could develop a few techniques to nudge practitioners towards more beneficial causal reasoning. To start, it is evident that, much like creativity, causal thinking, on several fronts, appears to become hindered by messages of higher external threats to the organisation. Therefore, in line with ST maximization efforts, facilitators should ensure workshop information gives the impression that there are fewer external, potential threats than other valued external factors. Threat-recognition and risk assessments are important, but should not dominate a management team's perspectives.

11.3.8 *Vantage Point*

Another cue to be aware of is in the developed of vantage points, as evidenced in the exploratory data presented in Appendix D from the Stage 3 studies. Most of the participants used a diagnostic mode of reasoning, which led to employing a form of conjunction fallacy to build clusters. Schoemaker (1993) actually discusses the potential benefits of employing a conjunction fallacy in SP. Considering that a host of other biased thinking may dictate the quality of practitioners' ST, and one of those appears to be normative thinking (Bradfield, 2008; Tetlock, 2005), Schoemaker suggests that the tendency to assume more causal sources for a single outcome may make unlikely scenarios appear more believable. Greater believability is correlated with great plausibility. Facilitators, therefore, may benefit from ensuring practitioners are not diminished or hindered in this seemingly natural effort.

11.3.9 *Causal Reasoning*

The conditions under which causal associations were generated would be classified as reactive, in that the procedure (i.e. Stage 3) dictates the action, rather than the action being spontaneous. This is important to understand because the efforts of spontaneous and reactive causal reasoning can be affected by different covarying factors. An interesting expansion with this theory is how reactive conditions affect the levels of implicit desire to create causal associations, and elaboration of associations. A series of studies over the past 30 years reveals two pertinent conditions that increase both desire to create associations and elaboration of those associations: *unexpected events and failure* (*see* Gendolla & Koller, 2001; Hastie, 1984; Kanazawa, 1992; Weiner, 1985). This is a particularly relevant behaviour for facilitators to be made aware. Within their reactive, choreographed SP setting, one of the desired outcomes is for practitioners to imagine plausible, but previously unimagined events – surprises. According to discoveries from cognitive psychology, it may be the case that the farther a scenario's elements stray from the norm, the more elaborative practitioners' thinking will become, in response. As well, imagining future failures within scenarios, either at some midpoint or endpoint, may also elicit greater complexity in causal reasoning. This could not only serve as a benefit for scenario practitioners, but if this is the case, it behooves the facilitator to guide practitioners into applying equally deliberative and complex causal reasoning to scenarios that reflect their norms and/or (expected) trends of progression.

11.3.10 *Matrix and the High-side Bias*

Stage 5 studies shows that the level of perceived external threats does not exhibit a strong anchoring effect on evaluative thinking, and therefore, may not be as big of a concern for facilitators to control for, once practitioners reach Stage 5 in the workshop. Particularly, if facilitators have maintained homogeneous conditions with information development through the first four stages, then one of their biggest concerns could be countering any high-side bias that may arise. Facilitators will also want to watch for related issues such as foresight bias and creeping determinism (Fischhoff, 1975; MacKay & McKiernan, 2010). One way to counter the high-side bias without limiting the practitioners' exploratory, evaluative thinking is to allow management teams to make their projections on the I/P matrix, first. This allows the facilitator to look for cues, then determine whether a bias is emerging or not through non-invasive observation. If a bias does emerge, the facilitator can step in *after* the clusters are fully plotted onto the matrix. Using Hegelian dialogic methods, management teams can then be tasked with explaining their logic behind each cluster's position. Adjustments should be encouraged through this effort, employing counterfactual thinking. If the bias persists, explicit awareness reinforcement techniques can be brought back into the process. Such an effort could be something as simple as highlighting the fact that the majority of the clusters are justified with logic that removes either most of the uncertainty (i.e. located above the x-axis) and/or assumes a closer relationship with the organisation (i.e. located to the right of the y-axis). Specific examples from their logic can be used to reflect their narrative back to the group. Once this bias is made salient to the practitioners, they could be asked to go through their logics again, to see if they may be assuming too confidence about the unknown future, and to deliberate over the I/P matrix a final time (Wright & Goodwin, 2002).

11.3.11 ***Mainstream Techniques***

The behaviours of participants across all empirical studies suggest a reality quite different from both the stated values of SP use in industry, as well as within the literature. Ramirez, et al. (2014) claim that the value of scenario work is a function of external threat. SP and ST are more valuable to practitioners (not necessarily to organisations) when threat levels are higher. Perceptions of value and threat guide much of the literature, whether implicitly or explicitly, and are the prevailing motivations evidenced in the early interviews of Chapter 4. However, this is a fuzzy area to make such assumptions. Furthermore the empirical studies conducted throughout this investigation show just the opposite may be true. If value is based on effectiveness of the process, and effectiveness is determined by the performance of the practitioners, then the data suggest that the best performance will be elicited from perceptions of lower external threats. Therefore, the final data-driven suggestion from this investigation is that the value of SP may be, indeed, a function of external threat, but in the opposite direction, and would be most beneficial if it were mainstreamed, and therefore practiced on a regular basis. These final thoughts comprise a series of recommendations that could not only be adopted into SP, but guide further empirical investigations to aid in developing a larger and more unified theory.

11.4 Concluding Summary

This investigation explored whether and how certain cognitive biases associated with ST have an effect on SP content. In the empirical discussion, it was elaborated how interdisciplinary work can provide new insights into business research by bringing in new language, techniques, and tools of other disciplines. Applying psychological theory against the backdrop of a business model provided the language and tools to map SP stages with dominating ST cognitions (i.e. the IL/ST framework). Using interdisciplinary efforts like the ones developed throughout this thesis, can bring new forms of methodology to future empirical studies. Through these collaborative efforts, the empirical studies revealed differing magnitudes of decision-making sensitivities to not just one bias (i.e. anchoring), but potentially a new bias (i.e. high-side). Even more importantly, participants were unaware of their biased decision-making, as revealed in part through their self-rated confidence scores. The current body of research adds to our understanding of how and in what manner scenario practitioners construct their views of the future. Research into cognitive biases helps show that any number of variables (e.g. media, business vignettes, personal research, and corporate records) can influence how people think about the future. SP is a practitioner-driven effort, and as a practice, can benefit from the field gaining greater understandings of just *how* practitioners are driven.

Translating discoveries to practice, this investigation focuses on how to apply the new body of knowledge to facilitator training, expertise, and technique. Having some form of expert knowledge in human thinking and reasoning can help facilitators identify patterns and sensitivities in scenario practitioners. Ensuring heterogeneity in group composition can help attenuate any number of limiting decision-making biases that could diminish the ultimate value of SP. Developing a repertoire of cues that identify reinforcement, countervailing, dialectic, and evaluative behaviours can afford facilitators opportunities to mitigate some of the negative outcomes of the more limiting decision-making efforts, while possibly supporting greater novel thinking efforts about the future. Through the regular use of such expert, interdisciplinary knowledge, facilitators can help scenario teams increase both the value and efficacy of SP at all levels.

Epilogue

Over the course of seven years, I explored the intuitive practice of scenario planning. Though a rich scholarship on the discipline continues to grow, there remains too few empirical investigations aimed at understanding scenario planning efficacy and impact. The dearth of scientifically rigorous work was the driving motivation behind my pursuits, and after seven years, I have started to understand only a small portion of the challenges that lie in attempting to empirically validate, map, and aid in developing a grounded theory for a practice that relies almost entirely on the unknown workings of the human mind. Much of my work took novel approaches to understanding the discipline and the main issues within. This thesis is the culmination of systematic efforts to not just fill some of the gaps in our knowledge, but provide robust evidence for new methodological designs for empirically approaching scenario planning and thinking. As much as this thesis provides answers to difficult questions, my hope is that it equally bolsters further novel investigations into scenario planning and thinking, with just as surprising discoveries.

References

Abelson, R. P. (1976). Script processing in attitude formation and decision-making. In J. S. Carroll, & J. W. Payne (Eds.), *Cognition and Social Behavior.* Hillsdale, NJ: Erlbaum.

Adame, B. J. (2016). Training in the mitigation of anchoring bias: A test of the consider-the-opposite strategy. *Learning and Motivation, 53*, 36-48. doi: 10.1016/j.lmot.2015.11.002.

Alexander, W., & Serfass, R. (1998). Beyond vision: Creating and analyzing your organization's quality future. *31*(7), 31-36. Retrieved from https://search.proquest.com/.

Almashat, S., Ayotte, B., Edelstein, B., & Margrett, J. (2008). Framing effect debiasing in medical decision making. *Patient Education and Counseling, 71*(1), 102-107. doi: 10.1016/j.pec.2007.11.004.

Amason, A. C. (1996). Distinguishing the effects of functional and dysfunctional conflict on strategic decision making: Resolving a paradox for top management teams. *Academy of Management Journal, 39*(1), 123–148. doi: 10.5465/256633.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures, 46*, 23-40. doi: 10.1016/j.futures.2012.10.003.

Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.). (1998). *Critical Realism.* London: Routledge.

Arden, R., Chavez, R. S., Grazioplene, R., & Jung, R. E. (2010). Neuroimaging creativity: A psychometric view. *Behavioural brain research, 214*(2), 143-156. doi: 10.1016/j.bbr.2010.05.015.

Aristotle. (1934). *Physics* (Vol. II). (P. H. Wicksteed, & F. M. Cornford, Trans.) Cambridge: Harvard University Press.

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist, 63*(7), 602–614. doi: 10.1037/0003-066X.63.7.602.

Asgary, A., & Levy, J. (2009). A review of the implications of prospect theory for natural hazards and disaster planning. *International Journal of Environmental*

Research, 3(3), 379-394. Retrieved from https://www.sid.ir/en/journal/ViewPaper.aspx?id=138223.

Axelrod, R. M. (Ed.). (1976). *The Structureof Decision: Cognitive Maps of Political Elites.* Princeton, NJ: Princeton University Press.

Baas, M., De Drue, C. K., & Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus? *Psychological bulletin, 134*(6), 779-806. doi: 10.1037/a0012815.

Balarezo, J., & Nielsen, B. B. (2017). Scenario planning as organizational intervention: An integrative framework and future research directions. *Review of International Business and Strategy, 27*(1), 2-52.

Balogun, J., Huff, A. S., & Johnson, P. (2003). Three responses to the methodological challenges of studying strategizing. *Journal of Management Studies, 40*(1), 197-223. doi: 10.1111/1467-6486.t01-1-00009.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1–26. doi: 10.1146/annurev.psych.52.1.1.

Bansah, A. K. (2016). *The effect of message credibility on attitude change as measured by argumentation style: In an introductory nutrition class.* Ohio: The Patton College of Education of Ohio University.

Banxia, S. L. (2017). Decision Explorer. Kendal, UK.

Bawden, R. (1998). The leadership revolution. *Keynote address to AusTAFE Regional Conference.* Ballarat, Victoria, Australia.

Bazerman, M., Giuliano, T., & Appelman, A. (1984). Escalation of commitment in individual and group-decision making. *Organizational Behavior and Human Performance, 33*, 141-152. doi: 10.1016/0030-5073(84)90017-5.

Becker, H. S. (1983). Scenarios: A tool of growing importance to policy analysts in government and industry. *Technological Forecasting and Social Change, 23*, 95-120. doi: 10.1016/0040-1625(83)90049-5.

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics, 117*(3), 871–915. doi: 10.1162/003355302760193913.

Bes, B., Sloman, S. A., Lucas, C. G., & Raufaster, E. (2012). Non-Bayesian inference: Causal structure trumps correlation. *Cognitive Science, 36*, 1178-1203. doi: 10.1111/j.1551-6709.2012.01262.x.

Bishop, P., Hines, A., & Collins, T. (2007). The current state of scenario development: An overview of techniques. *Foresight, 9*(1), 5-25.

Blalock Jr., H. M. (Ed.). (1985). *Causal Models in the Social Sciences.* New York: Routledge .

Bloom, M., & Menefee, M. L. (1994). Scenario planning and contingency planning. *17*(3), 223-230. doi: 10.2307/3380654.

Bodin, R., Chermack, T. J., & Coons, L. M. (2016). The effects of scenario planning on participant decision-making style: A quasi-experimental study of four companies. *Journal of Futures Studies, 20*, 21-40.

Boje, D. M. (2001). *Narrative methods for organizational & communication research.* London: Sage.

Boje, D. M., & Murnighan, J. K. (1982). Group confidence pressures in iterative decisions. *Management Science, 28*(10), 1187-1196. doi: 10.1287/mnsc.28.10.1187.

Boland, R. J., & Collopy, F. (2004). *Managing as Designing.* Stanford: Stanford University Press.

Bood, R., & Postma, T. (1997). Strategic learning with scenarios. *European Management Journal, 15*(6), 633-647. doi: 10.1016/S0263-2373(97)00047-9.

Botzung, A., Denkova, E., & Manning, L. (2008). Experiencing past and future personal events: Functional neuorimaging evidence on the neural bases of mental time travel. *Brain and Cognition, 66*(2), 202–212. doi: 10.1016/j.bandc.2007.07.011.

Bowman, G., MacKay, R. B., Masrani, S., & McKiernan, P. (2013). Storytelling and the scenario process: Understanding success and failure. *Technological Forecasting & Social Change, 80*, 735–748.

Boynton, M. H., Portnoy, D. B., & Johnson, B. T. (2013). Exploring the ethics and psychological impact of deception in psychological research. *35*(2), 7-13. PMID: 23672145.

Börjeson, L., Höjer, M., Drebord, K. H., Ekvall, T., & Finnveden, G. (2006). Scenario types and techniques: Towards a user's guide. *Futures, 38*, 723-739. doi: 10.1016/j.futures.2005.12.002

Bradfield, R. (2004). What we know and what we believe: Lessons from cognitive psychology. *Development, 47*(4), 35-42. doi:10.1057/palgrave.development. 1100093.

Bradfield, R. M. (2008). Cognitive barriers in the scenario development process. *Advances in Developing Human Resources, 10*, 198-215. doi: 10.1177/1523422307313320

Bradfield, R., Derbyshire, J., & Wright, G. (2016). The critical role of history in scenario thinking: Augmenting causal analysis within the intuitive logics scenario development methodology. *Futures, 77*, 56–66. doi: 10.1016/ j.futures.2016.02.002.

Bradfield, R., Wright, G., Burt, G., & Van der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures, 37*(8), 795-812. doi: 10.1016/j.futures.2005.01.003

Brewer, N. T., & Chapman, G. B. (2002). The fragile basic anchoring effect. *Journal of Behavioral Decision Making, 15*(1), 65-77. doi: 10.1002/bdm.403.

Brodwin, D. R., & Bourgeouis, L. J. (1984). Five steps to strategic action. In G. Carroll, & D. Vogel (Eds.), *Straetgy and Organization: A West Coast Perspective* (pp. 167-181). Boston: Pitman.

Brownstein, M. (2018). *The implicit mind: Cognitive architecture, the self, and ethics.* New York, NY: Oxford University Press.

Brownstein, M., & Madva, A. (2012). The normativity of automaticity. *Mind & Language, 27*, 410–434. doi: 10.1111/j.1468-0017.2012.01450.x.

Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(5), e1501. doi: 10.1002/wcs.1501.

Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation, 36*(3), 375-388. doi: 10.1177/1098214015581706.

Burgess, R. (1984). *In the Field: An Introduction to Field Research.* London: Unwin Hyman.

Burke, J. (1995). *Connections.* Boston: Little & Brown.

Burt, G., Wright, G., Bradfield, R., Cairns, G., & Van der Heijden, K. (2006). The role of scenario planning in exploring the environment in view of the limitations of PEST and its derivatives. *International Studies of Management & Organization, 36*, 50-76. doi: 10.2753/IMO0020-8825360303

Cairns, G., & Wright, G. (2018a). *Scenario Thinking* (2nd ed.). Switzerland: Palrave Macmillan.

Cairns, G., & Wright, G. (2018b). Why should the individual and organization practice scenario thinking? In *Scenario Thinking* (2nd ed., pp. 1-25. doi: 10.1007/978-3-319-49067-0.). Palgrave Macmillan.

Cairns, G., Goodwin, P., & Wright, G. (2015). A decision-analysis-based framework for analysing stakeholder behaviour in scenario planning. *European Journal of Operational Research, 249*(3), 1050-1062. doi: 10.1016/j.ejor.2015.07.033

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review, 16*, 330–350. doi: 10.1177/1088868312440047.

Castle, B. (2018, March 19). U.K. Reaches Brexit Transition Deal With E.U. *The New York Times*, pp. Retrieved from https://www.nytimes.com/2018/03/19/world/europe/uk-brexit-eu.html.

Chermack, T. J. (2002). The mandate for theory in scenario planning. *Futures Research Quarterly, 18*(2), 25–28. Retrieved from https://www.researchgate.net/publication/303817681_The_mandate_for_theory_in_scenario_planning

Chermack, T. J. (2003). *A Theory of Scenario Planning.* St. Paul, MN: Swanson & Associates.

Chermack, T. J. (2011). In *Scenario Planning in Organizations: How to Create, Use, and Assess Scenarios* (1st ed.). Oakland: Berrett-Koehler.

Chermack, T. J. (2018). An analysis and categorization of scenario planning scholarship from 1995-2016. *Journal of Futures Studies, 22*(4), 45-60. DOI:10.6531/JFS.201806.22(4).0004.

Chermack, T. J., & Lynham, S. A. (2002). Definitions and outcome variables of scenario planning. *Human Resource Development Review, 1*(3), 366-383. doi: 10.1177/1534484302013006

Chermack, T. J., & Nimon, K. (2008). The effects of scenario planning on participant decision-making style. *Human Resource Development Quarterly, 19*, 351-372. doi: 10.1002/hrdq.1245

Chermack, T. J., Van der Merwe, L., & Lynham, S. A. (2007). Exploring the relationship between scenario planning and perceptions of learning organization characteristics. *Technological Forecasting & Social Change, 38*(7), 379–390. doi:10.1016/j.techfore.2006.03.004.

Chermahini, S. A., & Hommel, B. (2010). The (b) link between creativity and dopamine: spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition, 115*, 458-465. doi: 10.1016/j.cognition. 2010.03.007.

Cheung, C. M., Sia, C. L., & Kuan, K. K. (2012). Is this review believable? A study of factors affecting the credibility of online consumer reviews from an elm perspective. *Journal of the Association for Information Systems, 13*(8), 618-635. doi: 10.17705/1jais.00305.

Cho, S. H., & Gower, K. K. (2006). Framing effect on the public's response to crisis: Human interest frame and crisis type influencing responsibility and blame. *Public Relations Review, 32*(4), 420-422. doi: 10.1016/j.pubrev.2006.09.011.

Cole, M., & Wertsch, J. V. (1996). Beyond the individual-social antinomy in discussions of Piaget and Vygotsky. *Human Development, 39*(5), 250-256. Retrieved from http://beyondbitsandatoms.org/readings/ cole_beyond_1996.pdf.

Collyns, N. (1994). Creating scenarios. In P. M. Senge, A. Kleiner, C. Roberts, R. B. Ross, & B. J. Smith (Eds.), *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning OrganIzation* (pp. 274-277). New York: Doubleday.

Cornelius, P., Van de Putte, A., & Romani, M. (2005). Three decades of scenario planning in Shell. *California Review of Management, 48*(1), 92-109. doi: 10.2307/41166329.

Crawford, M. M. (2019). The comprehensive scenario intervention typology. *Technological Forecasting and Social Change, 149*, 119748. doi: 10.1016/j.techfore.2019.119748.

Crawford, M. M., & Wright, G. (unpublished paper). The value of mass produced COVID-19 scenarios . pre-print doi: 10.31219/osf.io/t9krn

Creswell, J. (2009). *Research design: Qualitative, Quantitative, and Mixed Methods Approaches.* London: Sage.

Cummings, T. G., & Worely, C. G. (2001). *Organization Development and Change* (6th ed.). Cincinnati: South-Western College Publishing.

De Geus, A. (1997). *The Living Company.* Boston: Harvard University Press.

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social Psychology and Personality Compass, 8*, 342–353. doi: 10.1111/spc3.12111.

Deconinck, J., Boers, F., & Eyckmans, J. (2010). Helping learners engage with L2 words: The form–meaning fit. *Aila Review, 23*(1), 95-114. Retrieved from https://biblio.ugent.be/publication/2130071/file/6771743.

Denzin, N. (1989). *The Research Act: A Theoretical Introduction to Sociological Methods.* New Jersey: Prentice Hall.

Derbyshire, J., & Wright, G. (2017). Augmenting the Intuitive Logics scenario planning method for a more comprehensive analysis of causation. *International Journal of Forecasting, 33*(1), 254-266. doi: 10.1016/j.ijforecast.2016.01.004

Desjardins, J. (2018, March 3). 18 Cognitive Bias Examples Show Why Mental Mistakes Get Made. *Visual Capital*, pp. Retrieved from https://www.visualcapitalist.com/18-cognitive-bias-examples-mental-mistakes/.

Dewey, J. (1991). *The Collected Works of John Dewey.* (J. A. Boydston, Ed.) Carbondale: Southern Illinois University Press.

Dietrich, A. (2004). The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review, 11*, 1011–1026. doi: 10.3758/BF03196731.

Dilip, S. (2015). *The last mile: Creating social and economic value from behavioral insights.* Toronto: University of Toronto Press.

Dougherty, M. R., Gettys, C. F., & Thomas, R. P. (1997). The role of mental simulation in judgments of likelihood. *Organizational Behavior and Human Decision Processes, 70*(2), 135-148. doi: 10.1006/obhd.1997.2700.

Drach-Zahavy, A., & Erez, M. (2002). Challenge versus threat effects on the goal-performance relationship. *Organizational Behavior and Human Decision Processes, 88*(2), 667-682. doi: 10.1016/S0749-5978(02)00004-3.

Drew, A. (2018, April 30). Bringing the lab to the real world. *Association for Psychological Science*, pp. Retrieved from https://www.psychologicalscience.org/observer/bridging-the-lab-and-the-real-world.

Druskat, V. U., & Pescosolido, A. T. (2002). The content of effective teamwork mental models in self-managing teams: ownership, learning and heedful inter-relating. *Human Relations, 55*, 283-314. doi: 10.1177/0018726702553001

Duckett, D. G., Sutherland, L. A., Kyle, C., Boden, L. A., Auty, H., Bessell, P. R., & McKendrick, I. J. (2017). Scenario planning as communicative action: Lessons from participatory exercises conducted for the Scottish livestock industry. *Technological Forecasting and Social Change, 114*, 138-151. doi: 10.1016/j.techfore.2016.07.034

Ducot, G., & Lubben, G. J. (1980). A typology for scenarios. *Futures, 12*(1), 51-57. Retrieved from https://scholar.google.co.uk/

Duncan, N. E., & Wack, P. (1994). Scenarios designed to improve decision making. *Planning Review*, 18-46. doi: 10.1108/eb054470

Durance, P., & Godet, M. (2010). Scenario building: Uses and abuses. *Technological Forecasting and Social Change, 77*, 1488-1492. doi: 10.1016/j.techfore.2010.06.007.

Dutton, J., & Jackson, S. (1987). Categorizing strategic issues: Links to organizational action. *Academy of Management Review, 12*, 76-90. doi: 10.5465/amr.1987.4306483.

Effingham, N. (2009). Persistence, composition and identity. In R. L. Poidevin, P. Simons, A. McGonigal, & R. P. Cameron (Eds.), *The Routledge Companion to Metaphysics.* New York: Routledge.

Einhorn, H. J., & Hogarth, R. M. (1982). Predicitons, diagnosis, and causal thining in forecasting. *Journal of Forecasting*, 3-33. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a104372.pdf.

Emery, F. E., & Trist, E. L. (1965). The causal texture of organizational environments. *Human Relations, 18*(1), 21-32. doi: 10.1177/001872676501800103.

Epstude, K., Scholl, A., & Roese, N. J. (2016). Prefactual thoughts: Mental simulations about what might happen. *Review of General Psychology, 20*(1), 48-56. doi: 10.1037/gpr0000064.

EU, C. O. (2019, January 29). Brexit: Council (Article 50) adopts negotiating directives on the transition period. *Council of the European Union*, pp. Retrieved from https://www.consilium.europa.eu/en/press/press-releases/2018/01/29/brexit-council-article-50-adopts-negotiating-directives-on-the-transition-period.

Eysenck, H. J. (1993). Creativity and personality: Suggestions for a theory. *Psychological Inquiry, 4*, 147–178. doi: 10.1207/s15327965pli0403_1.

Fagley, N. S., Miller, P. M., & Jones, R. N. (1999). The effect of positive or negative frame on the choices of students in school psychology and educational administration. *School Psychology Quarterly, 14*(2), 148–162. doi: 10.1037/h0089002.

Fahey, L., & Randall, R. M. (1998). What is scenario learning? In L. Fahey, & R. M. Randall (Eds.), *Learning from the Future: Competitive Foresight Scenarios* (pp. 3-21). New York: John Wiley.

Feltz, D. L. (2007). Self-confidence and sports performance. In D. Smith, & M. Bar-Eli (Eds.), *Essential Readings in Sport and Exercise Psychology* (pp. 278–294. doi:10.1249/00003677-198800160-00016). Champaign, IL: Human Kinetics.

FEMA, F. E. (2018). *National Level Exercise 2018.* Department of Homeland Security. Washington, DC: FEMA.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General, 140*, 168-185. doi: 10.1037/a0022100.

Fiegenbaum, A. (1990). Prospect theory and the risk-return association: An empirical examination in 85 industries. *Journal of Economic Behavior & Organization, 14*(2), 187-203. doi: 10.1016/0167-2681(90)90074-N.

Fiegenbaum, A., & Thomas, H. (1988). Attitudes and the risk-return paradox: Prospect theory explanations. *Academy of Management Journal, 31*(1), 85-106. doi: 10.2307/256499.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299. doi: 10.1136/qhc.12.4.304.

Fischoff, B., & MacGregor, D. (1981). Subjective confidence in forecasts. *Perceptronics*, e47. doi: 10.1002/for.3980010203.

Fisher, K. L., & Statman, M. (2003). Consumer confidence and stock returns. *The Journal of Portfolio Management, 30*(1), 115-127. doi: 10.3905/jpm.2003.319925.

Fodor, E. M. (2010). Power motivation. In *Implicit Motivations* (pp. 3-29). New York: Oxford University Press.

Fontana, A., & Frey, J. (2005). The interview: From neutral stance to political involvment. . In N. Denzin, & Y. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (3rd ed., pp. 695-728). Thousand Oaks, CA: Sage.

Forsyth, D. (2010). *Group Dynamics.* Belmont, CA: Wadsworth.

Franco, L. A., Meadows, M., & Armstrong, S. J. (2013). Exploring individual differences in scenario planning workshops: A cognitive style framework. *Technological Forecasting and Social Change, 80*(4), 723-734.

Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences, 24*(1), 1-2. doi: 10.1016/j.tics.2019.10.007.

Fugelsang, J. A., & Dunbar, K. N. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia, 43*, 1204-1213. doi: 10.1016/j.neuropsychologia.2004.10.012.

Fullan, M., & Miles, M. (1992). Getting reform right: What works and what doesn't. *Phi Delta Kappan , 73*, 745-752. Retrieved from https://static1.squarespace.com/static/57c4731c893fc047731cea43/t/5946db65ff7c505f81a9939e/1497815910009/Getting++Reform+Right.pdf

Gawronski, B., Hofmann, W., & Wilbur, C. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition, 15*(3), 485–499. doi: 10.1016/j.concog.2005.11.007.

Gendler, T. S. (2008). Alief and belief. *Journal of Philosophy, 105*(10), 634–663. Retrieved from https://www.jstor.org/stable/20620132.

Gendolla, G. H., & Koller, M. (2001). Surprise and motivation of Causal search: How are they affected by outcome valence and importance? *Motivation and Emotion, 25*(4), 327-349. doi: 10.1023/A:1014867700547.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (Eds.). (2001). *The Analogical Mind: Perspectives from Cognitive Science.* Cambridge: MIT Press.

Gilovich, T., Griffin, D. W., & Kahneman, D. (2013). *Heuristic and Biases: The Psychology of Intuitive Judgment* (14 ed.). Cambridge: Cambridge University Press.

Gladwell, M. (2005). *Blink.* New York: Little, Brown and Company.

Glick, M., Chermack, T., Luckel, H., & Gauck, B. (2012). Effects of scenario planning on participant mental models. *European Journal of Training and Development, 36*(5), 488-507. doi: 10.1108/03090591211232066.

Godet, M. (1987). *Scenarios and Strategic Management.* London: Butterworth.

Godet, M. (1990). Integration of scenarios and strategic management: Using relevant, consistent and likely scenarios. *Futures, 22*, 730-739. doi: 10.1016/0016-3287(90)90029-H.

Godet, M. (2000a). How to be rigorous with scenario planning. *Foresight, 2*(1), 5-9. doi: 10.1108/14636680010802438.

Godet, M. (2000b). The art of scenarios and strategic planning: Tools and pitfalls. *Technological Forecasting and Social Change, 65*, 3-22. doi: 10.1016/S0040-1625(99)00120-1.

Godet, M. (2001). *Creating Futures: Scenario Planning as a Strategic Management Tool.* London: Economica.

Godet, M., & Roubelat, F. (1996). Creating the future: The use and misuse of scenarios. *Long Range Planning, 29*, 164-171. doi: 10.1016/0024-6301(96)00004-0

Gordon, T. J. (1994a). *Cross-impact method.* American Council for the United Nations University.

Gordon, T. J. (1994b). Trend impact analysis. In J. C. Glenn, & T. J. Gordon (Eds.), *Futures Research Methodology: The Millennium Project* (2 ed.).

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist, 72*(9), 861–871. doi: 10.1037/amp0000238861 .

Gregory, W. L., Cialdini, R. B., & Carpenter, K. M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so? *Journal of Personality and Social Psychology, 43*, 89-99. doi: 10.1037/0022-3514.43.1.89.

Grohs, J. R., Kirk, G. R., Soledad, M. M., & Knight, D. B. (2018). Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems. *Thinking Skills and Creativity, 28*, 110-130. doi: 10.1016/j.tsc.2018.03.003.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444-454. doi: 10.1037/h0063487.

Guilford, J. P. (1967). *The Nature of Human Intelligence.* New York: McGraw-Hill.

Gute, D., & Gute, G. (2015). *How creativity works in the brain.* Washington DC: National Endowment for the Arts.

Haeffner, M., Leone, D., Coons, L., & Chermack, T. (2012). The effects of scenario planning on participant perceptions of learning organization characteristics. *Human Resources Development Quarterly, 23*(4), 519-542.

Hagmayer, Y., & Fernbach, P. M. (unpublished paper). Causality in decision making.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General, 143*(3), 1369–1392. doi: 10.1037/a0035028.

Hall, J. (1971). Decisions, decisions, decisions. *Psychology Today. 5*(1), pp. 51-54. Retrieved from https://www.scirp.org/(S(351jmbntvnsjt1aadkozje))/reference/ReferencesPapers.aspx?ReferenceID=685705

Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S., & King, C. J. (2010). Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines. *Center for Studies in Higher Education*, 1-12. Retrieved from https://cshe.berkeley.edu/publications/assessing-future-landscape-scholarly-communication-exploration-faculty-values-and-needs.

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology, 46*, 44-56. doi: 10.1037/0022-3514.46.1.44.

Hastie, R. (2015). Causal thinking in judgments. In G. Keren, & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making* (Vol. 2, pp. 590-628). John Wiley & Sons.

Hawkins, H. A., & Chermack, T. J. (2014). The effects of scenario planning on antecedents to participant engagement at work. *Journal of Futures Studies, 18*(4), 77-92. Retrieved from https://jfsdigital.org/wp-content/uploads/2014/06/Articles5-Chermack.pdf

Heilman, K. M. (2005). *Creativity and the brain.* New York: Psychology Press.

Heugens, P. P., & Van Oosterhout, J. (2001). To boldly go where no man has gone before: Integrating cognitive and physical features in scenario studies. *Futures, 33*, 861-872. doi: 10.1016/S0016-3287(01)00023-4

Hocutt, M. (1974). Aristotle's four becauses. *Philosophy*, 385-399. doi: 10.1017/S0031819100063324.

Hodgkinson, G. P., & Clarke, I. (2007). Exploring the cognitive significance of organizational strategizing: A dual-process framework and research agenda. *Human Relations, 60*(1), 243-255. doi: 10.1177/0018726707075297

Hodgkinson, G. P., & Healy, M. P. (2008). Toward a (pragmatic) science of strategic intervention: Design propositions for scenario planning. *Organization Studies, 29*, 435–457. doi: 10.1177/0170840607088022

Hodgkinson, G. P., Bown, N. J., Maule, J., Glaister, K. W., & Pearman, A. D. (1999). Break the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal, 20*, 958-977. doi: 10.1002/(SICI)1097-0266(199910)20:10<977::AID-SMJ58>3.0.CO;2-X.

Hodgkinson, G., & Wright, G. (2002). Confronting strategic inertia in a top management team: Learning from failure. *Organization Studies, 23*(6), 949-977. doi: 10.1177/0170840602236014.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369–1385. doi: 10.1177/0146167205275613.

Holmes Jr, R. M., Bromiley, P., Devers, C. E., Holcomb, T. R., & McGuire, J. B. (2011). Management theory applications of prospect theory: Accomplishments, challenges, and opportunities. *Journal of Management, 37*(4), 1069-1107. doi: 10.1177/0149206310394863.

Honderich, T. (Ed.). (2005). *The Oxford Companion to Philosophy* (2nd ed.). Oxford: Oxford University Press.

Hoorens, S., Ghez, J., Guerin, B., Schweppenstedde, D., Hellgren, T., Horvath, V., . . . Kobzar, S. (2013). *Europe's societal challenges: An analysis of global societal trends to 2030 and their impact on the EU.* Brussels: RAND Europe.

Huff, A. S. (1990). *Mapping Strategic Thought.* Chichester, UK: Wiley.

Hume, D. (1748). *An enquiry concerning human understanding* (2nd ed.). (L. A. Selby-Bigge, Ed.)

Huss, W. R., & Honton, E. J. (1987). Scenario planning-What style should you use? *Long Range Planning, 20*(4), 21-29. doi: 10.1016/0024-6301(87)90152-X.

Hutchins, E. (1991). The social organization of distributed cognition. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 283–307). American Psychological Association.

Igartua, J. J., & Cheng, L. (2009). Moderating effect of group cue while processing news on immigration: Is the framing effect a heuristic process? *Journal of Communication, 59*(4), 726-749. doi: 10.1111/j.1460-2466.2009.01454.x.

Inayatullah, S. (2008). Six pillars: Futures thinking for transforming. *Foresight, 10*(1), 4-21. doi: 10.1108/14636680810855991

(2015). *Intelligence Community Directive 203: Analytic Standards.* Washington, DC: Office of the Director of National Intelligence.

International, S. (2019). *Scenario Planning.* California.

Israeli, N. (1941). The psychology of planning. *The Psychological Record, 4*(18), 254–256. doi: 10.1007/BF03395208.

Janis, I. L. (1971). Groupthink. *Psychology Today*, 84-90. Retrived from http://apps.olin.wustl.edu/faculty/macdonald/GroupThink.pdf.

Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes* (2nd ed.). London: Houghton Mifflin.

Janis, I. L., & Mann, L. (1977). *Decision Making : a Psychological Analysis of Conflict, Choice, and Commitment.* New York: Free Press.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition, 50*, 189-209. doi: 10.1016/0010-0277(94)90028-0.

Jones, P., & Roelopsma, P. (2000). The potential for social contextual and group biases in team decision-making: biases, conditions and psychological mechanisms. *Ergonomics, 43*, 1129-52. doi: 10.1080/00140130050084914.

Jun, E., Hsieh, G., & Reinecke, K. (2017). Types of motivation affect study selection, attention, and dropouts in online experiments. *The ACM on Human-Computer Interaction*, *56*, p. 15. doi: 10.1145/3134691.

Kahane, A. (1999). Scenarios for changing the world. In P. Senge, A. Kliener, C. Roberts, R. Ross, G. Roth, & B. Smith (Eds.), *The Dance of Change: The Challenges to Sustaining Momentumin Learning Organizations* (pp. 238-239). New York: Doubleday.

Kahn, H., & Weiner, A. J. (1967). *The year 2000: A Framework for Speculation on the Next Thirty-Three Years.* Macmillan.

Kahneman, D. (2011). *Thinking fast and slow.* London: Penguin Books.

Kahneman, D., & Klein, G. (2009). Condition for intuitive expertise. *American Psychologist, 64*(6), 515-526. doi: 10.1037/a0016755.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291. doi: 10.1142/9789814417358_0006

Kahneman, D., Slovic, S. P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Kanazawa, S. (1992). Outcome or expectancy? Antecedent of spontaneous causal attribution. *Personality and Social Psychology Bulletin, 18*(6), 659–668. doi: 10.1177/0146167292186001.

Kant, I. (1783/1994). *Prolegomena to Any Future Metaphysics.* Pearson.

Kapferer, J.-N. (1990). *Rumors: Uses, Interpretation and Necessity.* (B. Fink, Trans.) New Brunswick, NJ: Transaction.

Kaplan, R. S., & Norton, D. P. (2000). Having trouble with your strategy? Then map it. *Harvard Business Review*, 49-61. Retrieved from http://www.academia.edu/download/46833152/Kaplan_Norton_Balanced_Scorecard_-_3_articles.pdf#page=50.

Kaplowitz, N. (1990). National self-images, perception of enemies, and conflict strategies: Psychopolitical dimensions of international relations. *Political Psychology, 11*(1), 39-82. doi:10.2307/3791515.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*(4), 681-706. doi: 10.1037/0022-3514.65.4.681.

Kelly, D., & Roedder, E. (2008). Racial cognition and the ethics of implicit bias. *Philosophy Compass, 3*, 522–540. doi: 10.1111/j.1747-9991.2008.00138.x.

Kiely, K. (2014). Cognitive Function. In A. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. doi: 10.1007/978-94-007-0753-5_426). Dordrecht: Springer.

Kiesler, S., & Sproull, L. (1982). Managerial response to changing environment: Perspectives on problem sensing from social cognition. *Administrative Science Quarterly, 27*, 548-570. doi: 10.2307/2392530

Klein, H. E., & Linneman, R. E. (1981). The use of scenarios in corporate planning—eight case histories. *Long Range Planning, 14*(5), 69-77.

Kloss, L. (1999). The suitability and application of scenario planning for national professional associations. *Nonprofit Management & Leadership, 10*(1), 71-83. doi: 10.1002/nml.10106.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118. doi: 10.1037/0278-7393.6.2.107.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118. doi: 10.1037/0278-7393.6.2.107.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variations in a juror reasoning task. *Psychological Science, 5*, 289-296. doi: 10.1111/j.1467-9280.1994.tb00628.x.

Kuhn, K. M., & Sniezek, J. A. (1996). Confidence and uncertainty in judgmental forecasting: Differential effects of scenario presentation. *Journal of Behavioral Decision Making*, 231-247. doi: 10.1002/(SICI)1099-0771(199612)9:4<231::AID-BDM240>3.0.CO;2-L

Lamm, H., & Trommsdorff, G. (1973). Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review. *European Journal of Social Psychology, 3*(4), 361-388. doi: 10.1002/ejsp.2420030402.

Lazarus, R. S. (1999). *Stress and Emotion.* New York: Springer.

Lee, C. S., & Therriault, D. J. (2013). The cognitive underpinnings of creative thought: A latent variable analysis exploring the roles of intelligence and working memory in three creative thinking processes. Intelligence. *41*(5), 306-320. doi: 10.1016/j.intell.2013.04.008.

Lee, C., Tinsley, C. H., & Bobko, P. (2002). An investigation of the antecedents and consequences of group-level confidence 1. *Journal of Applied Social Psychology, 32*(8), 1628-1652. doi: 10.1111/j.1559-1816.2002.tb02766.x.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*(2), 149-188. doi: 10.1006/obhd.1998.2804.

Lichtenstein, S., & Fischoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*, 159-183. doi: 10.1016/0030-5073(77)90001-0.

Lindgren, M., & Bandhold, H. (2003). *Scenario Planning: The Link Between Future and Strategy.* Houndmills, UK: Palgrave Macmillan.

Lindgren, M., & Bandhold, H. (2009). *Scenario Planning - Revised and Updated: The Link Between Future and Strategy.* Spring Link. doi: 10.1057/9780230233584.

Linneman, R. E., & Klein, H. E. (1979). The use of multiple scenarios by US industrial companies. *Long Range Planning, 12*, 83-90. doi: 10.1016/0024-6301(79)90034-7.

Lintern, G. (2007). What is a cognitive system? *Fourteenth International Symposium on Aviation Psychology*, (pp. 398-402. Retrieved from https://corescholar.libraries.wright.edu/isap_2007/68). Dayton, OH.

Lowe, E. (2006). Endurantism versus perdurantism and the nature of time. *Rivista Di Filosofia Neo-Scolastica, 98*(4), 713-727. Retrieved from http://www.jstor.org/stable/43063748.

Lox, C. L. (1992). Perceived threat as a cognitive component of state anxiety and confidence. *Perceptual and Motor Skills, 75*(3), 1092-1094. doi: 10.2466/pms.1992.75.3f.1092.

MacKay, B., & McKiernan, P. (2010). Creativity and dysfunction in strategic processes: The case of scenario planning. *Futures, 42*(4), 271-281. doi: 10.1016/j.futures.2009.11.013.

Mackay, B., & McKiernan, P. (2018). *Scenario Thinking: A Historical Evolution of Strategic Foresight.* Cambridge: Cambridge University Press.

MacKay, B., & Tambeau, P. (2013). A structuration approach to scenario praxis. *Technological Forecasting and Social Change, 80*(4), 673-686. doi: 10.1016/j.techfore.2012.06.003.

MacKay, R. B., & McKiernan, P. (2004). The role of hindsight in foresight: refining strategic reasoning. *Futures, 36*, 161-179. doi:10.1016/S0016-3287(03)00147-2.

Madva, A., & Brownstein, M. (2018). Stereotypes, prejudice, and the taxonomy of the implicit social mind. *Noûs, 52*, 611–644. doi: 10.1111/nous.12182.

Malenka, D. J., Baron, J. A., Johansen, S., Wahrenberger, J. W., & Ross, J. M. (1993). The framing effect of relative and absolute risk. *Journal of General Internal Medicine, 8*(10), 543-548. doi: 10.1007/BF02599636.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs, 50*, 629–658. doi: 10.1111/nous.12089.

Marron, T. R., & Faust, M. (2018). Free association, divergent thinking, and creativity: Cognitive and neural perspectives. In O. V. R. E. Jung (Ed.), *The Cambridge Handbook of the Neuroscience of Creativity* (pp. 261-280). Cambridge University Press.

Marsick, V. J., & Watkins, K. (2003). Demonstrating the value of an organization's learning culture: the dimensions of the learning organization questionnair. *Advances in Developing Human Resources, 5*(2), 132 - 151. doi: 10.1177/1523422303005002002.

Martelli, A. (2001). Scenario building and scenario planning: State of the art and prospects of evolution. *Futures Research Quarterly, 17*, 57–70. doi: 10.1108/14636680710727516.

Mason, R. O. (1969). A dialectical approach to strategic planning. *Management Science, 15*, B403-B414. doi: 10.1287/mnsc.15.8.B403.

Mason, R. O., & Mitroff, I. I. (1981). *Challenging strategic planning assumptions* (1st ed.). New York: Wiley-Interscience.

Mayor, L. H., & Coleman, R. (2012). Source credibility and evidence format: Examining the effectiveness of HIV/AIDS messages for young African Americans. *Journal of Health Communication, 17*(515-531. doi: 10.1080/10810730.2011.635771).

McKiernan, P. (2017). Prospective thinking; scenario planning meets neuroscience. *echnological Forecasting and Social Change, 124*, 66-76. doi: 10.1016/j.techfore.2016.10.069.

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: a definition and theory. *Journal of Community Psychology, 14*, 6–23. doi: 10.1002/1520-6629(198601)14:1<6::AID-JCOP2290140103>3.0.CO;2-I

Mednick, S. A. (1962). The associative basis for the creative process. *Psychological Bulletin, 69*, 220–232. doi: 10.1037/h0048850.

Meissner, P., & Wulf, T. (2013). Cognitive benefits of scenario planning: its impact on biases and decision quality. *Technological Forecasting & Social Change, 80*, 801-814. doi: 10.1016/j.techfore.2012.09.011

Melchior, M., Caspi, A., Milne, B. J., Danese, A., Poulton, R., & Moffitt, T. E. (2007). Work stress precipitates depression and anxiety in young, working women and men. *Psychological Medicine, 37*(8), 1119-1129. doi: 10.1017/S0033291707000414.

Metzger, M. J., & Flanagin, A. J. (2015). Psychological approaches to credibility assessment online. In S. S. Sundar (Ed.), *The Handbook of the Psychology of Communication Technology* (1st ed.). New York: John Wiley & Sons, Inc.

Minichiello, V., Aroni, R., Timewell, E., & Alexander, L. (1990). *In-depth Interviewing: Researching people.* Hong Kong: Longman Cheshire Pty Limited.

Mitchell, R. B., Tydeman, J., & Georgiades, J. (1979). Structuring the future-application of a scenario-generation procedure. *Technological Forecasting and Social Change, 14*, 409-428. doi: 10.1016/0040-1625(79)90038-6

Mitroff, I. I. (1982). Dialectic squared: A fundamental difference in perception of the meanings of some key concepts in social science. *Decision Sciences, 13*(2), 222-224. doi: 10.1111/j.1540-5915.1982.tb00144.x.

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6(6): e1000097. doi:10.1371/journal.pmed1000097

Montibeller, G., & Belton, V. (2006). Causal maps and the evaluation of decision options - A review. *Journal of the Operational Research Society, 57*, 779-791. doi: 10.1057/palgrave.jors.2602214.

Moretto, L. (2012). *Exploring the Nature of the Relationship Between Decision-Driven Scenarios and Decision Confidence in High-Velocity Environments.* Washington, D.C.: ProQuest.

Moyer, K. (1996). Scenario planning at British Airways-A case study. *Long Range Planning, 29*(2), 172-181. doi: 10.1016/0024-6301(96)00005-2.

Mumford, M. D. (2003). Where have we been, where are we going? Taking stock in creativity research. *Creativity Research Journal, 15*(2-3), 107-120. doi: 10.1080/10400419.2003.9651403.

Neale, M. A., & Bazerman, M. H. (1985). The effects of framing and negotiator overconfidence on bargaining behaviors and outcomes. *Academy of Management Journal, 28*, 34–49. doi: 10.5465/256060.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220. doi: 10.1037/1089-2680.2.2.175.

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs and real-estate: An anchoring and adjust perspective in property pricing decisions. *Organisational Behaviour and Human Decision Processes*, 84-97. Retrieved from http://www.communicationcache.com/uploads/1/0/8/8/10887248/experts_amateurs_and_real_estate-_an_anchoring-and-adjustment_perspective_on_property_pricing_decisions.pdf.

Olson, M. A., & Fazio, R. H. (2008). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 19–63). New York, NY: Psychology Press.

Önkal, D., Sayim, K. Z., & Gönül, M. S. (2013). Scenarios as channels of forecast advice. *Technological Forecasting & Social Change, 80*, 772-788. doi: 10.1016/j.techfore.2012.08.015.

Pagani, M. (2009). Roadmapping 3G mobile TV: strategic thinking and scenario planning through repeated cross-impact handling. *Technological Forecasting and Social Change, 76*(3), 382-395. doi: 10.1016/j.techfore.2008.07.003.

Peirce, C. S. (1935). *Collected Papers of Charles Sanders Peirce, Volumes V and VI: Pragmatism and Pragmaticism and Scientific Metaphysics.* (C. Hartshorne, & P. Weiss, Eds.) Cambridge: Harvard University Press.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The meta-cognitive model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition, 25*, 657–686. doi: 10.1521/soco.2007.25.5.657.

Phadnis, S., Caplice, C., Sheffi, Y., & Singh, M. (2014). Effect of scenario planning on field experts' judgment of long-range investment decisions. *Strategic Management Journal*, 1401-1411. doi: 10.1002/smj.2293

Pighin, S., Byrne, R., Ferrante, D., Gonzalez, M., & Girotto, V. (2011). Counterfactual thoughts about experienced, observed and narrated events. *Thinking & Reasoning, 17*(2), 197–211. doi: 10.1080/13546783.2011. 561598.

Poon, J. C., Au, A. C., Tong, T. M., & Lau, S. (2014). The feasibility of enhancement of knowledge and self-confidence in creativity: A pilot study of a three-hour SCAMPER workshop on secondary students. *Thinking Skills and Creativity, 14*, 32-40. doi: 10.1016/j.tsc.2014.06.006.

Porter, M. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance.* New York: Free Press.

Proctor, R. W., & Capaldi, E. J. (Eds.). (2012). *Psychology of Science: Implicit and Explicit Processes.* New York: Oxford University Press.

Qualtrics. (2019). www.qualtrics.com. Available at: https://www.qualtrics.com. Provo, Utah, USA.

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/}.

Ram, C., & Montibeller, G. (2013). Exploring the impact of evaluating strategic options in a scenario-based multi-criteria framework. *Technological Forecasting and Social Change, 80*(4), 657-672. doi: 10.1016/j.techfore.2012.09.019

Ramírez, R., & Wilkinson, A. (2014). Re-thinking the 2X2 scenario method: Grid or frames? *Technological Forecasting and Social Change, 86*, 254-264. doi: 10.1016/j.techfore.2013.10.020

Rauth, I., Köppen, E., Jobst, B., & Meinel, C. (2010). Design thinking: An educational model towards creative confidence. *Proceedings of the 1st international conference on design creativity (ICDC 2010)*, Retrieved from https://www.designsociety.org/publication/30204/DS+66-2%3A+Proceedings+of+the+1st+International+Conference+on+Design+Creativity+%28ICDC+2010%29.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*(1), 1–75. doi: 10.1016/1041-6080(95)90031-4.

Ringland, G. (1998). *Scenario Planning: Managing for the Future.* Chichester: Wiley.

Rivers, A. M., & Sherman, J. (2018). *Pre-print*, Experimental Design and the Reliability of Priming Effects: Reconsidering the" Train Wreck". doi: 10.31234/osf.io/r7pd3

Rowe, G. (2010). *Assessment of the COT uncertainty framework from a social science perspective: A theoretical evaluation.* Report to the Food Standards Agency.

Rumsfeld, D. (2002, February 12). DoD News Briefing - Secretary Rumsfeld and Gen. Myers. (J. Miklaszewski, Interviewer)

Runco, M. A. (2014a). Biological perspectives on creativity. In *Creativity: Theories and Themes: Research, Development, and Practice* (2nd ed., pp. 69-108). London: Academic Press.

Runco, M. A. (2014b). *Creativity: Theories and Themes: Research, Development, and Practice* (2nd ed.). San Diego: Academic Press.

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66-75. doi: 10.1080/10400419.2012.652929.

Runco, M. A., & Albert, R. S. (1985). The reliability and validity of ideational originality in the divergent thinking of academically gifted and nongifted children. *Educational and Psychological Measurement, 45*(3), 483-501. doi: 10.1177/0013164485045003060.

Saebi, T., Lien, L., & Foss, N. J. (2017). What drives business model adaptation? The impact of opportunities, threats and strategic orientation. *Long Range Planning, 50*(5), 567-581. doi: 10.1016/j.lrp.2016.06.006.

Sanna, L. J. (1996). Defensive pessimism, optimism, and simulating alternatives: Some ups and downs or prefactual and counterfactual thinking. *Journal of Personality and Social Psychology, 71*(5), 1017–1020. doi: 10.1037/0022-3514.71.5.1020.

Schacter, D. L., & Addis, D. R. (2009). Remembering the past to imagine the future: A cognitive neuroscience perspective. *Military Psychology, 21*(1), S108-S112. doi: 10.1080/08995600802554748

Schacter, D., Benoit, R., De Brigard, F., & Szpunar, K. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of Learning and Memory, 117*, 14–21. doi: 10.1016/j.nlm.2013.12.008.

Schnaars, S. P. (1987). How to develop and use scenarios. *Long Range Planning, 20*(1), 105-114. doi: 10.1016/0024-6301(87)90038-0

Schnaars, S. P., & Topol, M. T. (1987). The use of multiple scenarios in sales forecasting: An empirical test. *International Journals of Forecasting, 3*, 405-419. doi: 10.1016/0169-2070(87)90033-1

Schoemaker, P. H. (1991). When and how to use scenario planning: A heuristic approach with illustration. *Journal of Forecasting, 10*, 549-564. doi: 10.1002/for.3980100602.

Schoemaker, P. J. (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic Management Journal, 14*, 193-213. doi: 10.1002/smj.4250140304

Schoemaker, P. J. (1995). Scenario planning: A tool for strategic thinking. *MIT Sloan Management Review, 36*(2), 25-50. Retrieved from https://www.ftms.edu.my/images/Document/MOD001074%20-%20Strategic%20Management%20Analysis/WK4_SR_MOD001074_Schoemaker_1995.pdf

Schoemaker, P. J. (2004). Forecasting and scenario planning: The challenges of uncertainty and complexity. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 247–296). Oxford: Blackwell.

Schoemaker, P. J., & Van der Heijden, C. (1992). Integrating scenarios into strategic planning at Royal Dutch/Shell. *Planning Review, 20*, 41-46. doi: 10.1108/eb054360

Schulthesis, O., & Brunstein, J. C. (Eds.). (2010). *Implicit Motives.* New York: Oxford University Press.

Schwartz, P. (1991). *The Art of the Long View: Scenario Planning, Protecting Your Company Against an Uncertain World.* New York: Doubleday.

Schweiger, D. M., Snadberg, W. R., & Ragan, J. W. (1986). Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, devil's advocacy, and consensus. *The Academy of Management Journal, 29*(1), 51-71. doi: 10.2307/255859.

Sedor, L. M. (2002). An explanation for unintentional optimism in analysts' earnings forecasts. *The Accounting Review, 77*(4), 731-753. doi: 10.2308/accr.2002.77.4.731.

Sherman, J. W., & Rivers, A. M. (2020). *Social priming: A dubious term.* Nature Publishing Group.

Shimizu, K. (2007). Prospect theory, behavioral theory, and the threat-rigidity thesis: Combinative effects on organizational decisions to divest formerly acquired units. *Academy of Management Journal, 50*(6), 1495-1514. doi: 10.5465/amj.2007.28226158.

Simon, H. A. (1972). Theories of bounded rationality. In C. B. McGuire, & R. Radner (Eds.), *Decision and organization: A volume in honor of Jacob Marschak.* Amsterdam: North-Holland Publishing Company.

Simpson, D. G. (1992). Key lessons for adopting scenario planning in diversified companies. *Planning Review, 20*(3), 10-48. doi: 10.1108/eb054355

Singer, J. L. (1999). Imagination. In M. Runco, & S. Pritzker (Eds.), *Encyclopedia of Creativity* (pp. 13-25). San Diego: Academic Press.

Skinner, C., Strecher, V., & Hospers, H. (1994). Physicians' recommendations for mammography: Do tailored messages make a difference? *American Journal of Public Health, 84*, 43-49. doi: 10.2105/AJPH.84.1.43.

Smircich, L., & Stubbart, C. (1985). Strategic management in an enacted world. *Academy of Management Review, 10*, 724-738. doi: 10.5465/amr.1985.4279096

Smith, S. M. (1997). Fixation, incubation, and insight in memory and creative thinking. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The Creative Cognition Approach* (pp. 135-156). MA: Springer.

Sniezek, J. A. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes, 52*(1), 124-155. doi: 10.1016/0749-5978(92)90048-C .

Sniezek, J. A., & Henry, R. A. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes, 43*(1), 1-28. doi.org/10.1016/0749-5978(89)90055-1.

Soste, L., Wang, Q. J., Robertson, D., Chaffe, R., Handley, S., & Wei, Y. (2015). Engendering stakeholder ownership in scenario planning. *Technological Forecasting and Social Change, 91*, 250-263. doi: 10.1016/j.techfore.2014.03.002

Spaniol, M. J., & Rowland, N. J. (2018a). Defining scenario. *Futures & Foresight Science, 1*(1), 1-13. doi: 10.1002/ffo2.3

Spaniol, M. J., & Rowland, N. J. (2018b). The scenario planning paradox. *Futures, 95*, 33-43. doi: 10.1016/j.futures.2017.09.006

Spreitzer, G. M. (1996). Social structural characteristics of psychological empowerment. *Academy Management Journal, 39*, 483–504. doi: 10.5465/256789

Staw, B., Sandelands, L., & Dutton, J. (1981). Threat rigidity effects in organizational behaviour: A multilevel analysis. *Administrative Science Quarterly, 26*, 501-524. doi: 10.2307/2392337.

Stirling, A. (2006). Analysis, participation and power: Justification and closure in participatory multi-criteria analysis. *Land Use Policy, 23*, 95–107. doi: 10.1016/j.landusepol.2004.08.010

Sukhera, J., & Watling, C. (2017). A framework for integrating implicit bias recognition into health professions education. *Academic Medicine, 93*(1), 35-40. doi: 10.1097/ACM.0000000000001819.

Sunstein, C. R. (2007). Group polarization and 12 angry men. *Negotiate Journal, 23*, 443-447. doi: 10.1111/j.1571-9979.2007.00155.x.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable* (2nd ed.). New York: Random House.

Tapinos, E. (2013). Scenario planning at business unit level. *Futures, 47*, 17-27. doi: 10.1016/j.futures.2012.11.009.

Tetlock, P. E. (2005). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52*(4), 700–709. doi: 10.1037/0022-3514.52.4.700.

Thomas, C. W. (1994). Learning from imagining the years ahead. *Planning Review, 22*(3), 6-10. doi: 10.1108/eb054462.

Tooby, J., Cosmides, L., & Barrett, H. C. (2005). Resolving the debate on innate ideas. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind: Structure and Content* (pp. 305-337). New York: Oxford University Press.

Torrance, E. P. (1966). *The Torrance Tests of Creative Thinking-Norms-Technical Manual Research Edition-Verbal Tests, Forms A and B-Figural Tests, Forms A and B.* Princeton: Personnel Press.

Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*(3), 411–435. doi: 10.1016/0010-0285(92)90013-R.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, New Series, 185*(4157), 1124-1131. Retrieved from https://www.jstor.org/stable/1738360.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 221*(4481), 453-458. doi: 10.1126/science.7455683.

Tversky, A., & Kahneman, D. (1982). Causal schemas in judgments under uncertainty. In D. Kahaneman, A. Tversky, & P. Slovic (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 117-128). Cambridge: Cambridge University Press.

Ülengin, F., Kabak, Ö., Önsel, S., Ülengin, B., & Aktaş, E. (2010). A problem structuring model for analysing transport and environment interactions. CiteSeerx pre-print, doi: 10.1.1.425.2663.

Van de Riet, O., Aazami, O., & Van Rhee, C. G. (2008). Scenario analysis and the adaptive approach: superfluous or underused in transport infrastructure planning? *2008 First International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future (INFRA)* (pp. 1-6). Rotterdam: IEEE.

Van der Heijden, K. (1994). Probabilistic planning and scenario planning. In G. Wright, & P. Ayton (Eds.), *Subjective Probability.* Chichester: Wiley.

Van der Heijden, K. (1997). *Scenarios: The Art of Strategic Conversation.* New York: John Wiley.

Van der Heijden, K., Bradfield, R., Burt, G., Cairns, G., & Wright, G. (2002). *The Sixth Sense: Accelerating Organizational Learning with Scenarios.* West Sussex: John Wiley & Sons.

Van Dessel, P. G. (2019). Does explaining social behavior require multiple memory systems? *Trends in Cognitive Sciences, 23*(5), 368-369. doi: 10.1016/j.tics.2019.02.001.

Van Notten, P. W., Rotmans, J., Van Asselt, M. B., & Rothman, D. S. (2003). An updated scenario typology. *Futures, 35*, 423-443. doi: 10.1016/S0016-3287(02)00090-3

Vanston Jr, J. H., Frisbie, W. P., Lopreato, S. C., & Boston Jr, D. L. (1977). Alternate scenario planning. *Technological Forecasting and Social Change, 10*(2), 159-180. doi: 10.1016/0040-1625(77)90043-9.

Vanston Jr., J. H., Frisbie, W. P., & Poston, D. L. (1975). *Technology Assessment of Portable Energy RDT & P.* Center for Engergy Studies. Austin: University of Texas.

Vanston, J. H., Frisbie, W. P., Lopreato, S. C., & Poston Jr., D. L. (1977). Alternate scenario planning. *Technological Forecasting and Social Change, 10*, 159-180. doi: 10.1016/0040-1625(77)90043-9

Volkery, A., & Riberio, T. (2009). Scenario planning in public policy: Understanding use, impacts and the role of institutional context factors. *Technological Forecasting & Social Change, 76*, 1198-1207. doi: 10.1016/j.techfore.2009.07.009

Wack, P. (1985a). Scenarios: Shooting the rapids. *Harvard Business Review*, 139-150. Retrieved from https://hbr.org/1985/11/scenarios-shooting-the-rapids

Wack, P. (1985b). Scenarios: Uncharted waters ahead. *Harvard Business Review*, 73-89. Retrieved from https://hbr.org/1985/09/scenarios-uncharted-waters-ahead

Wagenaar, W. A., & Keren, G. B. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. In E. Hollnagel, G. Mancini, & D. Woods (Eds.), *Intelligent Decision Support in Process Environments. NATO ASI Series (Series F: Computer and Systems Sciences)* (Vol. 21, pp. 87-103). Berlin, Heidelberg: Springer.

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children* (1st ed.). New York: Holt, Rinehart, & Winston.

Ward, E., & Schriefer, A. E. (1998). *Dynamic Scenarios: Systems Thinking Meets Scenario Planning.* (L. Randall, & R. Randall, Eds.) New York: John Wiley & Sons, Inc.

Wasserman, R. (2016). Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition. *Theories of persistence. 173*, pp. 243-250. Proceedings of the 41st Oberlin Colloquium in Philosophy: Metaphysics. Retrieved from http://www.jstor.org/stable/24704008.

Weick, K. E. (1995). The nature of sensemaking. In *Sense Making in Organizations.* London: Sage.

Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin, 97*(1), 74-84. doi: 10.1037/0033-2909.97.1.74.

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin, 142*(5), 472–497. doi: 10.1037/bul0000030.

Whittington, R. (2006). Completing the practice turn in strategy research. *Organization Studies, 27*(5), 613–634. doi: 10.1177/0170840606064101.

Wilson, I. (2000). From scenario thinking to strategic action. *Technological Forecasting and Social Change, 65*(3), 23-29. doi: 10.1016/S0040-1625(99)00122-5.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General, 125*(4), 387–402. doi: 10.1037/0096-3445.125.4.387.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101–126. doi: 10.1037/0033-295X.107.1.101.

Wright, G. (2014). *Exploring the international business environment.* University of Strathclyde. Glasgow: Strathclyde Business School.

Wright, G., & Bradfield, R. (2015). Exploring the International Business Environment. The University of Strathclyde Business School.

Wright, G., & Goodwin, P. (2002). Eliminating the framing bias by using simple instructions to 'think harder' and respondents with managerial experience: Comment on 'breaking the frame'. *Strategic Management Journal, 23*(11), 1059-1067. doi: 10.1002/smj.265.

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: enhancing the scenario method. *International Journal of Forecasting, 25*, 813–825. doi:10.1016/j.ijforecast.2009.05.019.

Wright, G., & Rowe, G. (2011). Group-based judgmental forecasting : An integration of extant knowledge and the development of priorities for a new research agenda. *International Journal of Forecasting, 27*, 1-13. doi: 10.1016/j.ijforecast.2010.05.012

Wright, G., Bradfield, R., & Cairns, G. (2013). Does the intuitive logics method – and its recent enhancements – produce "effective" scenarios? *Technological Forecasting & Social Change, 80*, 631–642. doi: 10.1016/j.techfore.2012.09.003

Wright, G., Van der Heijden, K., Burt, G., Bradfield, R., & Cairns, G. (2008). Scenario planning interventions in organizations: An analysis of the causes of success and failure. *Futures, 40*(3), 218-236. doi: 10.1016/j.futures.2007.08.019.

Yaniv, I. (2011). Group diversity and decision quality: Amplification and attenuation of the framing effect. *International Journal of Forecasting, 27*(1), 41-49. doi: 10.1016/j.ijforecast.2010.05.009.

Zarnoth, P., & Sniezek, J. A. (1997). The social influence of confidence in group decision making. *Journal of Experimental Social Psychology, 33*(4), 345-366. doi: 10.1006/jesp.1997.1326.

Zhang, Y., & Wildemuth, B. M. (2009). Unstructured interviews. In B. M. Wildemuth (Ed.), *Applications of Social Research Methods to Questions in Information and Library Science* (pp. 222-231). Denver, CO: Libraries Unlimited.

Zizek, S. (2004, May 21). *What Rumsfeld Doesn't Know That He Knows About Abu Ghraib.* Retrieved December 2016, from Lacan Dot Com: http://www.lacan.com/zizekrumsfeld.htm

Zuckerman, M., Eghrari, H., & Lambrecht, M. R. (1996). Attributions as inferences and explanations: Conjuction effects. *Journal of Personality and Social Psychology, 51*, 1144-1153. doi: 10.1037/0022-3514.51.6.1144.

Appendix A

Appendix A presents all the business vignettes tested for validation (see Chapter 6), and presented in the six studies reported in Chapters 7-9. There are three organisations profiled across the business vignettes. Vignette 1 profiles the University of Strathclyde (Scottish and UAE campuses). Vignette 2 profiles the local city zoo (The Edinburgh Zoo and The Emirates Park Zoo). Vignette 3 profiles a fictional organisation (Tammuz Confectionary Company).

Four different structures of Vignette 2 were used to test for priming validation. Structure I is the longest vignette, divided into two sections and heavily segmented. Structure II presents the same information as structure I, divided into two sections, but in paragraph form, without segments. Section 1 presents the framing manipulation (either high-threat or low-threat), goals, and timeline. Section 2 presents the background information for the organisation. Structure III is the shortest vignette and presents only two sentences from section 1. Structure IV is a modification of the previous three options and presents a full section 1 and a shortened section 2 (adapted from structure II), but in reverse order, where the background information is given first and the framing manipulation is given last.

This appendix is divided into three sub-sections – Vignette 1, Vignette 2, and Vignette 3. Each sub-section presents the different structures developed for that vignette. Vignette 1 was drafted into structure IV. Vignette 2 was drafted into structures 1-4. Vignette 3 was drafted into structures 4. Each structure presents the two different framing variations (high-threat and low-threat). In the following sub-sections, all vignettes are presented within a bordered table to show the exact information participants received, and all framing manipulations are **<u>bold and underlined</u>** within each vignette.

A.1 **Vignette 1 – University of Strathclyde**

Vignette 1 was used for participant samples in both the UK and UAE. The same two structures (high-threat and low-threat) were supplied to all participants. Vignette 1 was developed after priming validation tests confirmed that structure IV was the optimal structure for delivering the framing manipulation and used in the framing validation tests. Therefore, vignette 1 was only developed into structure IV for use in experimental studies.

A.1.1 ***Structure IV – High-threat (UK/UAE)***

(Section 1)

UNIVERSITY OF STRATHCLYDE

The University of Strathclyde is in the UK's top 20 research institutions, 7th for spin-out company creation, and 24th for employment post-graduation. Strathclyde has been partnering with international and local companies through their Research & Knowledge Exchange Services since 2013. The university has satellite locations in Malaysia, Switzerland, Greece, Bahrain, Dubai, Oman, and Abu Dhabi, and offers around 400 degree courses within business, engineering, science and humanities & social sciences. Strathclyde's main campus hosts just over 2,000 international students.

(Section 2)

GOALS

The University of Strathclyde is **under threat of departmental closures** due to **decreasing enrolment rates** since 2013, and **pressures from national standards** concerning accreditation and equality (Association to Advance Collegiate Schools of Business, European Foundation for Management Development, and Association of MBAs). Satellite campuses are either **threatening to close or become consolidated** into fewer colleges, thus **decreasing the number of available courses and degree paths**. Strathclyde's goals are to increase international student enrolment by 20%, and improve sustainable alumni career placements out-with the university.

A.1.2 ***Structure IV – Low-threat (UK/UAE)***

(Section 1)

UNIVERSITY OF STRATHCLYDE

The University of Strathclyde is in the UK's top 20 research institutions, 7th for spin-out company creation, and 24th for employment post-graduation. Strathclyde has been partnering with international and local companies through their Research & Knowledge Exchange Services since 2013. The university has satellite locations in Malaysia, Switzerland, Greece, Bahrain, Dubai, Oman, and Abu Dhabi, and offers around 400 degree courses within business, engineering, science and humanities & social sciences. Strathclyde's main campus hosts just over 2,000 international students.

(Section 2)

GOALS

The University of Strathclyde has been **holding steady enrolment rates** since 2013, **and holds a triple accreditation** through the Strathclyde Business School (Association to Advance Collegiate Schools of Business, European Foundation for Management Development, and Association of MBAs). **Satellite campuses and degree paths are sustaining**. Strathclyde's goals are to increase international student enrolment by 20%, and improve sustainable alumni career placements out-with the university.

A.2 **Vignette 2 – The Local Zoo**

Vignette 2 was used as the exemplar that was tested for priming and validation measures (Chapter 6). Therefore, structures 1-4 were developed for vignette 2 as part of the pre-testing efforts. Vignette 2 was used for participant samples in both the UK and UAE. The UK versions are presented first, followed by the UAE versions.

A.2.1 ***Structure I – High-threat (UK)***

(Section 1)

EDINBURGH ZOO

GOALS

Edinburgh Zoo is **under threat of closure** due to a **decreasing customer base** and **pressures from international animal rights councils**. The Zoo's goals are to increase their customer base by 20% by 2020 and increase their conservation efforts for endangered and threatened species.

(Section 2)

SOCIETY

The Edinburgh Zoo has a membership base of 13,000.

EDUCATION

Edinburgh Zoo has in-house education programmes that have reached almost 25,000 school children and the public at large to date. There are seasonal teaching and learning programmes throughout the year. Each programme is designed to present facts about the species' natural habitat, health, life style and cycle, dangers, and, when possible, to allow people to interact with the species.

The top education programmes:

- Beyond the Bees
- Meet the Meerkats
- Organised school visits
- Flora education programmes

TECHNOLOGY

Edinburgh Zoo has over 10,000 Facebook followers and 15,000 Twitter followers. The Zoo circulates e-newsletters to over 10,000 people quarterly, and is in the process of installing high-resolution cameras in 5 animal enclosures. This effort is to allow a 24-hour, live-stream online viewing of the animals.

INDUSTRY

Edinburgh Zoo receives funding from several sources. According to financial reports from 2013, the smallest percentage of funding came from government sources. The largest percentage came from park revenue.

EMPLOYMENT

The Edinburgh Zoo has nearly 200 employees, including scientists and doctors. A varying number of seasonal volunteers fill the remainder of the jobs at Edinburgh Zoo. This means that volunteers only help at the zoo when a demand is perceived, which can vary from season to season.

RESOURCES

The Edinburgh Zoo is on 82 acres of land. Animals are either rescued from dangerous environments or purchased from other zoos or wildlife reserves. The food required for the individual diets of the animals is grown within the zoo, purchased from local suppliers, and imported from the animal's natural habitat through third party suppliers.

DEMOGRAPHICS

PEOPLE

Scotland has a population of 5.3 million. 80,000 visitors are expected in 2015.

ANIMALS

Edinburgh Zoo houses 300 different species and almost 1,000 animals total.

- 470 birds representing 100 different species
- Domestic animals span 16 different species
- Aquatic animals represent more than 50 different species
- Reptiles represent 90 species
- Large mammals (large cats, primates, elephants etc.) comprise the remaining species

PLANTS

Edinburgh Zoo has a flora programme. The programme was designed to educate visitors and school children about the various plants each animal requires, local plants of the country, and propagation of endangered species. The variety and number of species is unknown.

ECONOMICS

The annual costs for running the Edinburgh Zoo in 2013 was £10.5. Standard ticket price is £17. Edinburgh Zoo has various product-related marketing deals with outside corporations. Product branding based sponsorship is offered.

FINANCIAL

In 2013, the Edinburgh Zoo received funding from several sources. The following list shows the percentages of the annual income each group supplied.

- Independent societies (21%)
- Corporate sponsorship (20%)
- Government councils (14%)
- Revenue generated in the park (45%)
    - General admission, Featured exhibits admission, Product sales, Food sales

ENVIRONMENT

The Edinburgh Zoo is located inside the city limits. It houses several enclosures, water attractions, a veterinary hospital, and research facilities. Edinburgh Zoo also supplies shelter and food for abandoned and stray animals. There is an isolation centre for infectious animals. A research facility runs together with the Edinburgh Zoo. University students and researchers conduct observational, sample, and data related projects on site. National centres that oversee disease control, species preservations, fertility, and well-being are all affiliated with the Edinburgh Zoo.

POLITICS

Scotland is a multi-party system.

- Scottish National Party (centre-left, pro-independence)
- Scottish Labour Party (centre-left, social democratic)
- Scottish Conservative Party (centre-right, conservative)
- Scottish Liberal Democrats (centrist, social liberal)
- Scottish Green Party (left-wing, eco socialist)

The loaning or purchasing of rare/endangered species from another government is sometimes used as a means to aid in positive relations between nations.

ENERGY

The Edinburgh Zoo's energy is supplied from several sources.

- Fossil fuels (36%)
- Natural gas (33%)
- Nuclear power (20%)
- Alternative (11%)
    - Hydro or renewable sources

Edinburgh Zoo hosts "Green Events" at various times during the year. These are open to the patrons of the Edinburgh Zoo. Green Events present information on alternative energy sources and plant life, as well as opportunities for interactive participation.

RELIGION

Scotland has a majority non-religious identifying population.

- No religion (43.6%)
- Church of Scotland (32.4%)
- Roman Catholic (15.9%)
- Islam (1.4%)
- Hinduism (0.3%)
- Buddhism (0.2%)
- Sikhism (0.2%)

A.2.2 ***Structure I – Low-threat (UK)***

(Section 1)

EDINBURGH ZOO

GOALS

Edinburgh Zoo has been **holding a steady customer base** since 2013, and **met the international requirements concerning animal welfare** as set by **international animal rights councils**. The Zoo's goals are to increase their customer base by 20% by 2020 and increase their conservation efforts for endangered and threatened species.

(Section 2)

SOCIETY

The Edinburgh Zoo has a membership base of 13,000.

EDUCATION

Edinburgh Zoo has in-house education programmes that have reached almost 25,000 school children and the public at large to date. There are seasonal teaching and learning programmes throughout the year. Each programme is designed to present facts about the species' natural habitat, health, life style and cycle, dangers, and, when possible, to allow people to interact with the species.

The top education programmes:

- Beyond the Bees
- Meet the Meerkats
- Organised school visits
- Flora education programmes

TECHNOLOGY

Edinburgh Zoo has over 10,000 Facebook followers and 15,000 Twitter followers. The Zoo circulates e-newsletters to over 10,000 people quarterly, and is in the process of installing high-resolution cameras in 5 animal enclosures. This effort is to allow a 24-hour, live-stream online viewing of the animals.

INDUSTRY

Edinburgh Zoo receives funding from several sources. According to financial reports from 2013, the smallest percentage of funding came from government sources. The largest percentage came from park revenue.

EMPLOYMENT

The Edinburgh Zoo has nearly 200 employees, including scientists and doctors. A varying number of seasonal volunteers fill the remainder of the jobs at Edinburgh Zoo. This means that volunteers only help at the zoo when a demand is perceived, which can vary from season to season.

RESOURCES

The Edinburgh Zoo is on 82 acres of land. Animals are either rescued from dangerous environments or purchased from other zoos or wildlife reserves. The food required for the individual diets of the animals is grown within the zoo, purchased from local suppliers, and imported from the animal's natural habitat through third party suppliers.

DEMOGRAPHICS

PEOPLE

Scotland has a population of 5.3 million. 80,000 visitors are expected in 2015.

ANIMALS

Edinburgh Zoo houses 300 different species and almost 1,000 animals total.

- 470 birds representing 100 different species
- Domestic animals span 16 different species
- Aquatic animals represent more than 50 different species
- Reptiles represent 90 species
- Large mammals (large cats, primates, elephants etc.) comprise the remaining species

PLANTS

Edinburgh Zoo has a flora programme. The programme was designed to educate visitors and school children about the various plants each animal requires, local plants of the country, and propagation of endangered species. The variety and number of species is unknown.

ECONOMICS

The annual costs for running the Edinburgh Zoo in 2013 was £10.5. Standard ticket price is £17. Edinburgh Zoo has various product-related marketing deals with outside corporations. Product branding based sponsorship is offered.

FINANCIAL

In 2013, the Edinburgh Zoo received funding from several sources. The following list shows the percentages of the annual income each group supplied.

- Independent societies (21%)
- Corporate sponsorship (20%)
- Government councils (14%)
- Revenue generated in the park (45%)
  - General admission, Featured exhibits admission, Product sales, Food sales

ENVIRONMENT

The Edinburgh Zoo is located inside the city limits. It houses several enclosures, water attractions, a veterinary hospital, and research facilities. Edinburgh Zoo also supplies shelter and food for abandoned and stray animals. There is an isolation centre for infectious animals.

A research facility runs together with the Edinburgh Zoo. University students and researchers conduct observational, sample, and data related projects on site. National centres that oversee disease control, species preservations, fertility, and well-being are all affiliated with the Edinburgh Zoo.

POLITICS

Scotland is a multi-party system.

- Scottish National Party (centre-left, pro-independence)
- Scottish Labour Party (centre-left, social democratic)
- Scottish Conservative Party (centre-right, conservative)
- Scottish Liberal Democrats (centrist, social liberal)
- Scottish Green Party (left-wing, eco socialist)

The loaning or purchasing of rare/endangered species from another government is sometimes used as a means to aid in positive relations between nations.

ENERGY

The Edinburgh Zoo's energy is supplied from several sources.

- Fossil fuels (36%)
- Natural gas (33%)
- Nuclear power (20%)
- Alternative (11%)
    - Hydro or renewable sources

Edinburgh Zoo hosts "Green Events" at various times during the year. These are open to the patrons of the Edinburgh Zoo. Green Events present information on alternative energy sources and plant life, as well as opportunities for interactive participation.

RELIGION

Scotland has a majority non-religious identifying population.

- No religion (43.6%)
- Church of Scotland (32.4%)
- Roman Catholic (15.9%)
- Islam (1.4%)
- Hinduism (0.3%)
- Buddhism (0.2%)
- Sikhism (0.2%)

A.2.3 ***Structure II – High-threat (UK)***

(Section 1)

EDINBURGH ZOO

GOALS
Edinburgh Zoo is **<u>under threat of closure</u>** due to a **<u>decreasing customer base</u>** and **<u>pressures from international animal rights councils</u>**. The Zoo's goals are to increase their customer base by 20% by 2020 and increase their conservation efforts for endangered and threatened species.

(Section 2)

The Edinburgh Zoo has a membership base of 13,000. Edinburgh Zoo has in-house education programmes that have reached almost 25,000 school children and the public at large to date. There are seasonal teaching and learning programmes throughout the year. Each programme is designed to present facts about the species' natural habitat, health, life style and cycle, dangers, and, when possible, to allow people to interact with the species. The top education programmes: Beyond the Bees, Meet the Meerkats, Organised school visits, Flora education programmes.

Edinburgh Zoo has over 10,000 Facebook followers and 15,000 Twitter followers. The Zoo circulates e-newsletters to over 10,000 people quarterly, and is in the process of installing high-resolution cameras in 5 animal enclosures. This effort is to allow a 24-hour, live-stream online viewing of the animals.

Edinburgh Zoo receives funding from several sources. According to financial reports from 2013, the smallest percentage of funding came from government sources. The largest percentage came from park revenue. In 2013, the Edinburgh Zoo received funding from several sources. The following list shows the percentages of the annual income each group supplied: Independent societies (21%), Corporate sponsorship (20%), Government councils (14%), Revenue generated in the park (45%) - General admission, Featured exhibits admission, Product sales, Food sales.

The Edinburgh Zoo has nearly 200 employees, including scientists and doctors. A varying number of seasonal volunteers fill the remainder of the jobs at Edinburgh Zoo. This means that volunteers only help at the zoo when a demand is perceived, which can vary from season to season.

The Edinburgh Zoo is on 82 acres of land. Animals are either rescued from dangerous environments or purchased from other zoos or wildlife reserves. The food required for the individual diets of the animals is grown within the zoo, purchased from local suppliers, and imported from the animal's natural habitat through third party suppliers.

Scotland has a population of 5.3 million. 80,000 visitors are expected in 2015. Edinburgh Zoo houses 300 different species and almost 1,000 animals total: 470 birds representing 100 different species, Domestic animals span 16 different species, Aquatic animals represent more than 50 different species, Reptiles represent 90 species, Large mammals (large cats, primates, elephants etc.) comprise the remaining species.

Edinburgh Zoo has a flora programme. The programme was designed to educate visitors and school children about the various plants each animal requires, local plants of the country, and propagation of endangered species. The variety and number of species is unknown.

The annual costs for running the Edinburgh Zoo in 2013 was £10.5. Standard ticket price is £17. Edinburgh Zoo has various product-related marketing deals with outside corporations. Product branding based sponsorship is offered.

The Edinburgh Zoo is located inside the city limits. It houses several enclosures, water attractions, a veterinary hospital, and research facilities. Edinburgh Zoo also supplies shelter and food for abandoned and stray animals. There is an isolation centre for infectious animals.

A research facility runs together with the Edinburgh Zoo. University students and researchers conduct observational, sample, and data related projects on site. National centres that oversee disease control, species preservations, fertility, and well-being are all affiliated with the Edinburgh Zoo.

Scotland is a multi-party system: Scottish National Party (centre-left, pro-independence), Scottish Labour Party (centre-left, social democratic), Scottish Conservative Party (centre-right, conservative), Scottish Liberal Democrats (centrist, social liberal), Scottish Green Party (left-wing, eco socialist).

The loaning or purchasing of rare/endangered species from another government is sometimes used as a means to aid in positive relations between nations.

The Edinburgh Zoo's energy is supplied from several sources: Fossil fuels (36%), Natural gas (33%), Nuclear power (20%), Alternative (11%), Hydro or renewable sources.

Edinburgh Zoo hosts "Green Events" at various times during the year. These are open to the patrons of the Edinburgh Zoo. Green Events present information on alternative energy sources and plant life, as well as opportunities for interactive participation.

Scotland has a majority non-religious identifying population: No religion (43.6%), Church of Scotland (32.4%), Roman Catholic (15.9%), Islam (1.4%), Hinduism (0.3%), Buddhism (0.2%), Sikhism (0.2%).

A.2.4 ***Structure II – Low-threat (UK)***

(Section 1)

EDINBURGH ZOO

GOALS

Edinburgh Zoo has been **holding a steady customer base** since 2013, and **met the international requirements concerning animal welfare** as set by **international animal rights councils**. The Zoo's goals are to increase their customer base by 20% by 2020 and increase their conservation efforts for endangered and threatened species.

(Section 2)

The Edinburgh Zoo has a membership base of 13,000. Edinburgh Zoo has in-house education programmes that have reached almost 25,000 school children and the public at large to date. There are seasonal teaching and learning programmes throughout the year. Each programme is designed to present facts about the species' natural habitat, health, life style and cycle, dangers, and, when possible, to allow people to interact with the species. The top education programmes: Beyond the Bees, Meet the Meerkats, Organised school visits, Flora education programmes.

Edinburgh Zoo has over 10,000 Facebook followers and 15,000 Twitter followers. The Zoo circulates e-newsletters to over 10,000 people quarterly, and is in the process of installing high-resolution cameras in 5 animal enclosures. This effort is to allow a 24-hour, live-stream online viewing of the animals.

Edinburgh Zoo receives funding from several sources. According to financial reports from 2013, the smallest percentage of funding came from government sources. The largest percentage came from park revenue. In 2013, the Edinburgh Zoo received funding from several sources. The following list shows the percentages of the annual income each group supplied: Independent societies (21%), Corporate sponsorship (20%), Government councils (14%), Revenue generated in the park (45%) - General admission, Featured exhibits admission, Product sales, Food sales.

The Edinburgh Zoo has nearly 200 employees, including scientists and doctors. A varying number of seasonal volunteers fill the remainder of the jobs at Edinburgh Zoo. This means that volunteers only help at the zoo when a demand is perceived, which can vary from season to season.

The Edinburgh Zoo is on 82 acres of land. Animals are either rescued from dangerous environments or purchased from other zoos or wildlife reserves. The food required for the individual diets of the animals is grown within the zoo, purchased from local suppliers, and imported from the animal's natural habitat through third party suppliers.

Scotland has a population of 5.3 million. 80,000 visitors are expected in 2015. Edinburgh Zoo houses 300 different species and almost 1,000 animals total: 470

birds representing 100 different species, Domestic animals span 16 different species, Aquatic animals represent more than 50 different species, Reptiles represent 90 species, Large mammals (large cats, primates, elephants etc.) comprise the remaining species.

Edinburgh Zoo has a flora programme. The programme was designed to educate visitors and school children about the various plants each animal requires, local plants of the country, and propagation of endangered species. The variety and number of species is unknown.

The annual costs for running the Edinburgh Zoo in 2013 was £10.5. Standard ticket price is £17. Edinburgh Zoo has various product-related marketing deals with outside corporations. Product branding based sponsorship is offered.

The Edinburgh Zoo is located inside the city limits. It houses several enclosures, water attractions, a veterinary hospital, and research facilities. Edinburgh Zoo also supplies shelter and food for abandoned and stray animals. There is an isolation centre for infectious animals.

A research facility runs together with the Edinburgh Zoo. University students and researchers conduct observational, sample, and data related projects on site. National centres that oversee disease control, species preservations, fertility, and well-being are all affiliated with the Edinburgh Zoo.

Scotland is a multi-party system: Scottish National Party (centre-left, pro-independence), Scottish Labour Party (centre-left, social democratic), Scottish Conservative Party (centre-right, conservative), Scottish Liberal Democrats (centrist, social liberal), Scottish Green Party (left-wing, eco socialist).

The loaning or purchasing of rare/endangered species from another government is sometimes used as a means to aid in positive relations between nations.

The Edinburgh Zoo's energy is supplied from several sources: Fossil fuels (36%), Natural gas (33%), Nuclear power (20%), Alternative (11%), Hydro or renewable sources.

Edinburgh Zoo hosts "Green Events" at various times during the year. These are open to the patrons of the Edinburgh Zoo. Green Events present information on alternative energy sources and plant life, as well as opportunities for interactive participation.

Scotland has a majority non-religious identifying population: No religion (43.6%), Church of Scotland (32.4%), Roman Catholic (15.9%), Islam (1.4%), Hinduism (0.3%), Buddhism (0.2%), Sikhism (0.2%).

A.2.5 ***Structure III – High-threat (UK)***

(Section 1)

EDINBURGH ZOO

GOALS

Edinburgh Zoo is **under threat of closure** due to a **decreasing customer base** and **pressures from international animal rights councils**. The Zoo's goals are to increase their customer base by 20% by 2020 and increase their conservation efforts for endangered and threatened species.

A.2.6 ***Structure III – Low-threat (UK)***

(Section 1)

EDINBURGH ZOO

GOALS

Edinburgh Zoo has been **holding a steady customer base** since 2013, and **met the international requirements concerning animal welfare** as set by **international animal rights councils**. The Zoo's goals are to increase their customer base by 20% by 2020 and increase their conservation efforts for endangered and threatened species.

A.2.7 ***Structure IV – High-threat (UK)***

(Section 1)

EDINBURGH ZOO

The Edinburgh Zoo is located inside the city limits. It houses 300 different species and almost 1,000 animals total. The Zoo has a membership base of 13,000, and in-house education programmes that have reached almost 25,000 school children and the public at large to date. Over the years they have branched into the social media sphere with Facebook, Twitter, and more. The Edinburgh Zoo has nearly 200 employees, including scientists and doctors. A varying number of seasonal volunteers fill the remainder of the jobs at Edinburgh Zoo. The annual costs for running the Edinburgh Zoo in 2013 was €20.5 million.

(Section 2)

GOALS

Edinburgh Zoo is **under threat of closure** due to a **decreasing customer base** since 2013, and **pressures from international animal rights councils** (i.e. World Association of Zoos and Aquariums, International Primate Protection League, Wildlife Conservation Society, amongst others). External affiliates are either **threatening to close** or **alter their market focus**. The Zoo's goals are to increase their paying customers (i.e. foot-traffic) by 20% by 2023, and increase their conservation efforts for endangered and threatened species.

A.2.8 ***Structure IV – Low-threat (UK)***

(Section 1)

EDINBURGH ZOO

The Edinburgh Zoo is located inside the city limits. It houses 300 different species and almost 1,000 animals total. The Zoo has a membership base of 13,000, and in-house education programmes that have reached almost 25,000 school children and the public at large to date. Over the years they have branched into the social media sphere with Facebook, Twitter, and more. The Edinburgh Zoo has nearly 200 employees, including scientists and doctors. A varying number of seasonal volunteers fill the remainder of the jobs at Edinburgh Zoo. The annual costs for running the Edinburgh Zoo in 2013 was €20.5 million.

(Section 2)

GOALS

Edinburgh Zoo has been **holding a steady customer base** since 2013, and **met the international requirements concerning animal welfare** as set by the **international animal rights councils** (i.e. World Association of Zoos and Aquariums, International Primate Protection League, Wildlife Conservation Society, amongst others). External affiliates are **maintaining as usual** with **the same market focus**. The Zoo's goals are to increase their paying customers by 20% by 2023, and increase their conservation efforts for endangered and threatened species.

A.3 **Vignette 3 – Tammuz Confectionary Company**

Vignette 3 was used for participant samples in the UK, US, and UAE. However, the same two structures (high-threat and low-threat) were supplied to all participants. Vignette 3 was developed after priming and validation tests confirmed that structure IV was the optimal structure for delivering the framing manipulation and used in the framing validation tests. Therefore, vignette 3 was only developed into structure IV for use in experimental studies.

A.3.1 ***Structure IV – High-threat (UK, UAE, US)***

(Section 1)

CONFECTIONARY COMPANY

Tammuz Confectionery Co. is an international confectionary producer. Their brand includes nearly 130 products: chocolate bars, truffles, pralines and filled chocolates, panned confections, hard candies, gummies, jellies, liquorice, chews, fruit snacks and jelly beans, and gum are their top-selling categories. They have 29 branches in several Eastern countries, the Mediterranean region, and Canada. Tammuz Confectionery Co. has over 500 employees, including confectionary chemists, and reported sales of $30.4 million in 2013.

(Section 2)

GOALS

The Tammuz Confectionery Co. is **under threat of closure** due to **decreasing net sales** since 2013, and **pressures from international human rights organisations** (i.e. Human Rights Watch, International Labor Rights Forum, Worker Rights Consortium, amongst others). International branches are **either threatening to close or become consolidated**, thus **decreasing the number** of processing and distribution centres. The Tammuz Confectionery Co. goals are to expand into the local market, increase their product sales by 20% by 2023, and increase their human-rights and ethical farming efforts for raw products (i.e. cacao, sugar).

A.3.2 ***Structure IV – Low-threat (UK, UAE, US)***

(Section 1)

CONFECTIONARY COMPANY

Tammuz Confectionery Co. is an international confectionary producer. Their brand includes nearly 130 products: chocolate bars, truffles, pralines and filled chocolates, panned confections, hard candies, gummies, jellies, liquorice, chews, fruit snacks and jelly beans, and gum are their top-selling categories. They have 29 branches in several Eastern countries, the Mediterranean region, and Canada. Tammuz Confectionery Co. has over 500 employees, including confectionary chemists, and reported sales of $30.4 million in 2013.

(Section 2)

GOALS

The Tammuz Confectionery Co. has been **holding steady in their net sales** since 2013. They have **received 3 out of 4 star ratings** from international human rights organisations (i.e. Human Rights Watch, International Labor Rights Forum, Worker Rights Consortium, amongst others). The Tammuz Confectionery Co. goals are to expand into the local market, increase their product sales by 20% by 2023, and increase their human-rights and ethical farming efforts for raw products (i.e. cacao, sugar).

Appendix B

B.1 Stage 3 Studies: Heterogeneous Templates

The following images are the templates presented to participants in both the UK and UAE that coincide with the different vignettes they were randomly assigned. For the heterogeneous templates, participants in both framing conditions (high-threat and low-threat) received the same template for their assigned vignette.

Figure B.1. Vignette 1 – UAE

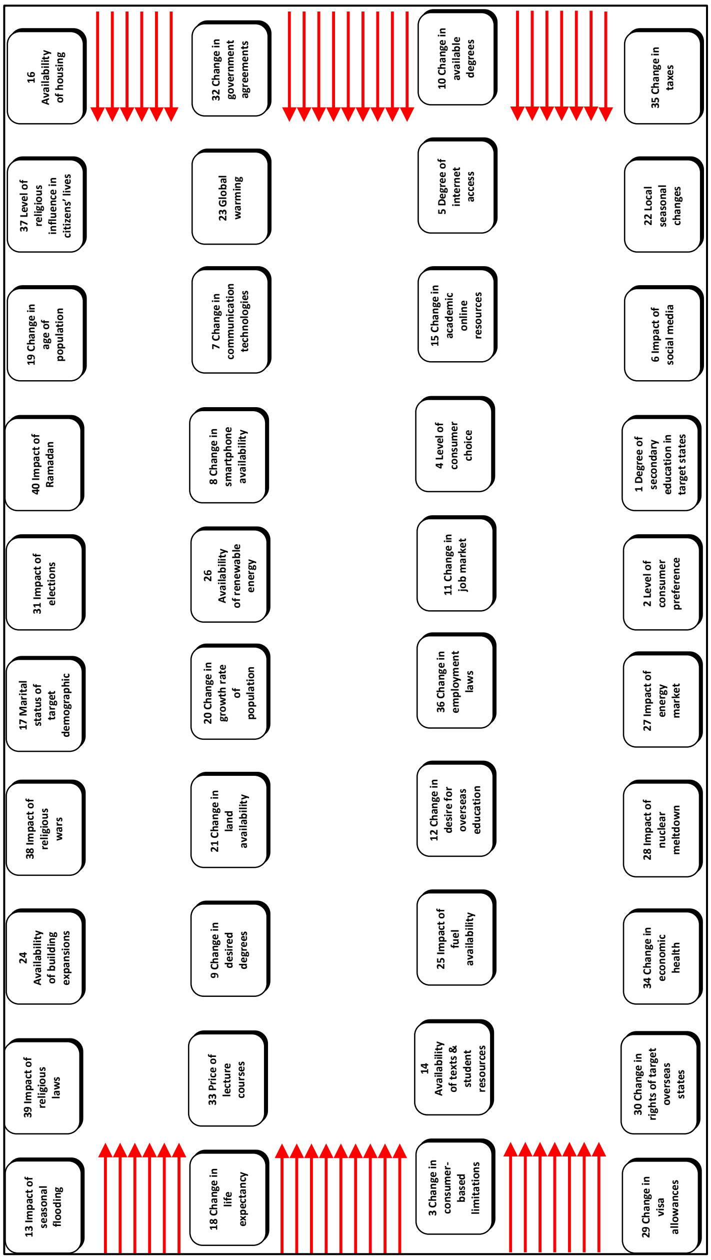


Figure B.2. Vignette 2 – UAE

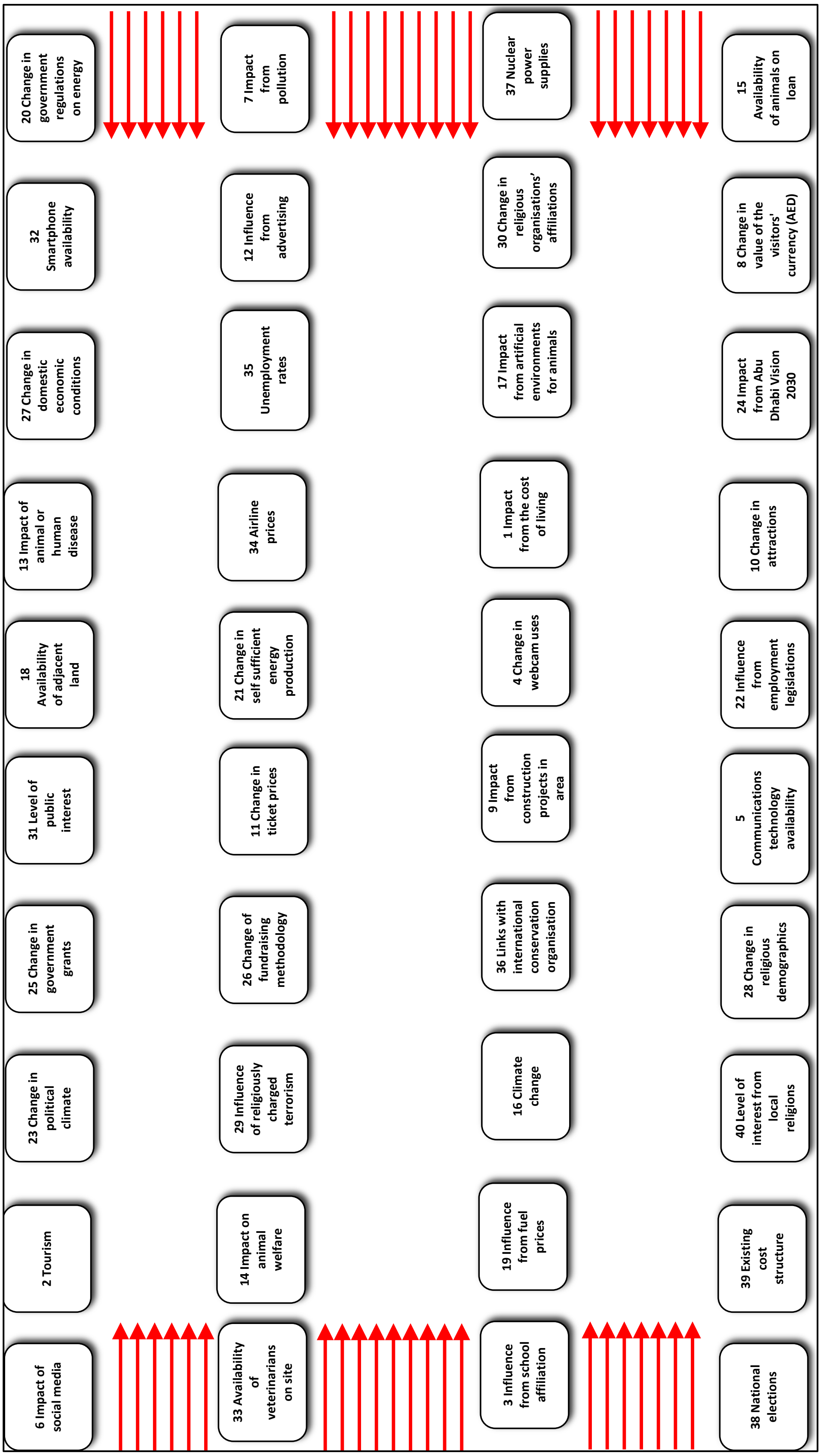


Figure B.3. Vignette 3 – UAE

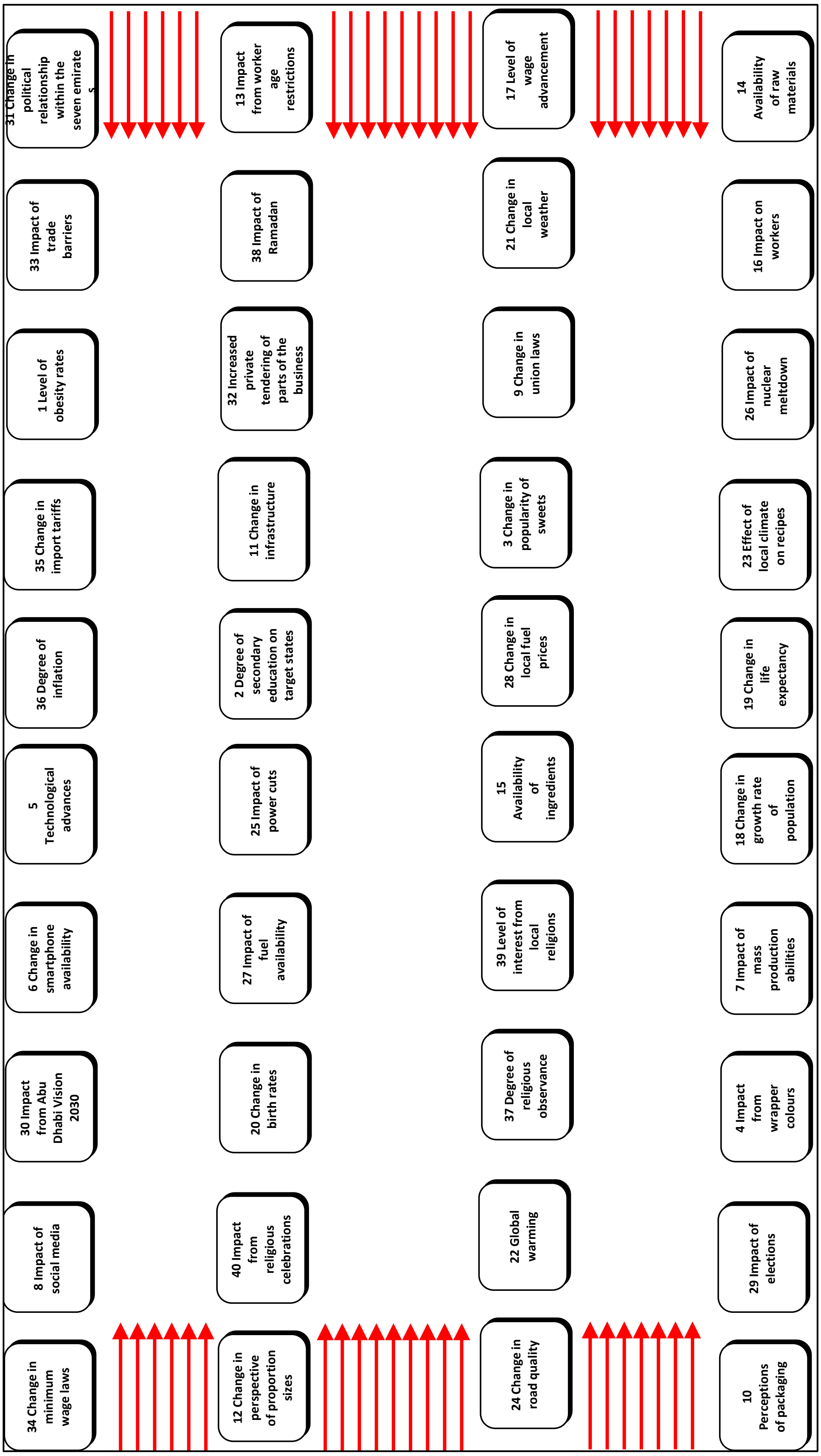
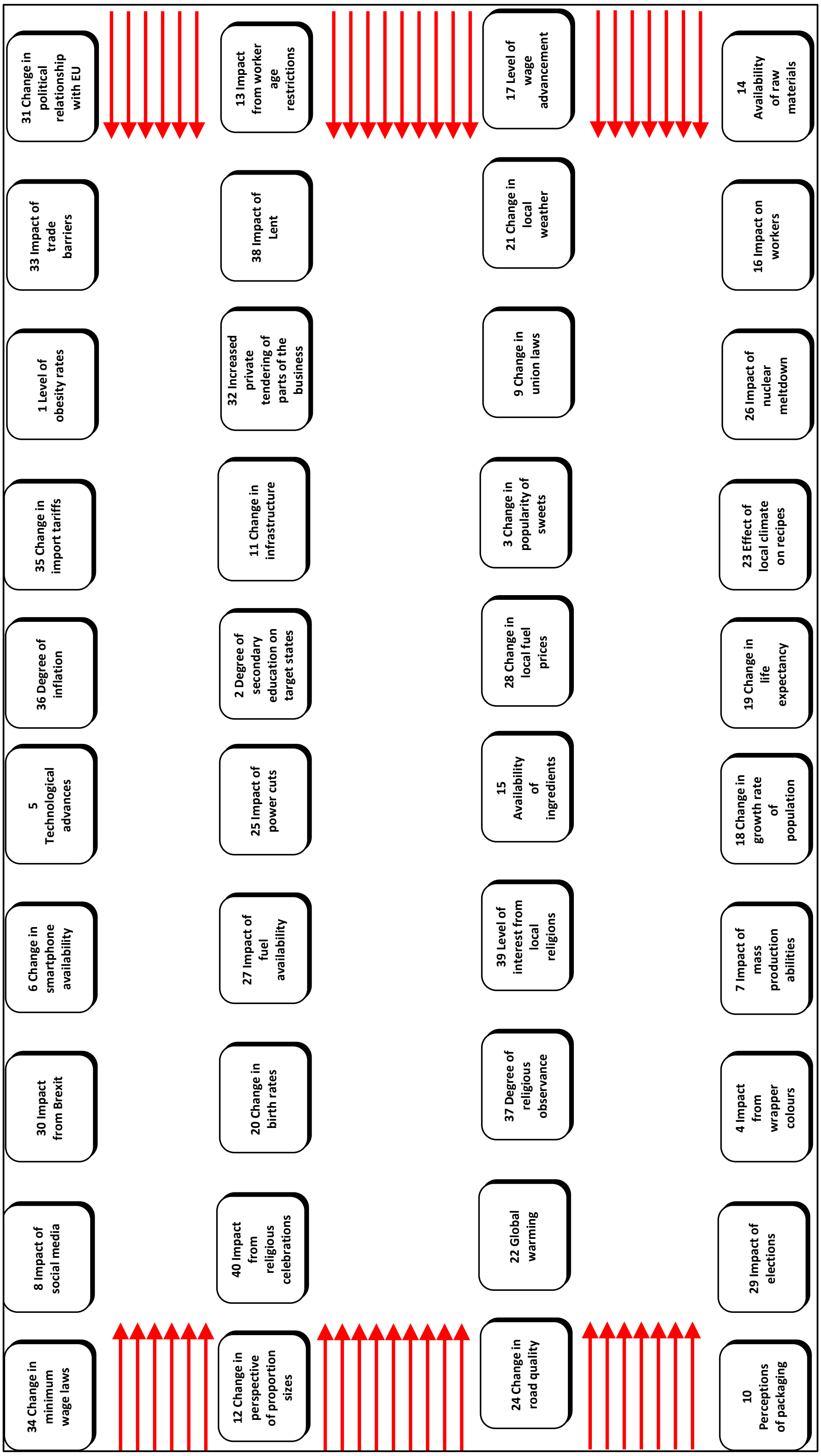


Figure B.4. Vignette 3 – UK



B.2 **Stage 3 Studies: Homogeneous Templates**

Only Vignette 2 was used for the homogeneous models of the study. Participants were located in both UK and UAE. The high-threat model resulted in the both locations receiving the same model. The low-threat model resulted in both locations receiving a slightly different model. The differences are highlighted with each low-threat model.

Figure B.5. Vignette 2 – high-threat (UK/UAE)

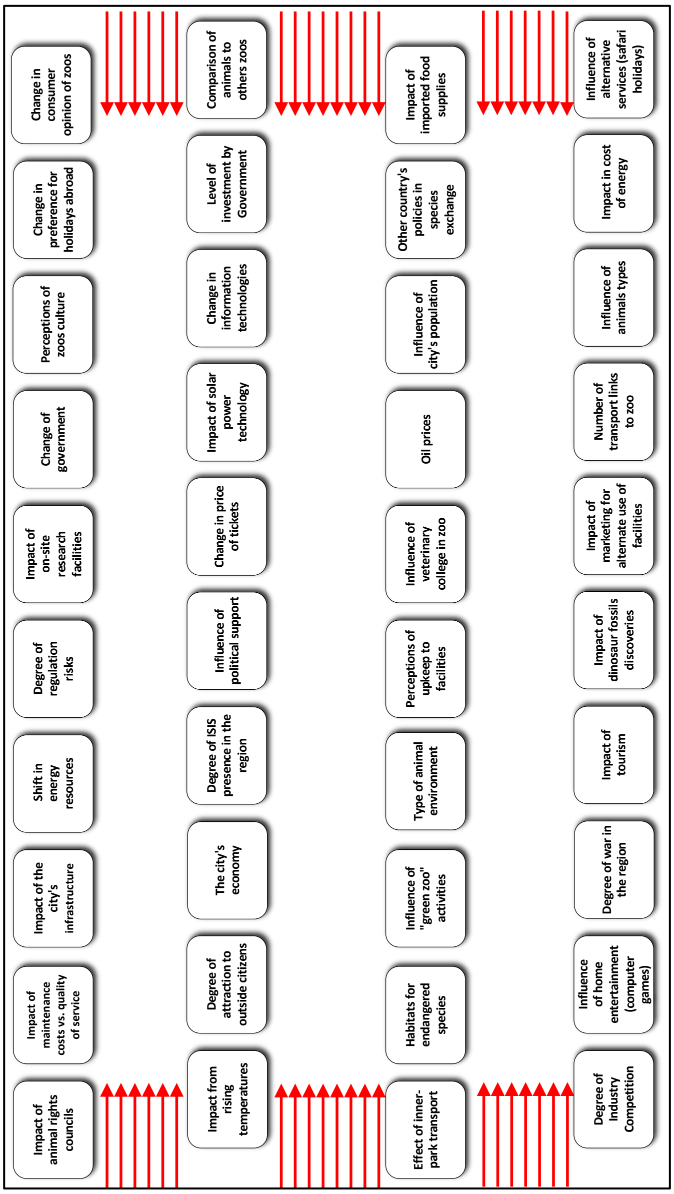


Figure B.6. Vignette 2 – low-threat – UAE

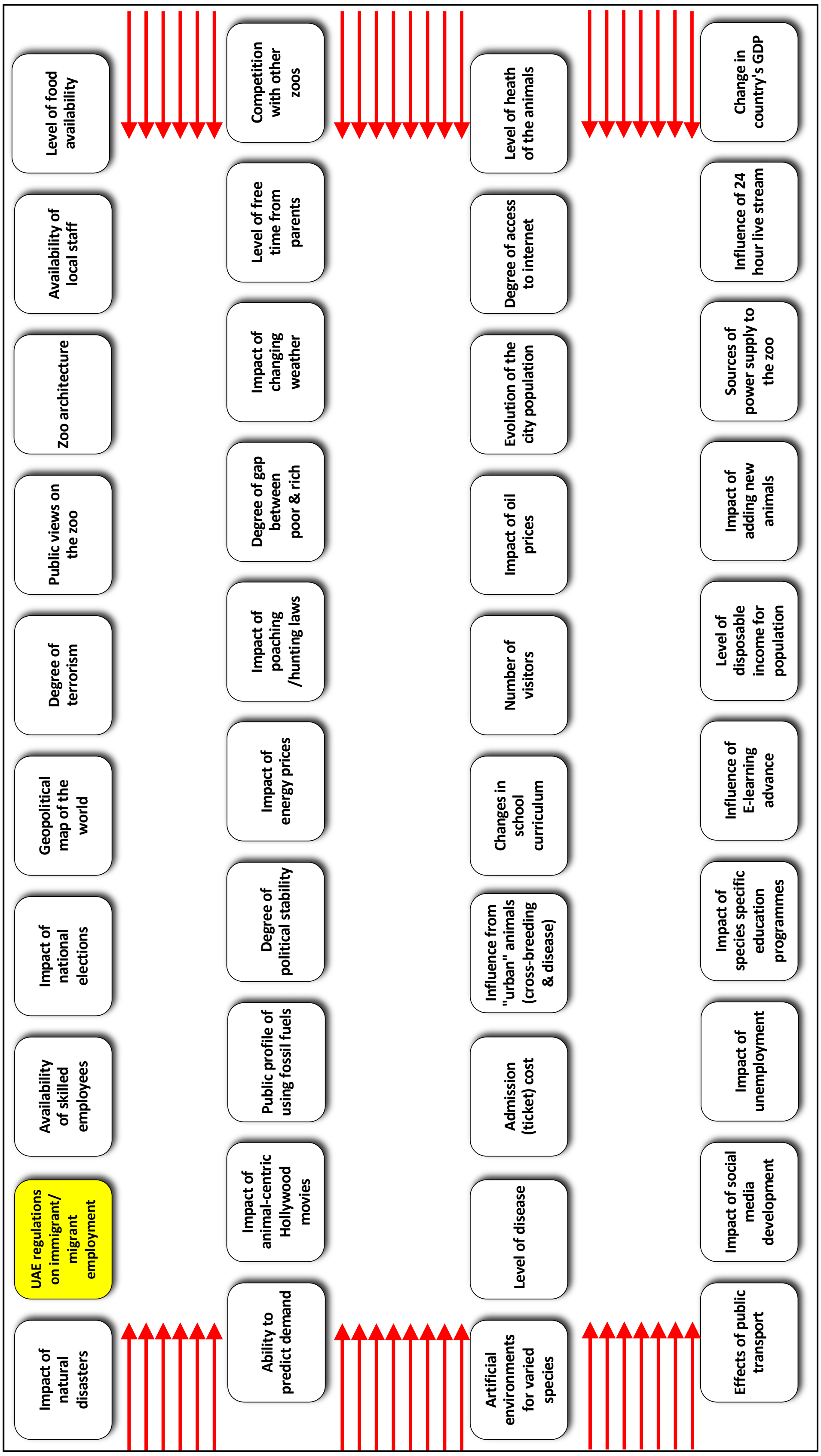
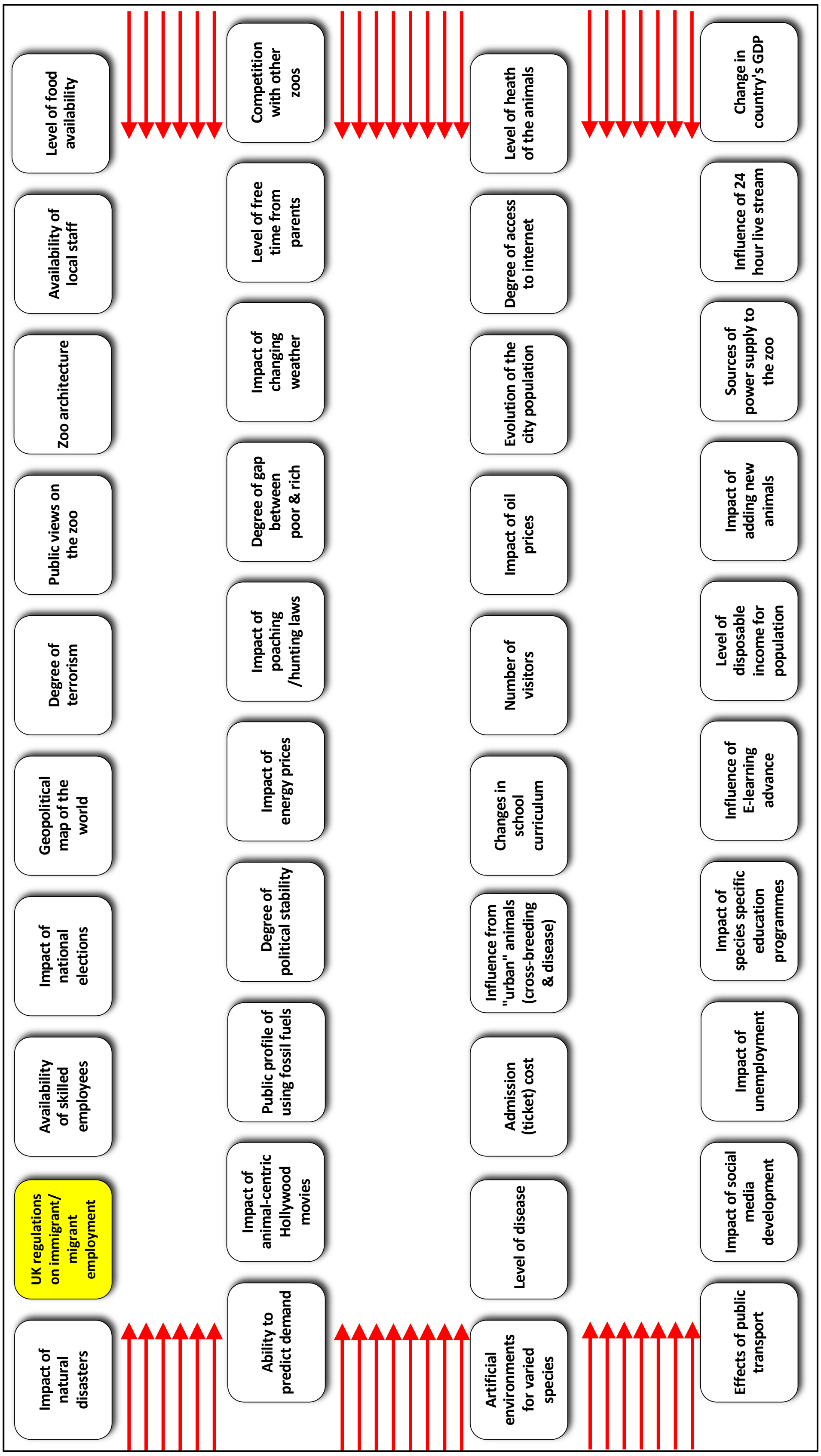


Figure B.7. Vignette 2 – low-threat – UK



Appendix C

Participant 108B's raw .pptx file converted into DE file for analysis.

Figure C.1. Original (.pptx)

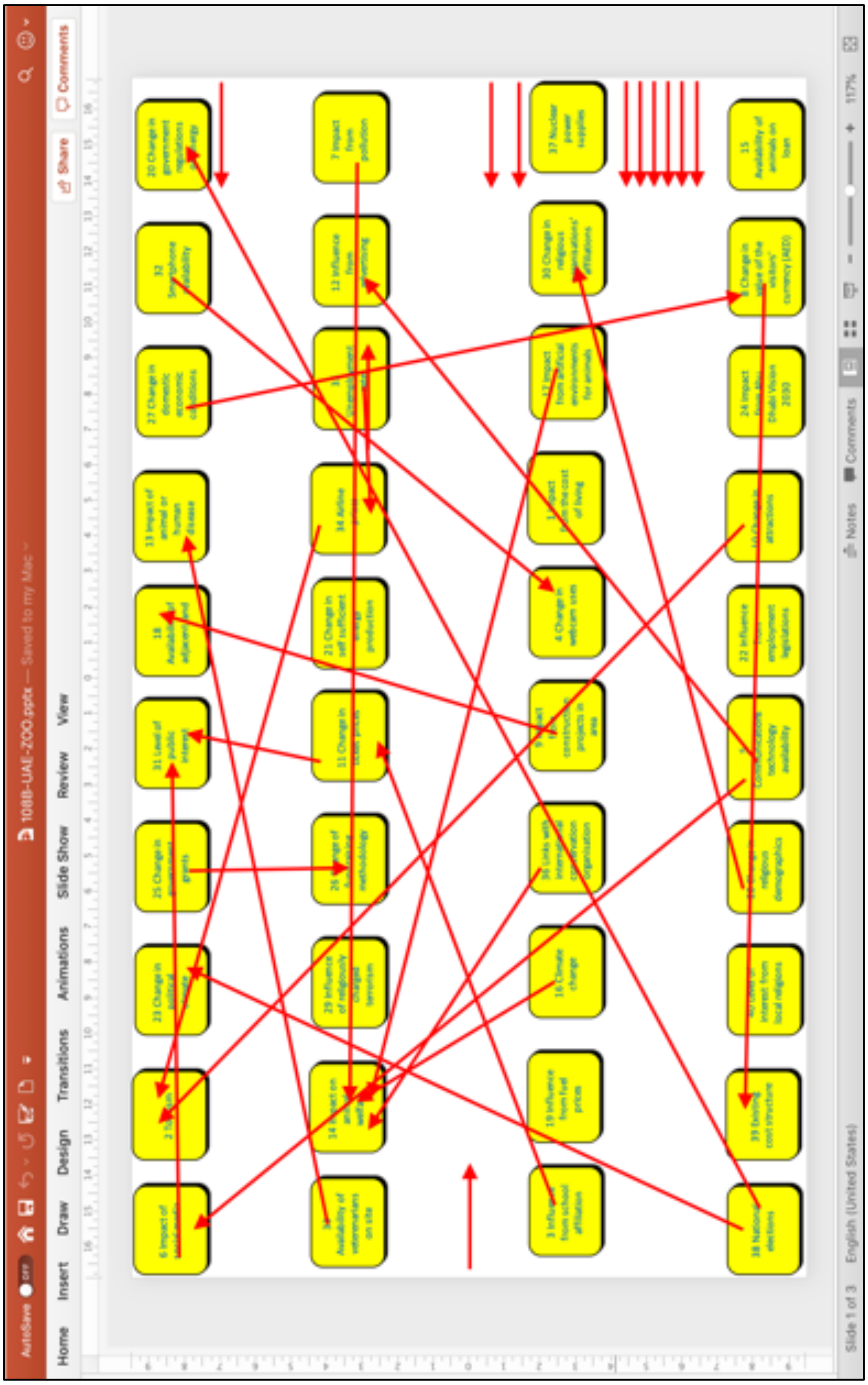


Figure C.2. Converted (DE file)

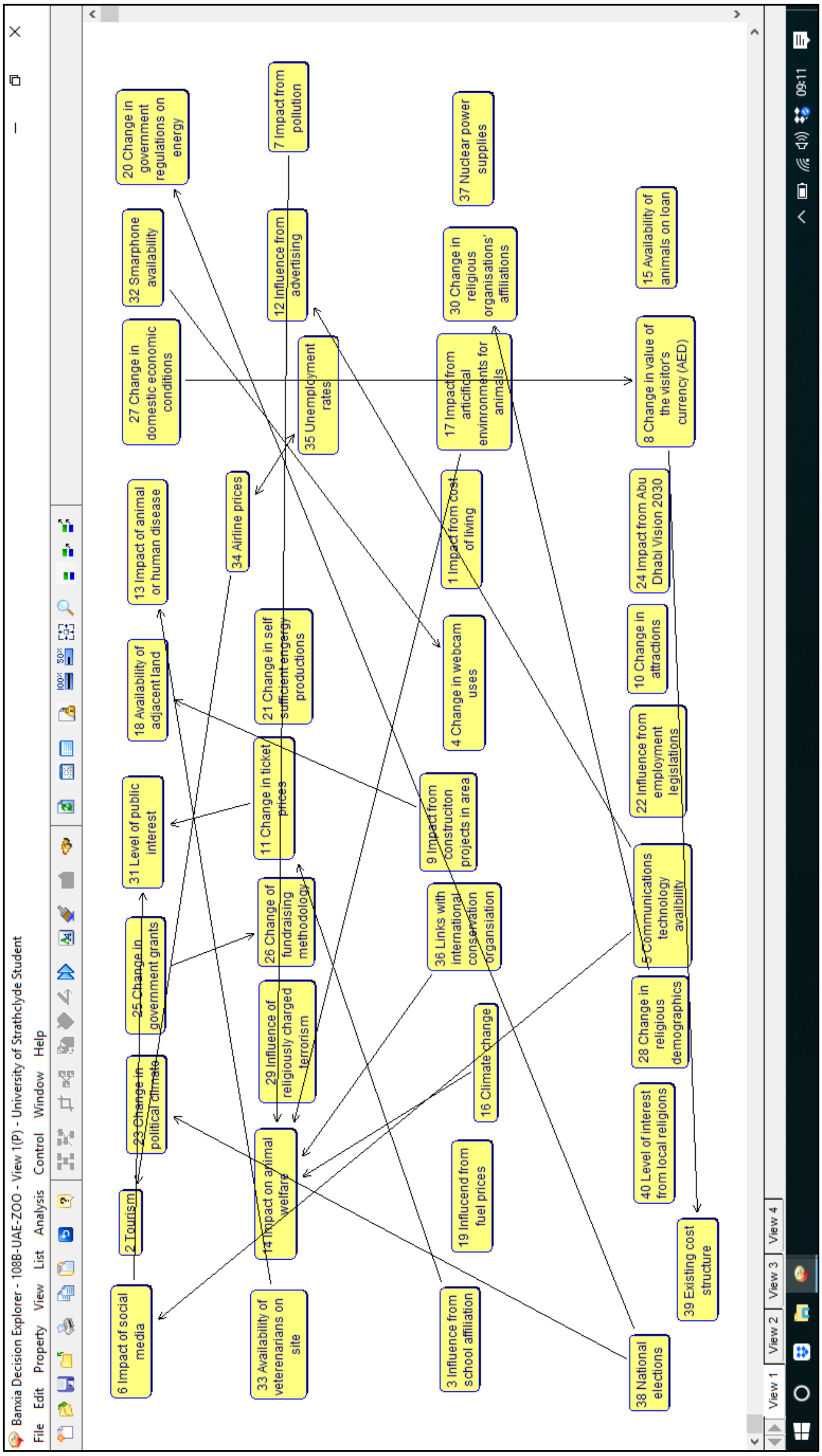


Figure C.3. Example output cluster2

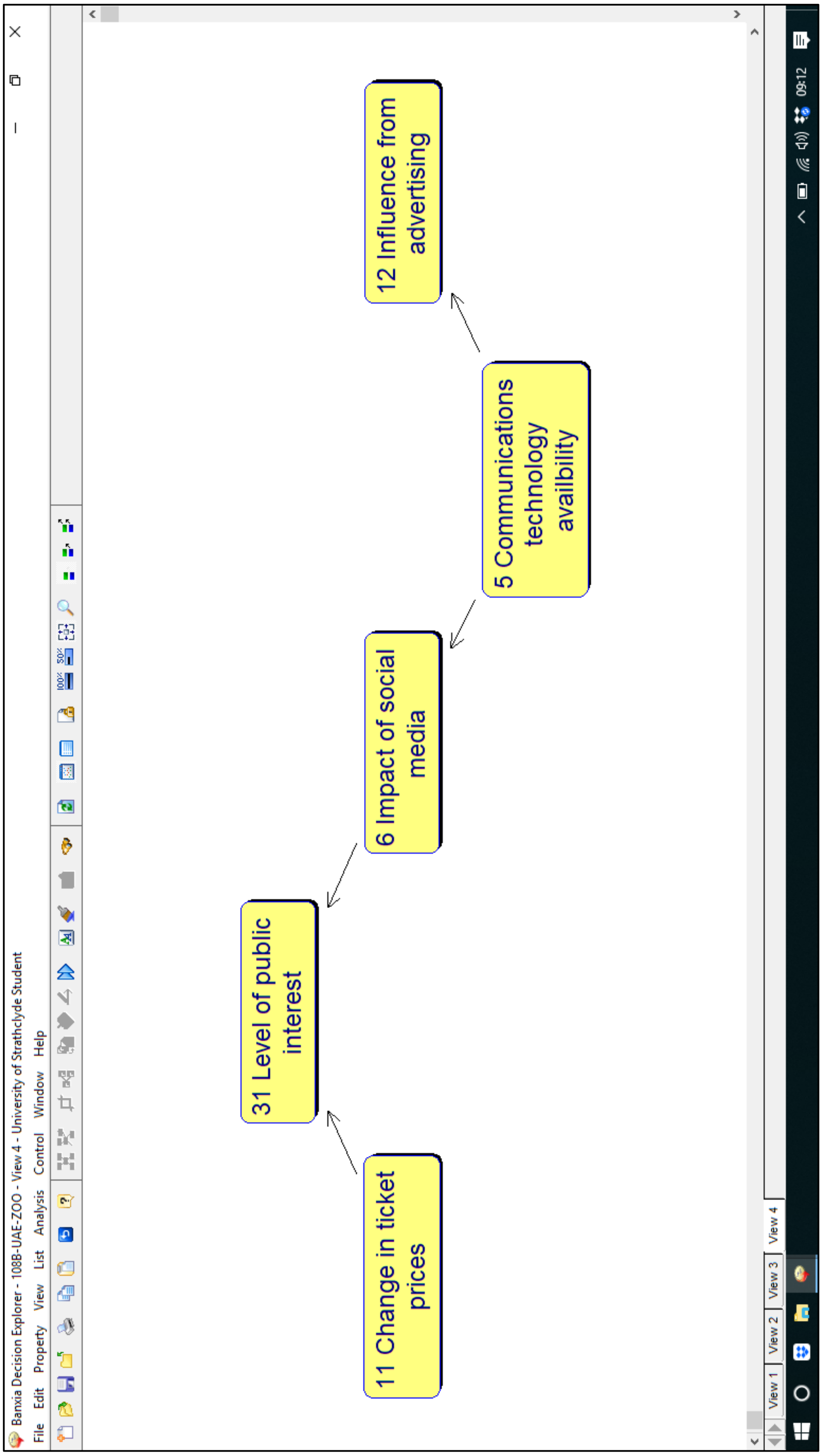
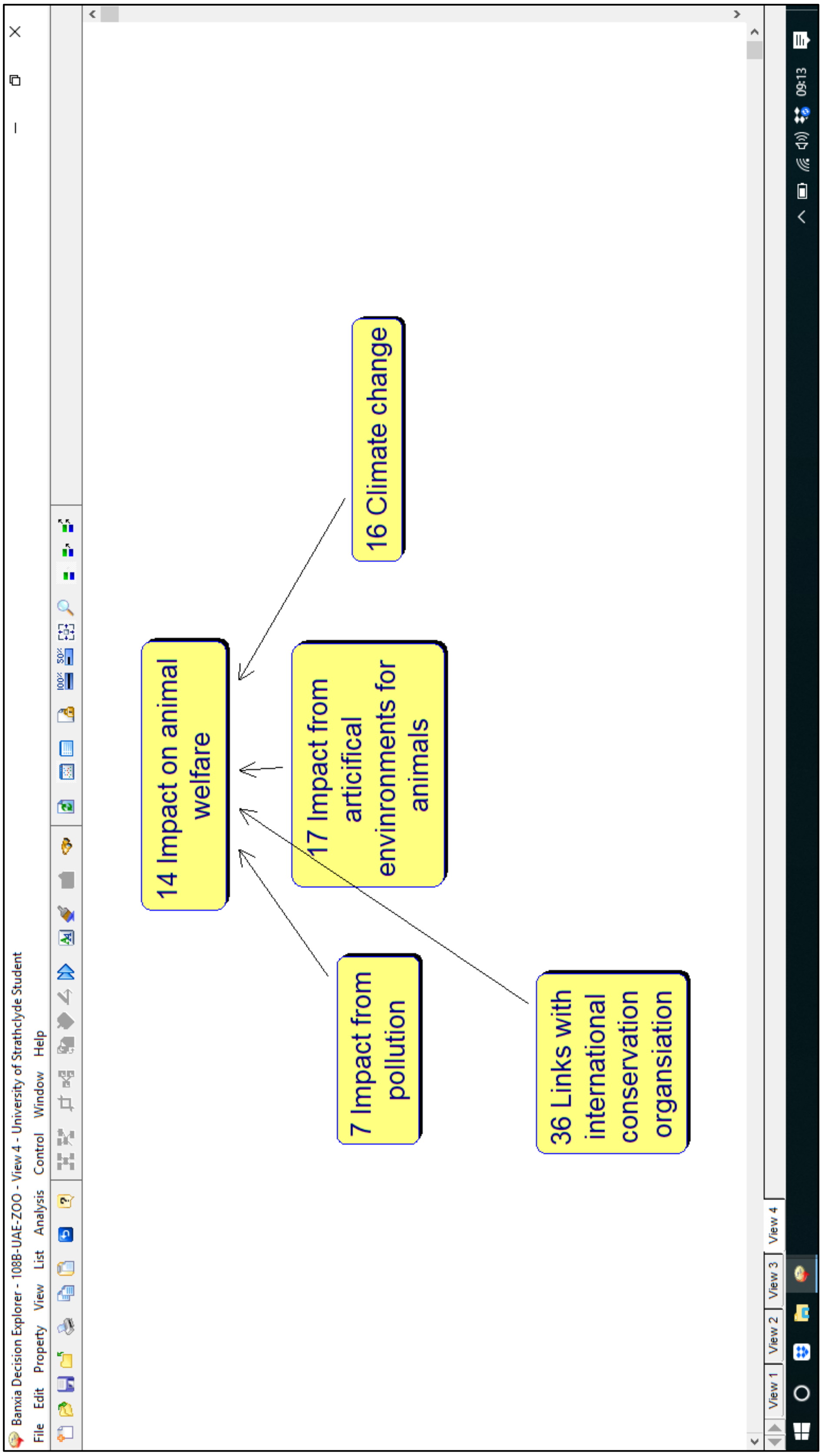


Figure C.4. Example output cluster4



Appendix D

Exploratory analyses from Stage 3 studies.

D.1 **Heterogeneous Study: Causal Relationships**

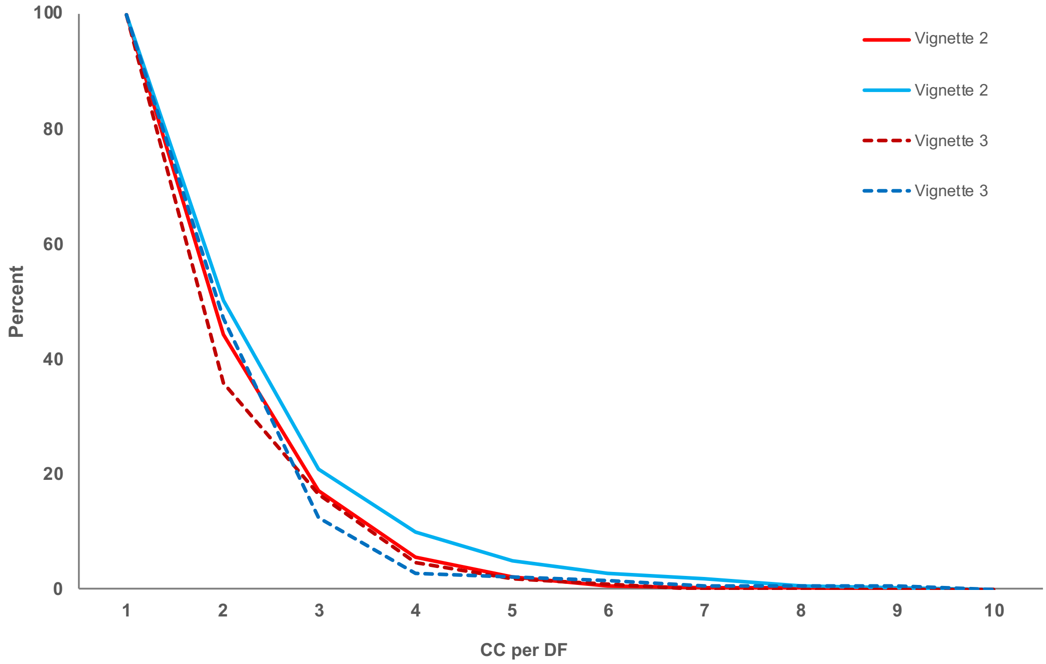
Each DF's quantity of CC (1-CC, 2-CC, 3-CC, etc.) is transposed into discrete proportions of the total number of CC by participant. To plot the results, the sum of DF (d) given their number of CC (c), from minimum to maximum CC are divided by the total d occurring within the condition, then multiplied by 100 to convert into a percentage. Each plot point is the next highest number of CC, denoted as $c + 1$ (see Equation D.1).

Equation D.1. Distributed percentage

$$f(d) = 100 * \left(\frac{d_c + (\sum d_{c+1})}{\sum d} \right)$$

Figure D.1 maps the distributed percentage of CC per *used* DF within the heterogeneous sample and shows that 100% of the used DF have at least one CC associated with them. This is unsurprising because all used DF must, by virtue of their inclusion in a cluster, have at least one CC with another DF. In all conditions for both vignettes, at least half of the DF have only one CC (vignette $2_{\text{high-threat}}$ = 56%, vignette $2_{\text{low-threat}}$ = 50%, vignette $3_{\text{high-threat}}$ = 64%, vignette $3_{\text{low-threat}}$ = 53%). The four curves reveal different concentrations of CC amongst the DF within the clusters, and therefore reveal part of their complexity behaviours. The sample that constructed the most complex clusters, of the four, are those who read vignette 2 with the low-threat condition. It is difficult to generalise too much from the behaviours shown in vignette 3 due to their low sample size. However, vignette 2 data aligns with Hypothesis 8 predictions.

Figure D.1. Distribution of causal connections and driving forces by condition by vignette



Note: The red lines represent the high-threat condition, and the blue lines represent the low-threat condition. The more simplistic the CC, the steeper the convex curve. The more complex the CC, the closer the curve reaches a concave orientation.

The direction of the relationship between DF is analysed to determine whether there is a prevalence of CC leading from DF (cause) or leading to DF (effect). Percentages are reported in Table D.1. The vignettes appear, at first glance to produce almost opposite behaviours, however it is important to remember that vignette 3 represents only four participants in each condition. Within vignette 2, there is no significant difference between conditions, illustrating anchoring no bias in cause-effect perspective. However, there is a significant effect within each condition. Both conditions produced more single CC from DF than single effect connection to DF. With at least half of the CC existing as only a single connection and the majority of those being causally oriented, the data support Fernbach's, et al. (2011) claim that people have a tendency to focus on single causal sources.

Table D.1. Direction of causal connections by condition for vignettes 2 and 3

Causal chain	Condition	Vignette	Percentage (%)
Cause	High-threat	2*	64.5
		3	45.6
	Low-threat	2**	68.3
		3	45.2
Effect	High-threat	2	35.5
		3	54.4
	Low-threat	2	31.7
		3	54.8

* $\chi 2(1{,}228) = 19.11, p < .000$

** $\chi 2(1{,}306) = 40.99, p < .000$

D.1.1 ***Predictive vs Diagnostic***

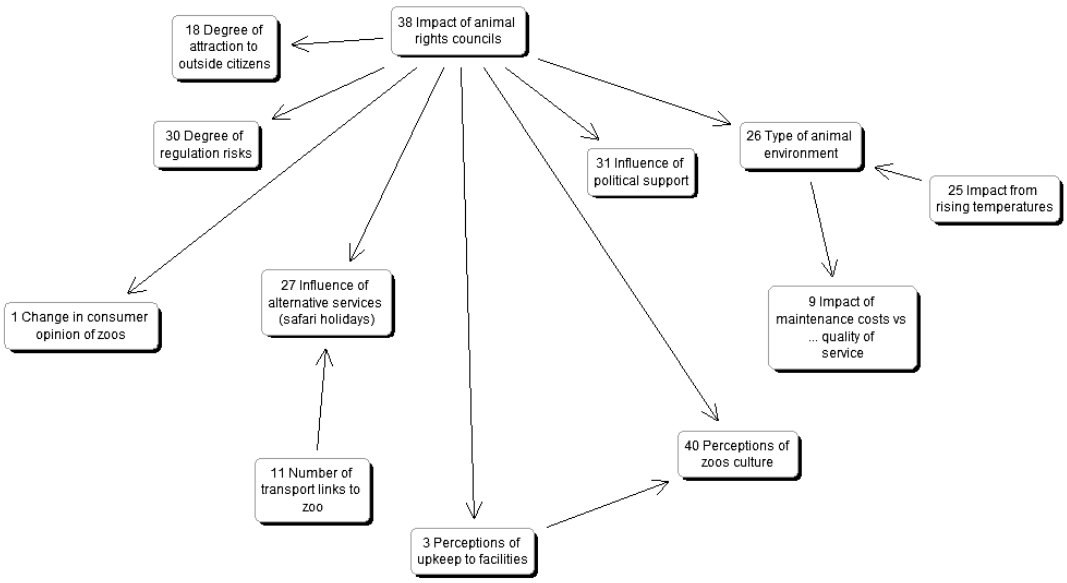
Two features of every clustering exercise are the designation of DF as starting causal factors (whether in the present or near future) or final effects of cluster chains (some distant future point). The instructions for the participants included both predictive (forward-casting) and diagnostic (backcasting) ST methods,

You are asked to show where you see **causal relationships** between the various influencing items. A causal relationship is when one influence has an impact on another. To help understand the beginning of a causal relationships, it helps to ask such questions as, *"What will happen if this occurs?"* or *"Why is this important?"* or *"What will this impact?"* To help understand the effect of a causal relationship, it helps to ask such questions as, *"How would this be achieved?"* or *"How was this caused?"* or *"What would have an impact on this?* To show a relationship, you will link an arrow from one influence to another. The direction of the arrow will show the direction of the causal effect (see Appendix C).

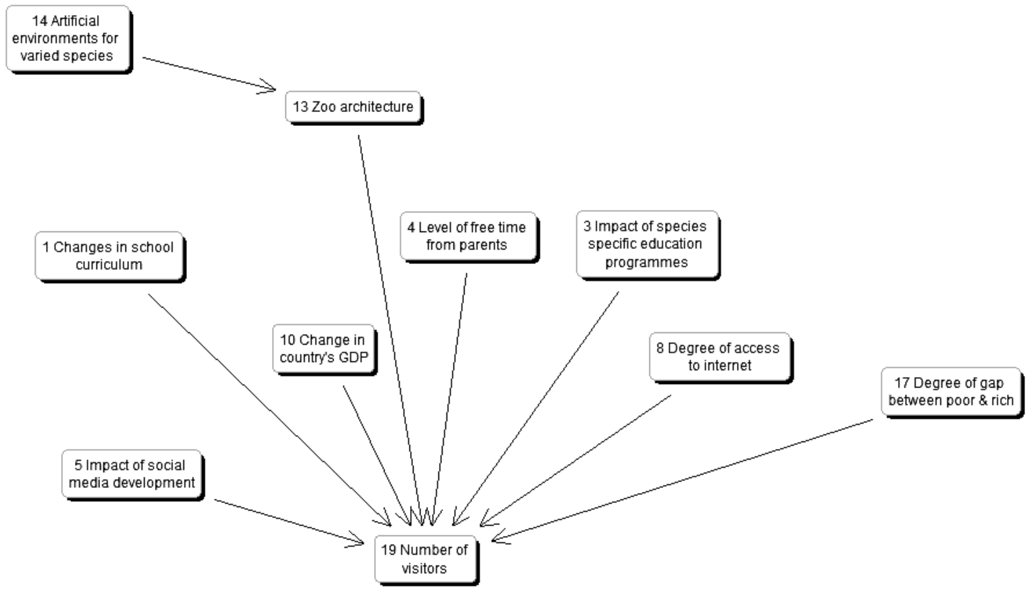
"Forward-casting scenarios look for the effects (future) of a suggested set of causes (past and present) and set the present as the starting point to the strategic conversation" (Crawford, 2019, p. 8). Clusters with a greater number of final effect DF could potentially reflect a predictive vantage point. Backcasting, or diagnostic, scenarios develop from the opposite direction, focusing on a future event and building a logical, storied, flow back to the present state to help determine the path(s) needed to reach the future event (Bishop, Hines, & Collins, 2007). Clusters with greater numbers of starting causal DF could be said to reflect a diagnostic mode of thinking. To illustrate the extreme limits of this kind of clustering, a predictive cluster would include a single DF as the causal source to several DF, and a diagnostic cluster would include a single DF as the final effected factor from several causal DF. Figure D.2 illustrates both extreme versions using real clusters constructed from two participants.

Figure D.2. Examples of predictive and diagnostic clusters

D.2.a. Predictive



D.2.b. Diagnostic



Analyses were conducted to determine whether participants used more predictive or diagnostic ST efforts. These DF do not include the DF within clusters that are both effected by an earlier DF and causes to a later DF, only the beginning and end of the cluster chains. The average number of beginning and ending DF were calculated per participant. Table D.2 presents these statistics. Within vignette 2, there is no significant difference between conditions by either the number of beginning (cause) DF nor ending (effect) DF. Participants in both conditions produced about the same average number. However, there are significant differences within condition. Regardless of framing message, participants designated more beginning DF, on average, than ending DF. This may reflect two possible issues. First, participants show a tendency towards diagnostic perspectives in their collective ST. Second, participants may be reflecting a conjunction fallacy, where more than one DF is considered a causal source, when only one would be sufficient. The issue of sufficiency, however, is a difficult reality to argue in SP. With high levels of uncertainty inherent in any foresight effort, what is a sufficient cause for one scenario may not be sufficient for another scenario. An attempt to partially explore this question is addressed in the qualitative analyses.

Table D.2. Beginning and ending driving forces by condition and vignettes 2 and 3

Causal chain	Condition	Vignette	M	SD	V
Beginning	High-threat	2*	2.70	1.37	1.87
(cause)		3	1.58	.37	.14
	Total		2.42	1.28	1.65
	Low-threat	2**	3.68	2.92	8.54
		3	1.75	.18	.03
	Total		3.43	2.80	7.84
Ending	High-threat	2	1.47	.33	.11
(effect)		3	3.31	3.80	14.41
	Total		1.93	1.91	3.64
	Low-threat	2	2.40	2.29	5.23
		3	1.97	.48	.24
	Total		2.34	2.14	4.58

*$p = .007$

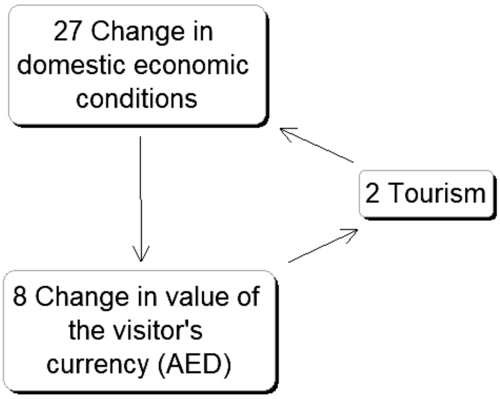
**$p = .002$

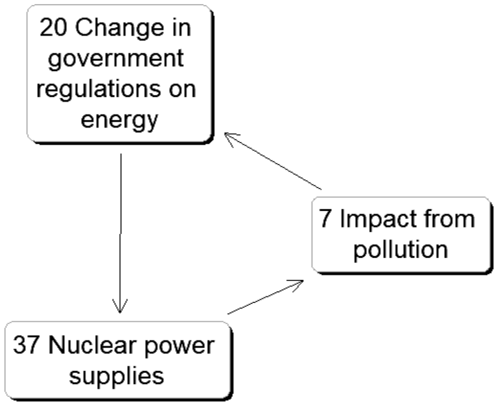
D.1.2 ***Transitivity & Loops***

The number of transitive and loop constructions is very few, as expected. Both relationships occur only within vignette 2, across both conditions. Only descriptive statistics are offered for these occurrences. Three participants within the high-threat condition, each produced a single 3-way loop. In Figure D.3.a, the loop is almost entirely focused on economic DF, Figure D.3.b shows a looping relationship between energy and political categories, Figure D.3.c is entirely focused on religious DF.

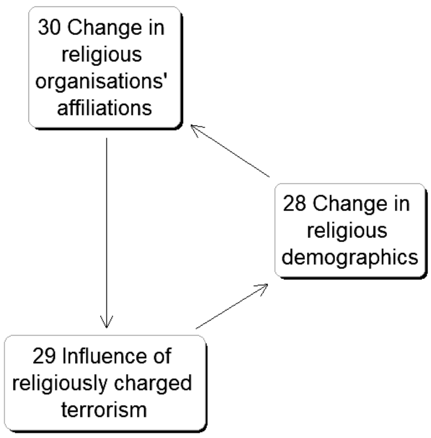
Figure D.3. Vignette 2 high-threat condition loops

D.3.a. Economic D.3.b. Energy & Politics





D.3.c. Religion

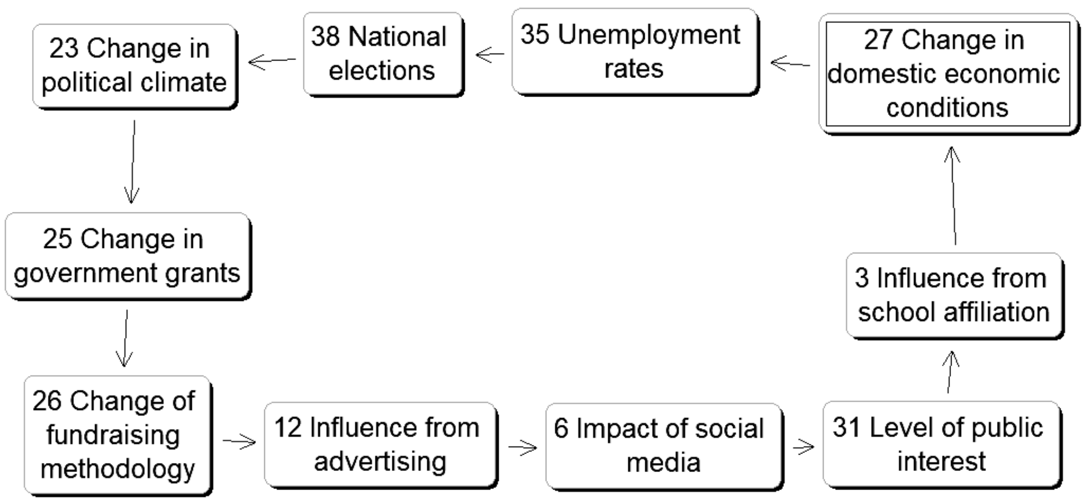


The low-threat condition produced quite a different outcome. Two participants created one loop, where one was a simple transitive relationship and the second was a very large loop across 10 DF. Figure D.4.a shows that the DF “Airline prices” and “Unemployment rates” are focused on economics, while Figure D.4.b shows a variety of DF and categories linked to the larger loop. Two other participants each created five loops, ranging from 2-5 DF within a loop. All transitive relationships are found within the larger loops, therefore one participant created three independent loops (Figure D.4.c.) and the other participant created two (Figure D.4.d). Figure D.4.c shows that the three loops overlap across the relationship between “Change in government regulations on energy” and “Impact from pollution”, revealing a focus on energy and political categories. Figure D.4.d. shows how the two loops overlap across a transitive relationship, just as the other participants'. However, this relationship is between "Airline prices" and "Tourism", revealing a focus on economics and demographics categories. Both conditions present a mix of DF and relations between them, though both conditions touch upon the same categories – energy, politics, and economics.

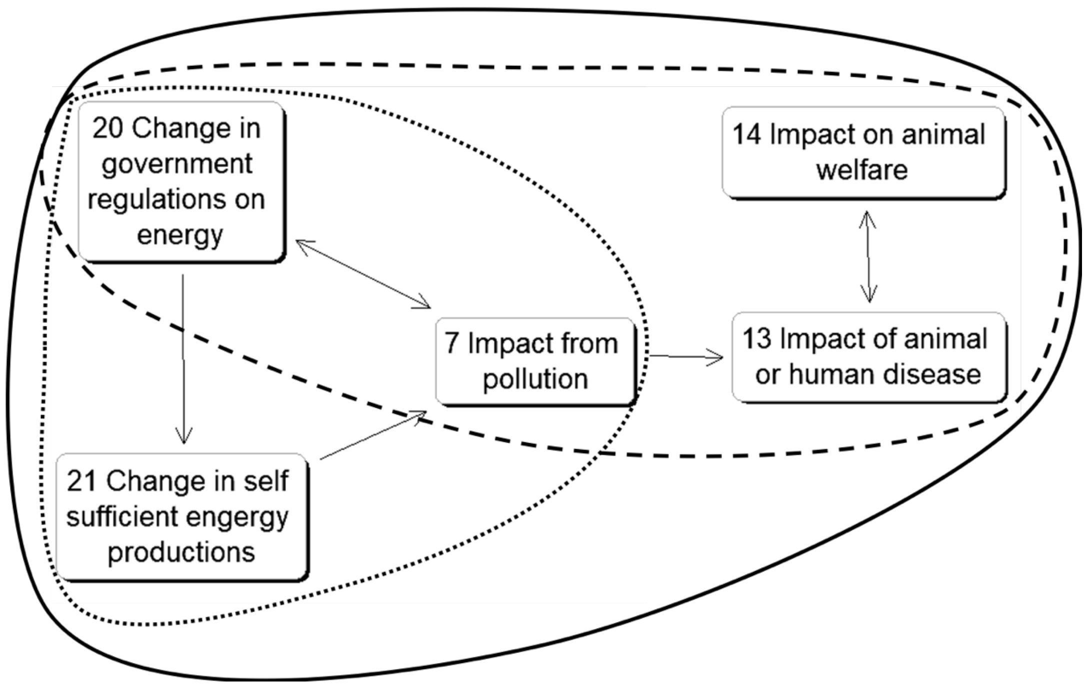
Figure D.4. Vignette 2 low-threat condition loops

D.4.a. Economics

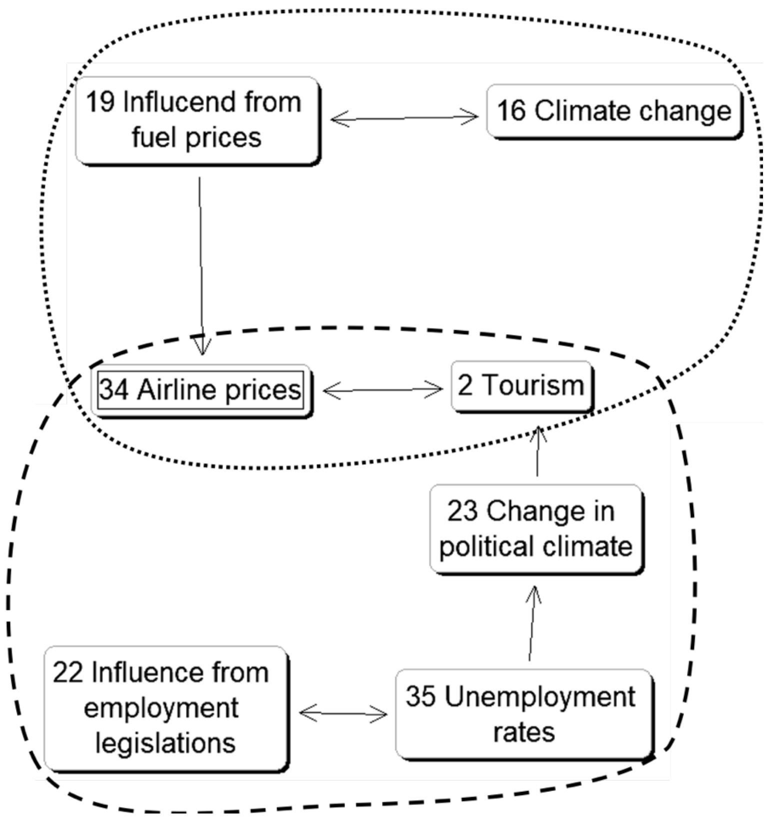
D.4.b. Mixture



D.4.c. Energy & Politics



D.4.d. Economics & Demographics



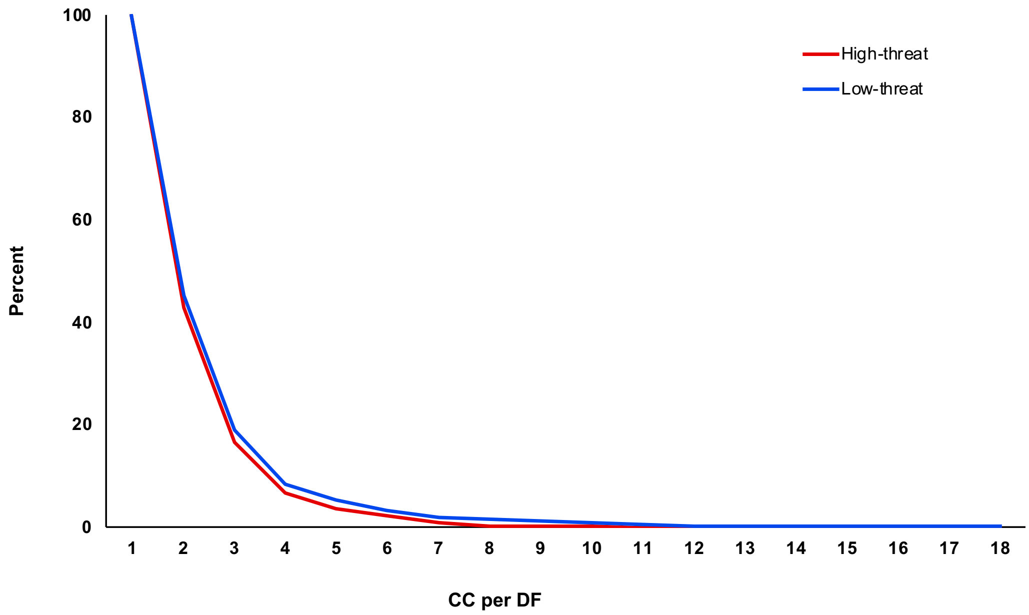
Note: Interestingly, the analyses did not determine a large loop for 8.6.d. as it did for 8.6.c. Loops are determine by both presence and strength of associations.

D.2 **Homogeneous Study: Causal Relationships**

The direction of the relationship between DF is analysed to determine whether there is a prevalence of CC leading from DF (cause) or leading to DF (effect). There is significance in direction of CC within both conditions, but the pattern of distribution is equivalent between them. Both conditions produced more single CC from DF than single effect connections to DF (high-threat = cause (58.8%), effect (41.2%); $\chi 2(1{,}233) = 7.22, p = .007$; low-threat = cause (62.3%), effect (37.7%); $\chi 2(1{,}329)19.94, p < .000$). More than half of the DF in each condition contain only a single association with another DF, and the majority of those connections are causally oriented (as opposed to effectually), which supports Fernbach's, et al. (2011) claim that people have a tendency to focus on single causal sources.

Figure D.5 maps the distributed percentage of CC per *used* DF within the homogeneous sample. The interesting story this figure reveals is that, within both conditions, just over half of the DF only have one CC (high-threat = 57%, low-threat = 55%), and a quarter have only two CC (both = 26%). The two conditions reveal an almost identical trend in distribution, up until CC become more complex (at ≥ 3 CC per DF). The high-threat condition produced fewer complex connections, topping out at 10 CC for a single DF, while the low-threat condition produced up to 18 CC for a single DF. Figure D.5 distribution plot reveals that even though cluster complexity takes a similar path, the low-threat condition created more complex CC between DF. This distribution pattern helps to illustrate that, at the group level, the low-threat condition lead to more clusters that focus heavily on an anchoring DF, whether as a cause, effect, or central factor. This analysis is important to reveal the difference between projected behaviours and observed behaviours.

Figure D.5. Distributed percentage of causal connections and driving forces

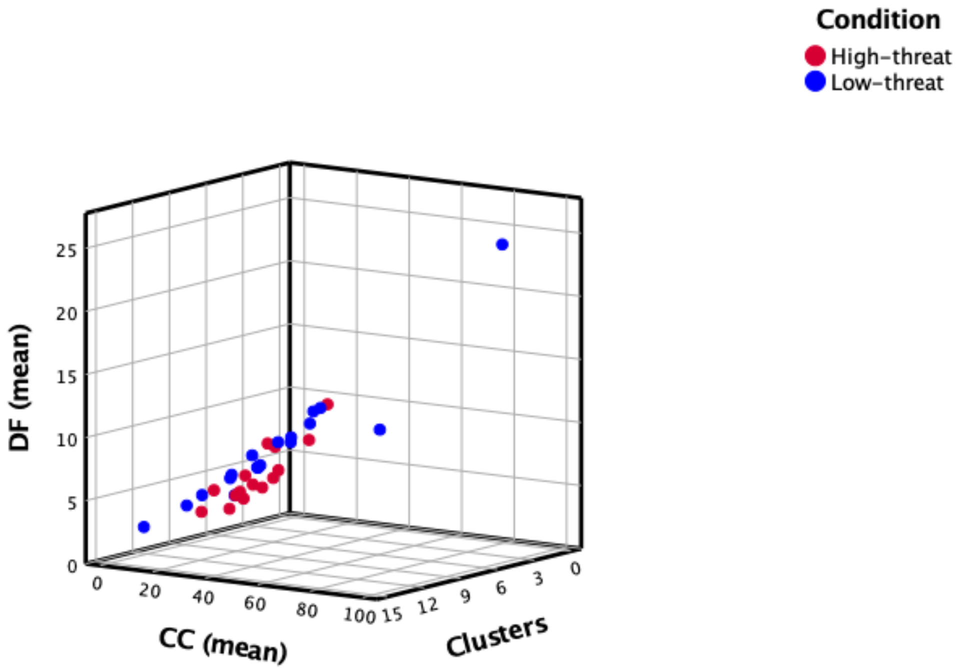


Note: The more simplistic the CC, the steeper the convex curve. The more complex the CC, the closer the curve reaches a concave orientation.

D.2.1 ***Cumulative Complexity***

Now it is time to reveal the full relationship of complexity. When the average number of DF and CC are given for each cluster, both conditions show that as the number of CC increase, pulling in more DF, the number of unique clusters diminishes (see Figure D.6). However the high-threat condition has a lower performance on all axes. What these analyses are all revealing is that the high-threat condition created fewer clusters, with fewer DF, that had fewer CC between them, compared to the low-threat condition.

Figure D.6. Complexity of clusters.



D.2.2 ***Predictive vs Diagnostic***

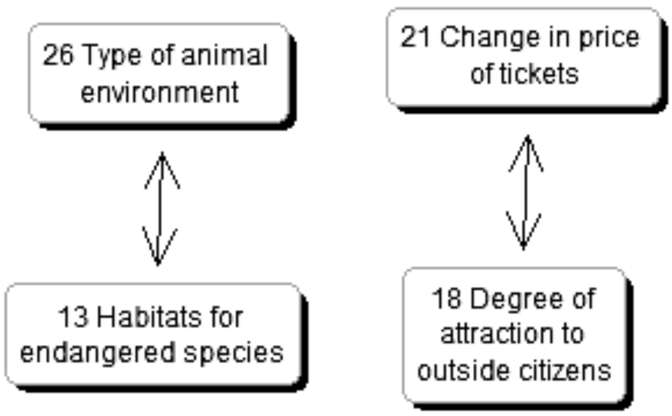
A paired samples t-test reveals a significant difference within both conditions. Significantly more DF were designated as causal factors in the cluster chain, compared to the lower number of final effects, in both conditions ($t_{high\text{-}threat}(14) = -2.10$, $p = .05$, CI = -4.98, .05; $t_{low\text{-}threat}(17) = -2.92$, $p = .009$, CI = -6.79, -1.10). There is a condition by construction interaction, as well, where the high-threat condition designated fewer causal DF compared to the low-threat condition ($M_{high\text{-}threat} = 10.73$, $SD = 4.09$, $V = 24.07$; $M_{low\text{-}threat} = 14.44$, $SD = 3.78$, $V = 14.26$; $F(1,31) = 6.03$, $p = .02$, $\eta^2 = .16$). The same trend is maintained for the number of DF designated as final effects of cluster chains, but the conditional differences are not significant ($M_{high\text{-}threat} = 8.27$, $SD = 4.15$, $V = 17.21$; $M_{low\text{-}threat} = 10.50$, $SD = 4.16$, $V = 17.32$). Both conditions appear to have used more diagnostic ST than predictive, to determine the logic between their CC. The low-threat condition, however, shows a greater degree of use with this vantage point.

D.2.3 ***Transitive & Loops***

The number of participants who created transitive and loop constructions is few, and most were removed from previous analyses due to being outliers. The high-threat condition includes three participants who created transitive and loop connections, but two participants are outliers. The remaining participant produced two transitive connections and no loops. They fall within the upper 75th percentile of cluster frequency and the 5th percentile of cluster size. Figure D.7 presents each transitive relationship, which both show a mix of categorical relationships.

Figure D.7. High-threat condition transitive relationships.

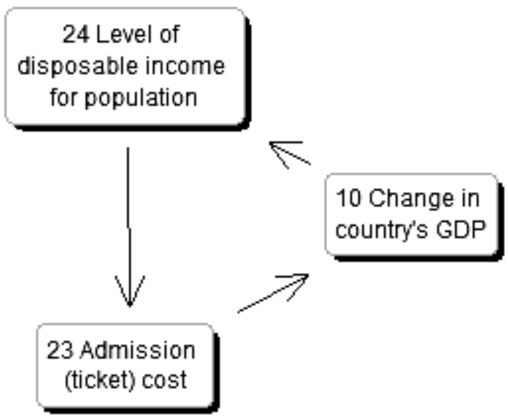
D.7.a. Environment & Resources D.7.b. Economics & Society

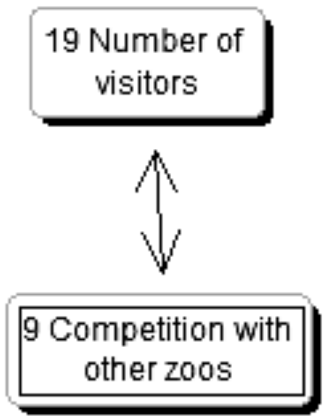


The low-threat condition includes four participants who created both transitive and loop constructions, but one is considered an outlier. Two participants each created one transitive relationship, and one participant created a small 3-way loop. Two participants fall within the upper 50th percentile of cluster frequency and the 30th percentile in cluster size. The anomaly in this group is the single participant who created a single transitive relationship, which occurs in the only cluster they created, that falls above the 99th percentile in size. Figure D.8 presents the transitive and loop relationships. Industry is slightly repetitive in focus between two participants' transitive relationships, and the looping relationship is within a single categorical focus. The remainder, however, is a mixture of other categories. In summary, there appears to be no trends of transitivity or loop constructions either between or within conditions. This is generally considered a beneficial outcome in SP. Looped causal relationships are difficult to work with, offer no clear resolution, and can confound future-oriented ST.

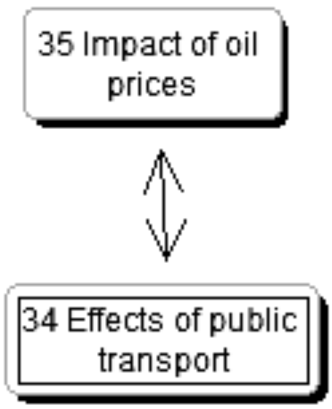
Figure D.8. Low-threat condition transitive and loop relationships.

D.8.a. Economics D.8.b. Demographics & Industry





D.8.c. Energy & Industry



D.3 Exploratory Stage 3 Discussion

Participants in both conditions appear to have favoured implicit levels of diagnostic assessments in their ST, which is surprising. Participants linked multiple alternative causes to DF in causal chains, which also reflects potential conjunction fallacies in use. Previous studies suggest that predictive order thinking is normatively stronger, and therefore more common, than diagnostic thinking (Bes, Sloman, Lucas, & Raufaster, 2012; Einhorn & Hogarth, 1982). Fernbach, et al. (2011) present a series of experimental studies that show not only a preference for predictive reasoning, but a neglect for finding alternative causes when reasoning predictively. The Fernbach, et al. experiments, however, show a greater effort in finding alternative causes in diagnostic reasoning. With the heterogeneous study participants showing a preference for reasoning diagnostically through their clustering efforts, the logic fits with other studies and models in that they showed greater efforts to find alternative causes.

Appendix E

Figure E.1. Heterogeneous clusters (1-10)

Figure. E.1.a

Cluster 1

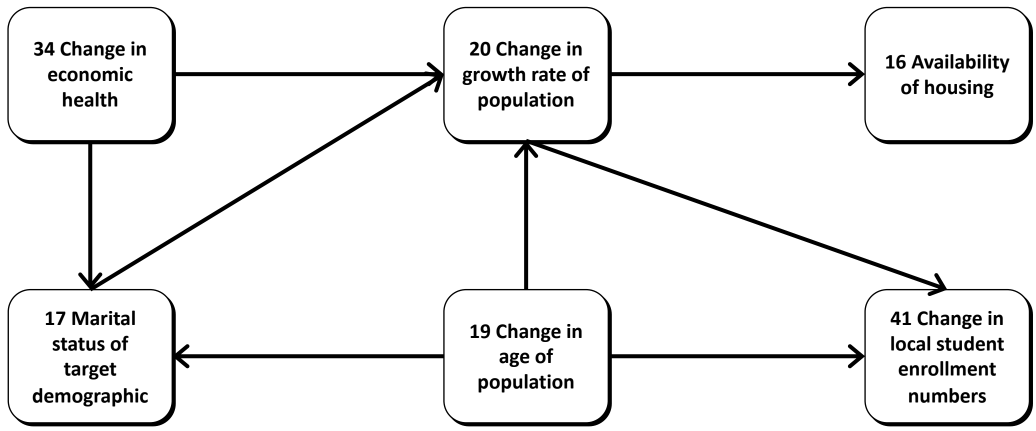


Figure. E.1.b

Cluster 2

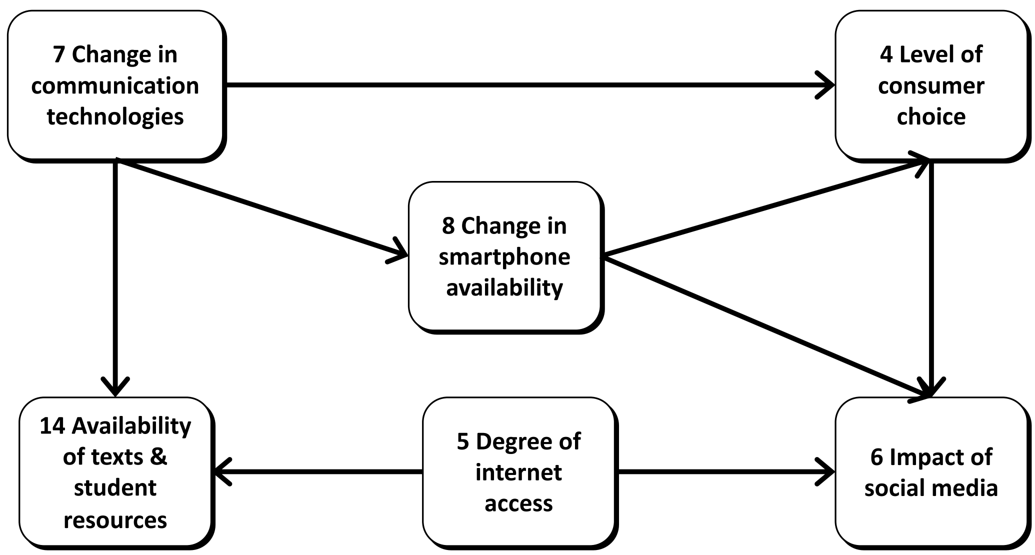


Figure. E.1.c

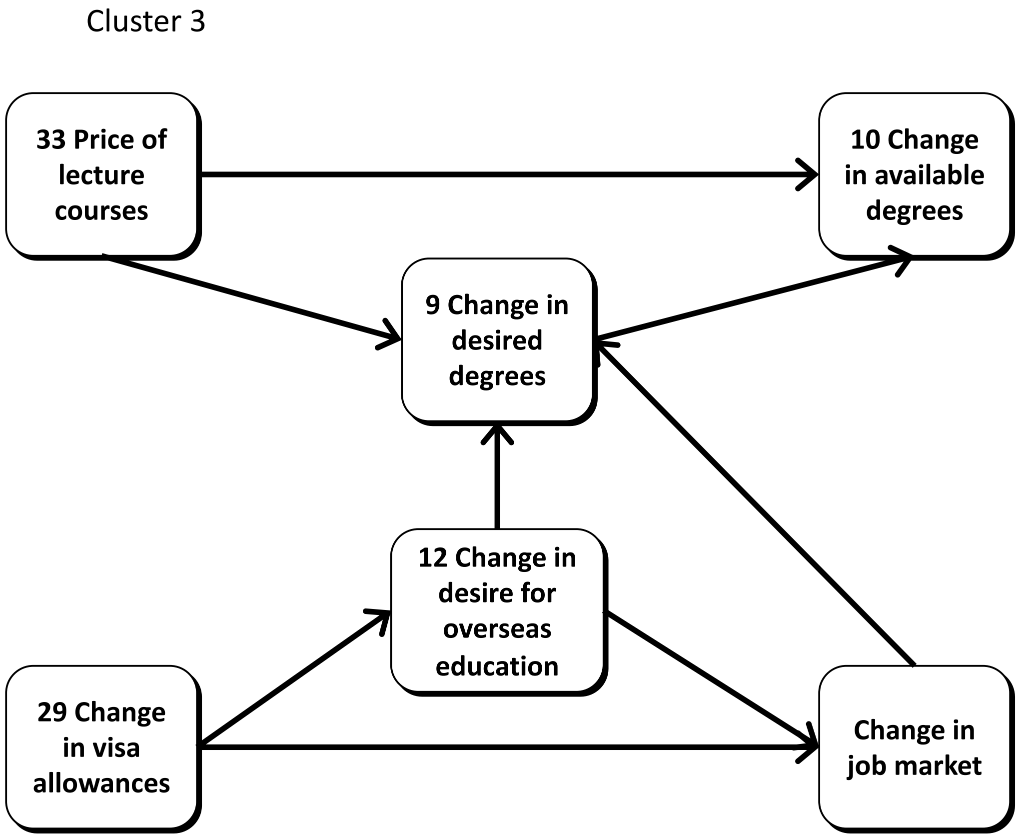


Figure. E.1.d

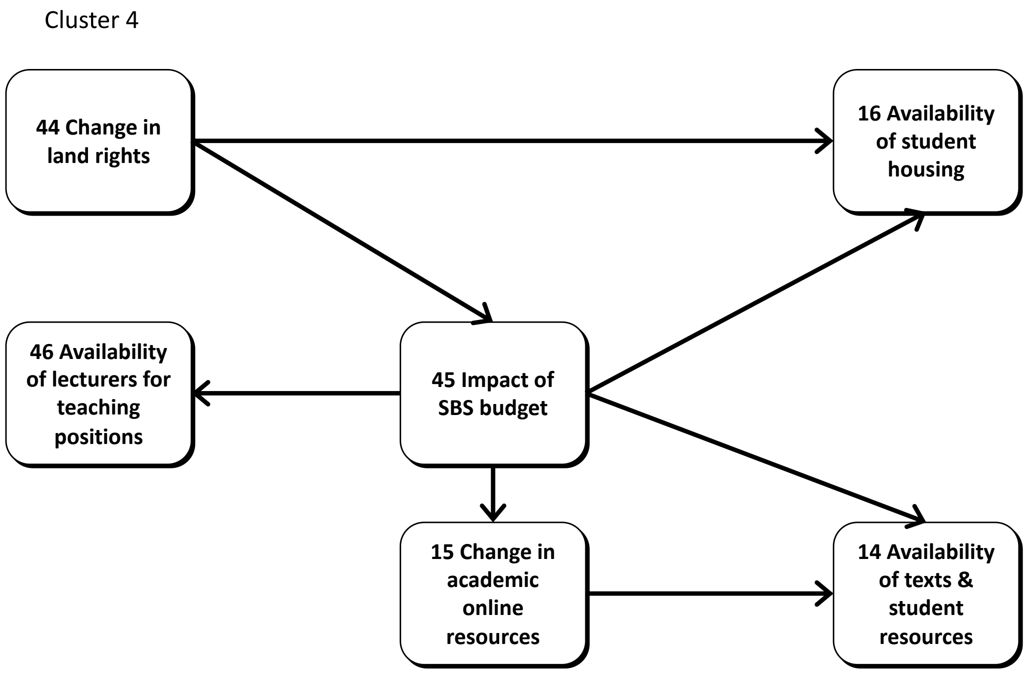


Figure. E.1.e

Cluster 5

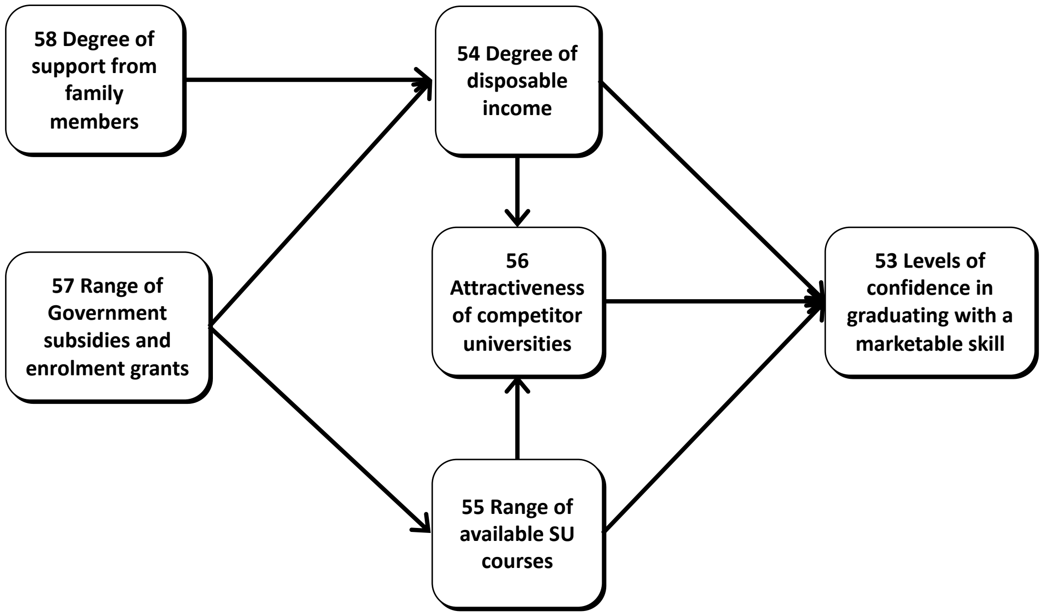


Figure. E.1.f

Cluster 6

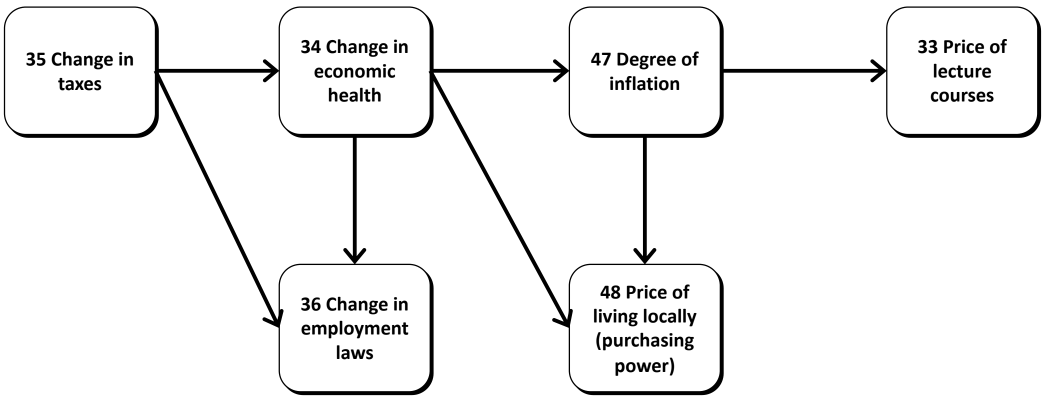


Figure. E.1.g

Cluster 7

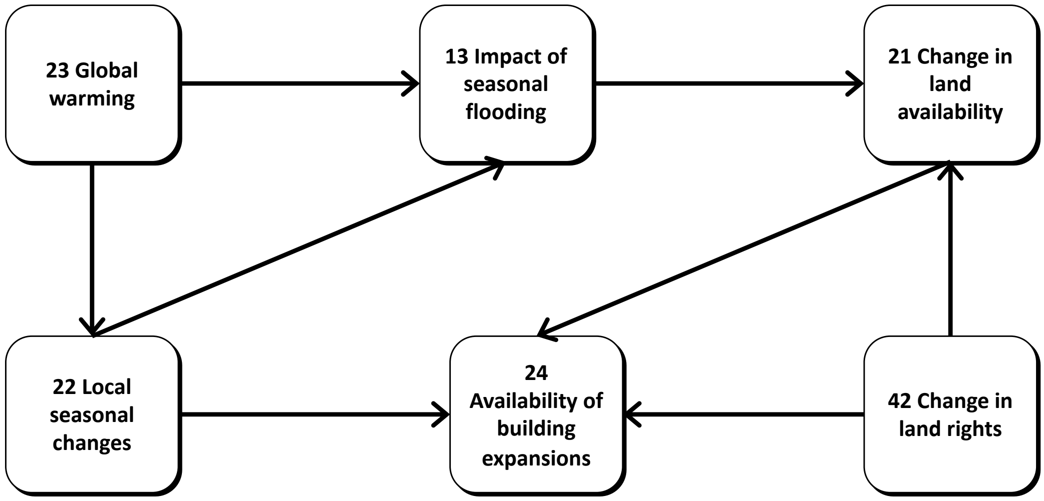


Figure. E.1.h

Cluster 8

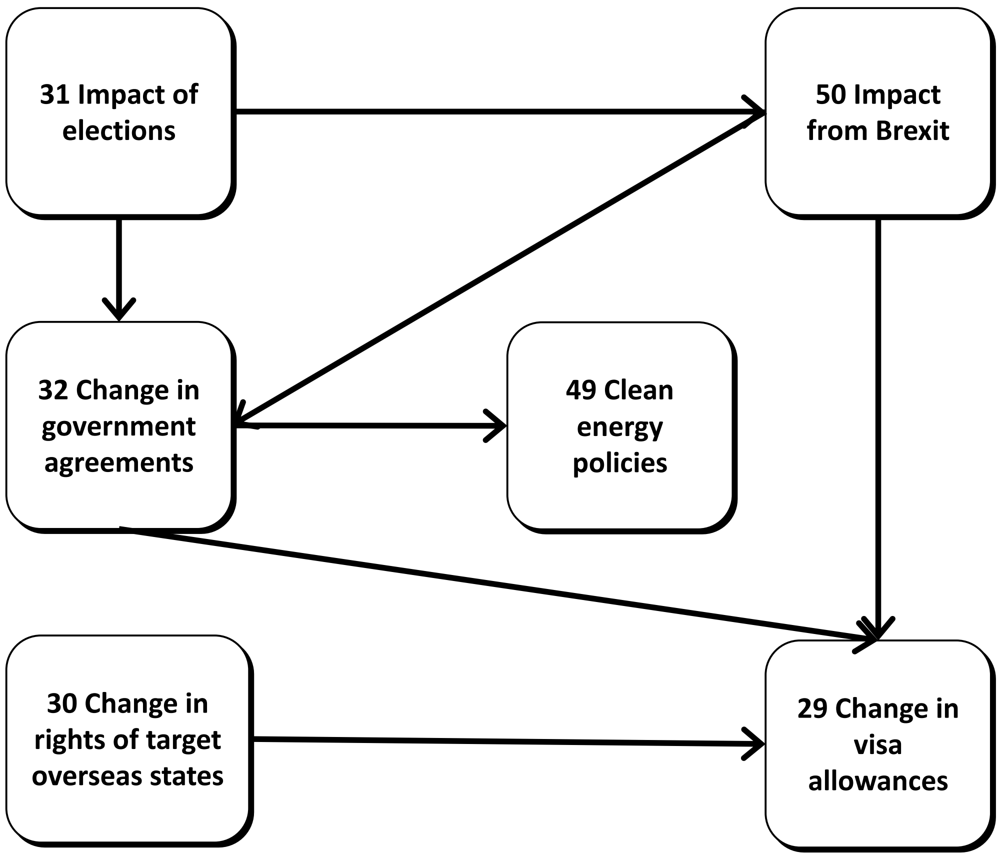


Figure. E.1.i

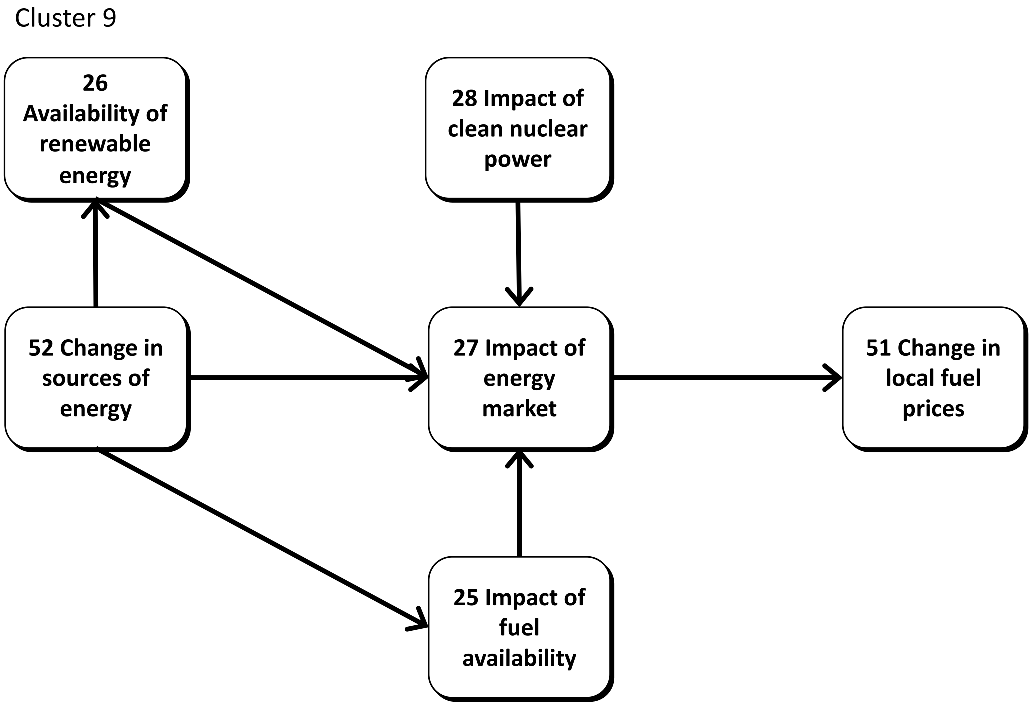
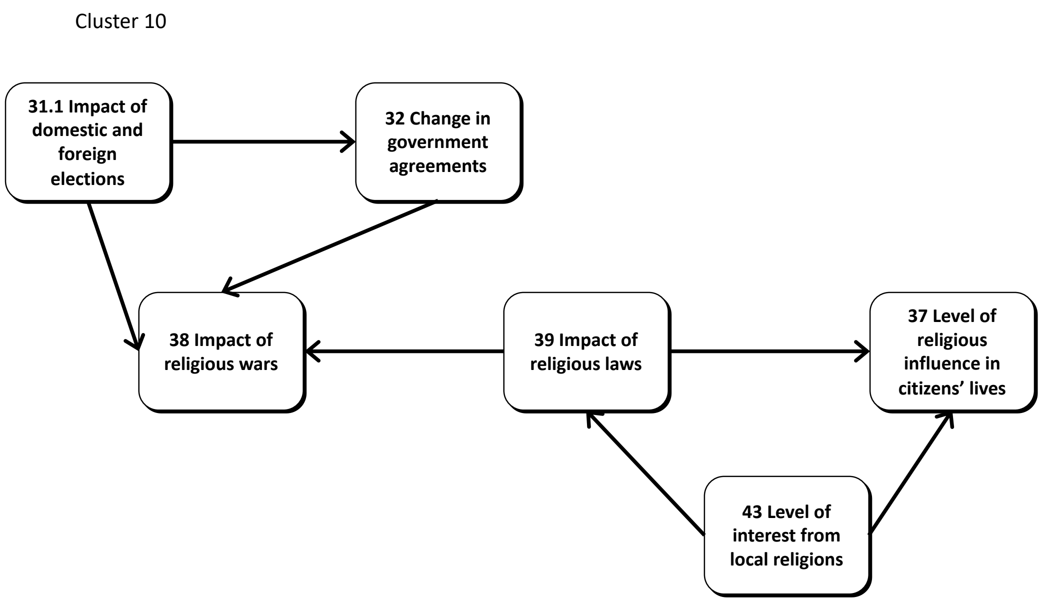


Figure. E.1.j



Appendix F

Figure F.1. High-threat homogeneous clusters (1-10)

Figure F.1.a

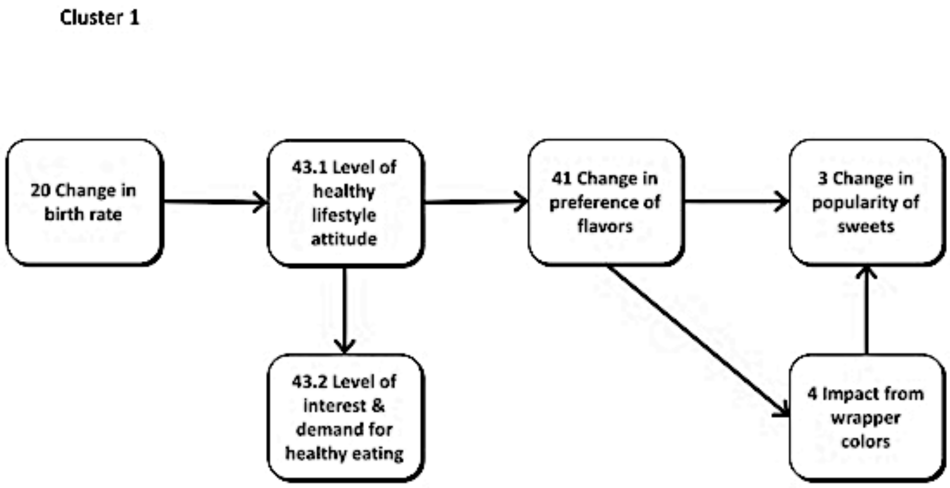


Figure F.1.b

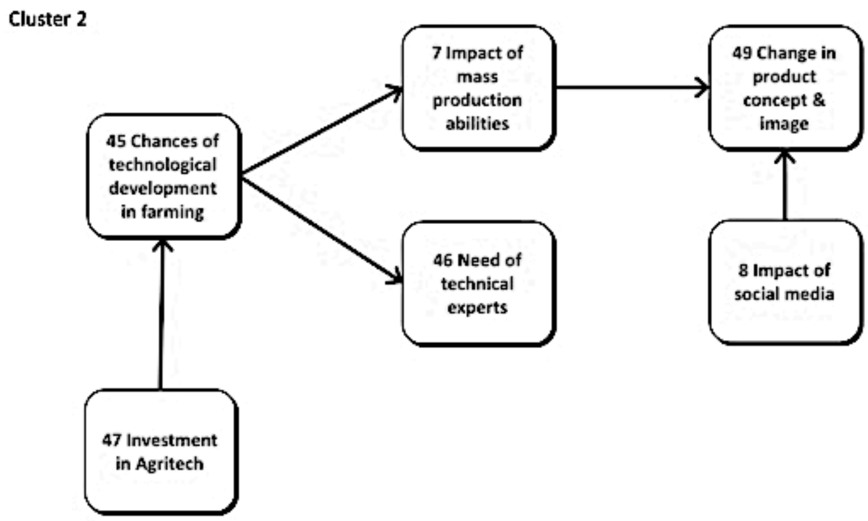


Figure F.1.c

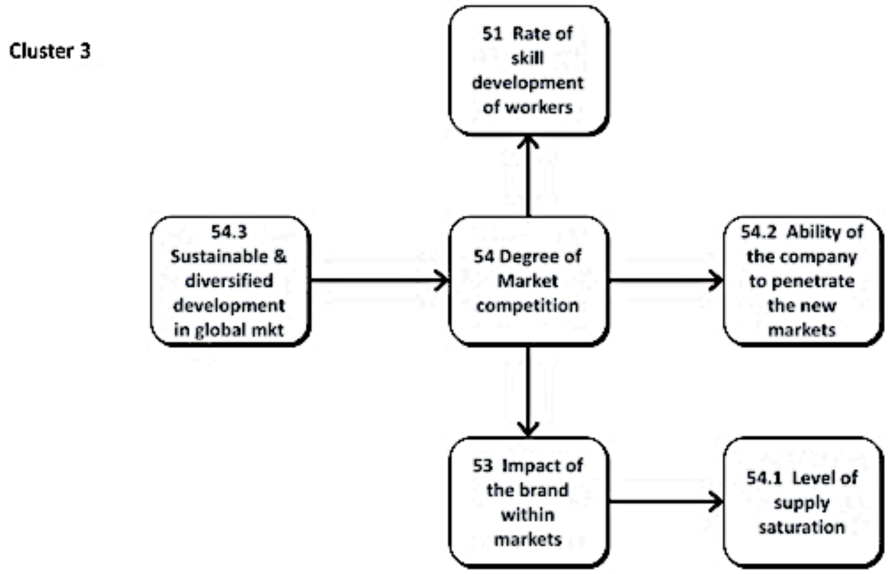


Figure F.1.d

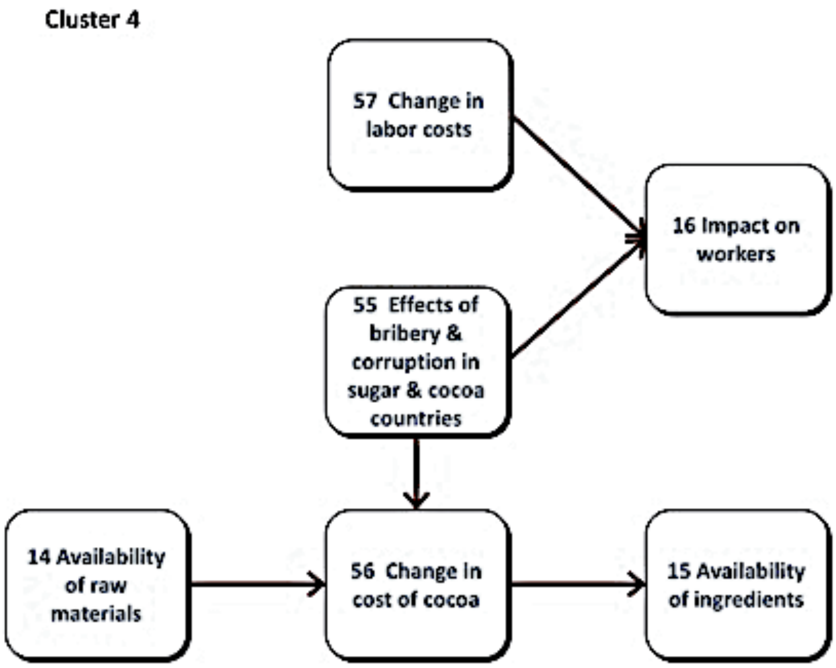


Figure F.1.e

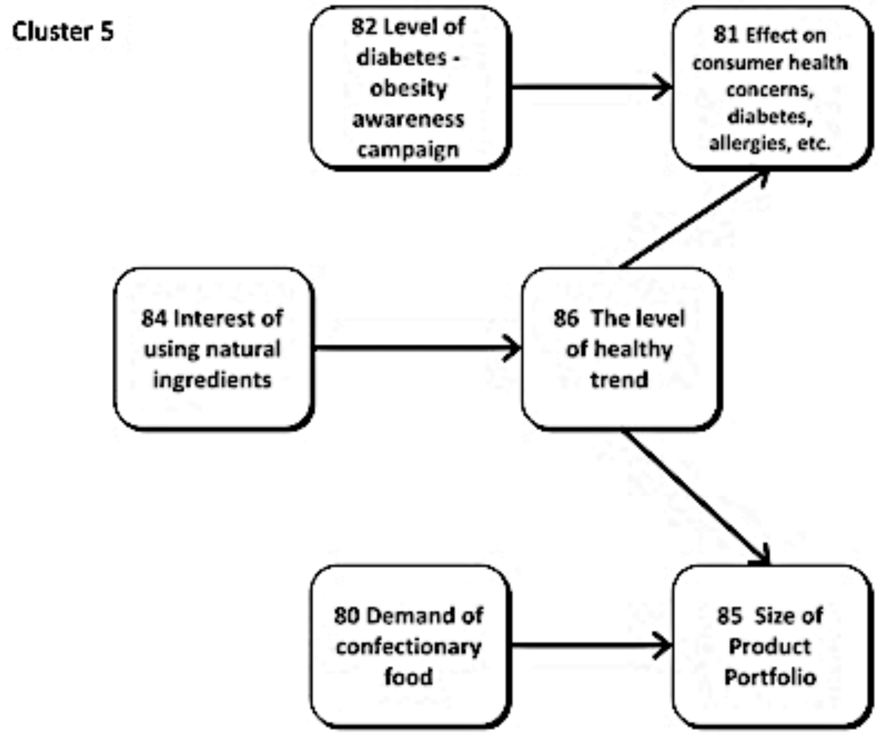


Figure F.1.f

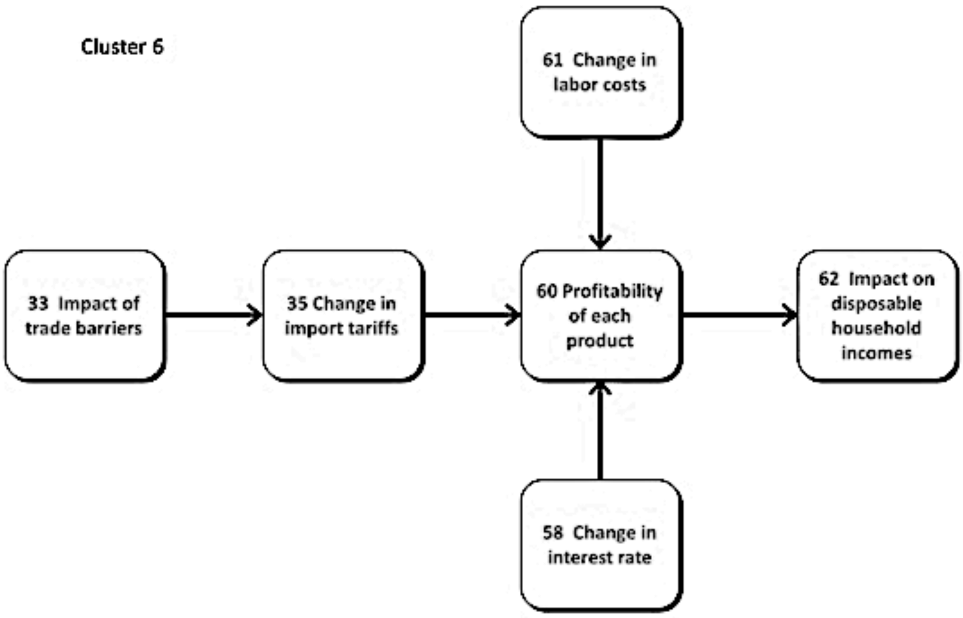


Figure F.1.g

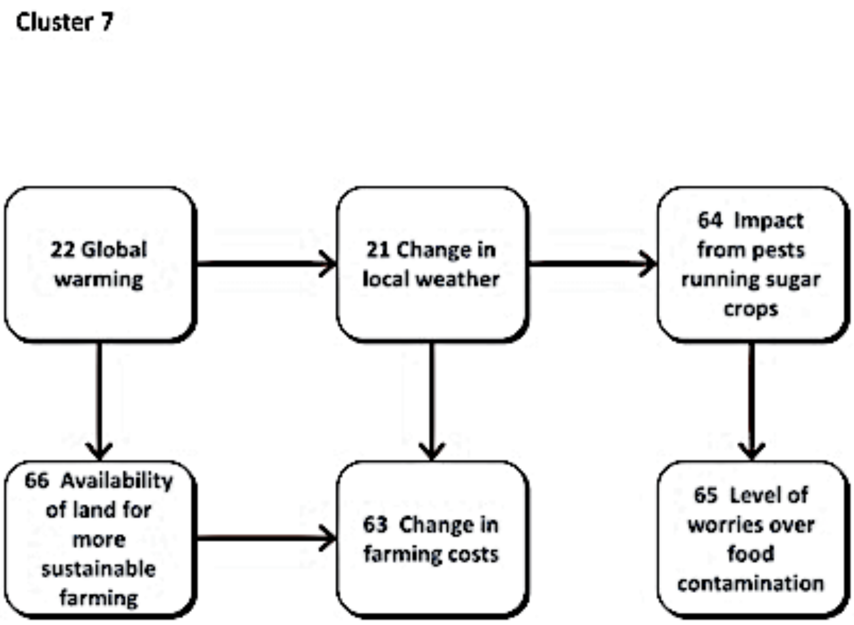


Figure F.1.h

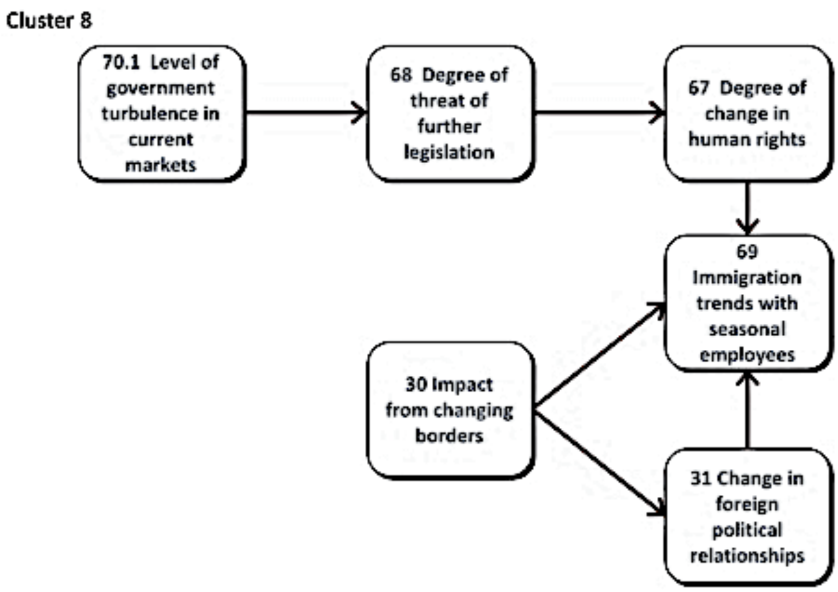


Figure F.1.i

Cluster 9

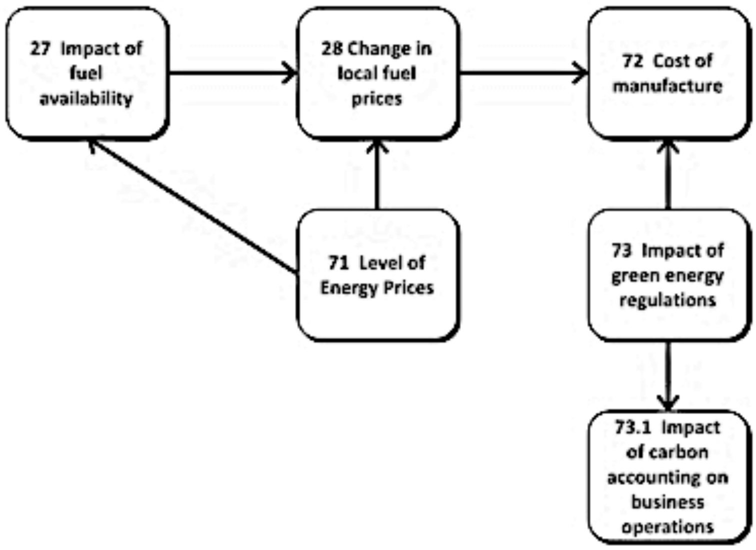


Figure F.1.j

Cluster 10

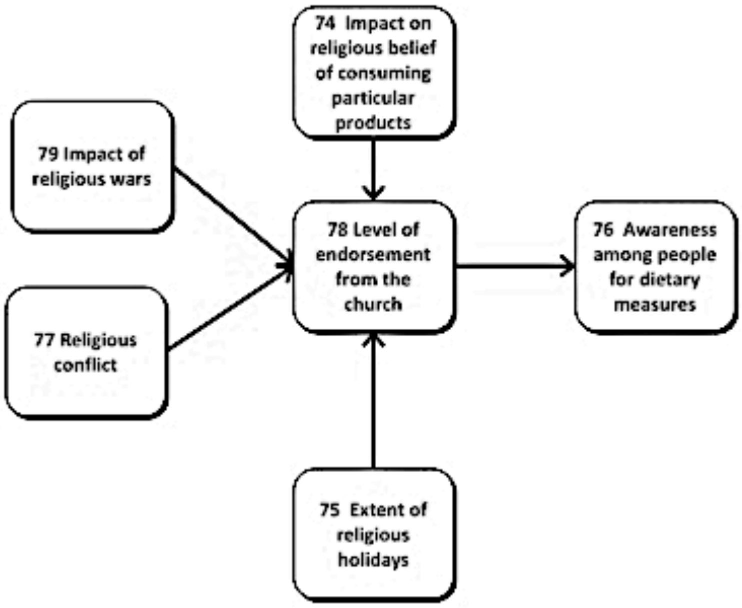


Figure F.2. Low-threat homogeneous clusters (1-10)

Figure F.2.a

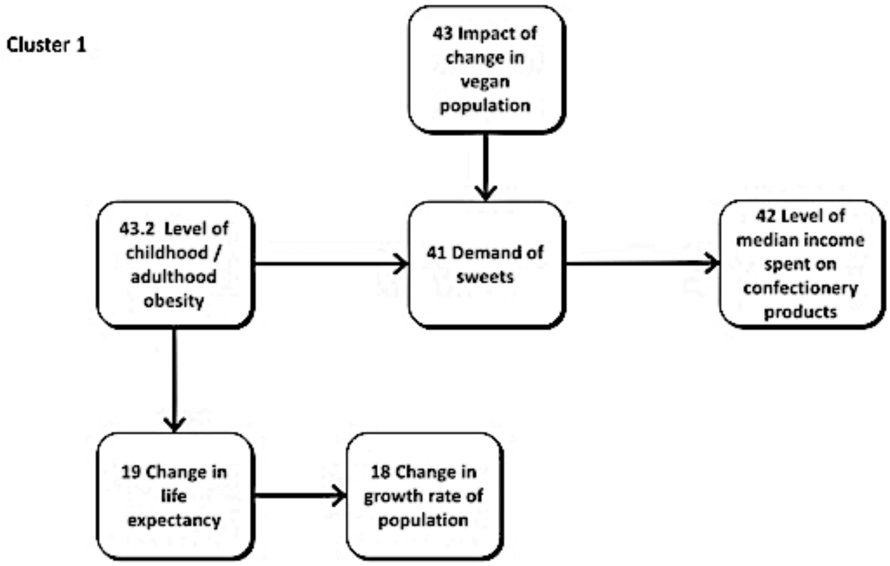


Figure F.2.b

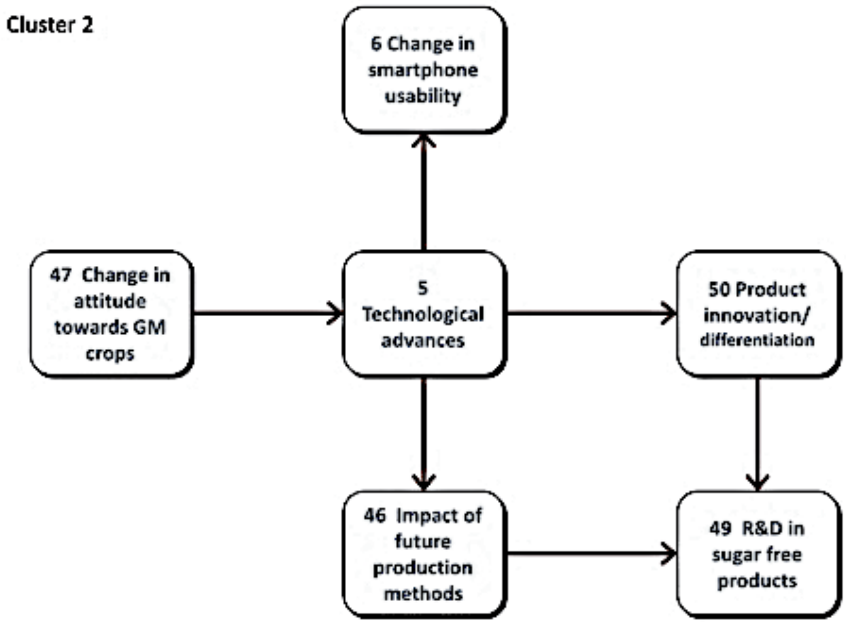


Figure F.2.c

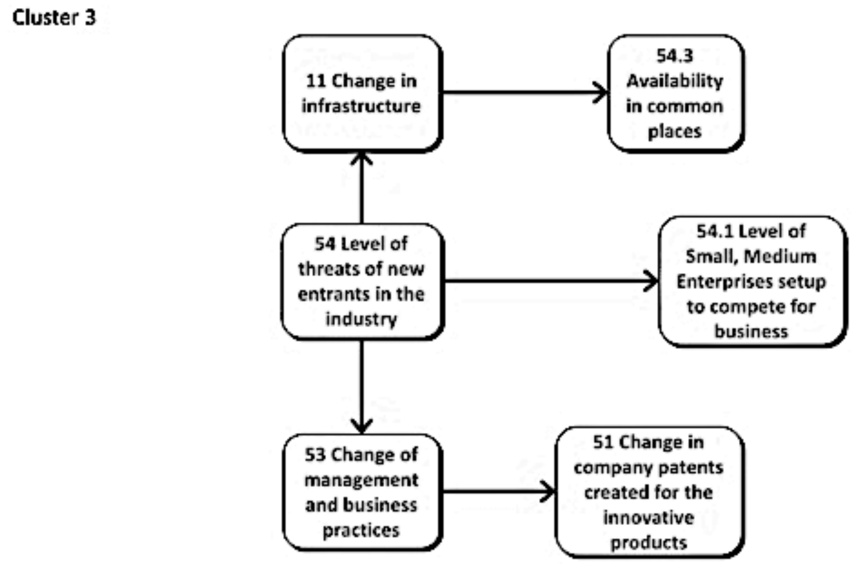


Figure F.2.d

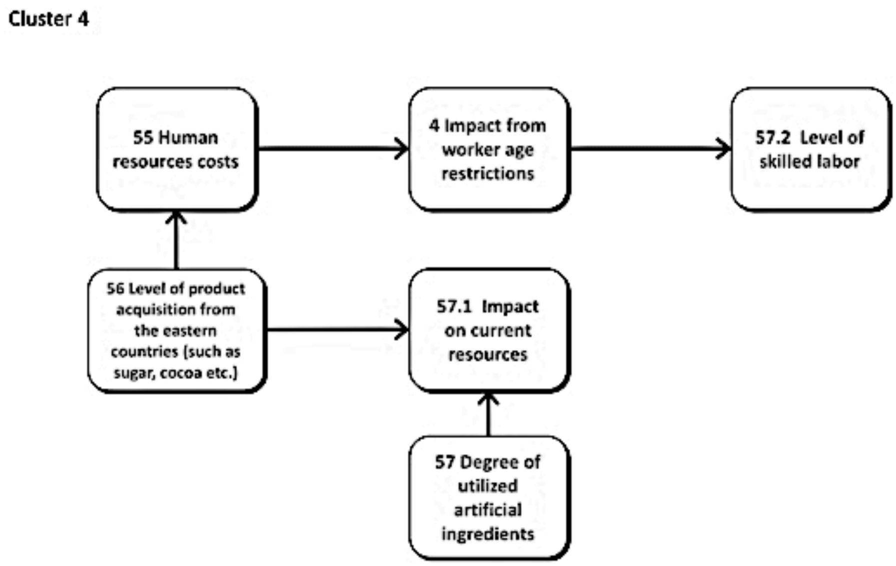


Figure F.2.e

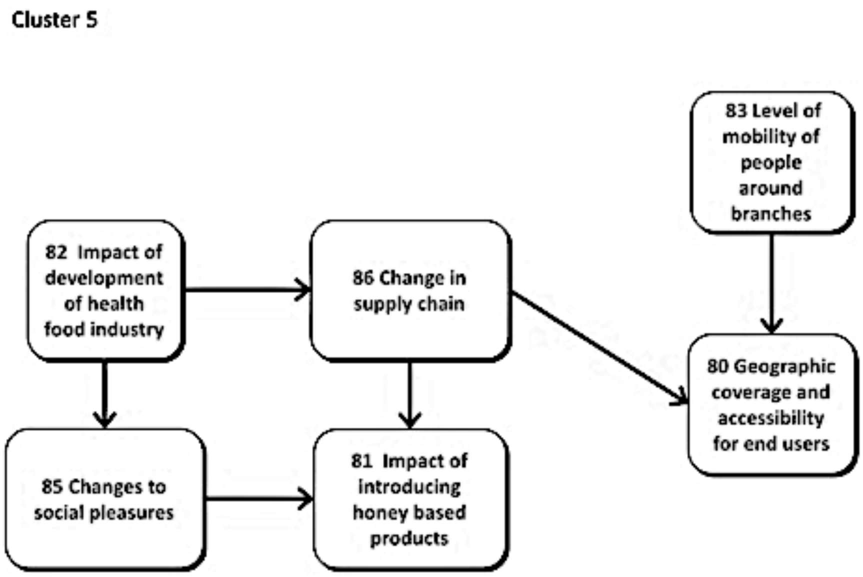


Figure F.2.f

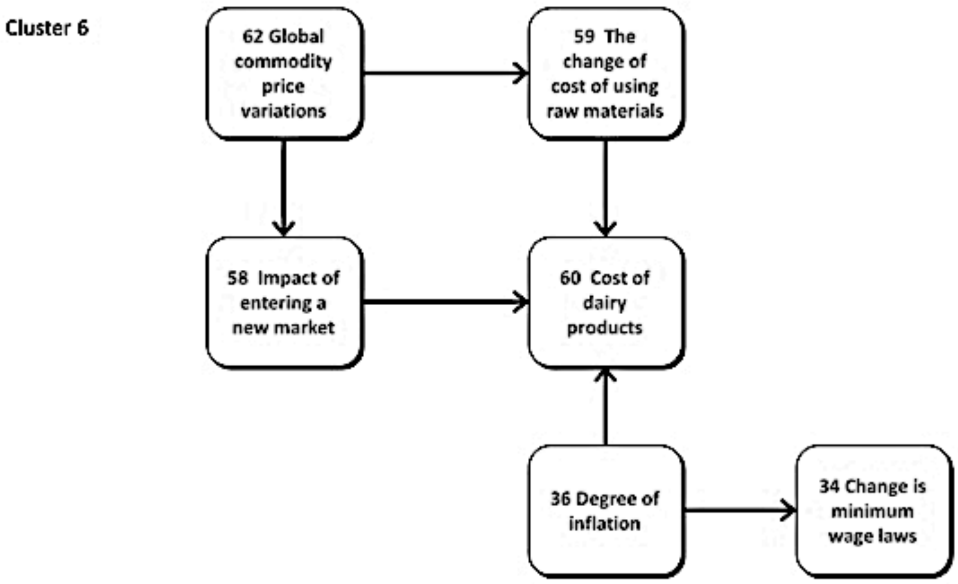


Figure F.2.g

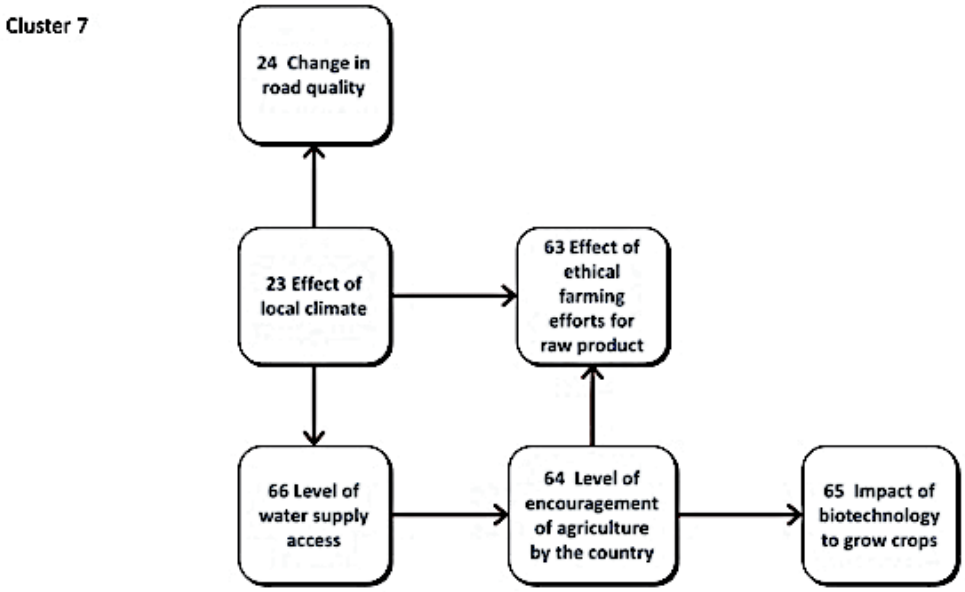


Figure F.2.h

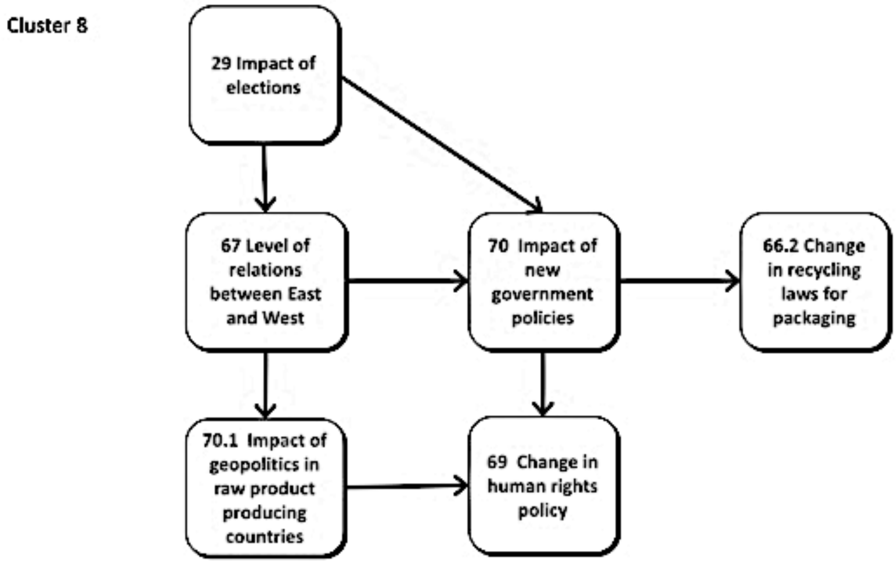


Figure F.2.i

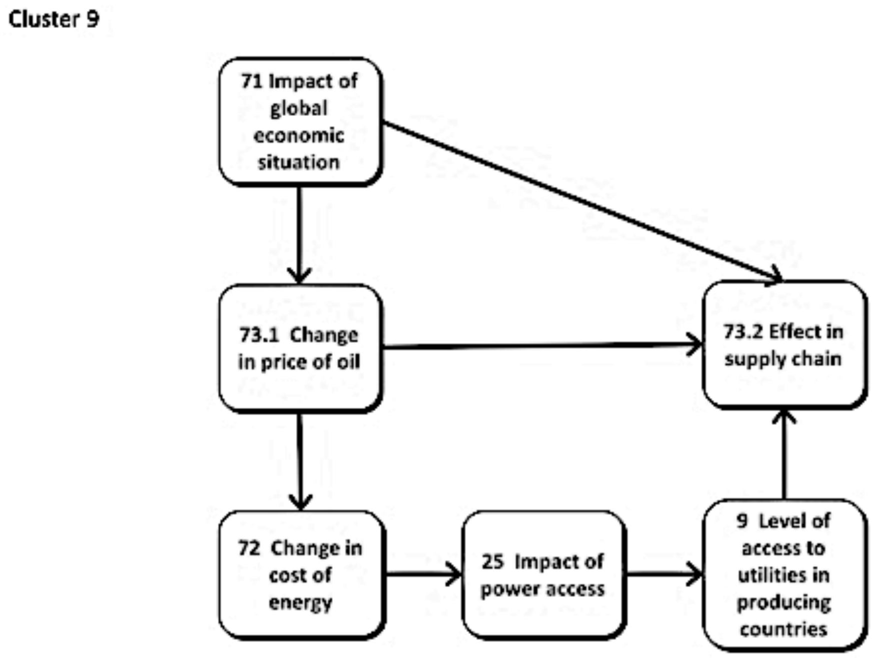
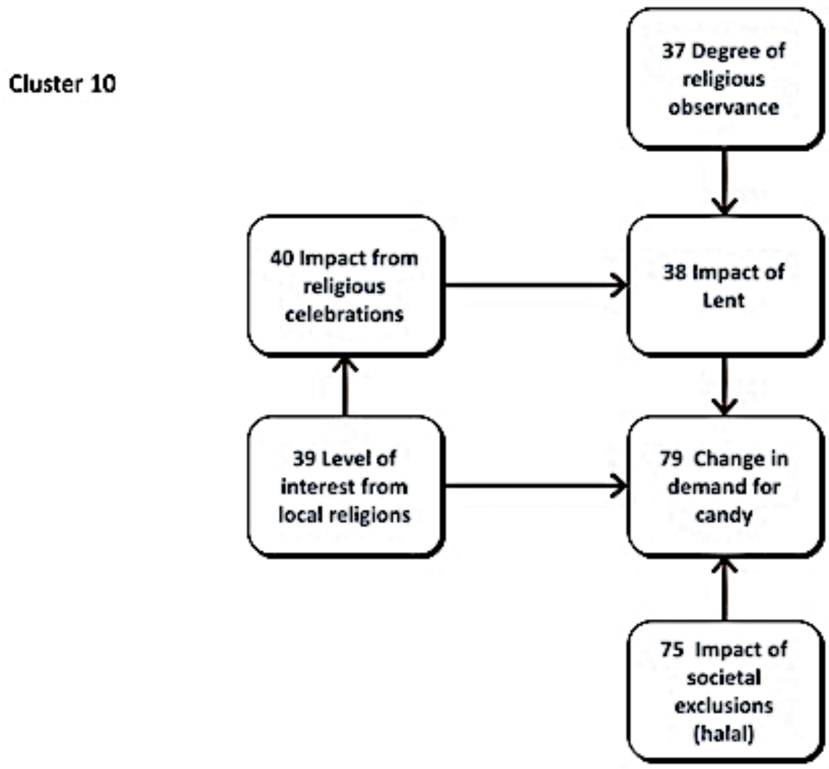


Figure F.2.j



Appendix G

In an effort to increase transparency within this thesis, the full collection of hypotheses are presented in their original language from the first submission and the few changes that were made to the hypotheses in response to examiners' specific requests for clarity and consistency in the second submission. Language is only altered where clarity was sought, to standardise the language, and fix grammatical mistakes (i.e. not all hypotheses). The only other alteration between the two submissions is the order the hypotheses are presented. The original order of the hypotheses presented in the first submission reflected the organic, emergent process of the investigation. However, this proved to be too difficult to follow, especially throughout the results sections, therefore some of the hypotheses are reordered to better reflect a logical flow of the chapter's arguments, analyses, and discussion. Of note, the original submission included a repeated presentation of hypothesis 7 and 12 in Chapter 8, which were combined into a single presentation in the final draft as hypothesis 11. The original hypothesis 11 was double numbered, therefore is now presented in accurate numerical order as hypothesis 12. Finally, the original hypothesis 15 in Chapter 9 presented two separate hypotheses and therefore was divided into hypotheses 13 and 14 in the present thesis draft. The hypotheses were *not* altered after data collection and neither analyses nor conclusions were altered between the first and second (final) submissions. The wording within the hypotheses are redrafted *only* to convey the logic more clearly, but does not fundamentally change the hypotheses. The first column presents the original order of the hypotheses. The second column presents the original wording of hypotheses. The third column presents the current order of the hypotheses. The fourth column presents the clarified language of the hypotheses.

Table G.1. Hypotheses across Section II, Chapters 7-9

Hypothesis (sub 1)	Submission 1	Hypothesis (sub 2)	Submission 2
1	Participants who read a high-threat business vignette will have a less creative output than participants who read a low-threat business vignette.	1	Participants who read a high-threat business vignette will generate lower creative output (fluency, flexibility, elaboration, and originality), than participants who read a low-threat business vignette.
2	Groups who read a high-threat business vignette will have a less creative output than groups who read a low-threat business vignette.	2	Groups who read a high-threat business vignette will create fewer new DF and eliminate fewer existing DF during the group-portion of SP, compared to groups who read a low-threat business vignette.
3	Groups who engage in dialectic inquiry will have a less creative output than groups who engage in consensus discussions.	3	Groups who engage in DI will create fewer new DF and eliminate fewer existing DF during the group-portion of SP, compared to groups who engage in CS.
4	Confidence will be lower for participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.	4	Average confidence scores will be lower for participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.
5	Participants will increase their confidence, overall, after participating in group discussions.	5	Participants will increase their confidence, overall, after participating in group sessions.

Table G.1. *(continued)*

6	Fewer causal associations between heterogeneous DF will be generated from participants who read a high-threat business summary compared to participants who read a low-threat business summary.	6	Fewer causal associations between DF will be generated from participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.
7	Confidence will be lower for participants who read a high-threat business summary compared to participants who read a low-threat business summary.	7	Fewer DF will be integrated into clusters, overall, from participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.
8	More independent clusters will be generated from participants who read a high-threat business summary compared to participants who read a low-threat business summary.	8	Clusters will, overall, be less complex from participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.
9	Fewer DF will be integrated into clusters, overall, from participants who read a high-threat business summary compared to participants who read a low-threat business summary.	9	More independent clusters will be generated from participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.
10	Clusters will be less complex from participants who read a high-threat business summary compared to participants who read a low-threat business summary.	10	There will be differences in focal concepts, by STIRDEEPER category, between threat conditions, which will reveal an anchoring bias.

Table G.1. *(continued)*

11	The type of potent concepts, by STIRDEEPER category, will reveal an anchoring bias.	11	Confidence will be lower for participants who read a high-threat business vignette compared to participants who read a low-threat business vignette.
11	Observed anchoring bias trends in the heterogeneous study will be intensified in the homogeneous study.	12	Observed anchoring bias trends in the heterogeneous study will be intensified in the homogeneous study.
12	Confidence will be lower from participants who read a high-threat business summary compared to participants who read a low-threat business summary.		
13	There will be a high-side bias across the impact hemispheres.	13	There will be an anchoring bias on the distribution of clusters, where participants presented with a high-threat business vignette will show a high-side bias on more clusters across the impact axis compared to participants presented with a low-threat business vignette.
14	There will be a high-side bias across the predictability hemispheres.	14	There will be an anchoring bias on the distribution of clusters, where participants presented with a high-threat business vignette will show a high-side bias on more clusters across the predictability axis compared to participants presented with a low-threat business vignette.

Table G.1. *(continued)*

15	There will be an anchoring bias on the distribution of clusters by category, where participants presented with a high-threat business vignette will show a high-side bias on more categories across both axes compared to participants presented with a low-threat business vignette.	15	Participants presented with a high-threat business vignette will group their clusters closer together (smaller spread) than participants presented with a low-threat business vignette.
16	Participants presented with a high-threat business vignette will group their clusters closer together (smaller spread) than participants presented with a low-threat business vignette.	16	Participants presented with a high-threat business summary will have lower confidence in their efforts compared to participants presented with a low-threat summary.
17	Participants presented with a high-threat business summary will have lower confidence in their efforts compared to participants presented with a low-threat summary.	17	There will be a high-side bias across the impact hemispheres.
		18	There will be a high-side bias across the predictability hemispheres.

"Do you mean to say that the story is finished?" said Don Quixote.

"As finished as my mother," said Sancho.